

CUESTIONES DE ASTRONOMÍA

Real Observatorio
Astronómico de Madrid



Ministerio de Fomento
Instituto Geográfico Nacional

CUESTIONES DE ASTRONOMÍA

Selección de artículos de divulgación
del Anuario del Real Observatorio de Madrid
publicados entre los años 2006 y 2016.



Edita: Centro Nacional de Información Geográfica (CNIG)
consulta@cnig.es

© de esta edición: Centro Nacional de Información Geográfica 2017

Instituto Geográfico Nacional <http://www.ign.es>
General Ibáñez de Ibero, 3, 28003 Madrid

NIPO: 162-17-016-4
DOI: 10.7419/162.04.2017

Cubierta: La luz y las ondas de choque originadas en el cúmulo NGC 602 de brillantes estrellas azules nacidas hace sólo 5 millones de años están abriendo una cavidad en el centro de la región de formación estelar, amontonando gas y polvo y desencadenando el nacimiento de nuevas estrellas en los bordes de la nebulosa N90. NGC 602 está situado en la Pequeña Nube de Magallanes (NGC 292), una galaxia enana muy cercana a la Vía Láctea. Cortesía de Hubble Space Telescope/NASA/ESA.

PRESENTACIÓN

El Anuario del Real Observatorio Astronómico de Madrid se viene publicando desde 1860, con alguna interrupción, correspondiendo la edición de 2018 al número 130 de la colección. Es tradicional que en él figuren un par de artículos de divulgación relativos a la astronomía en el sentido más amplio de la palabra (astrofísica, cosmología, instrumentación, su historia, su relación con la cultura y el arte), así como a temas afines. Los artículos están escritos por profesionales para un público interesado en la astronomía y la ciencia en general y, desde que empezó a publicarse en color, están profusamente ilustrados con fotografías, imágenes artísticas y diagramas. Un año después de su publicación en papel los artículos son accesibles en internet a fin de darles una mayor y continuada difusión. Los publicados desde 1994 en adelante pueden encontrarse en el enlace <http://astronomia.ign.es/web/guest/anuario-astronomico>.

El Comité Editorial del IGN/CNIG aprobó en 2016, por iniciativa del astrónomo Francisco Colomer, la publicación en formato electrónico de una selección de artículos de divulgación publicados en el Anuario en los últimos diez años, bajo el título *Cuestiones de Astronomía*. Con ello se trata de darles una nueva visibilidad, manteniendo su gratuidad. Los seleccionamos tratando de cubrir varias áreas temáticas. Los artículos no han sido modificados (salvo por la corrección de erratas y activación de enlaces) pero, dado que la astronomía y la tecnología evolucionan rápidamente, se ha considerado oportuno dar a los autores la posibilidad de complementar los artículos con un breve comentario de puesta al día. La recopilación consta de 17 artículos y 11 notas complementarias.

La maquetación de los artículos no ha sido modificada, por lo que su formato es análogo al del Anuario. Así mismo, se publica como dos archivos en formato PDF, uno adecuado para la navegación por el libro y su visualización en pantalla (incluso vertical), y otro que facilita la impresión en hojas A4 de las páginas que se considere oportuno.

La variedad de temas que se abordan, de manera independiente, en los distintos capítulos del libro muestran algunos de los muchos aspectos de la investigación actual en astronomía, a cual más fascinante. Confiamos en que cada lector encuentre artículos interesantes y que su lectura aumente su interés por los avances en la comprensión del Universo que continuamente nos depara esta bella ciencia.

Pere Planesas
Observatorio Astronómico Nacional

Mayo de 2017

Índice

I. Observando el cielo

1. La atmósfera y las estrellas. Las relaciones entre Meteorología y Astronomía. 9
2. El lado oscuro de la luz. 31
- 2b. Avances en la lucha contra la contaminación lumínica. 48

II. Los planetas

1. Las atmósferas de los planetas del sistema solar. 53
- 1b. Exploración de las atmósferas de Plutón, Marte y Júpiter. 76
2. Impactos sobre la Tierra. 79
- 2b. El meteorito de Chelyabinsk aviva la preocupación por los impactos. 109
3. La idea de planeta a lo largo de la historia. 111
- 3b. Más y más planetas. 133

III. El Sol y las estrellas

1. El Sol: nuestra estrella, nuestra energía. 137
2. ¿Cómo nació el Sol? Formación de estrellas tipo solar. 149
- 2b. Quince años de progreso ininterrumpido. 162
3. El coloquio de las nebulosas. 165

IV. El medio interestelar y las galaxias

1. El agua en el Universo. 187
- 1b. Avances debidos a Herschel y ALMA. 203
2. ¿Cuál es el destino de la Vía Láctea?. 205
3. El medio molecular en galaxias con brotes de formación estelar y en galaxias activas. 223
- 3b. La evolución de galaxias con brotes y activas. 254
4. La expansión acelerada del universo. 257

V. El tiempo y su medida

1. La medida del tiempo. 299
- 1b. Avances en la medida del tiempo. 326
2. La hora oficial en España y sus cambios. 329
- 2b. ¿Hora franquista, republicana, alemana, británica?. 359

VI. Retos tecnológicos de la astronomía actual

1. Navegación interplanetaria con VLBI usando las estrellas como referencia. 363
2. El Observatorio Espacial Herschel. 389
- 2b. Desarrollo de la misión y resultados de Herschel. 407
3. A la caza de las grandes ondas. 409
- 3b. A la caza de las grandes ondas diez más tarde. 429

Lista de autores

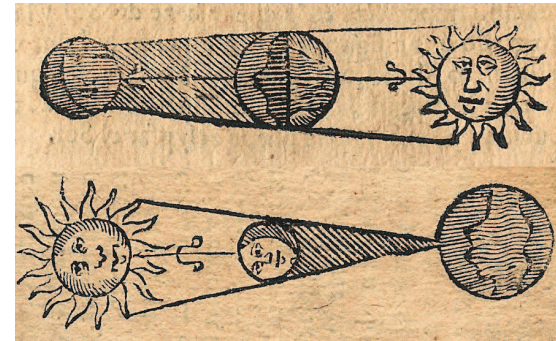
Alcolea Jiménez, Javier	VI.2 (2008)
Bachiller García, Rafael	III.1 (2009)
Bujarrabal Fernández, Valentín	III.3 (2014)
Cansado Auría, Alberto	I.1 (2012)
Colomer Sanmartín, Francisco	I.2 (2010)
Fuente Juan, Asunción	IV.1 (2012)
García Burillo, Santiago	IV.2 (2016)
Planesas Bigas, Pere	II.3 (2007), V.2 (2013), VI.2 (2008)
Rioja Capellán, María José	VI.1 (2006)
Solís García, Julio	II.1 (2015)
Tafalla García, Mario	III.2 (2002), IV.4 (2014)
Usero Villanueva, Antonio	IV.3 (2009)
de Vicente Abad, Pablo	II.2 (2013), V.1 (2011)

Créditos de las figuras

Las que encabezan los capítulos corresponden a libros antiguos de la Biblioteca del Real Observatorio Astronómico, cuyos autores son Pedro Apiano, Nicolás Copérnico y Lawrence Parsons.

Las que ilustran cada artículo han sido realizadas por el autor o, si no es así, se indica su procedencia.

OBSERVANDO EL CIELO



LA ATMÓSFERA Y LAS ESTRELLAS. LAS RELACIONES ENTRE METEOROLOGÍA Y ASTRONOMÍA

Alberto Cansado Auría

*Jefe del Servicio de Modelización de la Calidad del Aire
AEMET - Agencia Estatal de Meteorología*

Cuando hablamos de Meteorología, difícilmente nos paramos a pensar en la relación que mantiene con la Astronomía. A primera vista parecen ciencias muy diferentes y sin relación entre ellas. La una, es el arquetipo de la precisión absoluta: desde que Newton descubrió las leyes que regían el movimiento de los cuerpos celestes, con la poderosa Mecánica Celeste se puede predecir el instante preciso en el que se producirá un eclipse de Sol sobre cualquier punto de la Tierra o el momento en el que desde una localidad podemos ver cómo la Luna oculta una estrella. Mientras tanto, la otra trata de comprender los movimientos de una atmósfera caótica hasta el punto de que un simple aleteo de mariposa pueda llegar a provocar un tornado¹ y es incapaz de pronosticar con fiabilidad si hará Sol o lloverá más allá de cinco o seis días en el mejor de los casos.

¹El llamado “efecto mariposa” es una metáfora del concepto de sensibilidad a las condiciones iniciales que poseen algunos sistemas dinámicos. Edward Norton Lorenz (1917-2008) fue un meteorólogo americano que descubrió la sensibilidad de la atmósfera a las condiciones iniciales. En 1963 publicó sus conclusiones, en las que manifestaba que un meteorólogo le comentó, al conocer sus resultados, que eso significaba que el aleteo de una gaviota podría cambiar para siempre el tiempo (atmosférico). En sucesivas referencias, Lorenz cambió la gaviota por una mariposa. En 1972, tras fallar a la hora de elegir un título para su charla en la 139 Reunión de la Asociación Americana para el Avance de las Ciencias, Lorenz aceptó el título que le propuso Philip Merilees: “Does the flap of a butterfly’s wings in Brazil set off a tornado in Texas?”. A partir de entonces las referencias a los efectos del batir de alas de la mariposa han sido múltiples. Aunque la referencia al aleteo de la mariposa se han mantenido, la localización de la misma, sus efectos y el lugar en el que los efectos se sienten también han sido muy variados.

Y sin embargo, estas dos ciencias, aparentemente tan diferentes, están ligadas de forma estrecha y, además, sus caminos han circulado paralelos a lo largo de la mayor parte de la historia de la humanidad. A pesar de sus diferencias, sus relaciones son profundas.

La llegada del Neolítico, que no fue simultánea en todos los lugares pero que podemos situar hacia el año 9 500 A.C. en la zona de Oriente Medio, hizo que los hombres dejaran atrás la sociedad nómada de cazadores-recolectores para pasar a ser sedentarios en poblados donde llevaban a cabo actividades agrícolas y ganaderas. Al dar este paso, los ciclos naturales empezaron a cobrar gran importancia en la vida cotidiana de los hombres. Al ciclo diurno, que quizás hasta ese momento fuera el más importante puesto que era el que fijaba el periodo en el que el hombre podía realizar actividades, como la caza, con cierta seguridad, se añadieron otros como el ciclo anual que controlaba los trabajos agrícolas. Nuestros ancestros, que sin duda eran grandes observadores de la naturaleza, no pasaron por alto el hecho de que las regularidades estacionales climáticas, como los ciclos invierno-verano o periodos de lluvias frente a periodos secos, estaban ligadas a los ciclos astronómicos.

Me propongo repasar una serie de temas que nos las harán recordar.

El clima de la Tierra

Para empezar podemos recordar que la fuente de energía que origina los movimientos del aire en el seno de la atmósfera terrestre, y por lo tanto da lugar a la denominada circulación general atmosférica, es el Sol. Una estrella bastante vulgar de clase “G2 v” y sin ninguna característica que la haga extraordinaria a escala cósmica.

Al igual que el resto de los planetas del sistema solar, la Tierra debe su clima en gran medida a factores astronómicos. Entre esos factores astronómicos que condicionan el clima terrestre podemos citar la distancia de la Tierra al Sol y la inclinación del eje de nuestro planeta. Ellos son, junto a la composición de la atmósfera, los últimos responsables del clima terrestre. De hecho, la palabra griega de la que procede significa inclinación del Sol.

El clima está determinado por el balance de radiación del planeta, es decir, el equilibrio entre la radiación que alcanza la Tierra y la que emite. Existen tres vías para que este balance se altere: por un cambio en la radiación solar que llega a la tierra (hablaríamos de cambios en los parámetros orbitales de la Tierra o cambios en el Sol), por cambios en la fracción de radiación que es reflejada (cambios en el albedo causados por aumentos o disminuciones de la cobertura nubosa, alteraciones en la concentración de aerosoles o cambios en los usos del suelo) y, por último, por alteraciones en la radiación que el planeta emite (hablaríamos en este

caso de cambios en la composición de la atmósfera, por ejemplo en la concentración de gases de efecto invernadero). Los tres factores además están interrelacionados, de forma que existe una retroalimentación entre ellos.

Comencemos analizando los factores astronómicos que condicionan el clima. La inclinación del eje de rotación de la Tierra respecto a la eclíptica (actualmente $23^{\circ}26'$) provoca que a lo largo del año los hemisferios sur y norte tengan una iluminación diferente ocasionando la existencia de las estaciones. La situación se invierte a los seis meses, de forma que la fecha del solsticio que marca el comienzo del verano en el hemisferio norte es también la fecha en la que el invierno comienza en el hemisferio sur y viceversa. Sólo en dos momentos del año, cuando el Sol cruza el plano ecuatorial terrestre en los equinoccios, ambos hemisferios reciben la misma cantidad de luz.

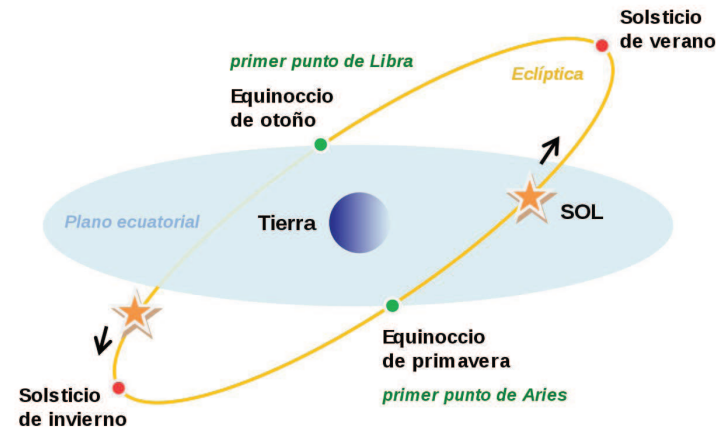


Figura 1: El Sol se mueve aparentemente durante el año sobre la eclíptica (línea amarilla). La eclíptica cruza el plano ecuatorial terrestre en dos puntos (los llamados primer punto de Aries y primer punto de Libra) y este hecho marca los equinoccios (igual duración del día y la noche).

La distancia Tierra-Sol es la adecuada para que, junto a una pequeña ayuda de los gases de efecto invernadero presentes en la atmósfera terrestre, nuestro planeta presente unas condiciones óptimas para que se haya desarrollado la vida tal y como la conocemos.

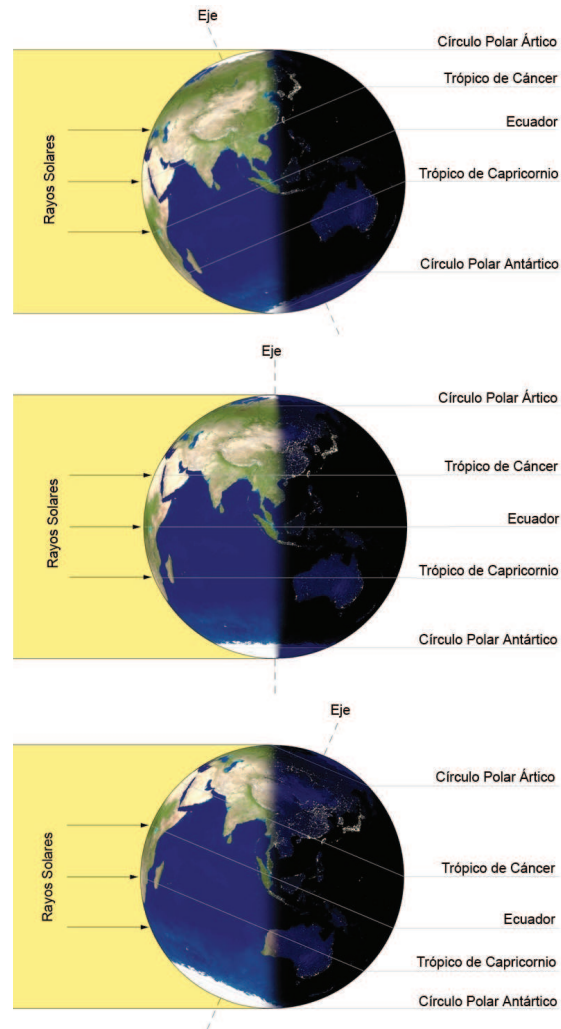


Figura 2: (Arriba) Durante el solsticio de verano (hemisferio norte), el Sol alcanza su máxima elevación y es el día de mayor duración del año. El hemisferio norte recibe la mayor parte de la radiación solar. (Centro) En los equinoccios la duración del día y la noche son prácticamente iguales. Ambos hemisferios están iluminados por igual. (Abajo) Durante el solsticio de invierno (hemisferio norte), el Sol culmina a la menor elevación del año y la duración del día es la más corta. El hemisferio sur recibe la mayor parte de la radiación solar.

Sin embargo, la excentricidad de la órbita de la Tierra en su movimiento de traslación alrededor del Sol no tiene prácticamente influencia en el clima de nuestro planeta ya que actualmente tiene un valor muy pequeño (0,017). De hecho, la diferencia entre las distancias en el perihelio (la máxima aproximación de la Tierra al Sol, 147 millones de kilómetros) y el afelio (el máximo alejamiento de la Tierra al Sol, 152 millones de kilómetros) es de sólo un 3 %. El perihelio tiene lugar los primeros días de enero, en el invierno boreal y el afelio a principios de julio, en el verano boreal.

Pero el clima no ha sido algo constante sino que ha habido variaciones muy significativas a lo largo de la historia de la Tierra. Son los cambios en esos factores astronómicos y sus interacciones con otros factores como el albedo o la composición atmosférica los que nos pueden ayudar a entender el clima en el pasado: la paleoclimatología trata de estudiar el clima de nuestro planeta a lo largo de la historia, sus variaciones y las causas de dichos cambios. Factores como la energía que llega del Sol, la situación astronómica, la distribución de masas de tierra y de agua y la composición de la atmósfera son importantes en paleoclimatología.

Las herramientas a las que recurre la paleoclimatología para averiguar el clima pasado son los denominados *proxies*. Los proxies (climatológicamente hablando) son características físicas del pasado que se han conservado hasta nuestros días en algún soporte y que, indirectamente, están relacionadas con algún elemento del clima. De esta forma, dichos proxies nos permiten reconstruir el clima de la antigüedad.

Existen diferentes proxies: Un análisis de los gases atrapados en un cilindro procedente de glaciares o de hielos polares (Groenlandia, Antártida) nos permitirá conocer cuál era la composición química de la atmósfera en épocas remotas. El espesor de cada capa de hielo de esos cilindros nos ayudará a estimar las precipitaciones anuales. Igualmente los árboles nos proporcionan información del clima antiguo: los vivos, de algunos cientos de años y los fosilizados extienden este periodo de forma muy importante.

Para ir más allá necesitaremos recurrir al análisis geológico de sedimentos o a medidas de algunos isótopos. El ^{18}O en el coral, por ejemplo, está relacionado con la temperatura del agua del mar. Por otra parte, los fósiles encontrados en los estratos de sedimentos marinos y lacustres nos permiten conocer la flora y fauna existente en diferentes épocas y nos ayudan a saber cuál fue el clima reinante en la época en la que vivieron. Así, si en un estrato encontramos restos fosilizados de flora y fauna tropical deduciremos que el periodo correspondiente fue cálido. Si, por el contrario, encontramos restos de coníferas y de animales como el mamut, concluiremos que el periodo fue frío.

De forma general podemos decir que la confianza y la precisión de estos registros disminuye conforme retrocedemos en el tiempo.

Se han buscado explicaciones a la existencia de periodos cálidos y fríos a lo largo de la historia de la Tierra. Kepler fue el primero en darse cuenta de que las órbitas que describen los planetas alrededor del Sol eran elipses y no circunferencias. Igualmente descubrió que cuando un planeta se encuentra más cerca del Sol (perihelio), su velocidad de traslación es máxima y cuando se encuentra más alejado (afelio) es mínima. Una de las primeras teorías fue postulada por el francés Joseph Adhémar, un matemático que hacia 1842 propuso que la razón por la cual la Antártida estaba cubierta de hielos era que el invierno es más largo en el hemisferio sur, puesto que el invierno austral coincide con la época del año en que la Tierra está en su afelio. De esta forma la suma de las duraciones de las estaciones frías (otoño e invierno) en el hemisferio sur es aproximadamente una semana más larga que la de las estaciones cálidas (primavera y verano). Lo contrario sucedería en el hemisferio norte.

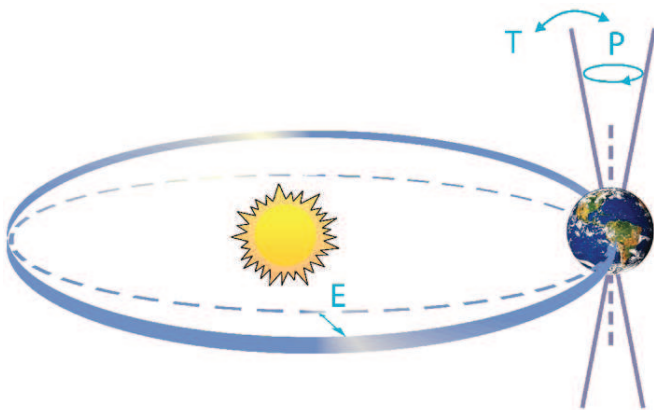


Figura 3: Esquema de los cambios en los parámetros orbitales terrestres. E denota cambios en la excentricidad por variaciones en el eje menor de la órbita, T cambios en la oblicuidad de la órbita y P en la precesión que cambia la dirección de la inclinación del eje. Fuente: IPCC Fourth Assessment Report: Climate Change 2007.

De acuerdo con Adhémar esta sería la razón por la cual la Antártida se encuentra cubierta de hielo mientras que el hemisferio norte disfruta de unas condiciones más templadas. Adhémar creía que la precesión de los equinoccios (y de los solsticios) haría que unas veces coincidiera con el afelio y otras con el perihelio y esto podría explicar los ciclos fríos que se correspondían con las glaciaciones. Alexander von Humboldt, en 1852 señaló un fallo en el razonamiento de Adhémar: es cierto que el verano es más largo en el hemisferio norte que en el hemisferio sur, pero

si esto sucede es porque la Tierra se halla en ese momento más alejada del Sol y recibe, por tanto, menor insolación. De esta forma, si hacemos un promedio de los periodos primavera-verano y otoño-invierno, ambos efectos se anularían y la insolación que recibirían ambos hemisferios sería idéntica. A pesar del fallo en la teoría que propuso Adhémar, ésta tuvo la virtud de poner el énfasis en las causas astronómicas de los cambios climáticos terrestres y de los ciclos de las glaciaciones.

El siguiente en dar un nuevo paso en la dirección correcta fue el escocés James Croll hacia 1860. Croll fue un científico autodidacta. Nacido en 1821, tuvo que dejar la escuela con 13 años para colaborar con su madre en el mantenimiento de su familia. Tras diversas vicisitudes acabó aceptando el puesto de conserje del Andersonian College and Museum ya que ese trabajo le permitía tener acceso a la biblioteca de la institución. A principios de la década de 1860 empezó a publicar ciertos trabajos científicos. Entre ellos se ocupó de la causa de las glaciaciones, tema éste de gran interés en los círculos académicos de la época. En 1876 publicó un libro llamado *Climate and Time* (El Clima y el Tiempo), en el que sintetizó diez años de investigaciones acerca de las glaciaciones y que le valió su aceptación como "Fellow of the Royal Society". En esta obra, Croll postulaba que las glaciaciones estarían causadas por la inclinación del eje terrestre, por el ciclo de precesión de los equinoccios que tiene un periodo de unos 26 000 años y por un tercer factor que sería el cambio de forma de las órbitas elípticas debida a las interacciones de los planetas del sistema solar. El astrónomo Urbain Le Verrier había probado que la excentricidad de las órbitas varía periódicamente con un periodo de unos 100 000 años. Pero como hemos visto el hecho de que la órbita sea más o menos excéntrica, no altera la cantidad de calor recibida por la Tierra. Croll pensó que el efecto importante aquí era el equilibrio entre el periodo de calor (verano) y frío (invierno). Según él, un invierno demasiado frío puede causar que se acumule mucha nieve, y esto dar lugar a que gran parte de la radiación que llega a la superficie terrestre se refleje e impidiera que se calentara. Sus ideas resultaron sorprendentemente correctas para la época y no entraban en contradicción con ningún dato conocido en ese momento. Sin embargo, a finales del siglo XIX los geólogos tuvieron evidencias de que el último periodo glacial había finalizado hacía sólo unos 10 000 años y su teoría se desechó. Su trabajo sirvió como punto de partida para que, ya entrado el siglo XX, Milankovitch propusiera una de las teorías más plausibles sobre el origen de los cambios climáticos en la Tierra: la conocida como "Teoría de Milankovitch".

Milutin Milankovitch fue un matemático serbio que a principios del siglo XX estudió la relación entre las variaciones conjuntas de los parámetros orbitales de nuestro planeta y los cambios climáticos que han tenido lugar a lo largo del tiempo. A los cambios que ya contenía la teoría de Croll (precesión de los equinoccios y variación de la excentricidad orbital) Mi-

lankovitch añadió un tercer factor: la variación en el ángulo de inclinación de la órbita terrestre (que varía entre un mínimo de $22^{\circ}1'$ y un máximo de $24^{\circ}5'$) con un periodo de 41 000 años. Milankovitch estudió cómo los cambios combinados de la excentricidad de la órbita terrestre, la inclinación del eje respecto a la eclíptica y la precesión axial influían en la radiación solar. La tarea no era pequeña en un momento en el que todavía no existían los ordenadores. La contribución clave de Milankovitch fue incorporar la idea del climatólogo Wladimir Köppen de que el factor desencadenante de una glaciación no era tanto una sucesión de inviernos crudos sino una reducción de la insolación en verano. Esta idea es considerada la clave de la teoría astronómica de las glaciaciones.

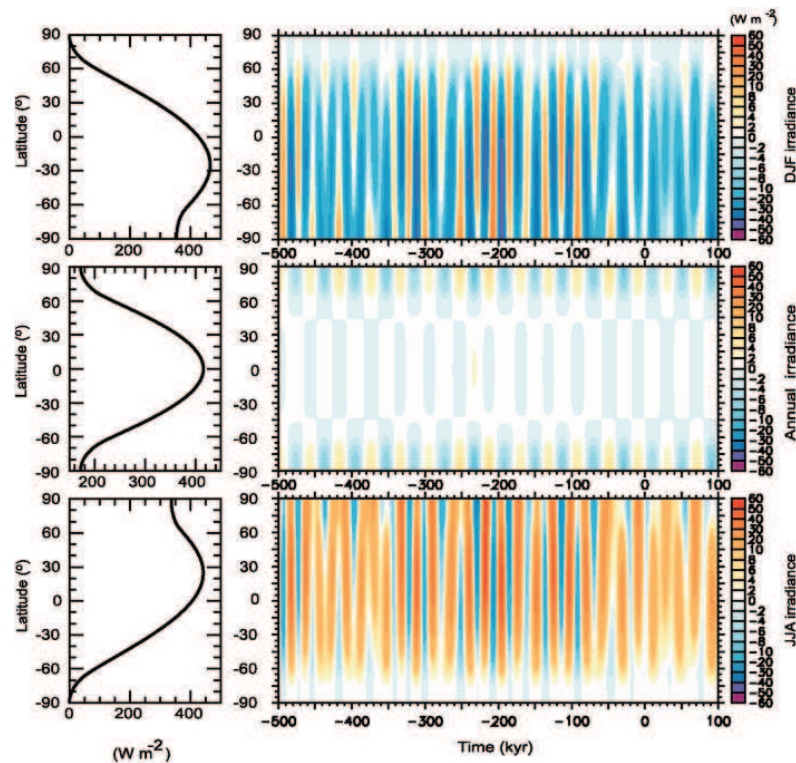


Figura 4: A la izquierda se presenta la distribución de radiación solar incidente actual para los meses de diciembre a febrero (arriba), media anual (centro) y de junio a agosto (abajo). A la derecha, se muestran las variaciones de la radiación respecto a la actual de los últimos 500 000 años y para los próximos 100 000. Fuente: IPCC Fourth Assessment Report: Climate Change 2007.

Hoy en día, existen evidencias de que las glaciaciones que han tenido lugar en los últimos 3 millones de años en ciclos regulares están ligadas a variaciones regulares con los cambios orbitales de la Tierra alrededor del Sol. Las últimas investigaciones publicadas en el Cuarto Informe de Validación del Panel Intergubernamental del Cambio Climático (*Climate Change 2007*) realizadas mediante simulaciones con modelos climáticos muestran que la radiación recibida en verano en las zonas continentales del hemisferio norte es crucial a la hora de desencadenar una glaciación. Cuando esta cae por debajo de un valor crítico (algunos autores sitúan el umbral en 40 W/m^2 menor que la actual a una latitud norte de 65° en verano) la nieve caída el invierno anterior no se derrite y permanece en el suelo. La capa de nieve empieza entonces a acumularse haciendo cada vez más difícil el deshielo y haciendo que la capa de hielo sea cada vez de mayor espesor y extensión.

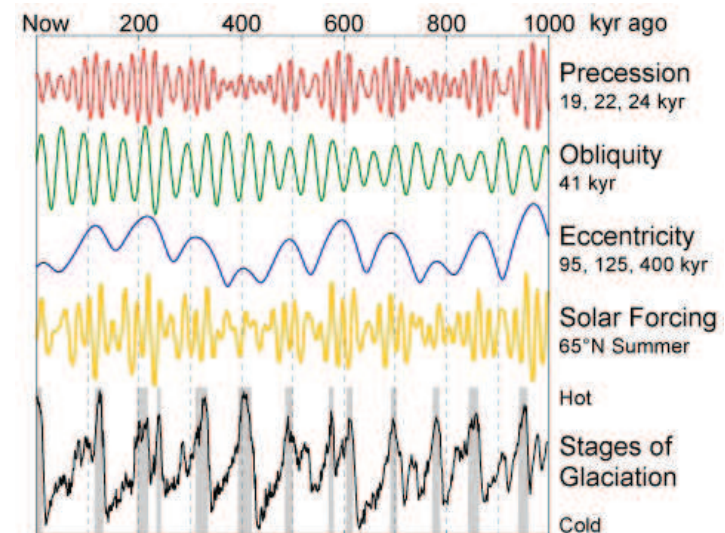


Figura 5: Efectos de la precesión, cambios en la excentricidad de la órbita y de la inclinación del eje de la Tierra en el forzamiento radiativo solar en verano a 65° Norte. Abajo se muestran los periodos cálidos y fríos. La escala de tiempo muestra miles de años.

Según los defensores de esta teoría, estos cambios darían lugar a oscilaciones periódicas (sinusoidales) de diferentes periodos que superpuestos serían los responsables últimos de los cambios por los que ha atravesado el clima terrestre a lo largo del tiempo.

El Sol y su influencia en el clima

La teoría de Milankovitch nos ha enseñado la importancia que sobre el clima tienen las variaciones estacionales de la energía que incide sobre nuestro planeta, incluso si la energía media a lo largo de un año no varía. ¿Qué pasaría si la cantidad de energía que llega de nuestra estrella varía? Sabemos que el Sol no es una estrella que se mantiene estática sino que está sometida a variaciones. Unas veces se encuentra calmado, sin apenas actividad, y en otras ocasiones presenta manchas y fulguraciones con tormentas solares que afectan incluso a las transmisiones radio y pueden llegar a dañar satélites artificiales. También sabemos que, normalmente, estos periodos de tranquilidad y actividad se repiten aproximadamente cada 11 años.

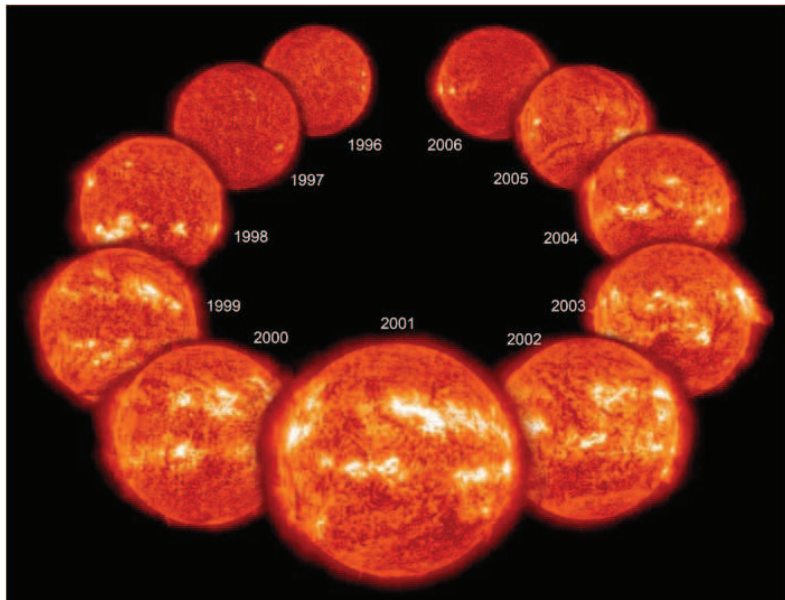


Figura 6: *El Solar and Heliospheric Observatory (SOHO)* que fue lanzado en Diciembre de 1996, cerca de un mínimo de actividad solar ha monitorizado la actividad del Sol. Desde entonces, ha sido testigo de un ciclo completo con un máximo en torno a 2001-2002 y un nuevo descenso de la actividad desde entonces.

La influencia del Sol sobre el tiempo meteorológico ha sido señalada en la literatura científica a partir de la segunda mitad del siglo XX. Al parecer, se conoce que la actividad de las borrascas que cruzan latitudes

altas ($> 60^\circ$ Norte) es mayor tras una serie de fulguraciones solares que tienen lugar en un breve espacio de tiempo. Este hecho, señalado por vez primera por Walter Orr Roberts para las borrascas que llegaban a Alaska desde el Pacífico, llamó la atención y empezó a ser investigado con mayor profundidad. Hoy en día los datos observados muestran que una actividad solar intensa sería capaz de inducir modificaciones en los sistemas de bajas presiones en desarrollo unos dos o tres días después.

Sin embargo la sospecha de que existe una relación entre el tiempo y la actividad solar viene de un poco más atrás. A pesar de no conocerse ningún mecanismo físico capaz de explicar la relación entre manchas solares y tiempo, remontándonos a la segunda mitad del siglo XIX, algunos científicos pretendían haber probado relaciones entre el ciclo solar y las cosechas inglesas o el nivel de las aguas en los lagos africanos, aunque siendo sinceros, es dudoso que dichas correlaciones fueran correctas.

A partir de 1970, analizando series históricas de observaciones meteorológicas, a las que se han aplicado técnicas estadísticas sofisticadas, han probado que en ellas existen ciclos de 11 años que pueden explicarse por la influencia de las variaciones de la actividad solar. A pesar de que el efecto es pequeño se ha mostrado que un Sol tranquilo implica una Tierra fría.

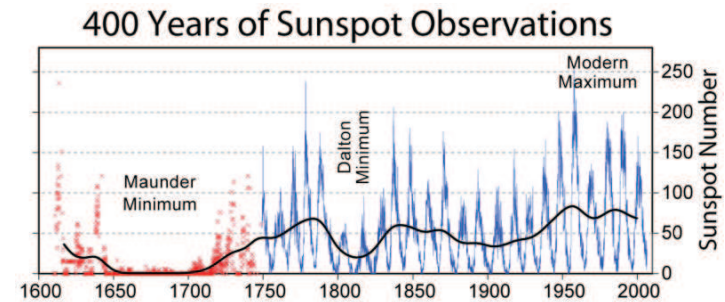


Figura 7: Actividad solar en los últimos 400 años. Desde 1749 existen observaciones mensuales que han sido incluídas. Para el resto del periodo existen observaciones esporádicas que han sido compiladas. Fuente: Solar Influences Data Analysis Center del Real Observatorio de Bélgica.

Una prueba más que apoya esta teoría sería la serie de temperaturas de Europa entre 1645 y 1715. Fueron años muy fríos en los que los canales y los ríos de Gran Bretaña y de Holanda se helaban. Incluso el mar Báltico llegó a helarse por completo permitiendo a los trineos viajar desde Polonia a Suecia. Después de conocer estos hechos no resulta sorprendente saber que esa época, que se conoce como la “pequeña edad del hielo”, fuera un periodo que se correlaciona perfectamente con aquel en el que el Sol

estuvo extremadamente calmado, sin apenas manchas ni actividad, en el denominado mínimo de Maunder.

En consecuencia, el análisis de las series históricas de datos muestran que existe una cierta correlación entre la temperatura y el número de manchas solares (que es la medida más común de la actividad solar). Además hay datos históricos que apoyan estas tesis como la existencia de la “pequeña edad del hielo” que tuvo lugar durante el mínimo de Maunder. Pero aún habría más, estudios climáticos realizados analizando los anillos de los árboles y que permiten remontarnos hasta mil años atrás han mostrado también periodicidades de 11 años que apoyarían también esta hipótesis reforzando su validez.

La observación astronómica y la meteorología

Existe una obvia relación entre meteorología y observación astronómica. Las condiciones meteorológicas resultan clave a la hora de realizar cualquier sesión de observación. Es evidente que la observación óptica resulta imposible en condiciones de cielo cubierto. Igualmente, las condiciones anticiclónicas que se prolongan en periodos largos de tiempo resultan perjudiciales, ya que la subsidencia del aire en la atmósfera ocasiona que grandes cantidades de polvo se inyecten en la atmósfera haciendo que se torne turbia. Lo mismo ocurre cuando el viento ocasiona intrusiones de polvo procedentes de zonas desérticas, que además de ser perjudiciales para la salud, ocasionan un aumento de la atenuación atmosférica.

Otro aspecto en el que se unen Meteorología y Astronomía, relacionado con el anterior, es la elección de sitios dedicados a la construcción de observatorios. La observación profesional, en la que se invierten millones de euros en equipos muy sofisticados, requiere lugares cuyas condiciones meteorológicas sean especiales. Se suelen buscar emplazamientos situados en zonas elevadas (como cimas montañosas), normalmente por encima de la capa límite planetaria² y en los que la humedad sea muy baja. Esas serían las condiciones óptimas para la observación en el rango óptico del espectro, que son las que encontramos, por ejemplo, en el Roque de los Muchachos

²Llamamos capa límite planetaria (conocida por sus iniciales en inglés: PBL) a la capa de la troposfera en la que el flujo se encuentra fuertemente influido por la interacción con la superficie de la Tierra y donde son importantes los fenómenos de turbulencia. En esta capa, variables como el viento, la temperatura y la humedad experimentan a menudo variaciones rápidas (turbulencia). Cuando las condiciones son de buen tiempo, la capa límite presenta un ciclo diurno: durante el día, el espesor de la capa límite aumenta y se producen movimientos verticales turbulentos muy importantes. Durante la noche el espesor disminuye, los movimientos verticales se limitan a una capa mucho más fina y aparece con frecuencia una inversión térmica (la temperatura en la superficie es más baja que en niveles más elevados). Normalmente, la contaminación y los aerosoles atmosféricos se encuentran confinados dentro de esta capa límite y por encima de ella se encuentran cielos libres de contaminación que resultan ideales para la observación astronómica.

en la isla de la Palma, Mauna Kea en Hawai o en la sede del Observatorio Europeo Austral situado del desierto de Atacama en Chile.

Pero quizás el aspecto que más limita la observación sea la misma existencia de una atmósfera como la terrestre. La luz es la materia prima del trabajo de los astrónomos y los astrofísicos. Salvo algunas excepciones como la astronomía de partículas (detectores de neutrinos, por ejemplo), los rayos cósmicos o los meteoritos que alcanzan la superficie de la tierra, toda la información que los astrónomos pueden obtener acerca de los objetos que pueblan el universo deben extraerla de la radiación electromagnética que nos llega y es de ahí de donde deben inferir los procesos físicos y químicos que intervienen en la formación y evolución de todos los objetos del cosmos y tratar de pronosticar cuál puede ser su futuro.

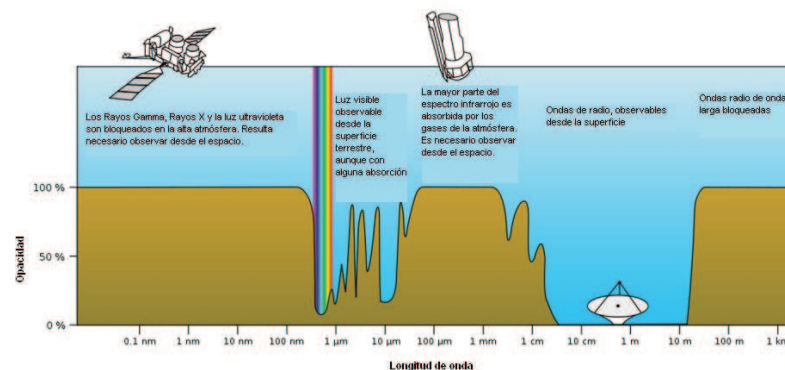


Figura 8: Transmisividad de la atmósfera (opacidad) para las diferentes longitudes de onda.

Esta luz (y hablo de luz en sentido extenso: cualquier radiación del espectro electromagnético) atraviesa el medio interestelar prácticamente sin verse modificada hasta llegar al tope de la atmósfera. Sin embargo, al atravesar los últimos kilómetros de su viaje hasta nosotros, se va a ver drásticamente alterada. Va a sufrir procesos de absorción, atenuación y dispersión por la existencia de la atmósfera. La presencia de determinadas moléculas en el aire que nos rodea hace que la observación astronómica desde la superficie terrestre sea posible sólo en un rango de frecuencias muy pequeño del espectro ya que el resto será absorbido en su totalidad o parcialmente por las diferentes moléculas que componen la atmósfera. Sólo algunas frecuencias especiales en la zona radio (ventana radio) y en una zona con longitudes de onda entre los 400 y los 700 nanómetros aproximadamente (ventana óptica) llegarán prácticamente de forma íntegra hasta la superficie del planeta³.

³También algunas frecuencias en el infrarrojo logran alcanzar la superficie.

Dicho de otra forma: aunque la extensión del espectro electromagnético es muy grande, la presencia de la atmósfera en nuestro planeta y su composición va a restringir de forma muy importante la cantidad de información que podemos recibir sobre lo que sucede en el Universo.

Durante miles de años, las observaciones astronómicas se habían realizado en la llamada ventana óptica del espectro: en longitudes de onda entre aproximadamente los 400 y los 700 nanómetros. No fue hasta 1931 cuando Karl Jansky demostró que la radiación a una longitud de onda de 14,6 metros, captada con una antena direccional que apuntaba a Sagitario debía proceder de una fuente de origen extraterrestre. En concreto se trataba de emisiones que llegaban del centro galáctico. Con este descubrimiento y el desarrollo posterior de la radioastronomía se amplió el rango de frecuencias que podían ser objeto de estudio.

Para explorar otras zonas del espectro, se hicieron algunos experimentos con instrumentos a bordo de globos que ascendían en la atmósfera pero fue con la llegada de la era espacial y el lanzamiento de satélites artificiales a bordo de los cuales se instalaban instrumentos de observación, cuando definitivamente se extendieron las observaciones a las zonas del infrarrojo, ultravioleta y, más tarde, incluso a rayos X y rayos gamma y ha añadido gran cantidad de información que va desde la estructura interna de las estrellas al origen del universo.

Vemos por lo tanto que, además de las influencias obvias de la meteorología en la capacidad de observación, la composición atmosférica condiciona de forma extraordinaria la cantidad de información que somos capaces de analizar desde la superficie terrestre. Y ha sido una limitación que sólo en el siglo pasado la Humanidad ha sido capaz de superar.

Por otra parte, las limitaciones meteorológicas no sólo se centran a la ventana óptica. En la propagación de las ondas de radio, el vapor de agua presente en la troposfera (0 – 10 km) juega un papel muy importante. Para complicar más todavía, el vapor de agua está muy mal mezclado en la atmósfera y tampoco puede ser medido con precisión. Existen técnicas sofisticadas de observación en radio como la interferometría de muy larga base (VLBI) en la que, para obtener buenos resultados, es necesario tener en cuenta parámetros como el vapor de agua precipitable de la columna atmosférica. Variaciones en la distribución del contenido de vapor de agua pueden originar fluctuaciones de fase que ocasionen una degradación de las medidas.

También la observación en el rango de la radioastronomía submilimétrica (300 μm - 1 mm, 300 GHz - 1 THz) se ve muy afectada por la humedad. En la figura 9 podemos ver como la presencia de humedad en la atmósfera afecta de forma dramática a la transmisión de la radiación en este rango de frecuencias. Esto hace que la radioastronomía submilimétrica solo sea posible en zonas en las que la humedad es extremadamente baja como la cima de las montañas.

Hoy en día, los modelos numéricos meteorológicos utilizan técnicas variacionales punteras como 4DVAR en los que se integran cientos de miles de observaciones tanto in-situ como procedentes de satélite para producir análisis y reanálisis que describen el estado de la atmósfera de la forma más precisa posible en un instante dado. Estos análisis y reanálisis hoy en día tienen una resolución espacial y vertical muy buena (que además está constantemente mejorando conforme aumenta la capacidad de cálculo) y podrían ser utilizados por los astrónomos profesionales como parámetros de entrada en el análisis de los datos obtenidos en sesiones de observación radio. Además los reanálisis pueden ayudar a determinar el emplazamiento óptimo de unas instalaciones científicas astronómicas dependiendo de las características que se busquen. Estos dos ejemplos serían un claro ejemplo de una posible colaboración futura entre las comunidades astronómicas y meteorológicas.

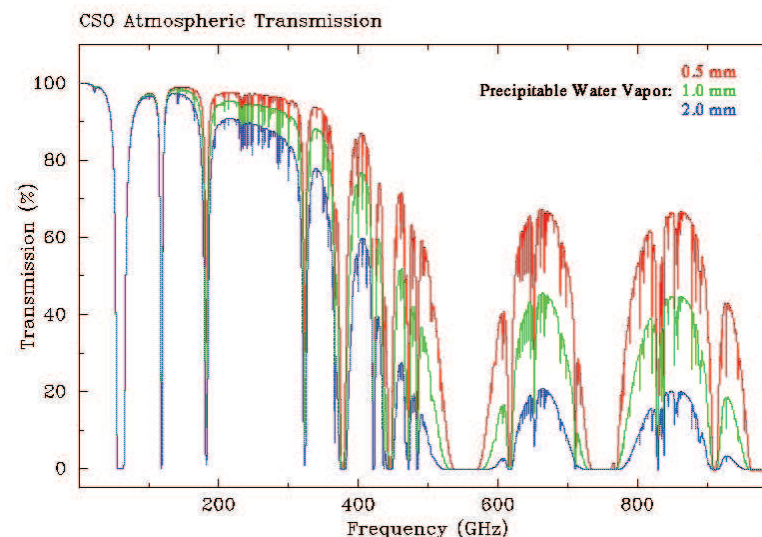


Figura 9: Transmisividad sobre Mauna Kea (Hawái) a determinadas frecuencias en función de la presencia de vapor de agua en la atmósfera. Fuente: Caltech Submillimeter Observatory.

Meteorología y Astronomía oficiales en España

Centrémonos ahora en la historia de las instituciones científicas españolas relacionadas con las dos ciencias que nos ocupan. La Astronomía también jugó un papel importante en el origen de la Meteorología oficial

en España. A la vuelta de su expedición para medir el grado de meridiano (1735-1744), Jorge Juan propuso al rey Carlos III la creación de un observatorio anexo a la Academia de Guardamarinas en la ciudad de Cádiz que ayudara a la formación de los oficiales a la vez que se atendiera las observaciones y el cálculo de efemérides astronómicas necesarias para la navegación. El Real Observatorio de la Armada de San Fernando (Cádiz) fue fundado en 1753 y, posteriormente, fue trasladado a su emplazamiento actual en San Fernando. Allí se empezaron a tomar medidas meteorológicas. Dichas observaciones meteorológicas se llevaban a cabo con el fin de corregir las medidas astronómicas, pero pronto se observó la utilidad que podían tener las variaciones de la presión atmosférica para predecir el estado del tiempo o la importancia de las variaciones de temperatura en las correcciones de las lecturas de los cronógrafos a bordo de los buques. Esto llevó al Observatorio a establecer un programa sistemático de observaciones meteorológicas en 1789, de modo que el observatorio dispone de una de las series climatológicas más extensas y continuas de España. A partir de 1870 se dio un nuevo impulso a las observaciones, bajo la dirección de D. Cecilio Pujazón. Se adquirieron nuevos aparatos de medidas y se sistematizó la publicación de las mismas dentro de los Anales que, a partir de entonces, se publicarían periódicamente. También se comenzó a organizar un Servicio Meteorológico Costero bajo Real Orden de 1876. Se proyecta una red de estaciones en puertos y observaciones desde buques, pero no se pudo llevar a cabo hasta 1884. Las actividades meteorológicas cesaron en 1888 tras la creación del Instituto Central Meteorológico y el inicio de sus trabajos.

Por otra parte, el Real Observatorio Astronómico de Madrid también realizó observaciones meteorológicas. El Observatorio fue fundado a finales del XVIII, al igual que el Real Observatorio de la Armada, a propuesta de Jorge Juan al rey Carlos III. En 1790 se iniciaron las obras del edificio que sería su sede en el Retiro, diseñado por Juan de Villanueva. Las primeras observaciones meteorológicas datan de esta primera época y eran responsabilidad de los profesores encargados de la Cátedra de Meteorología de la Escuela de Astronomía. El primero de estos profesores fue José Garriga que en 1794 escribió un Curso de Meteorología. Como el edificio de Villanueva se retrasaba, el director de la Escuela de Astronomía, Salvador Jiménez Coronado, decidió instalar un observatorio provisional en el Altillio de San Pablo, y allí se realizan también las medidas meteorológicas hasta que fueron interrumpidas en 1808 debido a la invasión francesa y la guerra de la Independencia. De estas medidas sólo se conserva el cuaderno original de 1803, el resto resultó destruido, pero por éste sabemos que se realizaban tres observaciones diarias a las 8 de la mañana, a las 2 del mediodía y las 10 de la noche y se medían la presión, la temperatura y otras variables como nubosidad y meteoros como las tormentas, nieblas, nieve o la lluvia pero no se anotaba la precipitación.

Jiménez Coronado residió en el extranjero entre 1785 y 1789, año en que regresó debido al estallido de la Revolución Francesa. Él había estado la mayor parte del tiempo en París, pero visitó otros observatorios europeos y allí pudo comprobar la importancia que en los ambientes científicos europeos se otorgaba a la observación meteorológica y la necesidad de disponer de observaciones simultáneas, con métodos uniformes y con instrumentos normalizados.



Figura 10: II Congreso Meteorológico Internacional (1879) al que asistieron los directores del Observatorio Astronómico y Meteorológico de Madrid y del Real Observatorio de San Fernando. Sentado, el segundo por la izquierda es Antonio Aguilar. De pie, detrás de Aguilar se encuentra Cecilio Pujazón.

Sin embargo, como ya se ha mencionado, la invasión francesa truncó de forma abrupta las actividades del observatorio. Tras años de abandono en 1835 se reanudaron las observaciones meteorológicas y en 1840 pasó a depender de la Dirección General de Estudios. Como únicamente se realizaban medidas meteorológicas, pasó a denominarse “Observatorio Meteorológico de Madrid”. Esta situación se prolongó hasta 1851, cuando la Real Orden de 24 de septiembre restableció el Observatorio Astronómico bajo la dirección de Antonio Aguilar y Vela, pero dispuso que el Observatorio también fuera meteorológico. Para dirigir estas actividades se propuso nombrar a un catedrático de Física de la Universidad de Madrid. Posteriormente, en 1858, se unificaron de nuevo ambas secciones y el segundo astrónomo, Miguel Merino, pasó a encargarse de las observaciones meteorológicas.

Otro hito importante fue el Real Decreto de 5 de marzo de 1860, en el que se organiza la recogida coordinada de datos de los observatorios

existentes y se crean otros nuevos, inicialmente bajo la dirección de la Junta General de Estadística del Reino. La realización física de las observaciones la solían realizar los catedráticos de Física de universidades e institutos o se encargaban a órdenes religiosas y debían transmitirse por telégrafo para su recopilación. En 1865, tras reformarse la Junta de Estadística, las actividades meteorológicas pasaron a depender del Ministerio de Fomento que las puso de nuevo bajo la dirección del Real Observatorio Astronómico y Meteorológico de Madrid. La recopilación de datos tenía únicamente un fin estadístico y climatológico, pero en ningún caso se realizaban predicciones. Esta época se prolongó hasta el 11 de agosto de 1887, fecha en la que se publicó el Real Decreto que creaba el Instituto Central Meteorológico con la misión de “calcular y anunciar el tiempo probable a los puertos y capitales de provincia” que ya estaba de acuerdo con la misión de los servicios meteorológicos europeos contemporáneos.

La idea de crear en España un servicio meteorológico similar a los existentes en otros países europeos la tenían hacía ya tiempo científicos como Antonio Aguilar y de Miguel Merino, su sucesor al frente del Observatorio Astronómico a partir de 1882. En 1879 el director del Observatorio Astronómico y Meteorológico de Madrid, Antonio Aguilar y el del Real Observatorio de San Fernando, Cecilio Pujazón, asistieron juntos al II Congreso Meteorológico Internacional. Probablemente, el interés por la creación del servicio meteorológico independiente, se hizo más fuerte a raíz de esta reunión, pero debido a reticencias de las autoridades no pudo lograrlo por lo que Aguilar se quejaría amargamente y murió en 1882 sin ver cumplido su deseo. A mediados de la década de los 80 solamente tres países europeos carecían de un servicio meteorológico dedicado a la predicción del tiempo: Grecia, Turquía y España.

Quien sí intervino en la creación del Servicio Meteorológico en España fue Francisco Giner de los Ríos. Giner de los Ríos no tenía relación alguna con la Meteorología, ni con la Astronomía, pero estaba interesado en cualquier actividad que supusiera el progreso del país. Convenció al Gobierno (del partido liberal) de la necesidad de disponer de un Servicio Meteorológico para informar y predecir el tiempo. Finalmente, como ya hemos dicho, en 1887 un Real Decreto creaba el Instituto Central Meteorológico. A continuación, se creó una comisión para definir las tareas y entre sus miembros estaba el director del Observatorio Astronómico de Madrid, Miguel Merino. A principios de 1888 la comisión emitió su informe y se preparó el programa para la oposición para el puesto de Director.

A dicha oposición se presentó Augusto Arcimis Wehrle, doctor en Farmacia, pero con inquietudes científicas muy amplias. Arcimis había viajado por Europa y conocía varios idiomas. En Londres se empezó a interesar tanto por la Meteorología como por la Astronomía. A partir de 1874 tuvo una intensa actividad astronómica. En su casa de Cádiz instaló

un completo observatorio meteorológico y astronómico. En 1878 publicó su obra “El Telescopio Moderno” en la que introducía técnicas de análisis espectral aplicadas a la Astronomía, por lo que puede ser considerado como un pionero de la astrofísica española, en un momento en que la astronomía oficial se dedicaba en exclusiva a una astronomía descriptiva y de posición. Fue un astrónomo amateur que publicó los resultados de sus observaciones en revistas científicas europeas como la revista de la *Società degli Spettroscopisti Italiani* o el *Monthly Notices of the Royal Astronomical Society* de la que fue nombrado *Fellow*. Luchó por sacar a la ciencia española, y específicamente la Astronomía, de su atraso tomando como ejemplo lo que sucedía en Europa. Se podría decir que fue un visionario que se percató de la necesidad de no perder el tren del progreso justo en el momento en que la industrialización más demandaba una modernización del país. Sus investigaciones y publicaciones le valieron un cierto reconocimiento internacional. Igualmente su defensa de las técnicas espectroscópicas para estudiar la composición química de las estrellas y su aplicación para el estudio de las estrellas fugaces, siendo esta una técnica novedosa que no había sido aplicada por la astronomía oficial, hizo que los astrónomos oficiales sintieran por él una cierta animadversión.

En 1884 se trasladó a Madrid para ser profesor en la Institución Libre de Enseñanza. Y fue Giner de los Ríos quien le convenció para opositar al cargo de Director del recién creado Instituto Central Meteorológico.

Arcimis ganó la oposición y fue nombrado el 19 de marzo de 1888. El primer trabajo de Arcimis consistió en encontrar un lugar adecuado para ser utilizado como sede del Instituto. El emplazamiento elegido fue el edificio conocido como “El Castillo” del Parque del Retiro y que fue cedido parcialmente por el Ayuntamiento de Madrid. Ese mismo año, Arcimis viaja a Inglaterra y Francia para adquirir diversa instrumentación de medida.

Tras diversos avatares, el gobierno conservador decretó su supresión que fue ordenada en abril de 1891 por Real Decreto que derogaba el de su creación. Las instituciones astronómicas (Real Observatorio de Madrid y Real Observatorio de San Fernando) apoyaron su supresión porque, si bien en principio apoyaron la creación del mismo, vieron como les restaba más competencias de las previstas. La supresión se prolongó durante algo más de un año. Tras su restablecimiento en julio de 1892, el 1 de marzo de 1893 el Instituto publicó su primer boletín. Su papel de pionero no se limitó a la Astronomía. En 1895 escribió un pequeño artículo llamado “La circulación atmosférica” en el que se hacía eco de las últimas corrientes en esta materia como por ejemplo los trabajos de Helmholtz en los que postulaba que la circulación debía tener en cuenta la turbulencia horizontal y vertical que condicionaban los cálculos de los flujos de calor y que se presentaban con gran detalle. De esta forma, se convirtió en el introductor en España de la Meteorología dinámica.

Durante algunos años las dificultades continuaron. A la falta de trabajadores (Arcimis, un ayudante y un ordenanza fueron el único personal del Instituto durante ese tiempo) se unió que la observación en provincias continuaron siendo responsabilidad de otros organismos (con instrumentos y métodos de medidas variados) y coordinados por el Observatorio Astronómico que se encargaba también de realizar estadísticas y publicar los resúmenes anuales. Por otra parte, continuó realizando observaciones.



Figura 11: Augusto Arcimis Wherle. Astrónomo amateur y primer director del Instituto Central Meteorológico.

En 1900, tras la supresión del Ministerio de Fomento, tanto el Observatorio Astronómico de Madrid como el Instituto Central Meteorológico pasaron a depender del Ministerio de Instrucción Pública. En 1904, ambos organismos confluyeron dentro del Instituto Geográfico para formar el Servicio Astronómico y Meteorológico. Paradójicamente esta integración fue beneficiosa para el Instituto Central Meteorológico ya que el reglamento despejó dudas y éste asumió todas las actividades relacionadas con la Meteorología desde 1906 que antes tenía asignadas el Observatorio

Astronómico, aunque persistieron los problemas con los observatorios ya que continuaron sin tener competencias sobre los observatorios y sobre los métodos de observación que seguían siendo heterogéneos y separó definitivamente los caminos de la astronomía y la meteorología oficial. Arcimis falleció el 18 de abril de 1910.

A su muerte, y dependiendo el Instituto Central Meteorológico del Instituto Geográfico, se nombró director a un Ingeniero Geógrafo: José Galbis Rodríguez. El progreso que experimentaron la aviación, por un lado, y las ciencias atmosféricas, por otro, proporcionaron un nuevo impulso al Observatorio Central Meteorológico (cambió de nombre en 1911). Se crearon redes pluviométricas y de aerología (con globos que se lanzaban desde el Observatorio Fabra de Barcelona, el de San Fernando, la Sociedad Oceanográfica de Coruña y el Observatorio del Ebro). Entre los años 1910 y 1915 se produjo un desarrollo muy importante del Observatorio Central Meteorológico, generando información y predicciones hasta 24 horas para diferentes zonas que se difundían vía radiotelegráfica. Asumió igualmente las tareas de observación e información aeronáutica de los aeródromos que iban surgiendo y pronto se hizo evidente que resultaba necesario disponer de personal en otras provincias. En 1920 se produjo una reorganización del Instituto, que pasó a denominarse Servicio Meteorológico Español. Se decidió entonces dar la posibilidad de crear centros regionales y observatorios que dependieran directamente del Servicio. Así, antes de 1925 funcionaban 13 observatorios que dependían directamente del SME, entre ellos Alicante, Coruña, Santander y Sevilla. En 1927 se creó el primer centro meteorológico regional en Zaragoza. A pesar de pertenecer al Instituto Geográfico, el Servicio Meteorológico ya había adquirido una cierta independencia funcional que se confirmó con la llegada de la República cuando en 1932 y mediante un Decreto se estableció definitiva y formalmente dicha independencia y puso fin a un largo periodo en los que la Meteorología y la Astronomía españolas habían caminado, administrativamente, de la mano.

Una estrecha relación

Hemos repasado en este trabajo las relaciones existentes entre dos ciencias que, en principio, parecen tener poco que ver. Pero son factores astronómicos los que condicionan los grandes rasgos de nuestro clima y es el Sol, la estrella más cercana a nosotros, el responsable de los movimientos atmosféricos y de la circulación general atmosférica. Por otra parte, la búsqueda de los mejores emplazamientos para la observación profesional requiere unas condiciones meteorológicas muy concretas y la composición química de nuestra atmósfera condiciona las zonas del espectro electromagnético en las que resulta posible realizar observaciones desde la superficie de nuestro planeta. También resulta necesario tener en

cuenta la presencia (variable) de vapor de agua en la atmósfera a la hora de hacer observaciones en la ventana radio.

Quizás los árboles no nos dejaban ver el bosque o quizás la especialización a la que está sometida hoy la Ciencia nos ha impedido saber hoy lo que nuestros antepasados han conocido durante miles de años: que los vínculos entre Astronomía y Meteorología son profundos y que estas dos ciencias han caminado de la mano a lo largo de toda la Historia de la Humanidad.

Referencias

A. Anduaga Egaña, *La Regeneración de la Astronomía y la Meteorología españolas: Augusto Arcimis (1844-1910) y el Institucionismo*, Asclepio Vol. LVII-2-2005 (2006).

R. Daley, *Atmospheric Data Analysis*, Cambridge Atmospheric and Space Science Series, Cambridge University Press (1991).

J. M. Giménez de la Cuadra, *La Meteorología en el Observatorio Astronómico de Madrid*, incluido en Doscientos años del Observatorio Astronómico de Madrid (1992).

J. Gribbin, *El clima futuro*, Biblioteca científica Salvat número 75, Editorial Salvat (1994).

Intergovernmental Panel on Climate Change Fourth Assessment Report: Climate Change 2007, Contribution of Working Group I: The Physical Science Basis, Cambridge University Press. http://www.ipcc.ch/publications_and_data/publications_and_data_reports.html.

J. Martín Dávila, J. Gárate Pasquín, A. Pazos García y M. Catalán Morollón, *La Geofísica en el Real Instituto y Observatorio de la Armada en San Fernando*, Física de la Tierra 18, páginas 119-135 (2006).

M. Palomares, *Los noventa primeros años*, incluido en El Instituto Nacional de Meteorología: Un reto tecnológico, Centro de Publicaciones del Ministerio de Medio Ambiente (2004).

A.R. Thompson, J.M. Moran, G.W. Jr Swenson, *Interferometry and Image Synthesis in Radio Astronomy*, Wiley-VCH Verlag (2004).

Caltech submillimeter observatory. http://www.submm.caltech.edu/cso/cso_submm.html.

EL LADO OSCURO DE LA LUZ

Francisco Colomer Sanmartín

*Jefe de Servicio de Efemérides y Hora
Observatorio Astronómico Nacional
Instituto Geográfico Nacional - Ministerio de Fomento*

La contemplación del cielo estrellado es probablemente uno de los mayores placeres que existen. Complejo y grandioso, nos envuelve y maravilla en sus detalles. Nos conmueve su belleza, y encogemos ante su visión, mientras imaginamos otros mundos, otras civilizaciones lejanas. O quizás no. Pues el cielo estrellado ya casi nadie lo ve.

Introducción

En los albores de la humanidad, los primeros seres humanos disfrutaban de la contemplación del cielo nocturno. Los cuerpos celestes formaron parte de sus vidas, y comprobaban posiblemente asombrados la aparición diaria del Sol, y los cambios en el cielo de la Luna, los planetas, cometas, etc. Tratando de comprender la existencia de los astros, y la suya propia, idearon teorías del mundo, dioses, mitos, héroes, ángeles y demonios. En todo el planeta, la visión del cielo estrellado ha intervenido en el establecimiento de las civilizaciones y de sus religiones. Había mucho que explicar, y pocos medios.

Durante siglos, el hombre ha continuado pudiendo contemplar el cielo. Ello ha sido fuente de inspiración cultural, en pintura, poesía, música... Su magnífica complejidad no ha pasado inadvertida a las mentes más brillantes. El desarrollo científico está íntimamente ligado a la constatación de que somos parte del Universo. Basta imaginar cómo sería nuestro mundo y nuestra civilización si viviésemos en un planeta cuya atmósfera no permitiese mirar más allá, quedando nuestra percepción del Universo acotada a nuestra pequeña Tierra.

Para millones de personas, en la actualidad, el mundo no es muy diferente de lo que acabamos de describir. Privados habitualmente de la contemplación del cielo nocturno por la contaminación lumínica que existe en las ciudades que habitamos, hemos perdido la consciencia de ser parte del Universo. Restringimos nuestros pensamientos a lo que es local e inmediato. La sociedad ha perdido una gran herramienta para mantener un estado de curiosidad que es básico para el progreso ordenado y sostenible.



Figura 1: “La noche estrellada”, por Vincent Van Gogh (1889), Museo de Arte Moderno de Nueva York. Muestra en la parte central del lienzo el pueblo de Saint-Rémy bajo un cielo arremolinado. El Museo de Orsay (París) alberga otro famoso cuadro, “Noche estrellada sobre el Ródano”, pintado por Van Gogh en 1888 durante su estancia en Arlés, en el que se aprecia la constelación de la Osa Mayor y otras estrellas, así como la incipiente contaminación lumínica producida por las lámparas del puerto.

¿Qué es la contaminación lumínica?

Se puede definir la contaminación lumínica como “la emisión de flujo luminoso de fuentes artificiales nocturnas en intensidades, direcciones, horarios o rangos espectrales innecesarios para la realización de las actividades previstas en la zona en la que se instalan las luces”¹.

¹Asociación contra la Contaminación Lumínica “Cel Fosc”, <http://www.celfosc.org>.

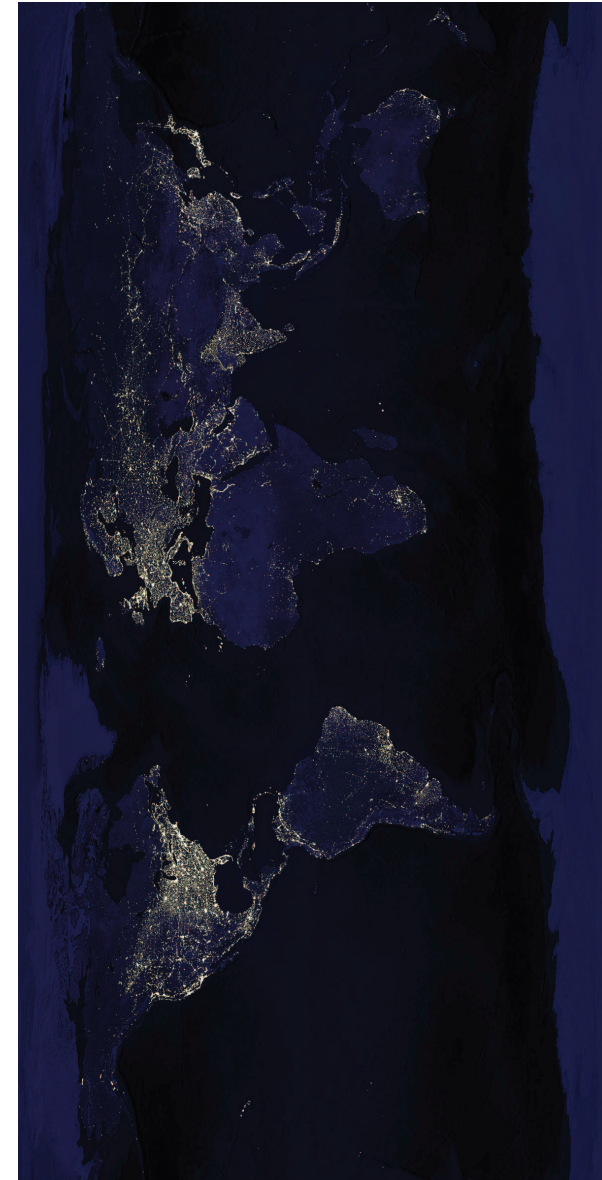


Figura 2: Distribución de la contaminación lumínica en la Tierra vista por los satélites del Defense Meteorological Satellites Program (DMSP) de EE.UU. (Cortesía NASA/GSFC y NOAA/NGDC).

Siendo obvio que la iluminación es imprescindible para la actividad humana, lo cierto es que por ahora, y principalmente en el mundo desarrollado, se ha identificado erróneamente el progreso con una mayor iluminación. Europa, Estados Unidos, o Japón son fácilmente reconocibles en fotos nocturnas del planeta, como lo son grandes ciudades como Sídney, Johannesburgo, Río de Janeiro... África, el oeste de China, los bosques canadienses, Siberia, la amazonia y las grandes llanuras australianas están a oscuras.

¿Pero realmente el progreso necesita de tanta iluminación? Sin duda, el hecho mismo de que veamos tanta luz desde el satélite implica que muchísimas luminarias iluminan el cielo, escapando por tanto su luz de los objetos que se pretende iluminar. Querer hacerlo a propósito sería un proyecto descomunal. Un exceso desmesurado e insostenible. Un derroche irracional y absurdo.

Nuestra preocupación, como profesionales de la Astronomía, es conseguir suficiente oscuridad en el cielo sobre nuestros instrumentos para permitir realizar estudios de gran calidad. Puede parecer una obsesión de unos pocos, un problema menor, resoluble por la instalación de dicha instrumentación en territorios “oscuros”. Dicha solución no siempre es realizable, no obstante, por la necesidad de compaginar la oscuridad del cielo con otros factores (como la altura para evitar la humedad de la atmósfera, o la necesaria cercanía a las infraestructuras de apoyo a la investigación).

La realidad de la iluminación de las ciudades europeas hace imposible compaginar la instalación de grandes infraestructuras de investigación astronómica en casi cualquier lado. En España, sin ir más lejos, la sobreiluminación de Madrid capital se aprecia a simple vista a más de 200 km de distancia, y produce un resplandor detectable por telescopios de mediano tamaño instalados incluso en Santander. Ni que decir tiene que ninguna zona de la península es completamente apta para la instalación de instrumentación astronómica avanzada.



Figura 3: Visión nocturna de Madrid y las poblaciones del Corredor...



Figura 4: Fotos tomadas por Todd Carlson el 14 de agosto de 2003, durante un apagón en la costa este de EE.UU. (izquierda). La visión directa de los astros desaparece en cuanto se repara la avería (derecha).

¿Pero cuál es la ventaja de iluminar inadecuadamente? Se considera que una mayor iluminación aumenta la seguridad. Sin embargo, en realidad aumenta sólo la sensación de seguridad. Los inconvenientes son numerosos: un consumo mayor e innecesario, un impacto negativo en el medioambiente (flora y fauna), y por supuesto, la pérdida de la visión del cielo nocturno para la mayoría de la población occidental.



... del Henares. Fotografía cortesía de Pedro-Luís Cuadrado Revuelta (pcuadrado@recol.es).

El proyecto StarLight y otras iniciativas

El 20 de Abril de 2007 se celebró en la isla de La Palma la “Conferencia Internacional en Defensa de la Calidad del Cielo Nocturno y el Derecho a Observar las Estrellas”². Considerando que el cielo nocturno es un recurso patrimonio de toda la humanidad, y que su observación ha representado una inspiración y elemento esencial en el desarrollo cultural y científico de todas las civilizaciones, se insta a las autoridades a promover acciones de protección del medio ambiente nocturno a nivel local, nacional e internacional, y a informar y sensibilizar a todos los agentes implicados y a la población general sobre la necesidad de dicha protección.

Desde entonces, cada 20 de abril se celebra la “Noche Mundial en Defensa de la Luz de las Estrellas”.

Como ésta, otras iniciativas se van proponiendo para aumentar la concienciación social sobre la realidad de la contaminación lumínica y la importancia de preservar la oscuridad del cielo. Durante 2009, Año Internacional de la Astronomía, se ha desarrollado en España un proyecto pilar denominado “Descubre el cielo oscuro”³, para tratar aspectos como el desarrollo de nuevas técnicas de iluminación y actividades diversas (como fiestas de estrellas, cuentacuentos, etc.). Los aficionados y la población general pueden también participar activamente y contribuir a diseminar la concienciación sobre la problemática de la contaminación lumínica realizando sencillas observaciones del cielo nocturno, en campañas como “Globe at Night”⁴ o el proyecto IACO⁵ de la Asociación Malagueña de Astronomía.

La contaminación lumínica también ha sido estudiada en congresos monográficos, como el celebrado en Eslovenia en 2007⁶, o el grupo de trabajo formado durante el Noveno Congreso Nacional del Medio Ambiente (CONAMA) celebrado en 2008⁷ en el que se discuten los aspectos científicos y legislativos relevantes a su prevención.

El mayor avance en este sentido ha sido la promulgación de la “Ley de Protección de la Calidad Astronómica de los Observatorios del Instituto de Astrofísica de Canarias” (Ley del Cielo) en 1988. Su principal impacto está en el control de la contaminación lumínica en el Observatorio del Roque de los Muchachos (isla de La Palma), la contaminación radioeléctrica,

atmosférica, y ecológica (control de rutas aéreas). La adaptación de luminarias con carácter retroactivo ha permitido disminuir al menos un 50 % los efectos de la contaminación lumínica en los alrededores del observatorio. En el proceso, los fabricantes certificaron más de un centenar de tipos de luminarias. En la actualidad, cualquier instalación precisa de un informe del IAC. La difusión de esta iniciativa ha permitido que otras comunidades autónomas regulen asimismo en el sentido de disminuir la contaminación lumínica; Cataluña fue pionera (2001), y posteriormente Baleares, Navarra, Cantabria y Andalucía. Algunos ayuntamientos han desarrollado ordenanzas municipales al respecto (Burgos, Santander, Sevilla, Córdoba, Tárrega, Castro-Urdiales, o sobre el entorno de la albufera de Valencia). Algunas de las normas publicadas no contemplan, desafortunadamente, la adaptación de instalaciones preexistentes.



Figura 5: Cielos anaranjados sobre Valencia, vistos desde L'Oceanografic. Fotografía de Francisco Colomer Sanmartín.

Este escenario contrasta, sobre todo, con la realidad de algunas grandes capitales como Valencia. Mientras el Plan de Eficiencia Energética del Ministerio de Fomento 2004-2012 propone un consumo medio de 75 kilovatios-hora por habitante y año⁸, en Valencia se consumen anualmente más de 127 kWh por habitante (a comparar con los 61,5 de Madrid

²Proyecto StarLight, <http://www.starlight2007.net>.

³http://www.astronomia2009.es/Proyectos_pilares/Descubre_el_Cielo_Oscuro.html, <http://www.darkskiesawareness.org>.

⁴<http://www.globe.gov/GaN/>, http://www.globe.gov/fsl/pdf/GaN/GaN2009ActivityPacket_Family_ES_web.pdf.

⁵“Investigación y Acción sobre el Cielo Oscuro”, <http://www.iaco.es/>.

⁶7th European Symposium for the protection of the night sky (Eslovenia, 5-6 Octubre 2007), <http://www.darksky2007.si/index.html>.

⁷<http://www.conama9.org/conama9/>, http://www.conama9.org/conama9/download/files/GTs/GT_LUZ/LUZ_final.pdf.

⁸http://www.mma.es/secciones/cambio_climatico/documentacion_cc/normativa_cc/pdf/plan_accion_2008_2012.pdf.

o los 57,4 de Barcelona). Ello tiene un impacto evidente también en los bolsillos de sus ciudadanos, que en 2007 gastaron 13 millones de euros en electricidad. Conviene recordar, no obstante, que si bien toda la contaminación lumínica implica gasto innecesario, el ahorro energético no es siempre síntoma de estar consiguiendo cielos más oscuros.

Impacto de la contaminación lumínica en Astronomía

La astronomía, profesional y aficionada, es la primera víctima de la contaminación lumínica. Los cuerpos celestes, incluso los más brillantes en origen, son detectados como relativamente débiles (a excepción del Sol) y su luz debe competir con la emitida por las actividades humanas, relativamente más intensas por estar mucho más cercanas. El cielo nocturno brilla principalmente por la luz de los astros, la luz zodiacal (polvo en el sistema solar) y el resplandor celeste (causado por átomos y moléculas en las capas altas de la atmósfera terrestre, que descargan radiación absorbida previamente). Sólo este último ya tiene un impacto sobre las observaciones; por ejemplo, una fuente de magnitud $V = 28$ se detecta con un telescopio de 8 metros (como un VLT) en 40 horas de observación, mientras que el HST (2,4 metros, en el espacio y por ello libre del impacto de la atmósfera) sólo precisa 4 horas. La contaminación lumínica, incluso si es pequeña, contribuye también a romper el delicado equilibrio que permita la observación de los objetos celestes desde la superficie de la Tierra.

Existen pocos estudios científicos que permitan estimar con precisión la cantidad de contaminación lumínica producida en cada lugar del planeta. Los trabajos de Cinzano, Elvidge⁹, o Falchi¹⁰ utilizan medidas de satélite para determinar el brillo del cielo nocturno, y desarrollan técnicas para prever la distribución y evolución de dicho brillo a cualquier longitud de onda y para diferentes condiciones atmosféricas. Según estos autores, dos tercios de la población mundial viven en zonas contaminadas (el 99 % en EE.UU.). Nuestra galaxia, la Vía Láctea, ya no es visible sobre los hogares del 50 % de los ciudadanos de la Unión Europea (el 67 % en EE.UU.), mientras que una sexta parte de los europeos (y el 40 % de los norteamericanos) no ven el cielo nocturno en absoluto. Este tipo de estudios, permitiendo cartografiar áreas muy extensas, deben realizar un modelado cuidadoso de la atmósfera para distinguir entre la luz que sale de ella (y es detectada por los satélites) y la que queda atrapada (y genera contaminación lumínica). Las medidas fotométricas realizadas desde tierra son el complemento ideal para dar exactitud a los estudios globales.

⁹P. Cinzano y C. D. Elvidge, "Night sky brightness at sites from satellite data", <http://sait.oat.ts.astro.it/MSAI740203/PDF/2003MmSAI..74..456C.pdf>, *Memorie della Società Astronomica Italiana*, V. 74, p. 456 (2003).

¹⁰P. Cinzano, F. Falchi, y C. D. Elvidge, "The first World Atlas of the artificial night sky brightness", *Mon. Not. R. Astron. Soc.* 328, 689-707 (2001).

El brillo del cielo nocturno puede medirse de forma cualitativa y sencilla, por ejemplo mediante la estimación visual de la magnitud límite. Es el caso del proyecto IACO, o del "Globe at night", en que los participantes comunican si llevan a ver o no estrellas de fácil identificación (como los asterismos de ciertas constelaciones) o cuentan el número de estrellas visibles en las áreas que dichos asterismos delimitan. Este proceso es útil cuando se recogen medidas de muchas personas distintas.

Para obtener medidas más precisas se requiere instrumentación adecuada. Se podría definir una magnitud asociada al brillo del cielo, con la dificultad de que es una fuente extensa mientras que las estrellas se ven puntuales: se considera por ello el brillo por unidad de arcosegundo cuadrado ($\text{mag}/\text{arcsec}^2$). Como la sensibilidad del ojo humano a diferentes longitudes de onda se parece bastante a uno de los filtros comúnmente usados en astronomía óptica (el filtro V), se define¹¹:

$$b = 10^{(12,60-m)/2,5}$$

$$m = 12,60 - 2,5 \log_{10} b$$

Así, un brillo de m magnitudes medido en el filtro V equivale de manera muy aproximada a b candelas por metro cuadrado¹².

Una fuente luminosa astronómica con un brillo aparente total de m magnitudes en la banda V, si se sitúa en el cenit, ilumina el terreno con una iluminancia horizontal (o densidad luminosa) i , medida en luxes, dada por: $i = 10^{(-13,97-m)/2,5}$

Por ello, la Luna llena en posición cenital ($m = -12$) y en una noche muy clara puede llegar a inducir una iluminancia en torno a 0,16 lux. Comparativamente, la densidad luminosa de las estrellas es de 0,0003 lux, y de un día nublado, 10 000 lux.

Para controlar el impacto de la contaminación lumínica sobre los estudios astronómicos, en primer lugar deben utilizarse luminarias que eviten por completo la emisión de luz sobre el horizonte. El límite adoptado en las mejores normativas de estados y regiones europeas es de 0 ($\pm 0,5$) cd/klm (candela por kilolúmen) sobre el plano horizontal. En España, sin embargo, las condiciones de brillo de las luminarias están recogidas en el "Reglamento de eficiencia energética en instalaciones de alumbrado exterior"¹³, en el que se acepta un valor diferente de cero del Flujo Hemisférico Superior (FHS) de las luminarias (en áreas o parajes que se suponen oscuros, zonas E1, como los observatorios astronómicos internacionales o los parques naturales, se acepta un FHS de hasta un 1 %).

¹¹Noveno Congreso Nacional del Medio Ambiente (CONAMA)

¹²Equivalencias: 1 lúmen (flujo luminoso) = 1 cd x sr (cd = candela = intensidad luminosa); 1 lux (densidad luminosa) = 1 lúmen/m².

¹³R.D. 1890/2008, de 14 de noviembre.

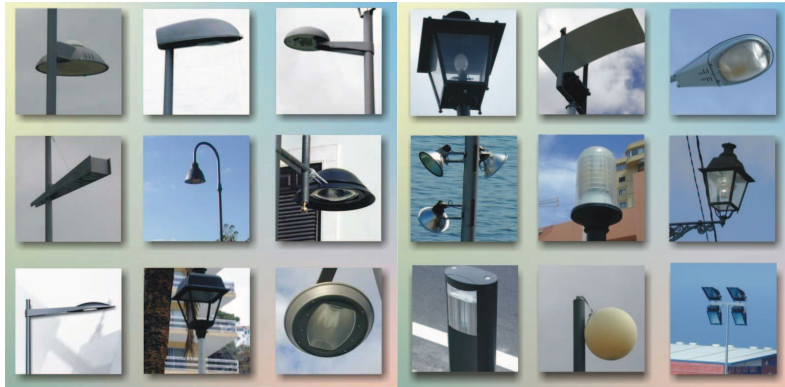


Figura 6: Ejemplos de luminarias adecuadas (izquierda) e inadecuadas (derecha) para combatir el problema de la contaminación lumínica. Cortesía de la Oficina Técnica para la Protección de la Calidad del Cielo (OTPC-IAC), <http://www.iac.es/otpc/OTPC-IAC>.

Hay que tener en cuenta que, contrariamente a lo que podría parecer, la luz emitida en ángulos pequeños y cercanos a la horizontal produce mucha contaminación lumínica. Dicha luz perniciosa es precisamente producida por las luminarias, y no por las superficies iluminadas. Debido al pequeño espesor de la atmósfera, la emisión de luz cerca de la horizontal se propaga más lejos y produce un efecto mucho mayor que la emisión a ángulos elevados, que se escapa al espacio. Es por ello que un valor pequeño del FHS no permite, en sí mismo, terminar con la contaminación lumínica.

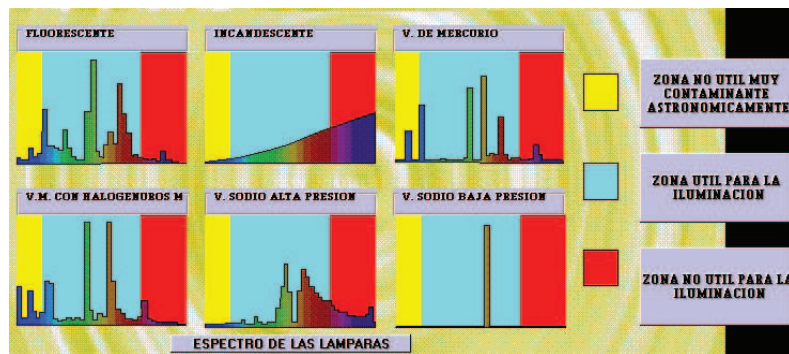


Figura 7: Espectro de varios tipos de lámparas. Cortesía: Oficina (OTPC-IAC).

Respecto a las lámparas que montan las luminarias, las menos perjudiciales son las de vapor de sodio de baja presión, que emiten toda su luz en una estrecha línea del espectro (a la longitud de onda de 589 nm), dejando limpio el resto del mismo. Estas lámparas son las más eficaces pues consumen, para un mismo paquete luminoso, 5 veces menos que las incandescentes, 2,2 menos que las de mercurio y 1,5 veces menos que las de vapor de sodio de alta presión y fluorescentes¹⁴. La concentración de la emisión en longitudes de onda determinadas permite construir dispositivos que filtren dicha emisión, recuperando la oscuridad del cielo.

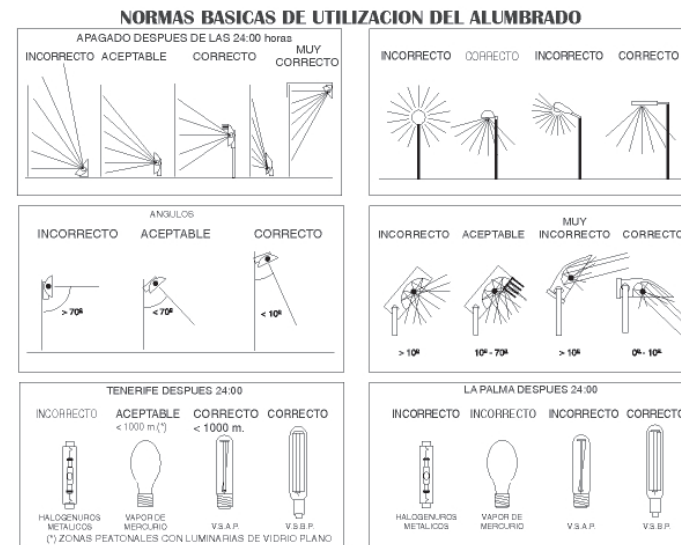


Figura 8: Orientación de las luminarias y tipos de lámparas, y su correspondiente producción de contaminación lumínica. Cortesía: Oficina (OTPC-IAC).

Impacto de la contaminación lumínica en otras áreas

Los seres vivos de nuestro entorno, y nosotros mismos, adaptamos nuestro ciclo vital a la disponibilidad o no de luz solar. La humanidad en su conjunto aprovecha la noche para descansar. La oscuridad, no obstante, tiene ventajas para numerosas especies que se sirven de ella para escapar de sus depredadores. Han evolucionado gracias a ella.

¹⁴Oficina Técnica para la Protección de la Calidad del Cielo (OTPC-IAC)

Recientemente, la luz artificial ha permitido ampliar la duración de los periodos de tiempo en que nuestras sociedades desarrollan su actividad. Ello también ha cambiado las condiciones físicas de los entornos en que viven muchas especies, demasiado rápido para que éstas puedan adaptarse. En muchos casos, este efecto no se ha contemplado como un problema sino como una oportunidad para realizar control de plagas, etc. Una grave simplificación que no es sino la punta del iceberg.

Algunas iniciativas dirigidas a reducir la contaminación lumínica para alcanzar beneficios medioambientales han estado ligadas a algún tipo de aprovechamiento como el ahorro energético o el turismo ecológico, no siendo capaces aún de comprender la necesidad de controlar la agresión que significa el uso de iluminación inadecuada, en cualquier entorno. Dos mensajes (el ahorro y la protección medioambiental) que deben llegar con igual fuerza al ciudadano.

La contaminación lumínica es un grave problema para los ecosistemas. La Unión Europea quizás comienza a darse cuenta, por lo que ha financiado proyectos bajo el paraguas del programa LIFE, como el proyecto Ecolight para proteger el entorno de la albufera de Valencia, capital de la excesiva e inadecuada iluminación en España, y probablemente en Europa. Con Ecolight se investigaron procedimientos de medida del impacto de la contaminación lumínica en flora y fauna (bioindicadores), y una metodología exportable a otros entornos a proteger.

La exposición a la luz durante la noche también tiene consecuencias en la salud de las personas, alterando la función del reloj biológico por su impacto en la producción nocturna de melatonina. Niveles menores a los habituales pueden provocar la aparición de tumores. Es fácil suponer que el aumento incesante de la iluminación nocturna tendrá también impacto, a largo plazo, en los demás seres vivos.

Es de señalar que la inhibición de producción de melatonina, por parte de la glándula pineal (conectada con los ojos), se produce por exposición incluso a periodos cortos de luz, fundamentalmente en la longitud de onda del color azul. Por ello debe evitarse salir de la oscuridad durante las últimas fases de la noche, cuando los niveles de melatonina no pueden ya restablecerse. La iluminación exterior que entra intrusa a las viviendas es especialmente dañina cuando incluye dichas longitudes de onda. Siendo también un antioxidante, la melatonina ayuda a envejecer mejor. Todo un cúmulo de ventajas a las que accedemos cuando no estamos expuestos a la contaminación lumínica.

Contaminación radioeléctrica

La emisión de ondas electromagnéticas a longitudes de onda distintas del visible también puede significar contaminación. En particular, las

interferencias de ondas radio son conocidas y sufridas por la población, significando una peor calidad de las telecomunicaciones, y por ello están reconocidas y reguladas.

La Ley General de Telecomunicaciones¹⁵ establece un marco de coordinación del uso del dominio radioeléctrico, así como la protección de los diversos servicios que precisan utilizarlo. En lo que compete a los observatorios astronómicos, la ley sigue lo dispuesto en la Recomendación ITU-R RA-769-2 de la Unión Internacional de Telecomunicaciones (UIT), que proporciona los “Criterios de protección para las mediciones radioastronómicas”. Esta recomendación fija los umbrales de radiación interferente en los observatorios a partir de los cuales se produce una degradación de la calidad de las observaciones radioastronómicas. Por ello, la Disposición adicional primera de la LGT establece las limitaciones aplicables a las construcciones, tendidos eléctricos, líneas férreas, y transmisores radioeléctricos (distancia y potencia) situados en las cercanías de los mismos. El espíritu de la Ley queda patente en la expresión “no podrán establecerse, por vía reglamentaria, limitaciones a la propiedad ni servidumbres que contengan condiciones más gravosas” que las indicadas.

Límites máximos de la emisión radioeléctrica interferente en las bandas de frecuencia protegidas para investigación radioastronómica		
Banda de frecuencias	Densidad de flujo de potencia dB (W/m ²)	Intensidad de campo eléctrico equivalente dB (V/m)
1400 - 1427 MHz	-180	-34,2
1610,6 - 1613,8 MHz	-181	-35,2
1660 - 1670 MHz	-181	-35,2
2690 - 2700 MHz	-177	-31,2
4990 - 5000 MHz	-171	-25,2
10,6 - 10,7 GHz	-160	-14,2
15,35 - 15,4 GHz	-156	-10,2
22,21 - 22,5 GHz	-148	-2,2
23,6 - 24 GHz	-147	-1,2
31,3 - 31,8 GHz	-141	4,8
42,5 - 43,5 GHz	-137	8,8
86 - 92 GHz	-125	20,8
Para todas las demás frecuencias de la banda, se establece una limitación de la intensidad del campo eléctrico de +88.8dB (mV/m), medida en la ubicación de la Estación de Radioastronomía.		

Como consecuencia, la protección radioeléctrica que disfruta el Centro Astronómico de Yebes (centro nacional de Radioastronomía, en Guadalajara), reflejada en la ORDEN CTE/1444/2003, de 22 de mayo (BOE del 4

¹⁵LGT, Ley 32/2003, de 3 de noviembre; BOE del 4 de noviembre de 2003.

de Junio de 2003)¹⁶ se limita principalmente a aquellas bandas de frecuencia en las que la Radioastronomía está asignada como servicio primario (definidas en el Cuadro Nacional de Atribución de Frecuencias, CNAF; actualizado en la orden ITC/3391/2007 de 15 de noviembre)¹⁷. En dichas bandas, la emisión de ondas radio no puede superar las intensidades reflejadas en la Tabla 1, con independencia de la ubicación del transmisor. A nivel práctico, no obstante, se establece una zona de coordinación de 20 km de diámetro alrededor del observatorio, en la que cualquier instalación emisora deberá demostrar que no interferirá. La Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información (SETSI) del Ministerio de Industria, Turismo y Comercio, debe velar por su cumplimiento.



Figura 9: Radiotelescopio de 40 metros de diámetro. Centro de Desarrollos Tecnológicos y Observatorio Astronómico Nacional. Instituto Geográfico Nacional, Yebes (Guadalajara).

La protección de los observatorios de radioastronomía conseguida a través de la legislación mencionada es claramente insuficiente. No sólo por-

¹⁶por la que se establecen limitaciones a la propiedad y servidumbres para la protección radioeléctrica del Centro Astronómico de Yebes.

¹⁷<http://www.mityc.es/telecomunicaciones/Espectro/Paginas/CNAF.aspx>.

que la investigación del Universo precise en ocasiones observar a frecuencias distintas de las indicadas, sino principalmente porque cualquier emisión intensa fuera de esas bandas (típica de los demás servicios de telecomunicaciones) puede ser detectada por la instrumentación ultrasensible del observatorio (grandes radiotelescopios, y receptores de muy bajo ruido y gran ganancia; figs. 9 y 10). Dispositivos como los UWB (“Ultra Wide Band”, como son los radares anticolidión de los automóviles de alta gama) llenan todo el espectro con radiación de baja potencia que, en las cercanías de un radiotelescopio, genera importantes interferencias. Mejoras imprescindibles en la ley nacional serán por ejemplo el establecimiento de “Zonas de Silencio Radioeléctrico” (como ya existen en otros países) alrededor de los observatorios, y la coordinación internacional con los operadores de transmisiones vía satélite para evitar la iluminación de los parajes donde se asientan los observatorios de radioastronomía más importantes.

La Fundación Europea para la Ciencia (ESF), mediante su Comité de expertos sobre Frecuencias para Radioastronomía¹⁸, vigila para que se legisle atendiendo las necesidades del estudio científico del cosmos, de manera que las frecuencias de interés se mantengan libres de interferencias.

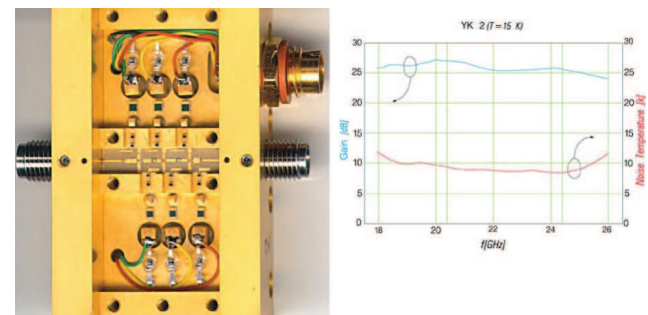


Figura 10: Amplificador criogénico ultrasensible de banda K, optimizado para cubrir la banda de 20,5 a 24,5 GHz para observaciones de las líneas del agua (22,3 GHz) y HC3N (23,7-24,1 GHz), que puede ser usado en la banda extendida de 18 a 26 GHz sin mucha degradación en los resultados de ruido y ganancia. El amplificador tiene tres etapas con dispositivos de fosforo de Indio (InP) hechos por TRW para CHOP (Cryogenic HEMT Optimization Program, JPL, EE.UU.). El diseño se realizó originalmente para el receptor de 22 GHz del radiotelescopio de 40 m del IGN en Yebes.

¹⁸CRAF, <http://www.astron.nl/craf/>

Asociaciones en defensa del cielo nocturno

La Asociación contra la Contaminación Lumínica Cel Fosc, es una organización española de ámbito nacional, dedicada expresamente al estudio, divulgación, asesoramiento y denuncia en materia de contaminación lumínica. Está integrada por científicos, ingenieros, educadores, periodistas y otros profesionales, y cuenta con una amplia participación de astrónomos aficionados, además de tener la adhesión expresa de algunas de las asociaciones astronómicas de más solera de nuestro país. La defensa de las condiciones naturales de oscuridad por medio del control del alumbrado nocturno, es uno de los principales objetivos de esta asociación. En particular, la defensa del cielo oscuro como patrimonio y derecho de todos, constituye uno de sus principales objetivos. Mantiene un grupo de discusión en internet¹⁹ al que pueden suscribirse gratuitamente todas las personas interesadas en el problema de la contaminación lumínica.

Objetivos similares se desarrollan en varias provincias. En Murcia, la “Campaña Cielo Oscuro”²⁰ pretende específicamente lograr que los ayuntamientos de dicha comunidad autónoma garanticen una adecuada iluminación y dispongan de ordenanzas municipales donde se establezcan las condiciones que deben cumplir las instalaciones de alumbrado exterior, y también conseguir la promulgación de una ley de ámbito autonómico contra la contaminación lumínica. Está adherida a la campaña “Energía Sostenible para Europa” (Sustainable Energy Europe Campaign), una iniciativa de la Comisión Europea para sensibilizar a la ciudadanía sobre la necesidad de cambiar el panorama energético en Europa.

En estrecha colaboración con la Asociación de Astrónomos Aficionados de la Universidad Complutense de Madrid (ASAAF-UCM), el Grupo de Cielo Oscuro de la Universidad Complutense de Madrid²¹ elabora mapas de iluminación de la UCM, realizando medidas de brillo del fondo de cielo, e informa acerca de los efectos de la contaminación lumínica, tratando de reducir su impacto y promoviendo una iluminación más eficiente y económica en general (y en la Ciudad Universitaria en particular).

A nivel internacional, la “International Dark-Sky Association” (Asociación Internacional para el Cielo Oscuro)²² promueve cambios en las legislaciones de los países que permitan la protección de los observatorios astronómicos, y la percepción que los ciudadanos pueden alcanzar del Universo, además del ahorro energético, financiero, y el uso racional de los recursos. Está respaldada por la Unión Astronómica Internacional (grupo de trabajo

“Controlando la contaminación lumínica” de la Comisión 50²³). Objetivos similares a los que persigue la activa asociación “CieloBuiro” en Italia²⁴.

Epílogo

En las últimas décadas se ha comprobado que la cantidad de iluminación producida por la sociedad, consecuencia del progreso deseado por todos, ha aumentado alarmantemente. Si bien dicha consecuencia es explicable, no es sin embargo inevitable. El progreso no necesita de más sino de mejor iluminación. El uso racional de los recursos, en este aspecto, tiene numerosos beneficios medioambientales, económicos, culturales, y para la investigación. La defensa del cielo nocturno tiene ventajas indudables. Permite admirar la belleza e inmensidad del cosmos, que debe estar accesible para todos los habitantes del planeta. No hacerlo sólo presenta desventajas. No esperemos más a plantear una solución consensuada, valiente, y eficaz.

¹⁹Grupo de discusión de Cel Fosc en internet, <http://es.dir.groups.yahoo.com/group/celfosc/>.

²⁰<http://www.um.es/cieloscurro/>.

²¹<http://www.ucm.es/info/Astrof/users/jaz/CL/index.html>.

²²<http://www.darksky.org/>.

²³http://www.iau.org/public_press/themes/light_pollution/.

²⁴Coordinamento per la protezione del cielo notturno, <http://www.cielobuiro.org>.

Avances y nuevos retos en la lucha contra la contaminación lumínica

La definición de contaminación lumínica más aceptada hoy día es un término genérico que indica la suma total de todos los efectos adversos de la luz artificial (Referencia 1).

La evolución del gasto eléctrico en alumbrado público para España ha crecido en los últimos 25 años a un ritmo del 4 %. El consumo energético es de 5,4 TWh/año, lo que supone un gasto de cerca de 950 millones de euros anuales, siendo el país con mayor consumo en alumbrado por habitante de la Unión Europea y el segundo en valores absolutos (2).

Impactos de la contaminación lumínica

En 2016 se ha publicado un nuevo atlas mundial de contaminación lumínica por Falchi y colaboradores (3). Según estos autores, más del 80 % del mundo y más del 99 % de las poblaciones de Estados Unidos y Europa viven bajo cielos contaminados por la luz. La Vía Láctea está oculta a más de un tercio de la humanidad, incluyendo el 60 % de los europeos y casi el 80 % de los norteamericanos. Además, el 23 % de las superficies terrestres del mundo en latitudes entre 75° N y 60° S, el 88 % de Europa y casi la mitad de los Estados Unidos experimentan noches contaminadas por la luz.

Con el fin de impulsar y coordinar la investigación interdisciplinar y la divulgación de este problema en nuestro país se constituyó en 2012 la Red Española de Estudios sobre la Contaminación Lumínica (REECL) (4). Esta red está integrada actualmente por catorce grupos pertenecientes a universidades, institutos, observatorios y planetarios e incluye tanto a astrónomos y físicos como a ingenieros, biólogos y científicos de la salud. La primera reunión formal de esta red tuvo lugar, precisamente, en el Real Observatorio de Madrid.

Desde la publicación del artículo en el Anuario 2010, las lámparas de descarga están siendo sustituidas por la industria masivamente en el alumbrado, tanto de interior como de exteriores, por una nueva categoría de lámparas de estado sólido que conocemos como LED (*light-emitting diode*) dentro de la cual existen diversos tipos: blanco frío, blanco neutro, blanco cálido, PC-ámbar... Muchos ayuntamientos han sustituido los alumbrados existentes por LED blanco frío, de forma precipitada y sin haber llevado

a cabo una evaluación de su impacto ambiental, lo que ha aumentado su impacto en el cielo y el medio nocturno y posiblemente también en la salud humana. Una solución son las lámparas PC-ámbar (*Phosphor-Converted*), que eliminan la emisión en el azul y longitudes de onda más cortas, y son las que se recomiendan a día de hoy desde el ámbito tanto astronómico como ecológico (5).

Legislación en España

La recopilación más completa está disponible en la web de la asociación Cel Fosc (6): las dos normativas de ámbito estatal son la Ley de Calidad del Aire y Protección de la Atmósfera (7), y el Real Decreto de Eficiencia Energética (8). Hay siete leyes autonómicas (Cataluña, Illes Balears, Navarra, Cantabria, Andalucía¹, Extremadura y Castilla y León), de las que sólo tienen actualmente desarrollo reglamentario tres (Cataluña, Navarra y Cantabria en parte). Los observatorios astronómicos en Canarias están protegidos por una normativa estatal especial (aunque no las islas Canarias completas) que sólo afecta en cuestión de contaminación lumínica a la isla La Palma y a la parte de Tenerife que se ve desde La Palma, y a ambas en contaminación radioeléctrica. Por último, la Ley de Evaluación Ambiental hace una mención que ampara la consideración de las afecciones por contaminación lumínica en proyectos y obras (9).

Contaminación radioeléctrica

La Ley General de Telecomunicaciones se modifica frecuentemente. En el momento de escribir esta actualización, está en vigor la Ley 9/2014 General de Telecomunicaciones, de 9 de mayo (10). La Disposición Adicional Segunda amplía el rango de frecuencias susceptible de solicitar protección, hasta los 270 GHz. Asimismo, el Cuadro Nacional de Atribución de Frecuencias, CNAF, ha sido modificado en 2013 (11).

Desarrollo del astroturismo, zonas Starlight en España y redes sociales

El astroturismo ha experimentado un gran desarrollo en los últimos años. Hay mucha información actualizada en la web de la Fundación Starlight (12).

La actividad en redes sociales es frenética... hay muchas iniciativas en

¹El decreto de la ley en Andalucía fue declarado nulo por el Tribunal Superior de Justicia de Andalucía y no está en vigor.

Facebook, Twitter, etc.; por ejemplo, recomendamos el grupo “Contaminación lumínica, cielos oscuros” en Facebook (13).

Referencias

Ref. 1: Vocabulario Internacional de Iluminación de la Comisión Internacional de la Iluminación (CIE). International Lighting Vocabulary (Standard CIE S 017/E:2011 ILV), <http://eilmv.cie.co.at/>

Ref. 2: Sánchez de Miguel, Alejandro (2016) *Variación espacial, temporal y espectral de la contaminación lumínica y sus fuentes: Metodología y resultados* [tesis doctoral], <http://eprints.ucm.es/31436/>

Ref. 3: F. Falchi, P. Cinzano, D. Duriscoe, C. C. M. Kyba, C. D. Elvidge, K. Baugh, B. A. Portnov, N.A. Rybnikova y R. Furgoni. *The new world atlas of artificial night sky brightness*, Science Advances, Vol. 2, no. 6, e1600377 (2016). <http://advances.sciencemag.org/content/2/6/e1600377.full>. El atlas se puede descargar o visualizar cómodamente en: <http://cires.colorado.edu/Artificial-light>

Ref. 4: Red Española de Estudios sobre la Contaminación Lumínica (REECL), <https://guaix.fis.ucm.es/splpr/>

Ref. 5: Espectros de los diversos tipos de luminarias: blanco frío, blanco neutro, blanco cálido, PC-ámbar. <http://www.papaqueeslavialactea.com/es/caracteristicas>

Ref. 6: Cel Fosc, Asociación contra la Contaminación Lumínica, <http://www.celfosc.org>

Ref. 7: Ley 34/2007, de 15 de noviembre, de calidad del aire y protección de la atmósfera. <http://www.boe.es/buscar/act.php?id=BOE-A-2007-19744>

Ref. 8: Real Decreto 1890/2008, de 14 de noviembre, por el que se aprueba el Reglamento de eficiencia energética en instalaciones de alumbrado exterior y sus Instrucciones técnicas complementarias EA-01 a EA-07. <http://www.boe.es/buscar/act.php?id=BOE-A-2008-18634>

Ref. 9: Ley 21/2013, de 9 de diciembre, de evaluación ambiental. <http://www.boe.es/buscar/act.php?id=BOE-A-2013-12913>

Ref. 10: Ley 9/2014, de 9 de mayo, General de Telecomunicaciones, <http://www.boe.es/buscar/act.php?id=BOE-A-2014-4950>

Ref. 11: Cuadro Nacional de Atribución de Frecuencias, CNAF, modificado en 2013 (B.O.E. de 9 de mayo 2013, actualizado B.O.E. de 16 de abril 2015). <http://www.minetad.gob.es/telecomunicaciones/Espectro/Paginas/CNAF.aspx>

Ref. 12: Fund. Starlight, <http://www.fundacionstarlight.org>

Ref. 13: Grupo “Contaminación lumínica – cielos oscuros” en Facebook: <https://www.facebook.com/groups/celfosc/>

Francisco Colomer Sanmartín, mayo de 2017

LOS PLANETAS



LAS ATMÓSFERAS DE LOS PLANETAS DEL SISTEMA SOLAR

Julio Solís García

Agencia Estatal de Meteorología

Estamos tan acostumbrados al azul del cielo, a las nubes blancas, a nuestros cúmulos, cirros, nieblas, heladas, lluvia, nieve, viento, etc., que muchas veces nos olvidamos de que por encima de nuestras cabezas, recorriendo el cielo entre las estrellas, nos acompañan esos magníficos luceros, nuestros parientes próximos, los planetas y alguno de sus grandes satélites. Ellos tienen también sus propias atmósferas, su particular geografía, su luz, sus nubes y vientos, que forman sus maravillosos y exóticos paisajes tan distintos de los nuestros.

Venus

Parece un planeta gemelo de la Tierra, por sus dimensiones (su radio es tan solo 285 km menor que el terrestre) y porque está rodeado de nubes, pero las apariencias engañan. Es un planeta hermoso cuando lo observamos a simple vista, con prismáticos o con telescopio. Después del Sol y la Luna es el objeto más brillante en nuestro cielo, lo que ha dado lugar en tiempos pasados a fantasías que describían a Venus, (que es por otro lado el planeta más cercano a nosotros, llegando a una distancia mínima de 38 millones de kilómetros, motivo por el que brilla tanto en el cielo, aparte de por su elevado albedo que le proporciona el permanente manto nuboso), como un paraíso con mucha agua, grandes y extensos bosques y un clima tropical, como pusieron de manifiesto en sus obras J.H.B. de Saint-Pierre o el premio Nobel S.A. Arrhenius, en las que manifestaban su creencia en que Venus era un lugar idílico y exótico, plagado de selvas y agua en abundancia. Con la llegada al planeta de las primeras sondas interplanetarias, la soviética Venera 1, y la norteamericana Mariner 2, se enterraron definitivamente todos aquellos bosques y selvas tropicales de ensueño.

En 1966 se envió el primer vehículo que pudo atravesar su densa atmósfera, y en 1970 se logró posar suavemente una sonda en el suelo de Venus (Venera 7), que pudo transmitir datos durante 23 minutos. Las posteriores sondas Venera 8, 9 y 10 fueron capaces de “sobrevivir” durante un tiempo más dilatado a los caóticos ataques de los componentes atmosféricos y al intenso calor superficial, enviando, además, las primeras fotografías de la superficie de ese mundo. En los años siguientes lograron posarse otras seis naves Venera, hasta completar una serie de 16, la última de las cuales llegó al planeta en 1983. Posteriormente se han lanzado más sondas interplanetarias, como las misiones Vega 1 y 2, quienes antes de continuar su trayectoria hacia el cometa Halley liberaron dos globos sonda que durante tres días estuvieron mandando datos de la atmósfera. Alguno de los módulos de la serie soviética Venera, mientras hacían su entrada en la atmósfera, detectaron, entre otros componentes, dióxido de carbono (96 %), nitrógeno (3 %), agua, gases sulfurosos (sulfuro de hidrógeno y sulfuro de carbonilo), y algunos gases nobles (argón, xenón, neón y helio).



Figura 1: La Tierra y Venus a la misma escala. (Cortesía de NASA)

La estructura nubosa de Venus es persistente, manteniendo al planeta siempre cubierto de nubes. Si fuéramos ascendiendo desde la superficie podríamos observar que la atmósfera es limpia y transparente hasta unos 30 km de altitud, aunque es necesario señalar que la presión y temperatura en superficie son tan altas, que el dióxido de carbono adquiere un aspecto fluido y pastoso, con una elevada eficiencia en la transmisión de calor, favoreciendo precisamente la uniformidad de esa elevada temperatura por todo el planeta. Entre los 32 y los 48 km iría apareciendo una bruma de dióxido de azufre y azufre polimerizado responsable de la reflexión de la radiación infrarroja hacia el suelo, y que se convertirá en una verdadera capa nubosa a partir de los 48 km, con un espesor de cuatro o cinco kilómetros, compuesta por partículas cristalinas, sales de cloro y azufre, y

por ácido sulfúrico. Por encima de esa capa nubosa aparecerían otras dos capas de nubes compuestas por una solución acuosa al 80 % de ácido sulfúrico, y pequeñas cantidades de ácido clorhídrico y ácido fluorhídrico, que irían perdiendo densidad y reduciendo el tamaño de las partículas con la altura, hasta sobrepasar los 67 km. Si seguimos subiendo nos encontramos finalmente con una capa brumosa de dióxido de azufre hasta los 90 km.

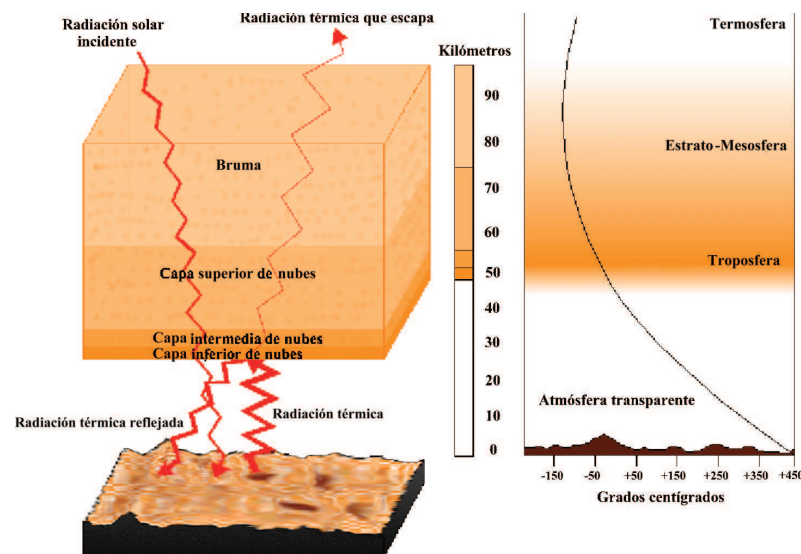


Figura 2: Corte de la atmósfera en Venus. (Imagen creada por Julio Solís)

Su velocidad de rotación es muy lenta, tarda nada menos que 243 días en rotar sobre sí mismo, y además lo hace en sentido contrario, es decir, en el sentido de las agujas del reloj mirándolo desde el polo norte, de modo que el sol en Venus sale por el oeste y se pone por el este (si es que pudiera verse la salida y puesta de sol, porque en realidad ese fenómeno no es observable desde la superficie al estar su cielo permanentemente nublado en todo el planeta). Tampoco existen estaciones similares a las de la Tierra, dado que su eje de rotación es casi perpendicular al plano de su órbita, la excentricidad de su órbita es muy pequeña y la uniformidad climática por toda la superficie es dominante, debida al fuerte efecto invernadero y al permanente manto de nubes.

Las condiciones climáticas en la superficie de Venus son radicalmente opuestas a las nuestras, con una temperatura ambiente por encima de los +460 °C casi constante por todo el planeta, da igual que nos situemos en

los polos o cerca del ecuador, o que sea de día o de noche, y una presión atmosférica en superficie de 93 200 hPa (90 veces la nuestra, es decir, la que podríamos encontrar en el fondo marino a una profundidad de un kilómetro), sequedad absoluta y presencia en la atmósfera de sustancias muy nocivas y corrosivas, como el ácido sulfúrico y el ácido clorhídrico, que darían lugar a verdaderos ríos de plomo, en el caso de que hubiera grandes cantidades de dicho metal en su superficie, corriendo en estado líquido sobre la superficie pedregosa. Todo ello sitúa a Venus en el primer lugar de la lista de planetas más cálidos del sistema solar, más aún que Mercurio.



Figura 3: Paisaje de Venus. (Imagen creada por Julio Solís)

La temperatura en el seno de las nubes oscila entre los +100 °C en la base de la capa más baja y unos +10 °C en la cima. En la parte superior de la capa nubosa se originan partículas de ácido sulfúrico que se precipitan hacia las partes inferiores, donde el fuerte calor las descompone en dióxido de azufre y agua, provocando su evaporación y vuelta a las alturas, por lo que aunque exista una continua lluvia, ésta no llega nunca al suelo (tenemos una permanente virga de ácido sulfúrico en Venus).

Curiosamente, las nubes de Venus deben su opacidad al gran espesor que poseen, y no a su densidad, son más bien una especie de neblina que permite una visibilidad de un kilómetro aproximadamente dentro de las propias nubes. No obstante, la circulación vertical de las partículas que componen las nubes genera un estado de frecuentes fenómenos eléctricos similares a los terrestres, con truenos y relámpagos que deben ser constantes, aunque dada la altura de las nubes no parece probable que alcancen nunca el suelo, al modo de nuestros conocidos rayos.

Los vientos en superficie son muy flojos o nulos, aunque pueden ejercer una fuerza considerable en objetos y obstáculos, debido a la enorme densidad del aire venusiano. Sin embargo la masa nubosa en su cima se mueve muy rápidamente, tanto que dan una vuelta al planeta en cuatro días (mientras el propio planeta tarda 243 días en girar sobre sí mismo). Este curioso fenómeno se conoce como “superrotación”, con vientos de más de 370 km/h, que arrastran las nubes de este a oeste. Entre el ecuador y las latitudes medias domina la mencionada superrotación con esos fuertes vientos zonales constantes, que van decreciendo globalmente dentro de las nubes, con la altura, hasta velocidades de 180 km/h. A partir de dichas latitudes medias, los vientos decrecen hasta hacerse nulos en los polos, donde se forma un inmenso vórtice. También se produce un movimiento meridional (norte-sur) en forma de célula de Hadley, que transporta el calor desde zonas ecuatoriales hacia las polares a unas velocidades muy débiles de unos 15 km/h. Todavía es una incógnita la verdadera causa de que un planeta que gira tan lento tenga vientos globales huracanados tan fuertes en la cima de su cubierta nubosa.

Las nubes son un factor determinante en el clima de Venus. La capa que envuelve al planeta deja pasar la mayor parte de la radiación solar, que calienta el suelo, pero es muy opaca a la radiación infrarroja, dejándola retenida entre el suelo y las nubes, provocando un recalentamiento de la superficie, caso extremo de efecto invernadero que debería ponernos en guardia respecto a lo que podría pasar en la Tierra en caso de aumentar descontroladamente la acumulación de gases como el dióxido de carbono. Si pudiéramos llegar a la superficie de Venus, y aguantar el ambiente corrosivo y las altísimas temperatura y presión, veríamos el suelo con una apariencia de desierto seco y rocoso, inmerso en una luz amarillenta-anaranjada. Debido a la alta presión y densidad podríamos apreciar fenómenos ópticos como la refracción múltiple, que dan lugar a sucesivas imágenes de un mismo objeto. Tendríamos la sensación de estar viendo el paisaje desde el fondo de una piscina o desde el fondo del mar, sin olvidar que la luminosidad en la superficie de Venus es incluso menor que la de un día nublado y tormentoso en la Tierra. No se verían nunca ni el Sol ni el cielo estrellado.

Tierra

La atmósfera de nuestro planeta no fue siempre como la conocemos ahora. Tras la desgasificación producida por la intensa actividad volcánica de una parte de los elementos volátiles químicamente ligados o absorbidos por el material condensado, que constituía la parte sólida del planeta, se liberaron cantidades muy importantes de vapor de agua, monóxido y dióxido de carbono, que dieron lugar a los océanos, lagos y ríos, conformando una “hidrosfera” cuyo carácter regulador de la temperatura resultó determinante para el clima terrestre, junto con su capacidad para transformar grandes

cantidades de dióxido de carbono, dióxido de azufre, ácido clorhídrico y ácido fluorhídrico en sales marinas y minerales. Su carácter reductor inicial se perdió debido a procesos de fotosíntesis bajo la acción de la luz solar, que genera oxígeno libre en la atmósfera, facilitando a su vez reacciones químicas que dan lugar al nitrógeno, gas inerte que no reacciona con las rocas superficiales y que por tanto quedó en la atmósfera, constituyendo el 78 % del total (en segundo lugar aparece el oxígeno con un 21 %, siendo el porcentaje de otros gases como el argón, dióxido de carbono, neón, vapor de agua, kriptón, metano, ozono, etc., residual, aunque cualitativamente determinante en la evolución climática).

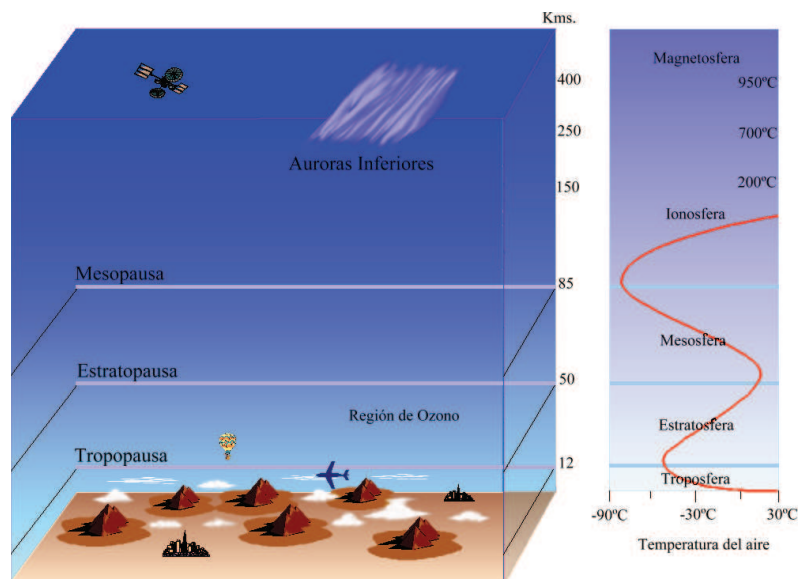


Figura 4: Corte de la atmósfera de la Tierra. (Imagen creada por Julio Solís)

La dinámica de la atmósfera terrestre tiene su “motor” en la energía que recibe del Sol, responsable último de todos los fenómenos meteorológicos desde un punto de vista energético, dando lugar a las borrascas y anticiclones, frentes, y todos los meteoros que nos resultan familiares. Ascendiendo desde la superficie, recorriendo la troposfera, que es el nombre que se le da a esta primera capa, la temperatura disminuye a razón de algo más de 6 °C por kilómetro, hasta alcanzar los -60 °C en la tropopausa (a unos 12 km de altitud), lugar en el que se entra en la siguiente capa, la estratosfera, en la que tiene lugar una inversión térmica, aumentando la temperatura con la altura hasta alcanzar los 0 °C en la estratopausa que da lugar a la tercera

capa atmosférica, la mesosfera, que se extiende desde los 50 a los 80 km aproximadamente y donde la temperatura vuelve a bajar con la altura hasta alcanzar los -100 °C en la mesopausa. A partir de ese punto se entra en la termosfera. El 80 % de los gases atmosféricos se concentra en la troposfera, que es donde se desarrollan todos los fenómenos meteorológicos, quedando el 20 % restante en la estratosfera, zona limpia y clara, sin nubes ni vientos destacables y con una visibilidad excelente.



Figura 5: Estratocúmulos y altocúmulos - mar de nubes. (Julio Solís)

En la estratosfera y en la termosfera se encuentran varias zonas de especial interés para la vida y para las comunicaciones, por un lado, a unos 25 km de altitud se encuentra una capa rica en ozono (ozonosfera) que nos protege de una parte importante de la radiación ultravioleta procedente del Sol, que afectaría muy negativamente el desarrollo de la vida si llegara directamente a la superficie. Por otro lado, entre los 90 y los 300 km se encuentra la ionosfera, zona donde las partículas del aire están ionizadas, comportándose como un espejo para las ondas de radio cortas y ultracortas, haciendo posible las telecomunicaciones entre lugares que quedan “invisibles” entre sí debido a la curvatura de la Tierra. También se producen en esta zona las espectaculares auroras polares, que no son otra cosa que resplandores de diversas formas y colores generados por la interacción del viento solar con el campo magnético terrestre, como si fuera un gigantesco tubo de neón.

Marte

Bastante más amigable que Venus, en Marte ya no tendremos que soportar presiones aplastantes, temperaturas abrasadoras o una atmósfera corrosiva. Su tamaño es la mitad del de la Tierra, cuenta con una masa 10 veces menor y la fuerza de la gravedad en su superficie es el 38% de la nuestra. Gira alrededor del Sol en 687 días y le separan del mismo 70 millones de kilómetros más.

Marte presenta estaciones, que duran el doble que en la Tierra, debido básicamente a la inclinación de su eje de rotación que es de 25° , también tiene casquetes polares y una atmósfera ligera. Otra similitud es la duración de sus días, con algo más de 24 horas de periodo de rotación. A pesar de todo, al carecer de vida (hasta donde somos capaces de saber, a día de hoy) y de otros elementos que puedan presentar cambios estacionales, salvo variaciones en las masas de hielo polar y el oscurecimiento de zonas de su superficie, esas estaciones no provocan demasiados cambios en el paisaje. La sensación que se podría tener sería parecida a la existente en un desierto pero con un clima ártico, 7 hPa de presión atmosférica en superficie, temperatura media de unos -50°C durante el día (casi siempre soleado), aunque en verano y en zonas ecuatoriales se pueden registrar hasta $+20^\circ\text{C}$, que llega a descender hasta -80°C durante la noche, y una tenue capa de ozono insuficiente para detener la radiación solar ultravioleta.

La atmósfera de Marte está constituida principalmente por dióxido de carbono (95,3 %), nitrógeno (2,7 %), argón (1,7 %), cantidades menores de agua, monóxido de carbono y oxígeno molecular, y vestigios de gases nobles como el neón, kriptón y xenón. En función de la estación del año marciano, hora del día y latitud, varía la concentración de vapor de agua en la atmósfera, aunque nunca da lugar a nubes de cierta consistencia como pudieran ser los cúmulos o estratocúmulos terrestres. La atmósfera más seca se ha detectado en las zonas polares durante el invierno, y la más húmeda también en zonas polares pero en verano. Las frecuentes tormentas de polvo reducen el contenido de vapor de agua atmosférico mientras tienen lugar.

En las primeras etapas de su formación, Marte poseía una envoltura gaseosa mucho mayor, que fue paulatinamente desapareciendo, y con ella el agua líquida superficial. De ahí que Marte no tenga océanos, ni lagos, ni corrientes de agua, y esté helado (aunque se han encontrado indicios muy claros de que en un tiempo pasado hubo líquidos, seguramente agua, corriendo por su superficie en forma de torrentes, ríos o lagos). Por tanto, en semejanza con Venus y la Tierra, Marte tampoco posee una atmósfera primigenia, sino secundaria, o sea, compuesta por gases que fueron liberados por su caliente interior, por actividad volcánica. Los últimos datos ofrecidos por los vehículos enviados a la superficie marciana (Curiosity, Phoenix, Odyssey, etc.) confirman que en el subsuelo existen grandes depósitos de agua helada mezclada con tierra y rocas, sobre todo cerca del polo sur, que

pudieran ser los restos de lo que fueron mares, lagos y ríos de hace miles de millones de años, cuando la atmósfera era más densa.

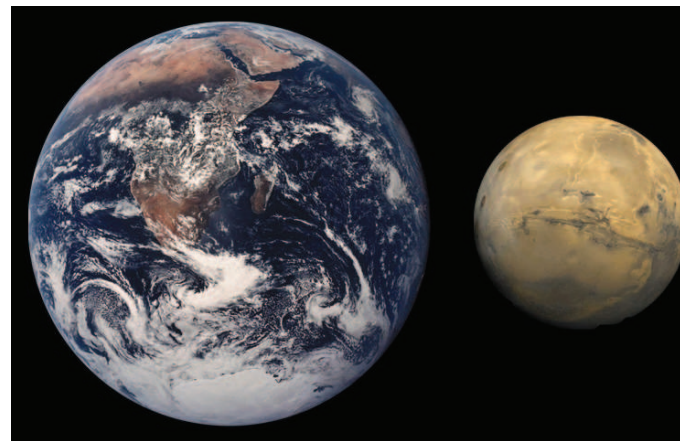


Figura 6: La Tierra y Marte a la misma escala. (Cortesía de NASA)

Contrariamente al proceso de calentamiento que ocurre en Venus, la tenue atmósfera de Marte no permite una acumulación de calor suficiente como para evitar el progresivo enfriamiento y formación de hielo en los polos (hielo seco, mayormente), estimándose en tan solo 5°C el aumento en la temperatura debido al efecto invernadero. La pequeña cantidad de oxígeno presente en la atmósfera imposibilita la formación de una verdadera capa de ozono, lo que permite la disociación del dióxido de carbono en monóxido de carbono y oxígeno en toda la atmósfera. El agua se disocia en hidrógeno atómico y en radical hidroxilo debido a los rayos UV solares, productos muy reactivos que pueden catalizar la recombinación del monóxido de carbono y del oxígeno para volver a dar dióxido de carbono, lo que mantiene la proporción de este compuesto en valores tan altos y constantes. Un fenómeno peculiar, exclusivo de Marte, es el flujo de condensación del dióxido de carbono desde el polo norte (que es calentado por los rayos solares, haciendo que el hielo se sublime), hacia el polo sur, donde se deposita en forma de hielo (se congela a -57°C). Las reacciones entre el hidrógeno atómico y el oxígeno molecular llegan a formar peróxido de hidrógeno, poderoso oxidante que juega un importante papel en la oxidación de los minerales de la superficie, dando a Marte ese tinte rojo tan característico.

A pesar de la pequeña cantidad de vapor de agua en la atmósfera marciana, se alcanza la saturación con mucha facilidad, dando lugar a diversos tipos de nubes bastantes parecidas a las existentes en la Tierra. Las nubes de Marte son amarillentas, blanquecinas y azuladas, las primeras deben estar

formadas por polvo superficial levantado por el viento, y siempre aparecen asociadas a las grandiosas tempestades de arena, tormentas de polvo tan habituales como únicas entre todos los cuerpos del sistema solar; las blancas se observan siempre en lugares próximos al “terminador” (línea que separa la zona iluminada de la oscura), es decir, al amanecer y al atardecer, dando la sensación de desaparecer con el calentamiento de la atmósfera en el transcurso del día. Las nubes azuladas siguen el mismo patrón de conducta que las blancas, que también se forman en las zonas polares. A pesar de la existencia de nubes, no llueve nunca, pues las mencionadas condiciones de presión y temperatura no lo hacen posible.

Entre los distintos tipos de nubes se encuentran las nubes convectivas, que se forman por calentamiento de la superficie durante el día a una altitud de entre 5 y 8 kilómetros, versiones reducidas y menos consistentes de nuestros conocidos cúmulos, en forma de pompa, que se forman en las altiplanicies ecuatoriales a partir del mediodía marciano. Cuando existen fuertes vientos y encuentran un gran obstáculo en su camino (una elevada cadena montañosa, por ejemplo), al rebasarlo se produce un movimiento ondulatorio que da lugar a las llamadas nubes ondulatorias, de bastante parecido a los altocúmulos lenticulares que vemos en la Tierra, siempre que se den las condiciones de humedad y temperatura necesarias. Si se da un ascenso forzado por una gran pendiente de enorme extensión, el aire, en su elevación, se satura y forma nubes orográficas, que son nubes aisladas, delgadas y uniformes, con similar aspecto a los estratos terrestres, y que se encuentran en las proximidades de las elevadas cumbres (recordemos que Marte posee las montañas más altas de todo el sistema solar). Las nieblas matinales parecen estar en equilibrio con la capa de escarcha de agua que cubre la superficie del planeta en extensas regiones, que con el calentamiento de los primeros rayos del sol se evapora (sublima), condensándose en el seno de la atmósfera y dando lugar a la niebla. Por la noche, cuando descienden las temperaturas de nuevo hasta valores de -120°C , aparece de nuevo la blanquecina escarcha sobre la superficie.

En las regiones polares, y en épocas invernales o de finales de otoño, donde pueden registrarse hasta -130°C , la temperatura de la atmósfera a grandes altitudes es suficientemente baja como para dar lugar a la formación de unas nubes blancas que pueden estar compuestas por cristales de hielo seco. Son nubes de dióxido de carbono, probablemente con aspecto parecido a los cirros terrestres, que desaparecen en primavera.

En Marte, los fenómenos erosivos están provocados por el viento y por las diferencias de temperaturas, tan notables a causa de una atmósfera tan liviana, y que han pulverizado las rocas cristalinas de la superficie, dando lugar a enormes campos de dunas y a una capa de polvillo que recubre la totalidad del planeta. Aparecen diferenciados dos regímenes de vientos, el de invierno en latitudes medias, en el que, como ocurre en la Tierra, se suceden vientos dominantes del oeste y corrientes en chorro a gran altura,

y el del verano ecuatorial, sometido únicamente al lento suceder de las largas estaciones marcianas, cuyo “motor”, al igual que en la Tierra, es la variación diurna en la insolación, y por tanto del calor suministrado al suelo por el Sol. El efecto combinado del calor y la topografía local, es la clave de los vientos en las zonas ecuatoriales, en donde los vientos dominantes son análogos a las brisas de montaña de la Tierra.



Figura 7: Superficie de Marte. (Imagen creada por Julio Solís)

El ciclo diario de calentamiento de la superficie, conlleva una notable oscilación en el régimen de vientos (efecto de marea) en la atmósfera, mucho más intenso que en nuestro planeta. Cerca de la superficie, el viento suele soplar con una velocidad media aproximada de 40 km/h, viento más que apreciable. Aún así, a distancias mayores del suelo, entre 50 y 100 km, por ejemplo, los efectos de marea son mucho más intensos, sirviendo de “mezclador” eficaz de los componentes atmosféricos. A consecuencia de la baja presión en la superficie, o mejor dicho de la baja densidad de su atmósfera, levantar partículas sólidas de polvo o arena del suelo precisa de vientos de más de 100 km/h; no obstante, una vez suspendidas en el aire pueden ya permanecer como tormentas o grandes tempestades de polvo durante largos periodos de tiempo, meses incluso, debido al efecto combinado de la baja gravedad en superficie ($1/3$ de la terrestre) y de su tenue atmósfera. En invierno, cerca de los polos, se producen fuertes vientos, con intensidad superior a 300 km/h, que suelen originar esas espesas tormentas de polvo locales que frecuentemente se esparcen por todo el planeta dando lugar a violentas tempestades globales que cubren a Marte de un opaco velo amarillo-rosáceo.

El paisaje típico de Marte nos muestra impresionantes cañones, elevados volcanes extinguidos, cráteres semejantes a los de la Luna y continuas tormentas de arena que nublarán su rojiza superficie muy pedregosa y escarchada (en Marte se encuentra el Monte Olimpo, la mayor elevación del sistema solar, con un diámetro de 600 kilómetros y una cima que se eleva 24 000 metros sobre su base). El cielo presenta un color asalmonado más o menos intenso dependiendo de la cantidad de polvo en suspensión, que deja ver el Sol (siempre que alguna nube o tormenta de polvo, no lo impidan) con un tamaño casi la mitad del que muestra desde la Tierra. El cielo de Marte debe mostrar un espectáculo nocturno extraordinario, pues su suave atmósfera permitirá contemplar el sistema Tierra-Luna, Venus y Mercurio con particular claridad y colorido, al igual que Júpiter con sus grandes satélites galileanos y Saturno con sus anillos. Fobos y Deimos, sus dos lunas, se verán como dos estrellas muy brillantes desplazándose por el cielo marciano aparentemente en sentido contrario la una respecto a la otra, debido a que mientras que el periodo orbital de Deimos es mayor que el día marciano, el de Fobos es mucho menor. No podrían verse “redondas” como nosotros vemos a la Luna, en primer lugar debido a su pequeño tamaño y a su forma irregular (Fobos: $27 \times 22 \times 18$ km y Deimos: $15 \times 12 \times 10$ km) que las asemeja a dos grandes “patatas” cósmicas.

Proseguimos nuestro viaje hacia las profundidades del sistema solar, cruzamos el cinturón de asteroides y empieza a dominar el fondo estrellado el planeta gigante que se quedó a un paso de ser una estrella, Júpiter.

Júpiter

Un poco más de masa y se hubiera convertido en el segundo sol del sistema solar, con su cohorte de satélites helados rebosantes de agua, que hubieran dado lugar a planetas con unas condiciones similares a las de la Tierra, como sería el caso de su luna Europa, de hecho Júpiter tiene una composición química muy similar a la de las estrellas. El más grande de entre los gigantes, con una órbita que le sitúa a 770 millones de km del Sol. Su enorme tamaño de 143 000 kilómetros de diámetro permitiría alojar en su seno a más de mil trescientos planetas como la Tierra. Tiene un sistema de anillos, nada comparable a los majestuosos de Saturno, pero que le confieren un detalle de “elegancia” al rey del sistema solar, que cuenta además con un número de satélites superior a 65.

A pesar de su enorme tamaño no es un planeta “perezoso”, es un mundo muy agitado y dinámico, gira sobre sí mismo en algo menos de 10 horas, lo que le produce un achatamiento por los polos, siendo su radio ecuatorial un 7 % mayor que el polar. Si nos zambulléramos en su atmósfera con la pretensión de posarnos en su superficie, al modo en que lo haríamos en los planetas terrestres (Mercurio, Venus, la Tierra y Marte), nuestro empeño sería vano, pues Júpiter no tiene una superficie sólida diferenciada.



Figura 8: Júpiter desde Ganímedes. (Imagen creada por Julio Solís)

La magnetosfera joviana es la más grande y poderosa de entre todos los planetas del sistema solar, con un campo magnético 10 veces superior al terrestre y con una extensión de 7 millones de kilómetros en dirección al Sol y cientos de millones de kilómetros en dirección opuesta (hasta llegar casi a la órbita de Saturno), provocando magníficas y continuas auroras polares. Por debajo de la enorme y densa cubierta nubosa que envuelve al planeta de manera permanente, y que tiene una profundidad estimada de 1 000 kilómetros, encontraríamos un fluido líquido de hidrógeno y helio, con trazas de carbono, nitrógeno y azufre, sometido a una presión en aumento que alcanza 3 millones de atmósferas a una profundidad de 18 000 kilómetros, donde el hidrógeno adquiere propiedades metálicas al transformarse en una mezcla conductora de electricidad, con protones y electrones sueltos inundando el fluido. La temperatura y la presión siguen aumentando conforme nos acercamos al núcleo sólido y rocoso de Júpiter, compuesto por metales y silicatos, donde se alcanzan unos 30 000 °C y millones de atmósferas de presión.

En la atmósfera joviana encontramos sobre todo hidrógeno (>87 %) y helio (>12 %), aunque existen compuestos, como los derivados del azufre y del fósforo, en mucha menor proporción, que otorgan el variado colorido de las colosales formaciones nubosas que rodean al planeta. Metano, amoníaco, agua, cianuro de hidrógeno, fosfina, monóxido de carbono, etano, acetileno, hidrosulfuro de amonio, y otros muchos compuestos que se generan en presencia de la radiación ultravioleta, y de los relámpagos y descargas eléctricas en el seno de las nubes de metano.

Nos encontramos con las cimas de las nubes en niveles de presión de 1 hPa, capa nubosa permanente que da lugar, en las profundidades de la densa atmósfera, a una espesa niebla de amoníaco y agua que cubre el pastoso océano global de hidrógeno molecular, sin olas ni mareas y con vientos en calma. Por encima de la capa superior de nubes, la atmósfera presenta un aspecto similar al terrestre, azulado, pero algo más oscuro debido a la menor luminosidad del Sol.

Como resulta difícil identificar algo a lo que pudiera llamarse “superficie” en Júpiter, por debajo de su capa de nubes, podemos establecerla en aquel lugar de aspecto mortecino y penumbroso, en casi total oscuridad tan solo rasgada por frecuentes relámpagos, que no siendo claramente ni océano ni atmósfera, donde se registran 20 atmósferas de presión y +140 °C de temperatura. A partir de esta referencia arbitraria, en sentido ascendente, tendríamos una troposfera que es donde se desarrollarían los fenómenos meteorológicos, y en su tropopausa (50 kilómetros por encima) alcanzaríamos temperaturas por debajo de -170 °C y presiones inferiores a 0,1 atmósferas. En la estratosfera, con un espesor superior a los 250 kilómetros, la temperatura aumenta hasta -80 °C, temperatura que se mantiene más o menos constante en dicha capa. A partir de 10^{-6} atmósferas entramos en la termosfera donde la temperatura aumenta con la altura hasta alcanzar los 1 000 °C.

A diferentes niveles de la atmósfera encontramos nubes de distintas coloraciones, las más frías y por tanto más altas son rojizas en su cima, tomando un color blanquecino o parduzco por las zonas medias, siendo las nubes más bajas de tonalidad azulada. Existen tres capas de nubes, la más alta contiene nubes de amoníaco cristalino, la intermedia compuesta por nubes de hidrosulfuro amónico, y la inferior caracterizada por nubes compuestas por cristales helados de agua. Por debajo de las nubes más altas tipo “cirros”, de amoníaco, deben encontrarse nubes convectivas que podrían dar lugar a chubascos (de amoníaco también).

Júpiter es un planeta que emite más energía de la que recibe del Sol, y por tanto, la energía liberada a nivel de la baja atmósfera parece ser la mayor causa de la agitación atmosférica; en cambio, la energía proveniente del Sol tiene un efecto menor en la dinámica atmosférica del gigante. El comportamiento de los gases atmosféricos más profundos se asemeja al de un líquido colocado en una cazuela puesta al fuego. El calor emanado de sus entrañas agita los gases generando un incesante movimiento ascendente/descendente, que mantiene al planeta siempre cubierto de nubes, en su mayoría de carácter convectivo.

La configuración nubosa de Júpiter presenta las características bandas paralelas al ecuador, debido sobre todo su rápida rotación, alternándose en colores claros y oscuros, a las que se denominó desde los tiempos de las primeras observaciones telescópicas, “zonas” y “cinturones” (o “bandas”). Las zonas son provocadas por una fuerte ascendencia de gas, y sus nubes

son de tipo cumuliforme, de amoníaco y tono claro, alcanzando las mayores alturas y temperaturas muy bajas, con vientos del oeste de hasta 500 km/h. En los cinturones, los movimientos son descendentes, y sus oscuras nubes de hidrosulfuro amónico son más oscuras, menos densas y alcanzan menor desarrollo vertical, con vientos de menor intensidad que en las bandas nubosas adyacentes (“zonas”), y que en este caso soplarían en dirección contraria (serían vientos del este). A latitudes por encima de los 55°, la disposición nubosa pierde su carácter zonal y se vuelve totalmente desordenada, rompiéndose en irregulares movimientos de torbellino.

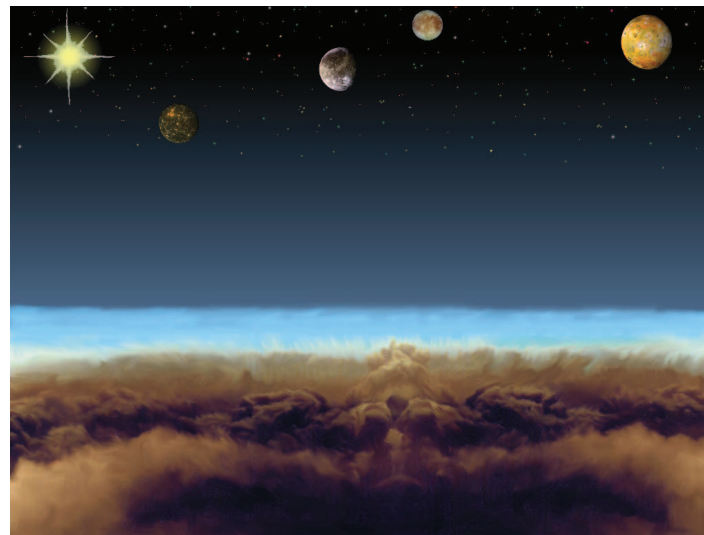


Figura 9: Vistas de Júpiter. Los tamaños angulares de los satélites galileanos no están a escala de sus diferentes distancias. (Imagen artística creada por Julio Solís)

Uno de los fenómenos meteorológicos más llamativos es la Gran Mancha Roja (GMR), que es un gigantesco anticiclón (óvalo de 12 000×30 000 kilómetros) situado en latitudes tropicales del hemisferio sur, que gira sobre sí mismo en unos cinco días, y tiene unos vientos periféricos huracanados de más de 400 km/h. También se observan vórtices que en su mayoría son blanquecinos y de carácter anticiclónico, estimándose que son formaciones de poca profundidad que no superan unos pocos cientos de kilómetros. Las estructuras atmosféricas descritas, bandas, zonas, vórtices, etc. son muy persistentes en el tiempo, dándose el caso por ejemplo de que la GMR se mantiene en su aspecto actual desde hace por lo menos 300 años.

Saturno

Otro gigante, solamente un poco menor que Júpiter, aunque si contamos su majestuoso sistema de anillos dejaría pequeño al rey de los planetas del sistema solar; sistema de anillos compuesto por ocho grupos diferenciados, situados en el plano ecuatorial del planeta, con un espesor aproximado de 1 kilómetro, y que se extienden desde 6500 km por encima de las nubes hasta casi los 480 000 km. Situado a 1 425 millones de km del Sol, tiene unas características únicas, gran achatamiento polar debido a su rápida rotación (de poco más de 10 horas), espectacular sistema de brillantes anillos, formados por trozos de hielo y rocas orbitando al planeta a 50 000 km/h, una densidad tan baja que le haría flotar en un barreño con agua lo suficientemente grande, y más de 60 satélites catalogados, entre los que se encuentra uno de gran tamaño y con una atmósfera importante (Titán).

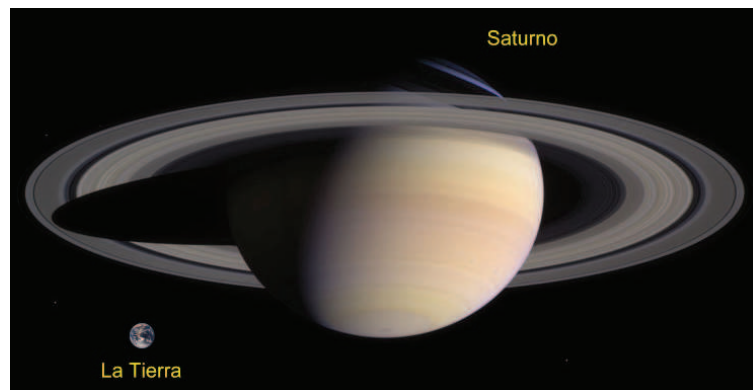


Figura 10: La Tierra y Saturno. (Cortesía de NASA)

Saturno está compuesto por hidrógeno (90 %), helio (5 %), y pequeñas cantidades de metano, vapor de agua, amoníaco, etano, propano, fosfina, monóxido de carbono, y acetileno. Tiene un núcleo sólido rodeado por una inmensa capa de hidrógeno líquido metálico, sobre la que se extiende a su vez la atmósfera de hidrógeno y helio, que alcanza unos 30 000 km, de espesor, y que presenta una estructura en bandas paralelas al ecuador semejantes a las de Júpiter pero menos marcadas y turbulentas, con velocidades del viento en las cimas de las nubes de hasta 1 500 km/h, del oeste, y sin que se aprecien grandes vórtices o remolinos.

El planeta se contrae lentamente generando un calor interno que, al igual que en el caso de Júpiter, es el motor principal de su dinámica atmosférica. Sus nubes, de amoníaco, hidrosulfuro de amonio, o agua, dependiendo

del nivel de presión, presentan unos colores tono “pastel” amarillento, que quedan semiocultas por una importante capa neblinosa uniforme de cristales de amoníaco, generada en la parte más alta de la atmósfera por fenómenos fotoquímicos. Se repiten los patrones observados en Júpiter respecto a la disposición nubosa en bandas paralelas al ecuador, con zonas (ascendencia) y bandas (subsistencia) debido al poderoso efecto Coriolis, hasta latitudes de $\pm 60^\circ$, dominando los torbellinos irregulares conforme nos acercamos a los polos.

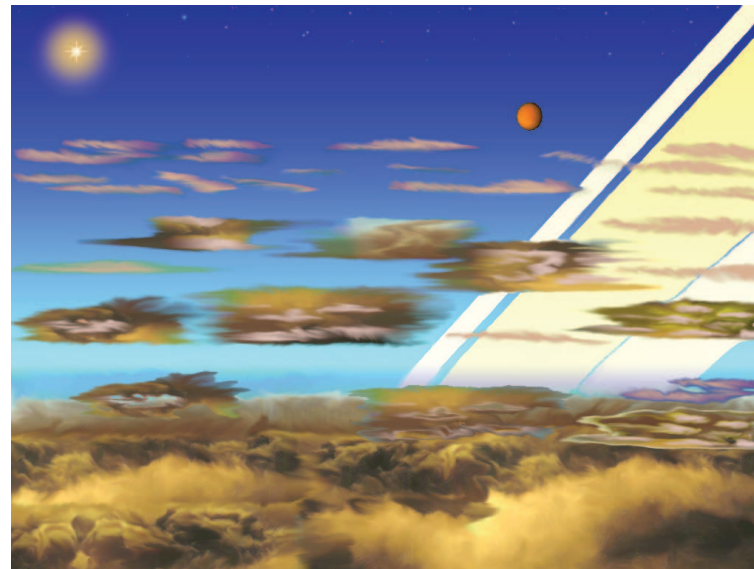


Figura 11: Vistas de Saturno. (Imagen creada por Julio Solís)

Tampoco encontraremos en Saturno una superficie sólida claramente diferenciada bajo las nubes. Los registros de las últimas sondas ofrecen temperaturas de -130°C a unos niveles de presión similares a los de la superficie de la Tierra, disminuyendo a razón de $0,85^\circ\text{C}$ por kilómetro, en sentido ascendente, hasta alcanzar -193°C a niveles de presión de 0,07 atmósferas. Por encima del manto de nubes se encuentra una ionosfera de hidrógeno ionizado, y una magnetosfera que aunque es tres veces menor que la de Júpiter supera en mucho a la terrestre, dando lugar a vistosas e intensas auroras polares. También se han encontrado en Saturno estructuras turbulentas a modo de grandes huracanes, fenómenos convectivos de tamaño colosal que atraviesan la bruma superior, sin el tamaño y espectacularidad de la GMR de Júpiter, pero que parecen tener un origen similar, siendo de color blanquecino, y con una vida media de pocos meses.

El satélite Titán

Titán es grande, tiene un tamaño intermedio entre Marte y Mercurio, y es el mayor de la numerosa cohorte de satélites de Saturno, y sobre todo el único con una atmósfera densa y homologable a la terrestre. El 14 de enero de 2005, la sonda Huygens logró posarse en la superficie de este gran satélite, tras separarse de la nave Cassini, ofreciendo imágenes y datos de la superficie y de la atmósfera, que nos han permitido conocer mucho mejor las características globales de este apasionante y helado mundo.

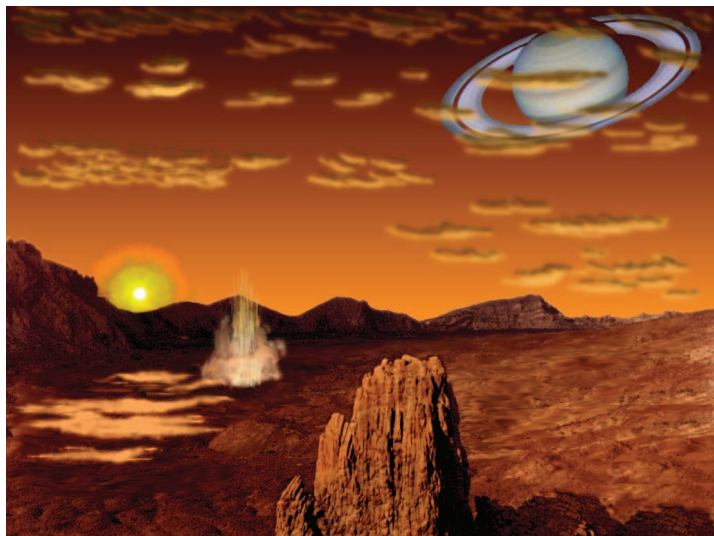


Figura 12: Superficie de Titán. (Imagen creada por Julio Solís)

Su atmósfera está compuesta de nitrógeno (95 %) y otros compuestos (5 %), mayormente hidrocarburos como el metano, y en menor medida etano, acetileno, metilacetileno, diacetileno, cianoacetileno, propano, dióxido y monóxido de carbono, cianógeno, cianuro de hidrógeno, helio y compuestos químicos complejos, resultado de la fuerte actividad fotoquímica en la atmósfera superior. Particularmente interesante resulta la presencia de cianuro de hidrógeno, nitrilo precursor de las purinas, constituyentes de los ácidos nucleicos presentes en las células vivas.

Titán es el único lugar del sistema solar, junto con la Tierra, donde se producen precipitaciones que dan lugar a ríos y lagos, aunque aquí los fenómenos meteorológicos no tienen su base en el agua, sino en el metano líquido, que es el componente principal de las nubes tormentosas de gran desarrollo vertical (hasta 35 kilómetros) que descargan importantes cantidades de precipitación. También existen nubes tipo cirros en la estratosfera

de Titán, a una altitud de entre 50 y 100 kilómetros, compuestas por cristales de hidrocarburos. Su temperatura ambiente media es de -180°C , con una superficie anaranjada y fría, arcillosa y blanda, con rocas dispersas y bloques de hielo, en la que puede haber actividad volcánica con erupciones de agua mezclada con amoníaco, y vapores brumosos próximos a ríos y lagos, o tras la lluvia, y donde la presión atmosférica registra 1 600 hPa. Se han detectado vientos dominantes del oeste, que en superficie tienen velocidades de entre 50 y 100 km/h, llegando hasta 200 km/h en zonas altas de la atmósfera, lo que unido a su rotación capturada o sincrónica, con un periodo de 16 días, pone de manifiesto un fenómeno de superrotación similar al de Venus, régimen dinámico cuyo origen sigue siendo una incógnita en ambos casos. Entre la niebla anaranjada y espesa en la alta atmósfera, y la lejanía al Sol, la luminosidad en un día cualquiera en Titán puede asemejarse a la crepuscular terrestre.

Su baja densidad media ($<2\text{ gr/cm}^3$) indica que debe estar compuesto por una mezcla de hielo y roca, con un manto arcilloso, y posiblemente con un océano de agua con amoníaco disuelto, y diversos hidrocarburos, a una profundidad de unos 100 kilómetros.

Urano

Al igual que Júpiter, Saturno y Neptuno, es un planeta gaseoso con un núcleo sólido y sin una superficie sólida diferenciada, con una permanente y densa cubierta nubosa. De aspecto verde azulado uniforme, en cuya superficie es difícil distinguir manchas o formaciones de cualquier tipo, es verdaderamente un planeta muy curioso, que tiene también anillos (nada parecidos a los espectaculares de Saturno), magnetosfera y un número importante de satélites (27), aunque ninguno tiene características destacables, y también experimenta un achatamiento por los polos debido a la rápida velocidad de rotación (lo hace en 17 horas), que además es retrógrada. Tarda 84 años en dar una vuelta completa alrededor del Sol, que destaca en su cielo (por encima de las nubes, claro) como una gran estrella brillante, perdida a 3 000 millones de kilómetros (su luz tarda 2 horas y media en llegar) y 400 veces menos luminoso que visto desde la Tierra; pero lo más llamativo es la gran inclinación de su eje de rotación (98°), que le convierte en un planeta que va “rodando” sobre su órbita con sus polos apuntando al Sol alternativamente durante 20 años cada uno, en cada vuelta alrededor del Sol, y tomando el aspecto de una diana visto desde el interior del sistema solar con los anillos como si fueran círculos alrededor del disco planetario.

Es el segundo planeta menos denso, después de Saturno, con un núcleo rocoso y un manto de hielos de agua, amoníaco y metano, aunque es importante señalar que este manto “helado” no está compuesto de hielo en sentido convencional, sino que es un fluido sometido a enorme presión,

caliente y denso con una conductividad elevada, y se parecería más a un océano espeso. Por encima tiene una atmósfera de hidrógeno molecular (83 %), helio (15 %), metano (2 %), amoníaco (0,01 %) y otros elementos en menor cantidad (etano, acetileno, monóxido de carbono, sulfuro de hidrógeno...). En esta atmósfera, la más fría del sistema solar (-224°C) encontramos varias capas nubosas de cristales de metano en zonas de mayor altitud, y de agua en zonas más bajas, con vientos que pueden llegar hasta los 1 000 km/h.

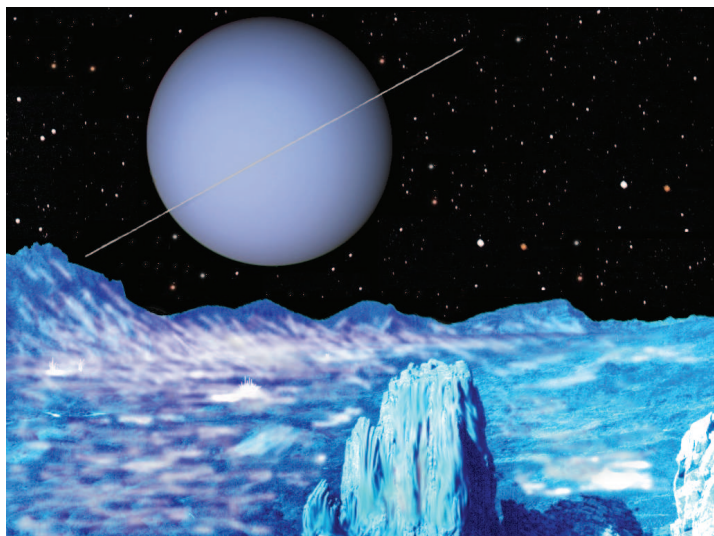


Figura 13: Urano desde Titania. (Imagen creada por Julio Solís)

Los sensores de la sonda espacial Voyager 2, que ha analizado la atmósfera desde las cercanías del planeta, la han escudriñado hasta una profundidad de 300 kilómetros por debajo de lo que se ha dado en llamar arbitrariamente “superficie”, o sea, el nivel de presión en su atmósfera de 1 bar, donde midió presiones de 100 000 hPa a una temperatura de $+50^{\circ}\text{C}$. Dicho análisis ha determinado la presencia de una troposfera desde los -300 kilómetros hasta los $+50$ kilómetros (tomando como referencia el nivel mencionado de 1 bar), donde la temperatura disminuye con la altitud hasta los -224°C en la tropopausa. No disponemos de datos observacionales a lo largo de un año completo uraniano, pero se da por hecho que existen cambios estacionales apreciables en su troposfera, que es la capa más dinámica de la atmósfera, con presencia de nubes de agua en los niveles inferiores, encontrando nubes de hidrosulfuro amónico, de amoníaco, sulfuro de hidrógeno, según se asciende, y nubes de metano llegando a la tropopausa.

Por encima tenemos la estratosfera, donde la temperatura aumenta con la altitud hasta valores de $+500^{\circ}\text{C}$ (4 000 kilómetros por encima de la “superficie”). En las zonas más bajas de la estratosfera se condensa una mezcla de hidrocarburos (metano, acetileno y etano) en forma de niebla o bruma que seguramente es la causante del aspecto liso que se aprecia desde el exterior. La capa más exterior es la termosfera/corona que se extiende hasta los 50 000 kilómetros, manteniendo de manera uniforme la temperatura alcanzada en la estratopausa.

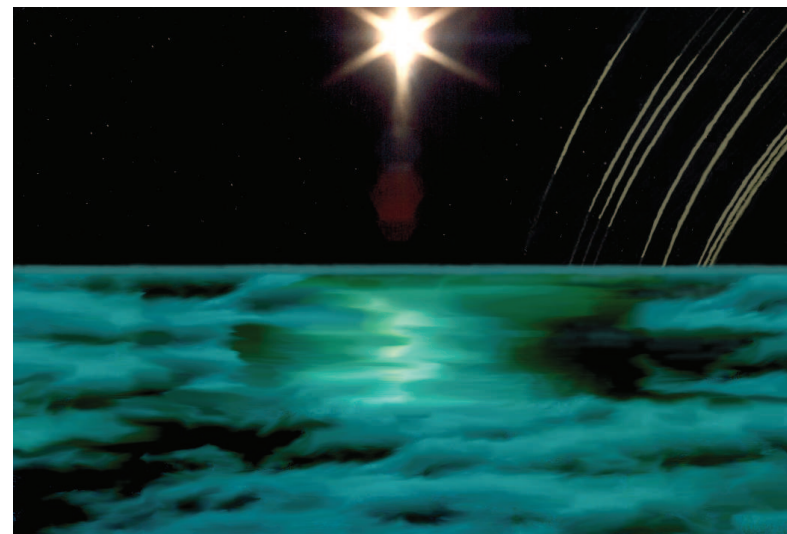


Figura 14: Vistas de Urano. (Imagen creada por Julio Solís)

Neptuno

Algo más pequeño que Urano pero muy parecido en aspecto y composición, también tiene anillos y un menor número de satélites (14), aunque entre ellos destaca Tritón, algo más pequeño que nuestra Luna, pero mayor que cualquiera de los de Urano. Tarda nada menos que 165 años en recorrer su órbita alrededor del Sol, y 16 horas en rotar sobre sí mismo.

Su estructura interna es similar a la de Urano, con un núcleo sólido rocoso envuelto por una especie de mar de agua y amoníaco, o mejor dicho un fluido de gran conductividad eléctrica, y sometido a grandes presiones, y temperaturas de varios miles de grados, y una atmósfera externa con nubes de hidrógeno, helio y metano. También se ha detectado un campo magnético que al igual que el de Urano está muy inclinado respecto a su eje de rotación.

Como está tan lejos del Sol (4 500 millones de kilómetros), recibe muy poco calor, por lo que se dan temperaturas en su “superficie” inferiores a los $-200\text{ }^{\circ}\text{C}$, a pesar de lo cual el planeta genera calor interno a expensas de una lenta contracción, de manera similar a Júpiter y Saturno, lo que sirve de motor para su dinámica atmosférica, dando lugar a manchas, remolinos, y estructuras nubosas muy activas, dispuestas también en bandas paralelas al ecuador que recuerdan más a los dos planetas gigantes citados que a Urano. Su atmósfera está compuesta por hidrógeno (84%), helio (12%), metano (2%), amoníaco (0,01%), y cantidades menores de etano y acetileno, mostrando un color azulado más intenso que el de Urano. La disociación del metano atmosférico, causada por la radiación solar, da lugar a una finísima bruma de hidrocarburos en las capas altas.

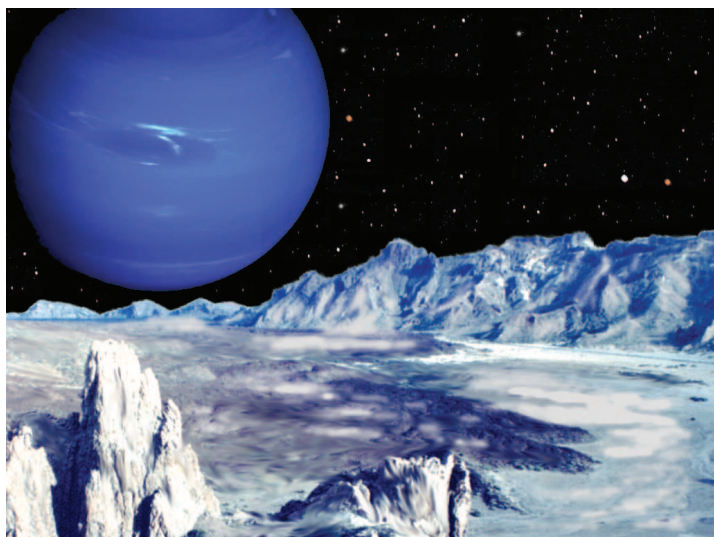


Figura 15: Neptuno desde la superficie de Tritón. (Imagen creada por Julio Solís)

La nave Voyager 2, cuando llegó a Neptuno en agosto de 1989, descubrió una gran mancha oscura, similar a la GMR de Júpiter, con un tamaño de $12\,000 \times 8\,000$ kilómetros, un gigantesco huracán que presenta formaciones nubosas tipo cirros de color blanquecino en sus bordes, donde se producen vientos de hasta $2\,000\text{ km/h}$, que le convierten en el planeta con los mayores registros de velocidad del viento. Se catalogaron varios remolinos ovalados más pequeños, con estructura similar, que mostraban las cimas de enormes nubes de carácter convectivo con capas de cirros de metano de color blanco, unos 50 kilómetros por encima. Estos “cirros” son descompuestos, mediante procesos fotoquímicos, en hidrocarburos más pesados que el

metano. Con todo, la atmósfera de Neptuno tiene mayor transparencia que la de Júpiter y Saturno.



Figura 16: Vistas de Neptuno. (Imagen creada por Julio Solís)

Bibliografía y consultas

J. Solís García, G. P. Arranz, 1994, “Cielos Exóticos”, Ed. Equipo Sirius.

R. Millar y W. K. Hartmann, 1983, “Viaje Extraordinario. Guía turística del sistema solar”, Ed. Planeta.

M. Marov, 1985, “Planetas del sistema solar”, Ed. Mir.

J. Solís García, G. P. Arranz, 1992, “Atmósferas y cielos planetarios”, *Tribuna de Astronomía* n° 80/81.

Investigación y Ciencia, colección TEMAS, números 12, 15, 26 y 45 (años 1998 a 2006).

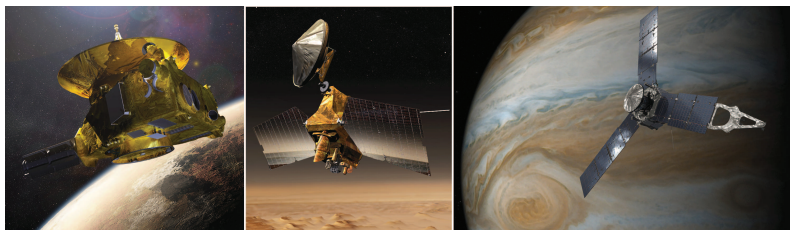
NASA <http://www.jpl.nasa.gov/>.

ESA <http://www.esa.int/ESA>.

<http://www.astromia.com>.

Exploración de las atmósferas de Plutón, Marte y Júpiter

El desarrollo científico y el trabajo de los investigadores no descansa, sobre todo en el ámbito planetario, ofreciéndonos continuamente descubrimientos en nuestro sistema solar con frecuencia sorprendentes y que nos van mostrando, con cada nuevo dato, un escenario más completo de nuestro sistema planetario. Desde la publicación del artículo **Las atmósferas de los planetas del sistema solar** en 2015 se han seguido haciendo progresos en el conocimiento de las atmósferas de los planetas del sistema solar y de algunos de sus mayores satélites, pero con el que se ha dado un paso de gigante ha sido con Plutón, planeta enano que parecía anodino y sin importancia, y que sin embargo, con la llegada de la nave *New Horizons* en el verano del año 2015, ha mostrado una dinámica geológica y meteorológica inesperada. De ser un débil puntito sin importancia en nuestro cielo, se ha puesto de manifiesto como un planeta complejo y dinámico, un mundo helado que ha mostrado una atmósfera más consistente de lo esperado y una dinámica superficial que no ha dejado de sorprender desde la llegada de los primeros datos e imágenes.



Representaciones artísticas de las sondas *New Horizons*, *Mars Reconnaissance Orbiter* y *JUNO*. (Cortesía de NASA)

En el momento de la publicación de este trabajo, Plutón no mereció la consideración suficiente como para incluirlo junto a los otros planetas del Sistema Solar con atmósferas significativas, pero seguramente a día de hoy merecerá un hueco entre sus hermanos mayores. En la dirección https://repositorio.aemet.es/bitstream/20.500.11765/5466/1/Astronomia_plutonSolis.pdf puede encontrarse un

breve artículo con los últimos descubrimientos sobre el planeta enano, a modo de anexo al trabajo que nos ocupa publicado en el Anuario del Observatorio Astronómico de Madrid para el año 2015.

Respecto al resto de cuerpos analizados en el artículo, no se han producido avances tan significativos respecto a lo que ya se conocía, salvo en el caso de Marte y Júpiter. Del primero se sigue estudiando y analizando la ingente cantidad de información suministrada por los últimos ingenios robotizados que escudriñan la superficie, el subsuelo y la atmósfera marciana, como el orbitador *Mars Reconnaissance Orbiter* de la NASA, con su sistema óptico de alta resolución HiRISE, y otros vehículos y plataformas que desde la superficie siguen enviando incansablemente datos, registros y fotografías de enorme valor. Respecto a Júpiter, recibió la visita exitosa de la sonda *JUNO* de la NASA durante el pasado verano (julio de 2016), después de cinco años de iniciada la misión. La sonda ha quedado en órbita y está previsto que ofrezca datos hasta febrero del año 2018. La comunidad científica tiene grandes esperanzas en esta misión, ya que *JUNO* será capaz de facilitar enorme información sobre la atmósfera de Júpiter, su campo magnético, sus auroras y su núcleo sólido. *JUNO* terminará su misión al penetrar en la atmósfera del gigante gaseoso desintegrándose durante la entrada, no sin antes habernos enviado durante su misión las imágenes de mayor resolución que hayamos visto nunca de Júpiter, incluso de sus zonas polares.

Julio Solís García, enero 2017

IMPACTOS SOBRE LA TIERRA

Pablo de Vicente Abad

*Observatorio Astronómico Nacional
Instituto Geográfico Nacional - Ministerio de Fomento*

Abstract

The Earth receives around 30 000 tons of material from meteoroids every year. Most of them cause meteor showers, some, fireballs and a few of them meteoroids that impact on the ground. The Earth is also subject to big impacts from Near Earth Objects larger than 100 m. In the past these impacts have probably carried life to our planet but also destruction and large extinctions of species. The possibility of large impacts that could put into risk our civilization causes great worry. Some programs are currently underway to detect, monitor and compute the trajectory of Potentially Hazard Objects. Two web pages list and classify the impact risk by Near Earth Asteroids.

There are some ideas to shift the trajectory of hazard asteroids but all of them require spacecraft technology currently not available. For the time being only a very few spacecraft missions have been launched to land on an asteroid or shoot a comet with a small impactor. In the future a big asteroid or comet will impact the Earth and by then, human beings should have developed and tested a technology to avoid it. It will probably not happen soon, but we should be ready.

Introducción

El 22 de febrero de 2012, un telescopio robótico del Observatorio de Mallorca detectó un nuevo asteroide de unos 45 metros de diámetro. La noticia saltó a los medios 12 días después y en algunos periódicos se

podieron leer entonces titulares advirtiendo de que un pequeño asteroide pasaría muy cerca de la Tierra un año después.

La posibilidad de que objetos procedentes de fuera de la atmósfera terrestre caigan sobre la superficie de la Tierra causa gran preocupación en la población. Los amantes de los comics saben que los invencibles y valientes galos de la aldea de Asterix sólo temen una cosa: que el cielo caiga sobre sus cabezas. Se trata de una preocupación comprensible. Dado que este asunto recibe bastante atención haremos un recorrido por lo que sabemos en la actualidad sobre impactos de objetos extraterrestres sobre nuestro planeta, comenzando por los más pequeños y finalizando por los de mayor tamaño. También describiremos como se estima el riesgo de colisión, los daños causados y finalmente las medidas que se pueden tomar para evitar impactos peligrosos.

Meteoros, meteoroides, bólidos y meteoritos

La palabra *meteoro* en castellano se utiliza para referirse a fenómenos atmosféricos, como la precipitación, suspensión o depósito de partículas sólidas y líquidas, produzcan o no efectos ópticos o eléctricos. Dentro de esta clasificación entran las nubes, la lluvia, el granizo, las heladas, las ventiscas, los truenos, las auroras y otros muchos eventos. Sin embargo la palabra *meteoro* en inglés sólo se refiere al fenómeno astronómico conocido vulgarmente como *estrella fugaz*. En astronomía, un *meteoro* es el fenómeno luminoso que produce un *meteoroides* cuando atraviesa la atmósfera terrestre. La mayor parte de los meteoros que se observan proceden de partículas del tamaño de un grano de sal o menores y su intensidad es considerable. Una partícula del tamaño de un guisante produciría una estrella fugaz espectacular y extraordinariamente brillante.

El nombre de *meteoroides* se aplica a las “pequeñas” partículas que pueblan el espacio interplanetario. La Unión Astronómica Internacional (UAI) considera meteoroides a un objeto sólido que se mueve en el espacio interplanetario y que es menor que un asteroide y mayor que un átomo. No se trata de una definición muy exacta, porque ambos límites son bastante difusos. Se ha propuesto que se consideren meteoroides los objetos mayores de 100 micras (por debajo se considera *polvo interplanetario*) y menores que 10 m. Sin embargo, recientemente se han descubierto y bautizado asteroides de 10 m de tamaño por lo que el límite superior se podría rebajar hasta 1 m. Todavía no existe acuerdo sobre estos valores.

La composición de los meteoroides no es uniforme y su densidad es muy variable. Los menos densos son porosos, compuestos de fragmentos de polvo con hielo y con densidades entre 0,3 y 0,8 g/cm³ (la mitad que la del hielo), y los más densos son rocas de hierro y níquel. La composición se puede estudiar a partir de su trayectoria en la atmósfera y del espectro de luz que emiten.

La Tierra recibe aproximadamente 30 000 toneladas de polvo interplanetario y meteoroides al año. Esto se debe a que en su movimiento de traslación alrededor del Sol colisiona con ellos. Algunos meteoroides son restos de cometas, fragmentos de asteroides o simplemente residuos de la formación del sistema solar. En una noche oscura y despejada, se pueden ver unos diez meteoros por hora habitualmente distribuidos temporalmente de manera irregular. El mecanismo que produce el fenómeno luminoso es la compresión del aire y la elevada velocidad de los meteoroides. Cuando un objeto viaja por la atmósfera el aire frente a él se comprime y se calienta. Si la velocidad del objeto es elevada, la temperatura del aire sube hasta miles de grados y se ilumina. Cuando el aire se calienta, la temperatura del meteoroides también lo hace, hasta que habitualmente se volatiliza. Por esta razón las estrellas fugaces suelen ser visibles durante un máximo de uno o dos segundos. Estos eventos ocurren en la termosfera, a partir de unos 120 km por encima de la superficie terrestre, donde la densidad es bastante baja pero suficiente para calentar y detener estas partículas. A una altura de 80 km la mayor parte de los meteoroides se ha volatilizado.

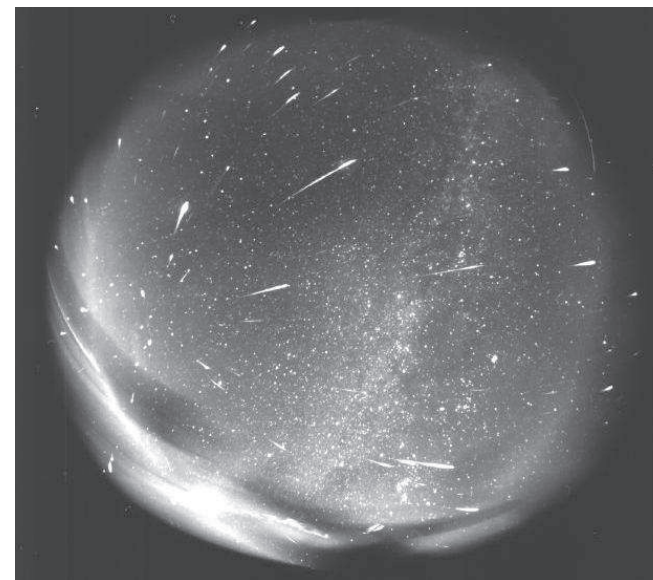


Figura 1: Lluvia de meteoros de las Leónidas en 1998. Se observaron 156 objetos durante 4 horas de tiempo de exposición. Fotografía de J. Tóth del observatorio de Modra.

La velocidad de los meteoroides oscila entre los 11 km/s, 30 veces la velocidad del sonido, y 72 km/s. Estas velocidades son una combinación

de la velocidad de traslación de la Tierra en torno al Sol (30 km/s), de la velocidad de los meteoroides en el espacio y la gravedad de la Tierra. Los meteoroides que proceden de cometas conservan parte de la velocidad de traslación del cometa progenitor y por ello pueden alcanzar los 42 km/s. Si el sentido de la trayectoria es opuesto al de traslación de la Tierra se obtienen velocidades máximas de 72 km/s en la atmósfera terrestre. La velocidad media de los meteoroides que atraviesan la órbita de la Tierra es de unos 20 km/s.

Cuando los meteoroides aparecen agrupados se les denomina *enjambres* y en ese caso, normalmente, están asociados a los restos de un cometa progenitor cuya órbita atraviesa la de la Tierra. Son ampliamente conocidos como *lluvias de estrellas* y se producen en ciertos días del año. En estos casos los meteoroides parecen proceder de un lugar del cielo conocido como *radiante*. Las lluvias de meteoroides reciben el nombre de la constelación en la que se encuentra el radiante. Una de las lluvias de estrellas más famosas son las Perséidas, cuyo máximo se produce entorno al 12 de agosto pero que se extienden unos días antes y después del máximo. Las Perséidas se originan por la traza del cometa 109P/Swift-Tuttle, que tiene un periodo de 133 años y cuya última aparición se produjo en 1992.

Los meteoros cuya magnitud aparente es superior a -4 (un número más negativo indica mayor brillo) reciben el nombre de *bolas de fuego* o *bóolidos*. Una magnitud de -4 es el límite para poder ver a simple vista un astro durante el día. El planeta más luminoso, Venus, puede llegar a tener magnitud $-4,9$. Por tanto los bóolidos son más brillantes que casi todos los planetas. Cuando la magnitud aparente del bólido es -14 o superior suelen producirse explosiones en el trayectoria generando sonidos audibles. Los superbóolidos, de magnitud -17 , son visibles incluso desde satélites en órbita en torno a la Tierra. El aspecto de un bólido es fácilmente reconocible, tiene aspecto ahusado con un brillo considerable en la cabeza y una traza en el cielo que puede durar unos segundos. A veces se observan nódulos brillantes a lo largo de la trayectoria. Obviamente los bóolidos se corresponden con objetos de tamaños superiores a los que producen los meteoros y son menos frecuentes. Suelen ser objetos individuales atravesando la atmósfera de la Tierra y por tanto no están asociados a lluvias de estrellas.

En los bóolidos la compresión del aire y su elevada temperatura, cerca de $5\,000\text{ }^{\circ}\text{C}$, hace que los objetos, además de calentarse intensamente, sufran aplanamiento por el frente de choque. En la mayor parte de los casos y dependiendo de la densidad del objeto este se fragmenta en centenares o miles de partes todas viajando a enormes velocidades dentro de la atmósfera y sufriendo un calentamiento enorme que las fracciona hasta que finalmente en décimas de segundo todas ellas se volatilizan. Esto es una explosión. La energía liberada por la explosión de un objeto de 1 m de tamaño es de varios centenares de toneladas de TNT (no llega al

kilotón). La altura en la atmósfera a la que se produce la volatilización del meteoroides es variable y depende de su composición. Así, los meteoroides con metales se volatilizan a alturas más bajas porque son más densos. La frecuencia de los bóolidos es difícil de estimar pero puede estar en unos 50 al año para objetos de 1 m y uno al año para objetos de 5 m.

Los meteoroides cuyo progenitor es un asteroide suponen sólo un 5 % del total; por tanto la mayor parte de estos son de origen cometario. De hecho se han observado bóolidos con magnitud -15 de origen cometario que no han generado meteoritos.



Figura 2: Un bólido sobre el estado de Nevada, Estados Unidos, el 22 de Abril de 2012. Imagen publicada por NASA en su página web y fotografiada por Lisa Warren.

Cuando el meteoroides es muy grande, superior a 10 m, algunos de los fragmentos que se originan pueden ralentizarse antes de explotar. Al decelerarse sobreviven a su paso por la atmósfera e impactan en el suelo. En estos casos reciben el nombre de *meteorito*. Los meteoritos que impactan contra el suelo habitualmente, entre unos pocos kg y 7 000 kg, reducen considerablemente su velocidad debido a la capa de aire que atraviesan, hasta alcanzar una velocidad terminal de unos pocos centenares de kilómetros por hora, entre 320 y 650 km/h. La velocidad terminal es resultado del equilibrio entre la atracción gravitatoria, que acelera el meteorito, y la viscosidad del aire, que lo frena, ya que la velocidad cósmica de entrada desaparece al atravesar la atmósfera.

Para que un bólido dé lugar a uno o varios meteoritos se tienen que cumplir varias condiciones:

- La magnitud del bólido debe ser superior a -8 o -10 .
- El progenitor del meteoroides debe ser un asteroide con material denso.

- El meteoróide debe entrar en la atmósfera con una velocidad relativamente baja.

Los meteoroides cuya masa es superior a 10 toneladas (diámetros superiores a 1,5 o 2 m), reducen su velocidad original fuera de la atmósfera (velocidad cósmica) hasta el 6 % y por tanto impactan el suelo con velocidades entre 1,2 y 4,3 km/s. Si la masa es 1 000 toneladas (diámetros entre 5 y 10 m), la velocidad cósmica sólo se reduce hasta el 70 % y si supera las 100 000 toneladas la velocidad de impacto es la velocidad cósmica, y por tanto es como si no existiera atmósfera.

La posibilidad de que un meteorito mate una persona es muy remota y de hecho hasta la fecha no se conoce ningún caso documentado. Hay casos registrados de personas que han resultado heridas por meteoritos como resultado de rebotes después del impacto original, pero nunca directamente alcanzadas. El caso más conocido fue el de Ann Hodges una estadounidense en Alabama que en 1955 sufrió el impacto de un meteorito de 4 kg de peso que atravesó el tejado de su casa. El meteorito le produjo abrasiones en la cadera y en una mano.

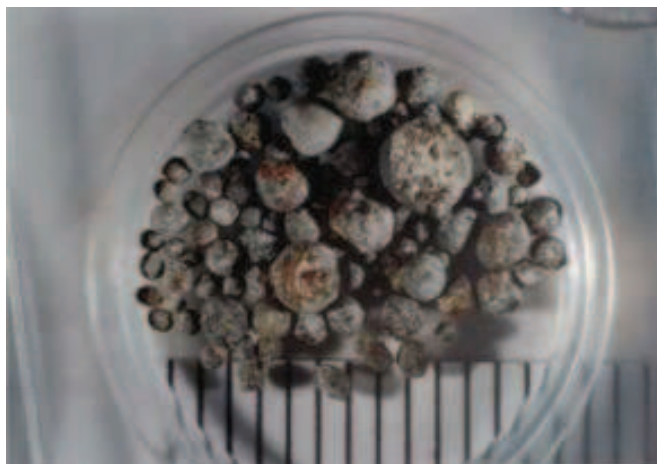


Figura 3: Cóndrulos de condrita Bjurböl. Imagen tomada de Wikimedia.

Cerca del 84 % de los meteoritos que caen a la Tierra son condritas, meteoritos rocosos no metálicos cuyos componentes no se han fundido a pesar de haber alcanzado temperaturas altas. Proceden de asteroides y son una muestra excelente para comprender la química primitiva del sistema solar y su edad, ya que, en muchos casos, la no fusión de sus elementos

los ha preservado tal y como se formaron hace miles de millones de años. Su principal constituyente son silicatos. Se caracterizan por estar formados por pequeñas esferas, conocidas como cóndrulos, de diámetros inferiores al milímetro formados por minerales diversos. Las condritas proceden de las primeras etapas de la formación del sistema solar, y su crecimiento se produjo a partir de granos de polvo de la nube molecular progenitora. Las condritas se clasifican en función del grado de alteración sufrido, por altas temperaturas o por acción del agua. El 4 % de las condritas caídas sobre la Tierra son carbonáceas, es decir contienen compuestos de carbono y se cree que se formaron en las partes más alejadas del sistema solar. Este tipo de condritas son especialmente importantes porque contienen agua, o elementos alterados por ella, y compuestos orgánicos como aminoácidos. Se cree que los compuestos orgánicos en el sistema solar son abundantes y pudieron jugar un papel importante en la aparición de la vida en la Tierra.

El 16 % restante de los meteoritos procede de asteroides diferenciados, es decir que han sufrido fusión en su interior. En estos el hierro y otros metales pesados se concentra en el interior, y los basaltos y silicatos en el exterior. Existe un grupo extraordinariamente pequeño y raro de meteoritos que proceden de la Luna, otros planetas rocosos y sus lunas, que se habrían formado por el impacto de un gran meteoróide sobre su superficie y han viajado hasta la Tierra. Hace años surgió una gran polémica con el meteorito ALH 84001 procedente de Marte porque se observaron estructuras que se interpretaron como residuos fosilizados de bacterias. Todavía no se ha confirmado ni desmentido el origen biológico de dichas estructuras.

El estudio de los bólidos resulta muy interesante porque si su trayectoria se registra desde diferentes puntos del cielo se puede determinar su órbita y el origen que da lugar al meteoróide. En el año 1997 se estableció en España la Red de Investigación sobre Bólidos y Meteoritos (SPMN). Se trata de un proyecto interuniversitario, en el que intervienen astrónomos profesionales y aficionados, que realiza observaciones continuas de actividad meteórica que ha dado lugar a resultados importantes. En este momento cuentan con catorce estaciones con cámaras CCD¹ de todo el cielo o/y cámaras de vídeo para la detección de meteoros y bólidos que cubren todo el territorio español. Los resultados más interesantes de los últimos años han sido el estudio del meteorito de Villalbeto de la Peña el 4 de enero de 2004 y el de Puerto Lápice el 10 de mayo de 2007. La SPMN determinó a partir de la descripción de los testigos, vídeo y fotografías la órbita del primero y el origen del segundo (el asteroide Vesta).

Las pocas órbitas que se han determinado de los meteoritos caídos hasta la fecha en la Tierra indican que todos ellos proceden del cinturón de asteroides entre las órbitas de Marte y Júpiter.

¹Dispositivo electrónico que capta la luz con sensibilidad y permite formar una imagen.

Cometas y asteroides: origen de meteoroides y meteoritos

Como ya hemos comentado en la sección anterior la mayor parte de los meteoroides son de origen cometario y la mayor parte de los meteoritos son de origen asteroidal, por lo que resulta interesante saber algo más sobre sus progenitores.

Los cometas son cuerpos menores compuestos de hielo y rocas que orbitan en torno al Sol en diferentes trayectorias. Muchas de las trayectorias son elípticas muy excéntricas por lo que se acercan mucho al Sol y tienen un periodo muy largo. Los cometas se ven en el cielo con aspecto difuso y una cola de menor brillo en dirección opuesta al Sol. La atmósfera que envuelve el núcleo se llama coma o cabellera.

Los cometas siempre han llamado la atención del hombre por su súbita aparición, corta permanencia, aspecto característico y por su facilidad de identificación en el cielo nocturno. Antes de la existencia de los telescopios se registraban unos 10 cometas por siglo. Casi siempre estuvieron asociados a creencias supersticiosas y habitualmente eran signos de mal augurio.



Figura 4: Imagen del cometa Halley tomada por el observatorio de Yerkes en 1910.

En 1705 Edmund Halley utilizando la teoría de la gravitación, recientemente formulada por Isaac Newton, publicó un trabajo sobre cometas una de cuyas mayores contribuciones fue una tabla con las órbitas de 24 cometas registrados entre 1337 y 1698. De ellos 22 correspondían a objetos

diferentes y uno se repetía 3 veces. Predijo que se trataba de un cometa con un periodo de 76 años que, cuando se confirmó su siguiente aparición una vez que Halley ya había muerto, recibiría su nombre.

En la actualidad (junio de 2012) se han clasificado unos 4 557 cometas, de los que unos centenares son de periodo corto y unos 1 500 son los conocidos como “rascasoles”, por acercarse mucho al Sol. Se cree que el periodo de algunos cometas se acorta por la influencia gravitatoria de los planetas gigantes. Esta modificación en la trayectoria se suele producir si el cometa pasa “cerca” del planeta. La UAI ha establecido un método para denominarlos, compuesto de un número que indica el año del descubrimiento, una letra que indica el medio mes de dicho descubrimiento y un segundo número que indica el número de orden del descubrimiento. Todo ello va precedido de una letra que especifica el tipo de cometa: P de periódico o C de no periódico. Si el cometa es periódico, y tras su segundo paso por el perihelio, se coloca un número delante de la letra indicando el número de orden del descubrimiento. Por ejemplo el cometa Halley sería, según esta denominación, 1P/1682 Q1. Los cometas más famosos se suelen denominar sin el año de descubrimiento, como por ejemplo el 109P/Swift-Tuttle que mencionamos anteriormente como progenitor de las Perséidas.

En 1860 W. Huggins, también británico, comenzó el estudio de la química de los cometas al obtener por primera vez un espectro de varios de ellos. Unos años más tarde se descubrió la relación entre la órbita de los cometas, los restos dejados por ellos y las lluvias periódicas de meteoros. En 1950, el astrónomo norteamericano F. Whipple, propuso un modelo de cometa en el que éste estaba compuesto por un núcleo de hielo y rocas en rotación en torno a sí mismo. Esta hipótesis explica la presencia del halo y la cola y la aceleración y deceleración en función de la rotación en torno a sí mismo. El núcleo del cometa está rodeado de una atmósfera llamada cabellera que resulta de la sublimación de sus materiales a medida que se acerca al Sol. La presión de radiación del Sol, es decir sus fotones, hace que restos de polvo y rocas del cometa se separen de éste ligeramente, formando una cola de polvo que se sitúa casi a lo largo de su órbita. Esta cola es visible porque las partículas que la componen reflejan la luz del Sol. Debido al tamaño de los granos de polvo la luz reflejada está levemente enrojecida. La cabellera sufre la acción del viento solar y genera otra cola formada por gas ionizado, que apunta siempre en la dirección opuesta al Sol y suele tener un color azulado. Estas dos estructuras, el gas ionizado y el polvo y rocas forman las dos colas de un cometa. A veces las dos colas están claramente diferenciadas. La cola ionizada puede llegar a medir decenas de millones de kilómetros, la cabellera un millón de kilómetros y su núcleo unos pocos kilómetros.

La primera misión espacial a un cometa fue la International Comet Explorer (ICE) que pasó cerca del cometa 21P/Giacobini-Zinner en 1985 y realizó observaciones de 1P/Halley en 1986. Hubo 5 misiones (dos

japonesas, dos rusas y una europea, la Giotto) que exploraron el 1P/Halley entre los años 1992 y 1998. En 1998 se lanzó la misión Deep Space 1 que pasó junto a los cometas 107P/Wilson-Harrington y 19P/Borelly en 2001. Wilson-Harrington es un objeto interesante pues últimamente se comporta más como un asteroide que como cometa. 19P/Borelly es un cometa de muy corto periodo (6,8 años).



Figura 5: Las dos colas del cometa Hale-Bopp: la azulada procede de gas ionizado, la blanca procede del polvo. Imagen de APOD (Astronomy Picture of the Day) publicada en diciembre de 2000. La imagen fue tomada por John Gleason en marzo de 1997.

La Agencia Espacial Europea (ESA) lanzó la misión Rosetta en 2004 con el objetivo de situarse junto al cometa 67P/Churyumov-Gerasimenko y enviar un módulo que aterrice sobre su núcleo en 2014. La misión continuará estudiando el cometa hasta 2015, año en el que se producirá su paso por el perihelio.

La NASA lanzó en 1999 y en 2005 las misiones espaciales Stardust y Deep Impact, para estudiar los cometas 81P/Wild 2 y 9P/Tempel 1 respectivamente. La primera situó una sonda en la cabellera del cometa y capturó muestras de material que luego trajo a la Tierra. La segunda misión consistió en el lanzamiento de un proyectil sobre el cometa para

luego observar la composición del gas y polvo que se originó. A partir de este experimento se consiguió la primera evidencia de que la superficie del cometa estaba cubierta por una fina capa de hielo. Así mismo también se observó la presencia de material orgánico que se supone estuvo presente en la mezcla prebiótica de la Tierra. La misión Stardust también permitió descubrir glicina, un aminoácido utilizado por los seres vivos para crear proteínas.

La presencia de compuestos orgánicos y agua en cometas permite aventurar que los cometas jugaron un papel fundamental en la aparición de la vida en la Tierra y que la vida en el universo puede ser más común de lo que sospechamos.

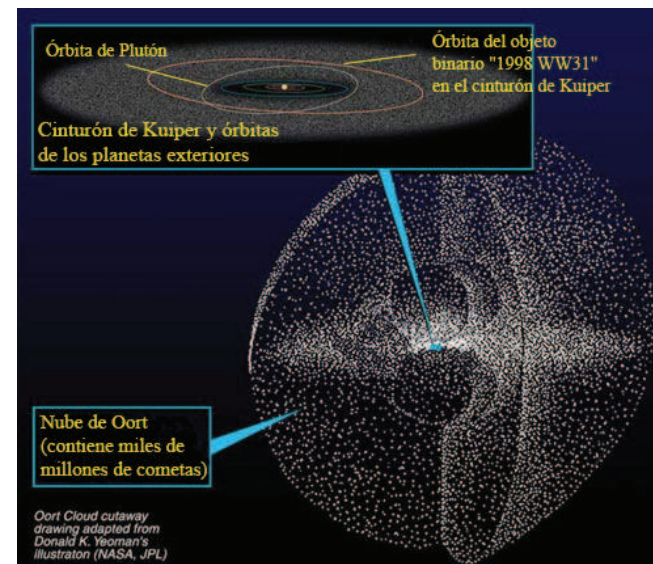


Figura 6: Representación artística de la nube de Oort y el cinturón de Kuiper, por D.K. Yeoman de NASA/JPL.

Se cree que los cometas de periodo largo (más de 200 años) se originan en la Nube de Oort, una nube esférica en torno al Sol a una distancia entre 2 000 y 100 000 unidades astronómicas, en el exterior del sistema solar. Se sabe que las órbitas de los cometas son inestables por su pérdida de material y la influencia gravitatoria de los planetas del sistema solar, y por tanto no se han podido formar en dichas órbitas. El astrónomo holandés Jan Oort postuló la existencia de un repositorio donde se originan, en la parte externa del sistema solar, que recibe el nombre de nube de Oort. Se cree que la nube de Oort podría albergar miles de millones de cometas, y que se

trata de un remanente del anillo protoplanetario que se formó alrededor del Sol hace 4 500 millones de años. Los cometas de periodo corto se originan en el cinturón de Kuiper, situado en la eclíptica y más allá de Plutón, entre 30 y 100 unidades astronómicas.

Los asteroides son cuerpos menores, de tamaño menor que los planetas, en órbita alrededor del Sol, habitualmente situados en la parte interior del sistema solar. Los cuerpos menores más allá de la órbita de Júpiter suelen recibir el nombre de centauros², troyanos neptunianos, y objetos transneptunianos. La mayor parte de los asteroides conocidos están situados en el cinturón entre Marte y Júpiter. Se trata de objetos que no lograron crear un planeta durante la formación del sistema solar. Estos cuerpos son acelerados por la gravedad de Júpiter y su alta velocidad impide que tras una colisión se agreguen. Sin la presencia de Júpiter, las colisiones serían más lentas permitiendo agregamientos de los trozos y probablemente, tras un tiempo, la formación de un cuerpo mayor. Algunos fragmentos, resultado de las colisiones entre asteroides, se desplazan con el tiempo hacia los planetas interiores. Los asteroides cercanos a la Tierra o NEAs constituyen una familia especialmente importante por su probabilidad de impactar con ella como veremos más adelante.

Los troyanos son asteroides asociados a planetas que no colisionan con ellos porque orbitan en torno a los puntos de Lagrange L4 y L5. Los puntos de Lagrange son lugares de equilibrio gravitatorio en la órbita de un planeta, de modo que los objetos en dichos puntos orbitan en torno al Sol manteniendo la misma posición relativa respecto del planeta. Existen cinco puntos de Lagrange en cada órbita, y L4 y L5 son puntos de equilibrio estable.

Los asteroides se pueden clasificar según su composición en carbonáceos, rocosos y metálicos. Los asteroides grandes, con un diámetro promedio superior a 100 km, han sufrido diferenciación química. Es decir, su interior se fundió y permaneció así durante millones de años, permitiendo que el hierro y el níquel se desplazaran al centro y los materiales menos densos hacia el exterior. Se estima que existen entre 1 y 2 millones de asteroides con tamaños superiores a 1 km. Los asteroides compuestos de hierro de tamaños inferiores a 1 km suelen proceder de colisiones entre dos asteroides, de los que uno al menos ha sufrido diferenciación química.

Ha habido algunas misiones espaciales dedicadas (parcial o completamente) a estudiar algunos asteroides. Las primeras se limitaron a fotografiarlos, en su viaje hacia otros destinos, como la Mariner 9, que fotografió las lunas de Marte, Phobos y Deimos, probablemente dos asteroides capturados por el planeta. Las formas de estos se revelaron irregulares, como las de las patatas. El primer asteroide verdadero fotografiado fue el 951

²Los centauros tienen órbitas inestables que cruzan o han cruzado las de los planetas gigantes.

Gaspra por la sonda Galileo en su viaje hacia Júpiter en 1991. Dos años más tarde pasó junto al asteroide 243 Ida.



Figura 7: El asteroide 951 Gaspra desde la sonda Galileo en octubre de 1991. Imagen de NASA.

La primera misión específicamente diseñada para ir a un asteroide fue la Near Shoemaker de NASA en 1997. Esta misión voló a 1 200 km del asteroide 253 Mathilde, más tarde orbitó en torno a 433 Eros y aterrizó con éxito sobre su superficie. En el año 2005 la misión japonesa Hayabusa aterrizó sobre el asteroide 25143 Itokawa y tomó muestras que trajo de vuelta a la Tierra.

Otras misiones posteriores han sido Rosetta, ya mencionada antes, de la ESA, que visitó en 2008, el asteroide 2867 Steins y en 2010, 21 Lutetia, el segundo mayor asteroide visitado hasta la fecha. La misión Dawn de NASA llegó en julio de 2011 a 4 Vesta, el mayor asteroide visitado hasta la fecha. Vesta tiene una forma casi esférica pero irregular, con un diámetro promedio de 525 km. Dawn ha estudiado Vesta durante un año. Se sabe que está compuesto por un núcleo de hierro y su superficie está cubierta de cráteres. Dos de ellos, situados en el polo sur, son especialmente grandes y como resultado del choque que los originó se produjeron ondulaciones en la superficie que recorren la superficie del asteroide. Dawn ha abandonado Vesta en julio de 2012 en dirección a Ceres, un planeta enano de hielo y rocas. Con un diámetro promedio de 950 km, es el cuerpo más grande del cinturón de asteroides. La llegada de la sonda se producirá en 2015.

Se han programado dos misiones espaciales que en el futuro irán a asteroides para tomar muestras que serán enviadas de vuelta a la Tierra. Se trata de la misión japonesa Hayabusa 2 prevista para 2015 y la misión OSIRIS-REx de NASA para el año 2016.

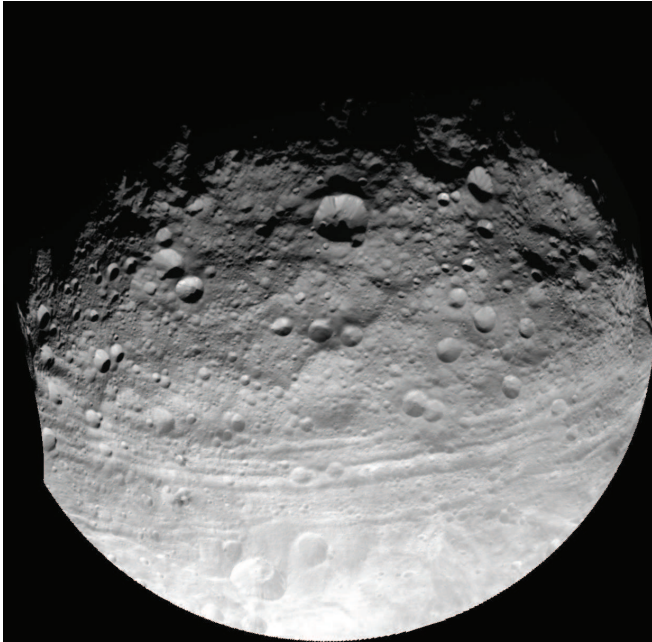


Figura 8: El asteroide Vesta desde la sonda Dawn de NASA en 2011. En la superficie se pueden apreciar cráteres y en el ecuador ondulaciones, probablemente consecuencia de un choque con un asteroide menor contra uno de sus polos. Imagen de NASA.

Grandes impactos conocidos: cráteres en la Tierra

La Tierra desde su formación ha sufrido millones de impactos de meteoritos sobre su superficie. Durante la formación del sistema solar el bombardeo fue continuo y, si la hipótesis del Gran Impacto es correcta, un evento fue especialmente relevante. Hace unos 4 500 millones de años un cuerpo del tamaño de Marte colisionó con la Tierra. Los fragmentos vaporizados se insertaron en la órbita de la Tierra, quedaron atrapados por su campo gravitatorio y se agregaron para formar la Luna. El calor generado produjo una salida de gases de la corteza terrestre que pudo resultar decisiva en la formación de la atmósfera e hidrosfera del planeta.

Tras la formación de la Luna hay un periodo que va desde hace 4 100 millones de años hasta hace 3 800 millones de años conocido como el Gran Bombardeo Tardío, en el que se produjeron numerosos cráteres de impacto en la Luna. Por inferencia se cree que lo mismo ocurrió en Mercurio, Venus, la Tierra y Marte. De hecho se estima que el número de impactos en la Tierra es superior al recibido por la Luna debido a la diferencia de tamaño y masa entre ambas. Es probable que este periodo esté asociado al desplazamiento de las órbitas de los planetas gigantes gaseosos hacia el Sol. En su desplazamiento dispersaron cuerpos del cinturón de asteroides y del cinturón de Kuiper que adquirieron órbitas excéntricas que atravesaban las de los planetas interiores y dieron lugar al bombardeo.

Tras ese periodo aparece la vida en la Tierra. De hecho los primeros fósiles están datados hace 3 500 millones de años, pero se tienen evidencias de que la vida surgió antes. Durante el bombardeo de cometas y asteroides la temperatura en la superficie era tan elevada que impedía la formación de suficiente agua y moléculas orgánicas para que prosperase la vida. Se piensa que tras el Gran Bombardeo Tardío los impactos continuaron de modo menos intenso y que los cometas suministraron agua y moléculas orgánicas a la Tierra.

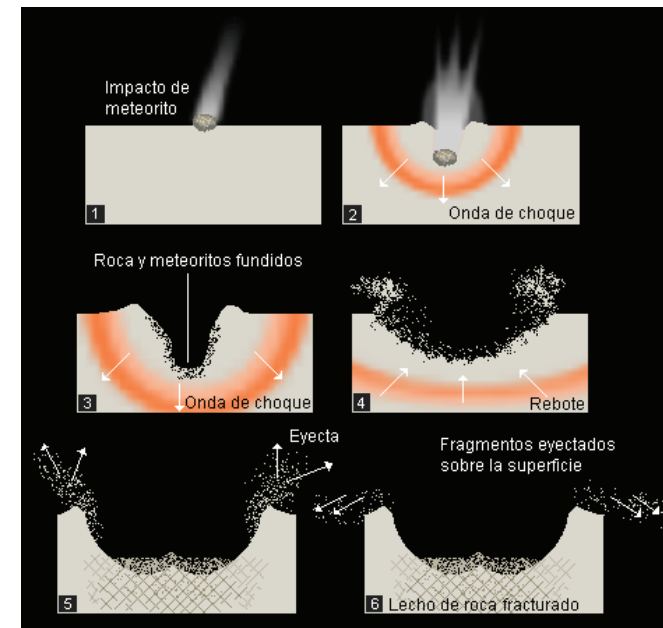


Figura 9: Secuencia en la formación de un cráter de impacto. Imagen de L.M. Benítez tomada de Wikipedia.

Cuando un meteorito o asteroide impacta sobre la superficie de un planeta produce un cráter. Se forma una onda de choque sobre el terreno y un agujero en cuya superficie se funden las rocas y el meteorito. La onda de choque rebota y eyecta material por los bordes elevando estos hasta generar un pared. En algunos casos, en los cráteres superiores a 4 km de diámetro, se crean un pico o anillos en el centro y el lecho de roca bajo el cráter se fractura.

La Tierra es un planeta geológicamente activo y esa dinámica ha borrado las huellas de los impactos. A pesar de ello se tienen evidencias de algunos pocos cráteres en su superficie. En la actualidad se mantiene una base de datos (Earth Impact Database) que clasifica los cráteres de impacto en confirmados, probables con diferentes grados de probabilidad y descartados. En la actualidad hay 178 cráteres confirmados y unos 650 probables. La antigüedad de los cráteres confirmados varía desde hace 2400 millones de años hasta la actualidad y con diámetros entre los 160 km y los pocos metros. El mayor cráter confirmado es el Vredefort en Sudáfrica, de 160 km de diámetro entre su bordes y una antigüedad de 2023 millones de años. Fue causado por un asteroide de 5 a 10 km de diámetro del tipo de las condritas. El segundo cráter confirmado en tamaño es el Chicxulub, en la península del Yucatán con un diámetro de 150 km y una antigüedad de 65 millones de años. Se trata de un cráter muy famoso sobre el que nos extenderemos en la siguiente sección.

Uno de los cráteres más icónicos es el Cráter Barringer en Arizona, Estados Unidos. Se produjo por el impacto de un asteroide de hierro-níquel de unos 45 m de diámetro hace 50 000 años. Todavía existe hierro de origen meteórico mezclado con roca en sus paredes. Se trata probablemente del cráter mejor conservado en la superficie terrestre, a pesar de que se formó un lago en su interior y su profundidad decreció con la sedimentación. Tras el final de la glaciación, el clima en la zona cambió (originalmente era un bosque) y se secó preservando su existencia.

El cráter de más reciente formación confirmado es el de Carancas en Perú de 13 m de diámetro y 4,5 m de profundidad. Se formó en el año 2007 tras el impacto de un meteorito de tipo condrita de unos 3 m de tamaño. El cráter hizo aflorar agua con arsénico que produjo efectos nocivos en la población cercana.

Finalmente mencionaremos también el evento de Tunguska acaecido el 30 de junio de 1908 en Siberia. Se trató de una explosión aérea de unos 10 a 15 megatones a unos 8 km de altura producida por un bólido de unos 50 a 80 m. Incendió y derribó árboles en un área de unos 50 km de diámetro. Sólo los árboles que estaban en la vertical de la explosión quedaron en pie. Originalmente se propuso que la explosión se produjo por la desintegración de un cometa compuesto de hielo y roca. Sin embargo en la actualidad, y tras comprobar que los modelos numéricos predicen que un cuerpo de baja densidad con esa energía cinética no puede atravesar la troposfera, la

hipótesis más aceptada es que fue un cuerpo de origen asteroidal el causante de la explosión.



Figura 10: Imagen aérea del Cráter Barringer en Arizona, Estados Unidos.

La tasa de formación de cráteres depende del flujo de asteroides sobre la Tierra. Existen dos métodos para calcularlo. El método astronómico consiste en medir el número y brillo de asteroides cercanos, obtener el diámetro de estos cuerpos a partir de su brillo y calcular la fracción de ellos que cruza la órbita y posición de la Tierra en un intervalo de tiempo. El método geológico consiste en utilizar el número, tamaño y edad de los cráteres en una área determinada de la Tierra. Estableciendo una relación entre el tamaño del cráter y el diámetro del proyectil se puede obtener el flujo de estos últimos. También ha sido habitual extrapolar el recuento de cráteres en la Luna y su ritmo de formación, y obtener un valor para la Tierra.

Hasta hace poco se pensaba que el ritmo de formación de cráteres en la Luna era probablemente constante desde hace 3 000 millones de años y con bastante certeza desde hace 400 millones de años. Sin embargo recientemente, y utilizando el método geológico descrito más arriba, se han obtenido conclusiones diferentes estudiando los cráteres con una antigüedad inferior a 500 millones de años:

- La tasa de formación de cráteres en la Tierra cambió hace 125 millones de años aumentando bruscamente en un factor 2 a 4. Se ha postulado que dicho aumento se produjo como consecuencia de una

colisión entre dos cuerpos de 60 y 160 km de diámetro promedio en el cinturón de asteroides hace 160 millones de años. Los fragmentos se desplazaron lentamente hasta la órbita de la Tierra aumentando el flujo de impactos durante 100 millones de años. El máximo se produjo 40 millones de años después de la colisión. Este aumento del flujo de asteroides es el origen más probable del asteroide que generó el cráter Chicxulub.

- La atmósfera es un excelente escudo para detener proyectiles de tamaños inferiores a 80 m. La protección es parcial para proyectiles entre 80 m y 800 m de tamaño.
- El método astronómico subestima el número de cráteres de diámetros superiores a 10 km y antigüedad inferior a 125 millones de años en un factor 4 y sobrestima el número de cráteres de tamaños inferiores a 4 km en un factor 10.

La figura muestra el número de cráteres formados en los últimos 125 millones de años en áreas seleccionadas de la Tierra en función del diámetro del cráter y del proyectil.

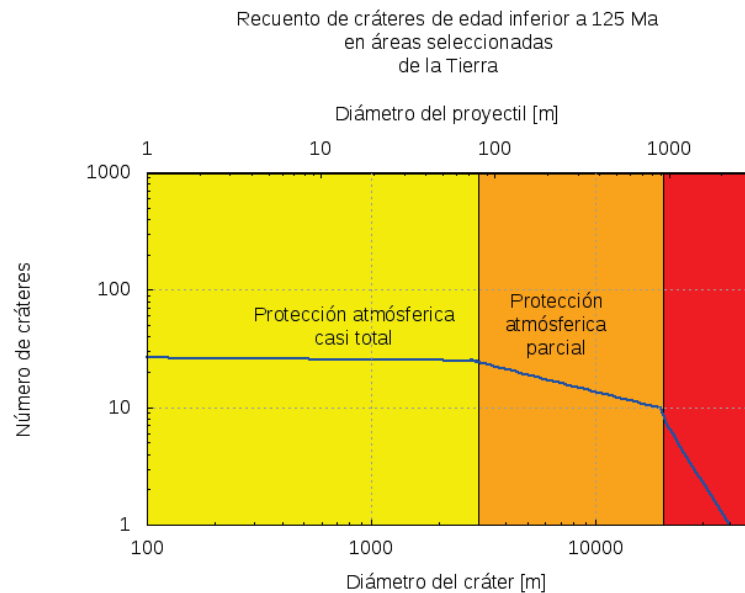


Figura 11: Número de cráteres de edad inferior a 125 millones de años en áreas seleccionadas de la Tierra. La curva de la izquierda es casi plana por el efecto protector de la atmósfera. Las curvas son aproximaciones a los valores reales (discretos).

Chicxulub: la extinción de los dinosaurios

Desde hace tiempo se conoce la existencia de una fina capa de 1 cm de grosor llamada Cretácico-Terciario (K-T en inglés) en los sedimentos correspondientes al fin de la era Mesozoica, hace 65 millones de años, fecha que coincide con la extinción masiva de los dinosaurios. En 1980 L. Álvarez, W. Álvarez, F. Asaro y H. Michels descubrieron que dicha capa tenía un alto contenido en iridio. Dado que dicha concentración es mucho mayor en los asteroides que en la Tierra, propusieron que se debía al impacto de un asteroide cuyos restos se extendieron por toda la superficie terrestre. Durante 12 años la hipótesis fue muy discutida, pero al cabo de este tiempo se descubrió la existencia de un cráter (Chicxulub) en la península del Yucatán con un diámetro similar al estimado por L. Álvarez y colaboradores. Desde su descubrimiento la explicación del impacto ha sido considerada la más plausible y se cree que el cráter Chicxulub es la prueba y la huella de ese evento apocalíptico.

Hace 65 millones de años un asteroide de 10 km de diámetro, ligeramente más grande que el Everest, se aproximó a la Tierra. Al entrar en la atmósfera creó un inmenso frente de choque calentando la atmósfera en muchos kilómetros a la redonda. Su luminosidad era probablemente similar a la del Sol e incendió y comprimió brutalmente todo aquello que estaba bajo él. El asteroide cayó al mar, en el golfo de México, y parte de su energía cinética se empleó en transformar el agua en vapor. Como la profundidad en el golfo de México es pequeña, la velocidad no disminuyó apreciablemente y el asteroide golpeó fuertemente la plataforma continental.

La colisión con la corteza terrestre produjo una tremenda explosión que envió rocas fundidas y vapor de agua a varios kilómetros de altura y distancia. La columna que se originó brillaba tanto como el Sol. Además el impacto produjo un intenso terremoto que provocó la muerte de los seres vivos en un radio de centenas de kilómetros. La onda de choque en el aire se propagó a gran velocidad creando un sonido atronador que posiblemente dejó sordos a muchos animales. Finalmente el choque con el mar produjo un gigantesco tsunami de centenas de metros de altura y una velocidad de 600 km/h que arrasó las costas cercanas y se introdujo en el interior devastando todo a su paso.

Todos los residuos que saltaron despedidos hacia arriba lo hicieron a unas velocidades enormes, de varios kilómetros por segundo. Al volver a caer se calentaron y se convirtieron en una lluvia de fuego en un radio de miles de kilómetros alrededor. Prácticamente todo el planeta se incendió. Debido a los incendios toda la atmósfera se oscureció durante un periodo de varios meses produciendo después un descenso dramático de la temperatura en lo que se conoce como *invierno nuclear*.

El impacto se produjo sobre una zona arcillosa con minerales. El frente de choque del impacto creó nitratos que a su vez formaron ácido nítrico

que cayó sobre la superficie terrestre. También cloro y otros compuestos químicos del asteroide ascendieron hasta las capas altas de la atmósfera destruyendo la capa de ozono. Se produjo la desaparición del 75 % de la vida marina y terrestre. Supuso el final de los dinosaurios y la oportunidad para el florecimiento de los mamíferos.

Este escenario apocalíptico lo produjo un cuerpo menor, de sólo 10 km de diámetro. En el espacio interplanetario existen objetos mayores y en algún momento de la historia de la Tierra volverán a impactar sobre ella. Es simplemente una cuestión de tiempo.

NEOs y los programas de observación

Existe un documental televisivo que recrea la caída del asteroide que provocó el cráter Chicxulub y los acontecimientos posteriores hasta un año después. Su visión resulta muy instructiva y a su finalización uno se pregunta si, en el caso de una colisión similar en el futuro, la humanidad tendría una oportunidad para sobrevivir a ese cataclismo o mejor aún, si se podría evitar. ¿Qué podemos hacer? Obviamente lo primero es investigar; y esto ya ha comenzado. Sabemos mucho más de lo que conocíamos hace 40 años. Después es necesario observar cuerpos potencialmente peligrosos y catalogarlos. Finalmente hay que desarrollar una tecnología que desvíe los asteroides más peligrosos antes de que impacten sobre la Tierra.

En la actualidad existen varios programas de observación de cuerpos menores cercanos a la Tierra, los NEOs. NEO es un acrónimo cuyo significado es Near Earth Object (Objeto cercano a la Tierra). Los NEOs son asteroides o cometas desestabilizados por otros planetas cuyas órbitas pasan cercanas a la trayectoria de la Tierra. Como ya se mencionó en la sección sobre cometas y asteroides, son cuerpos menores que se pueden considerar restos del proceso de formación del sistema solar y cuya investigación proporciona mucha información sobre su origen. La mayor parte de los NEOs son NEAs (Near Earth Asteroids), asteroides cercanos a la Tierra, y sólo una pequeña parte son cometas. La UAI considera NEAs a los cuerpos menores cuya distancia al Sol es inferior a 1,3 unidades astronómicas (UA) y por tanto a 0,3 UA de la Tierra. Los NEAs se dividen en tres grupos:

- Los tipo Atón, cuyo semieje orbital es inferior al de la Tierra, y ocasionalmente cruzan su órbita.
- Los tipo Apolo, con semiejes orbitales superiores a la Tierra pero que pueden cruzar su órbita. Potencialmente son los más peligrosos.
- Los tipo Amor, cuyo semieje orbital es superior al de la Tierra y menor que el de Marte.

Los PHA (Potentially Hazard Asteroids), acrónimo que significa Asteroides Potencialmente Peligrosos, son un subgrupo de los NEAs, cuyas órbitas son similares a la de la Tierra (fundamentalmente son asteroides tipo Apolo), y son tan grandes que la atmósfera de nuestro planeta no puede actuar de escudo volatizándolos. En caso de impacto producirían daños locales, regionales o globales sobre la superficie de la Tierra. La distancia a partir de la cual un NEA es considerado un PHA es cuando su órbita le acerca a la Tierra a una distancia inferior a 7,5 millones de kilómetros, unas 20 veces la distancia a la Luna. El diámetro mínimo a partir del que se considera peligroso es 140 m.

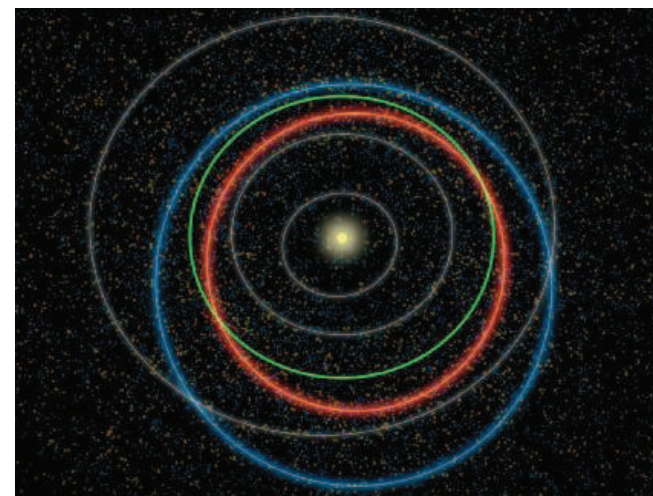


Figura 12: Órbitas de un NEO en azul y un PHA en naranja. La órbita de la Tierra está pintada en verde. La órbita del PHA cruza la de la Tierra dos veces por año. También se han pintado otros NEOs que aparecen como puntos en la imagen. Imagen propiedad de NASA.

De acuerdo con la UAI, con fecha de agosto de 2012, se han descubierto 892 NEAs con diámetros superiores a 1 km, lo que supone un 92 % de una población estimada en 966 con estas características. De los NEAs antes mencionados, 158 se han clasificado como PHAs. Se conocen 8 101 NEAs con diámetros inferiores a 1 km, de los que 1 167 son PHAs.

La mejor estimación del número de PHAs hasta la fecha (mayo de 2012) la ha realizado la misión espacial WISE de la NASA que, a partir de observaciones en el infrarrojo, ha proporcionado datos de un centenar de PHAs. Esta pequeña muestra permite realizar extrapolaciones a toda la población de PHAs existentes. Se cree que existen unos 4 700 PHAs

(1 500 arriba o abajo). Hasta la fecha se han detectado con los programas de observación entre un 20 % y un 30 % de estos objetos. WISE también ha permitido descubrir que existen más PHAs en órbitas con muy poca inclinación respecto de la de la Tierra de lo que se pensaba anteriormente. Probablemente esto se deba a que muchos de estos objetos proceden de colisiones entre dos asteroides en el cinturón entre Marte y Júpiter. Los cuerpos más masivos están más cerca de la eclíptica y sus fragmentos permanecen en órbitas de poca inclinación.

El cuadro siguiente resume el número de NEAs estimados y descubiertos hasta el 12 de agosto de 2012, clasificados por tamaño e incluye el porcentaje de NEAs descubiertos respecto de los estimados.

NEAs descubiertos, estimados y fracción en función del diámetro			
diámetro	observados	estimados	observados/estimados
>1 km	892	966 ± 45	92 %
1 km - 140 m	4 401	~ 14 000	31 %
140 m - 40 m	2 140	285 000	1 %
40 m - 1 m	1 560		

Cuadro resumen proporcionado por la UAI en agosto de 2012.

La observación individual de los NEOs se hace habitualmente desde telescopios en tierra. Es necesario que los telescopios dispongan de CCDs con gran amplitud de campo y un movimiento rápido que permita seguirlos. Cuanto más cerca está un NEO de la Tierra, más rápido parece su movimiento respecto del fondo de estrellas. La sensibilidad de las CCDs es importante ya que permite la detección de NEOs débiles o/y lejanos.

Existen varios programas específicos de observación y detección de NEOs en diferentes observatorios astronómicos. Los programas que han producido el mayor número de resultados son el Catalina Sky Survey de la Universidad de Arizona, el programa de monitorización LINEAR proyecto conjunto de la fuerza aérea norteamericana y NASA, el patrullaje Pan-STARRS de la universidad de Hawai y el proyecto NEO-WISE de la misión espacial WISE mencionada anteriormente.

En el año 1996 se creó la Spaceguard Foundation con delegaciones en varios países con el objetivo de promover y coordinar el estudio y detección de NEOs e informar sobre su potencial peligro para la Tierra. El sistema Spaceguard agrupa a un conjunto de observatorios en todo el mundo. Unos pocos centros se dedican al descubrimiento de objetos nuevos y la mayor parte a su seguimiento posterior. En España la sede está en el Observatorio de Mallorca.

Cuando un telescopio descubre un NEO remite la información al MPC (Minor Planet Center) o Centro de Planetas Menores en Boston, Estados Unidos. Los programas de cálculo de órbitas NEO de NASA/JPL y

NEODYs-2, proyecto italo-español financiado parcialmente por la Agencia Espacial Europea, analizan los datos, les asignan una posibilidad de impacto y publican los resultados en una página web.

Este artículo de divulgación comienza con el descubrimiento de un NEA de tipo Apolo por uno de los telescopios del Observatorio de Mallorca. Se trata del asteroide 2012 DA14 de 45 m de diámetro que cruza el plano de la órbita de la Tierra dos veces por año. Según el Minor Planet Center, NEODYs-2 y Sentry el 15 de febrero de 2013, a las 19:27 UTC (tiempo universal coordinado) y con un error de 23 minutos, pasará a una distancia de la Tierra de 30 000 km, es decir más cerca que los satélites geoestacionarios. Este paso tan cercano a la Tierra modificará su órbita de modo impredecible.

Las tablas de riesgo de impacto. Las escalas de Turín y de Palermo

El programa NEO de JPL mantiene una tabla centinela (Sentry) con información sobre NEAs. En ella sólo se incluyen objetos con probabilidad de impacto en los próximos 100 años. La tabla incluye entre otros datos la probabilidad de impacto, el momento estimado y los daños según las escalas de Turín y Palermo. La tabla se puede consultar en la web de NASA/JPL.

La tabla de riesgo de NEODYs-2 contiene información similar aunque calculada de un modo diferente e incluyendo objetos con probabilidad de impacto hasta 2080. También se utiliza la escala de Palermo y Turín para evaluar la probabilidad de impacto y los daños estimados. En ambas tablas se puede obtener información adicional de cada objeto, como por ejemplo sus parámetros orbitales, magnitud, albedo, etc, pulsando sobre su nombre.

El hecho de que un objeto aparezca en las tablas sólo indica que su órbita puede cruzar la de la Tierra en un futuro. Existe una amplia gama de trayectorias que puede seguir y al menos una de ellas puede ser peligrosa. Los objetos aparecen en la tabla cuando se observan por primera vez. En estos casos su órbita futura no está bien determinada. Lo más probable y lo que suele ocurrir constantemente, es que al realizar nuevas observaciones del objeto, la trayectoria quede mejor definida y la órbita del objeto deje de cruzar la de la Tierra con lo que el objeto desaparece de las listas. Existen objetos que sólo se observan una vez y luego se “pierden”. Estos objetos se mantienen en la tabla a la espera de poder ser observados en el futuro.

La escala de Turín (o Torino) es una escala de daños cualitativa de 0 a 10 que tiene en cuenta la probabilidad de impacto y la energía que liberaría dicho impacto. El valor 0 indica que no hay ninguna posibilidad de daños, 1 indica que la probabilidad de impacto es muy baja y que observaciones futuras probablemente recalificarán la probabilidad de impacto a 0. Los

valores entre 2 y 4 constituyen la zona amarilla y requieren la atención de los astrónomos para reevaluar la probabilidad de impacto, los valores entre 5 y 7 constituyen la zona naranja o de amenaza y en ella se incluyen los objetos que pueden producir con gran probabilidad una catástrofe regional o global. En estos casos es necesario crear planes de contingencia. Los valores entre 8 y 10 constituyen la zona roja y de colisión asegurada. El 8 se asigna a fenómenos de destrucción locales que en caso de caída en el mar generarían tsunamis. Se producen una vez cada 50 a 1 000 años. El 9 produciría una catástrofe de carácter regional con grandes tsunamis si cae en el océano. Un evento como este estaría producido por un asteroide de un tamaño de 200 m y se producen una vez cada 10 000 o 100 000 años. El valor 10 se asigna a eventos que producen catástrofes climáticas globales y pueden amenazar el futuro de la civilización como la conocemos hasta ahora. En general se puede considerar que un asteroide de un 1 km de diámetro o superior producirá daños a escala global. Este tipo de eventos se produce aproximadamente cada 100 millones de años.

La escala de Palermo es una escala desarrollada por especialistas para ordenar el riesgo de impactos, teniendo en cuenta fechas, energías y probabilidades de estos. Se trata de una herramienta de comunicación entre astrónomos y no está diseñada para el gran público. La escala compara la probabilidad de un impacto potencial con el riesgo promediado para objetos del mismo tamaño o mayor a lo largo de los años hasta la actualidad. Se trata de una escala logarítmica en la que un valor de 0 indica que la probabilidad de impacto es igual a la promedio, -2 indica que es 1 % inferior a la promedio y +2 que es 100 veces superior a la promedio. En el cálculo se tiene en cuenta la probabilidad de que se produzca el evento dividido por la probabilidad anual de un impacto similar o mayor y por el tiempo que resta hasta el impacto. En la práctica valores por debajo de -2 se refieren a eventos que no precisan atención, entre -2 y 0 necesitan una monitorización cuidadosa, y superior a 0 indica eventos que deberían causar cierta preocupación. La suma de las probabilidades de los diferentes posible impactos es lo que se conoce como valor acumulado en la escala de Palermo.

**Objetos con valor 1 en la escala de Turín
(según tablas Sentry y NEODyS-2)**

NEA	año de impacto	tamaño (km)	Palermo	Turín
2011 AG5	2040-2047	140	-1.00	1
2007 VK184	2048-2057	130	-1.56	1

La tabla cambia a lo largo del tiempo incluyendo y retirando objetos.

En la fecha de redacción de este artículo, agosto de 2012, sólo figuran 2 objetos que en la escala de Turín tengan un valor superior a 0, los asteroides 2011 AG5 y 2007 VK184. La probabilidad de impacto de ambos es pequeña y se sitúa entre los años 2040 y 2057. De cualquier modo la

órbita se recalculará cuando sean visibles de nuevo en los años 2014 y 2013 respectivamente. De acuerdo con la escala de Turín ambos objetos tienen asignado un 1, y de acuerdo con la de Palermo, tiene asignado un máximo de -1 y -1,6 respectivamente. Apophis es un NEA de 270 m del que se habló en los medios de comunicación recientemente. En este momento tiene asignado un valor de 0 en la escala de Turín. La tabla Sentry contiene en este momento 402 objetos y la NEODyS-2 335.

Medidas de protección. Ciencia ficción y realidad

Hace unos años se estrenó una película de Hollywood cuyo argumento gira en torno a un asteroide que se acerca a la Tierra en rumbo de colisión. Un equipo de astronautas, mineros petrolíferos y militares salvan heroicamente al planeta Tierra en una misión espacial que viaja al asteroide donde hacen detonar una bomba nuclear en su interior. Probablemente este es el tipo de solución en el que pensaría mucha gente ante esta circunstancia: emplear la herramienta más poderosa de destrucción de que dispone la humanidad. Sin embargo es muy posible que no fuera lo más adecuado o que simplemente que no funcionara.

Ya sabemos que hay asteroides rocosos y metálicos. Un asteroide rocoso pequeño con una densidad de 2 o 3 gramos por centímetro cúbico probablemente se fragmentara con una explosión. Sin embargo es muy improbable que una bomba nuclear destruyera un asteroide compuesto de hierro; como mucho lo calentaría. Incluso los asteroides de menor densidad, casi porosos, como el 253 Mathilde, serían difícil de destruir.

Se ha experimentado disparando proyectiles contra materiales rocosos y porosos. En el primero de los casos el asteroide se fragmenta, pero en el segundo no ocurre nada apreciable. Es equivalente a disparar contra un saco de arena en el que el proyectil es absorbido sin daños internos. En el caso de un pequeño asteroide rocoso que se pudiera romper, la multitud de fragmentos generados extenderían el daño a un área más grande en la Tierra, incluso aunque dichos fragmentos fueran menores. Por otra parte, muchos de los asteroides contienen compuestos de cloro y bromo que podrían destruir la capa de ozono. Romperlos en pedazos no evitaría ese daño en la atmósfera de la Tierra. En vista de estos efectos adversos, si la fragmentación resulta peor que un sólo impacto, este último escenario sería preferible, ya que al menos se podría determinar con mejor precisión su trayectoria.

Existen ideas más sutiles que emplear la fuerza bruta para evitar el impacto de asteroides contra la Tierra. Revisemos algunas de las propuestas hechas hasta la fecha.

Se puede disparar un proyectil, no nuclear, contra un asteroide con el objetivo de cambiar su trayectoria, no de fragmentarlo. En el año 2005

NASA, durante la misión Deep Impact, envió un proyectil de 370 kg a 10 km/s sobre el núcleo del cometa Tempel 1. Se cree que tras el impacto se formó un cráter de 150 m de diámetro y 30 m de profundidad. Esta misión, muy compleja técnicamente porque el cometa se mueve a varios km/s, demostró la viabilidad de acercarse a cuerpos menores y disparar proyectiles sobre ellos. Tras el impacto, equivalente a unas 5 toneladas de TNT, el núcleo del cometa emitió gases durante 13 días. Se desconoce cuál fue el cambio producido en la trayectoria del cometa, pero probablemente fue prácticamente inapreciable dadas las dimensiones del proyectil, menores de 1 m, y las del núcleo del cometa, unos 7 km aproximadamente.



Figura 13: Imagen del núcleo del cometa Tempel 1, 67 segundos después de recibir un impacto de la misión Deep Impact de la NASA. Imagen propiedad de NASA.

La Agencia Espacial Europea está estudiando esta posibilidad y ha puesto en marcha el estudio de una misión llamada Don Quijote. Esta misión estaría compuesta por dos módulos: Sancho orbitaría en torno al asteroide para obtener información de él, e Hidalgo sería el proyectil de unos 530 kg. Antes del impacto Sancho se retiraría a una distancia de seguridad y después de él volvería a acercarse para estudiar sus efectos.

Más tarde Sancho enviaría una sonda sobre el cráter para conocer las propiedades de la superficie tras la colisión.

Otra de las propuestas consiste en detonar una bomba en la cercanía del asteroide de modo que se genere una cantidad de calor suficiente que vaporice los gases en su superficie. Estos se expulsarían, como en el caso de los gases de un motor, proporcionando al asteroide un impulso que lo desviaría de su trayectoria. El impulso sería bastante pequeño y dependería del tamaño del asteroide, pero tras varias explosiones, es posible que la trayectoria de la órbita se desplazara lo suficiente para evitar la colisión con la Tierra. Este tipo de solución funcionaría con todo tipo de asteroides independientemente de su densidad. El principal requisito es que esta operación habría que realizarla con varios años de antelación para que el cambio en la órbita fuera apreciable. Probablemente diez años sería el mínimo y 100 años sería óptimo. El principal problema es que es necesario detectar con mucha antelación el asteroide y, si estos son pequeños, esto es especialmente difícil.

Desviar un asteroide para evitar una primera colisión con la Tierra, como en las propuestas anteriores, sería un éxito pero puede suponer colocarlo en rumbo de colisión en ulteriores pasos cerca de la Tierra. Para evitar que esto ocurra sería óptimo controlar su órbita, algo técnicamente más complejo.

Una de las opciones contempladas consiste en aterrizar un cohete o nave espacial sobre la superficie del asteroide, anclarlo a este “boca abajo” y encender un motor que le aplique un impulso en la dirección adecuada. La idea es sencilla pero llevarla a cabo puede resultar muy complicado técnicamente. La superficie del asteroide puede ser porosa, llena de escombros que se deshacen o metálica y muy dura. Anclar el satélite puede resultar especialmente difícil. Además hay numerosos asteroides que giran caóticamente por lo que además de complicar el aterrizaje, una vez conseguido no se podrían encender los motores continuamente; sólo en el momento en el que la superficie donde se ha producido el aterrizaje se orienta en la dirección adecuada.

Si resulta difícil aterrizar, ¿por qué no “aparcar” una nave espacial junto al asteroide? La gravedad del asteroide tirará de la nave espacial, y esta del asteroide. Para evitar que ambos colisionen es necesario encender los motores y aplicar un impulso pequeño que contrarreste la gravedad. Al mismo tiempo el cohete debe empujar suavemente en la dirección deseada. Probablemente para esta operación sea necesario al menos un par de cohetes situados en direcciones diferentes. Esta técnica es muy sencilla de aplicar pero también requiere ponerla en práctica varios años antes del paso del asteroide junto a la Tierra. Si el asteroide pasa periódicamente cerca de la Tierra, habría que hacerlo una o varias pasadas antes del pase peligroso que impactaría en la Tierra. Aunque parezca sorprendente puede que el tirón que es necesario aplicar no sea muy grande, especialmente si de lo que se trata es de desviar el satélite del área que conduce a una colisión

en una órbita posterior. Si el desvío hay que hacerlo inmediatamente antes de la colisión, la tracción gravitacional a emplear debe ser mucho mayor.

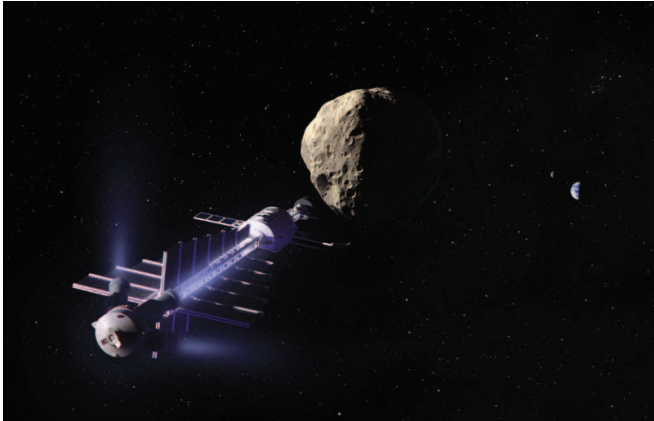


Figura 14: Tractor gravitatorio para arrastrar un asteroide fuera de su órbita. Se trataría de una nave espacial de 20 toneladas alimentada con un reactor nuclear. Emplearía motores iónicos para moverse y arrastrar el asteroide de modo continuo. Imagen de Dan Durda (FIAAA, Fundación B612).

Alternativamente se podría emplear una nave espacial más ligera que lanzara continuamente iones sobre el asteroide que a su vez transmitirían su impulso al asteroide. Es un equivalente al de la tracción gravitatoria comentada antes.

Poner en práctica estas ideas está fuera del alcance de la tecnología actual. Impactar un proyectil es probablemente lo único que funcionaría ahora. Una mejora de la tecnología actual y la posibilidad de realizar pruebas nos permitiría modificar la órbita de un PHA. Es más, sería incluso posible colocar un asteroide en órbita alrededor de la Tierra para poder extraer sus minerales. Quizá este último objetivo fuera un aliciente para la inversión de capital privado en un proyecto tecnológico de este tipo.

Conclusión

La investigación de los cuerpos menores del sistema solar es importante; no sólo porque nos permite conocer mejor el origen del sistema solar, sino también porque nos proporcionará información sobre el origen de la vida en la Tierra y la posibilidad de que se produzca fuera de esta. Además su estudio nos permite evaluar mejor el peligro potencial de impactos de

asteroides o cometas sobre la Tierra. Conocer los asteroides y cometas, su composición, órbita y origen son fundamentales.

Resulta irónico pensar que la vida en la Tierra probablemente comenzó con el bombardeo de cometas y asteroides sobre la Tierra y que esta puede desaparecer por el impacto devastador de uno de estos cuerpos. Los asteroides y cometas son portadores de vida, y de muerte, y a veces simplemente cambian el rumbo de la primera. Como ya hemos visto en el caso del impacto que originó el Chicxulub, tras un impacto, la vida no desaparece totalmente. Hace 65 millones de años se extinguieron la mayor parte de las especies, pero surgió una oportunidad para otras, como los mamíferos. La raza humana probablemente debe su existencia, entre muchos otros factores, a ese impacto y a la posterior desaparición de los dinosaurios.

Sin embargo un eventual impacto que ahora acabe con la civilización tal y como la conocemos es un motivo serio de preocupación para la humanidad. Es seguro que un asteroide o un cometa de tamaño grande impactarán de nuevo sobre la Tierra en el futuro. Lo que desconocemos es cuándo. Mientras tanto es importante continuar con los programas de observación, detección y predicción de órbitas de cuerpos menores, y desarrollar una tecnología capaz de desviar los asteroides potencialmente peligrosos. Esta tecnología requiere inversión y misiones espaciales.

Para saber más: bibliografía

Para saber más sobre bólidos y meteoritos, incluso para informar sobre avistamientos de los primeros y hallazgos de los segundos, puede consultar la web: <http://www.spmn.uji.es/> de la Red de Investigación sobre Bólidos y Meteoritos y la Organización Internacional de Meteoros: <http://www.imo.net/>. Le animamos a que informe de estos eventos si ha tenido la fortuna de presenciarlos. Su contribución será utilizada y agradecida.

El libro “Asteroids III” (2002) de la Universidad de Arizona editado por W. Bottke, A. Cellino, P. Paolicchi, y R.P. Binzel es una obra de referencia actualizada sobre asteroides. En él se puede encontrar una sección con varios capítulos sobre impactos de asteroides. También se puede encontrar una estimación de la cantidad de material interplanetario recogida por la Tierra a lo largo del año.

La hipótesis de que la formación de cráteres no ha sido constante ha sido formulada por S.N. Ward y S. Day en la revista “Earth, Planets and Space” (2008). La variabilidad en el flujo de asteroides había sido propuesta un año antes por W. Bottke, D. Vokrouhlický y D. Nesvorný en un artículo de Nature en 2007. Estos últimos autores son los que postulan que dicho evento se puede deber a la colisión de dos cuerpos en el cinturón de asteroides.

La base de datos de cráteres de impacto se encuentra en la página <http://www.passc.net/EarthImpactDatabase/>. También en <http://impacts.rajmon.cz/> hay una base de datos creada por el grupo de estudios de campo de impactos.

En la web <http://neo.jpl.nasa.gov/neo/> la NASA mantiene información sobre NEOs. La tabla Sentry de riesgo de impacto de JPL se puede consultar en <http://neo.jpl.nasa.gov/risk/> y en <http://newton.dm.unipi.it/neody2/index.php?pc=4.1> la de NEODYS-2.

La Agencia Europea del Espacio (ESA) mantiene información sobre NEAs en: <http://neo.ssa.esa.int/web/guest/neo-chronology>.

El meteorito de Chelyabinsk aviva la preocupación por los impactos

El artículo **Impactos sobre la Tierra** se redactó a finales del año 2012 y se publicó en el año 2013 y, como si sus palabras fueran una premonición, ese mismo año la Tierra recibió el impacto de un meteorito en la región rusa de Chelyabinsk. El 15 de febrero de 2013 un asteroide de unos 16 m de diámetro y unas 10 000 toneladas de masa se internó en la atmósfera terrestre y tras 32 segundos de vuelo se estrelló en el lago helado Chebarkul dejando unos 4000 kg de restos. Las explosiones en su descenso por la atmósfera causaron roturas de cristales y como consecuencia unos 1400 heridos. Los últimos segundos de la trayectoria fueron grabados por varias personas desde diferentes localizaciones. Se trata del impacto de un asteroide mejor documentado y retransmitido hasta la fecha. Ese mismo día se esperaba el paso cercano del asteroide 2012DA14 pero la aparición de este meteorito, indetectado al aproximarse desde la dirección del Sol, fue completamente inesperada y avivó la preocupación y la necesidad de la detección temprana de objetos cercanos a la Tierra (NEOS). Desde entonces los medios prestan especial atención a la aproximación de este tipo de asteroides y en algunos casos generan, injustificadamente, cierta alarma social dando pábulo a noticias no contrastadas científicamente.

En los últimos años se han propuesto misiones para la detección temprana de objetos cercanos a la Tierra y el estudio de sus órbitas y propiedades, así como posibles medidas de deflexión para impedir que impacten sobre nuestro planeta. Incluso se ha discutido sobre un hipotético impacto de un objeto sobre la Tierra. En los años 2013 y 2015 se celebraron las conferencias bianuales de defensa planetaria de la Academia Internacional de Astronáutica que incluyen un escenario realista en el que la probabilidad de impacto de un asteroide sobre la Tierra es alta. El objeto de dicho escenario consiste en estudiar los procesos de toma de decisiones, la información a recopilar y las acciones a llevar a cabo. Las siguientes conferencias, en 2017 será la próxima, seguirán realizando estos ejercicios prácticos para prepararse para un futuro.

En este momento (feb. 2017) se han detectado 15 688 asteroides cercanos a la Tierra mayores de 140 m de diámetro, de los que 875 son de más de 1 km de diámetro. El número de detecciones ha aumentado a gran ritmo desde el año 2000 por lo que en un futuro cercano se conocerá un porcentaje muy elevado de dichos objetos y sus órbitas cercanas a la Tierra.

Algunos de los proyectos más atractivos de detección de asteroides son aquellos que se realizan desde el espacio, en particular desde un satélite situado en una órbita interior a la de la Tierra que permita descubrirlos de espaldas al Sol. De entre estas misiones de detección temprana propuestas recientemente destaca la de la Fundación B612 que propone la puesta en órbita de Sentinel, un satélite infrarrojo que permitiría detectar objetos potencialmente peligrosos trabajando conjuntamente con el Large Synoptic Survey Telescope. Esta misión, originalmente propuesta para 2012, todavía no tiene financiación. JPL ha propuesto y está estudiando la misión NEOCam, también un telescopio infrarrojo inspirada en Sentinel, cuyo objetivo es no sólo el descubrimiento de asteroides sino también el estudio de sus propiedades. Finalmente mencionaremos una misión de alerta temprana, denominada SODA, propuesta por la Academia de Ciencias Rusa (RAN) cuyo objetivo sería la detección, desde el punto lagrangiano L1 y de espaldas al Sol, de objetos de más de 10 m de diámetro con una antelación de 4 horas en el hemisferio diurno de la Tierra.

En los últimos años se han realizado diversos estudios para la deflexión de asteroides con proyectiles cinéticos, nucleares, velas solares, tractores gravitatorios e incluso enjambres de nanoobjetos de poca masa y alta velocidad golpeando el asteroide que se desea desviar. Una de las misiones más atractivas para estudiar la deflexión de asteroides es AIDA. Se trata de una misión conjunta de las agencias espaciales norteamericana, NASA, y europea, ESA, en el sistema Didymos. Didymos es un sistema doble compuesto de un asteroide central de unos 800 m de diámetro en torno al que orbita una luna de 150 m de diámetro. La misión, compuesta de dos sondas, consiste en lanzar una de ellas contra el asteroide más pequeño y observar, desde la otra sonda, la alteración de su trayectoria en torno a su asteroide mayor. El lanzamiento está previsto para diciembre de 2020 pero el proyecto, que se discutió en el consejo de ministros europeo a finales de 2016, está de momento en fase de diseño preliminar. En caso de aprobarse, el encuentro con Didymos se produciría en 2022. Otra misión interesante a la que conviene prestar atención es OSIRS-Rex, lanzada en septiembre de 2016. Esta misión alcanzará el asteroide Bennu en 2018, tras pasar por el punto L4, y traerá muestras de vuelta a la Tierra en 2023. En su paso por L4 intentará descubrir posibles asteroides troyanos de la Tierra.

Recientemente JPL ha actualizado su página web sobre NEOS, haciéndola más accesible y sencilla de consultar. Se trata de un lugar excelente donde los lectores interesados pueden encontrar una gran cantidad de información actualizada y fiable y que muestra de forma comprensible las aproximaciones recientes y futuras de este tipo de objetos. El autor recomienda consultar dicha página, especialmente cuando se producen noticias alarmistas en los medios de comunicación. Se trata de <http://cneos.jpl.nasa.gov/stats/>.

Pablo de Vicente, febrero 2017

LA IDEA DE PLANETA A LO LARGO DE LA HISTORIA

Pere Planesas

*Observatorio Astronómico Nacional
Instituto Geográfico Nacional - Ministerio de Fomento*

Abstract

As the 26th IAU General Assembly unfolded last August the internal debate on a definition for what a planet is reached the mass media. The approved resolution was reasonable, given the constraints, but not everybody liked or understood it. In this paper that discussion is put into historical perspective and in the context of the current knowledge of the solar system, that has changed considerably in the last few decades. Unfortunately the new vision of the solar system has not reached the common culture: for many people the solar system consists of the Sun, nine planets, scores of moons and a few comets. They are not aware of the several hundred thousand objects of different size, orbit or composition that populate the solar system and that have been classified in a larger variety of object families, like that of the dwarf planets.

Introducción

En el mes de agosto de 2006 se ha producido un debate entre los centenares de astrónomos reunidos en Praga con ocasión de la 26^a asamblea general de la Unión Astronómica Internacional (UAI¹) que ha trascendido las reuniones de trabajo allí celebradas y ha tenido repercusión en los medios de comunicación de masas, incluido internet. Posiblemente el hecho de que haya tenido lugar en un mes vacacional en los países del

¹<http://www.iau.org/>

hemisferio norte, mes en que escasean las noticias que habitualmente acaparan tales medios, ha influido en su tratamiento casi diario e incluso preferente. Además, el sujeto de la discusión no era un abstracto concepto cosmológico o astrofísico sino uno de cultura general. Tras la pretensión de definir por primera vez lo que es un planeta, realmente se discutía cómo tratar toda una familia de objetos descubiertos a partir de 1992 en el sistema solar, situados más allá del planeta gigante Neptuno. Se trata de objetos más grandes que los asteroides, pero menores que la Luna, que orbitan alrededor del Sol en trayectorias excéntricas e inclinadas con respecto a la órbita de la Tierra; en definitiva, se trata de objetos similares a Plutón. ¿Había que designarlos como una nueva clase de objetos, incluyendo en ella el hasta ahora llamado planeta Plutón? O bien, para preservar el estatus de éste, ¿convenía ampliar la clase de objetos a los que llamamos planetas?

Aun cuando ésta pueda parecer una discusión original, como veremos más a abajo ya se ha producido en otras ocasiones en astronomía. Tal tipo de discusión es habitual entre los científicos de un determinado campo y se produce cuando el aumento de conocimiento sobre un determinado aspecto de estudio (por ejemplo, los astros del sistema solar) provoca el replanteamiento y posible modificación de un concepto que hasta entonces parecía bien establecido. Un concepto en permanente discusión es el de la vida. Ello ya ha ocurrido en el pasado con el concepto de planeta y volverá a ocurrir en el futuro, quizás pronto pues en mi opinión el debate se ha cerrado precipitadamente. Aunque ello tampoco es ningún problema: los astros sujeto de la discusión seguirán siendo como son independientemente de cómo sean llamados y su estudio no se verá afectado por un simple convenio de nomenclatura. El querer pasar en una semana de un convenio a una definición rigurosa posiblemente ha sido la causa de la acritud de las discusiones y de la modestia del resultado conseguido.

Las discusiones habidas en los medios de comunicación también han revelado que en la cultura general el sistema solar se reduce al Sol, los nueve planetas, sus lunas y algunos cometas. Actualmente se han identificado y numerado más de cien mil astros en el sistema solar, la mayoría pequeños objetos agrupados en distintas familias en función de su tamaño, su órbita o sus características físicas. Una de las de más reciente creación es la de los planetas enanos (denominación fallida), actualmente constituida por Plutón, el asteroide Ceres y el objeto transneptuniano Eris.

El sistema solar en la antigüedad

Una inspección continuada del cielo ha revelado a la mayoría de pueblos antiguos que puede imaginarse constituido por una esfera giratoria estrellada, frente a la que se mueven varios astros a menudo considerados, por tal peculiaridad, personificaciones de dioses. Destacan dos: el Sol, por su tamaño claramente apreciable y su brillo, y la Luna, por su tamaño, su

variación de aspecto y la rapidez con que se mueve. Además hay otros astros, de aspecto puntual o parecido a las estrellas fijas en la bóveda celeste, pero que destacan por su movimiento con respecto de ella y sus épocas de visibilidad e invisibilidad. Salvo la Luna, que recorre la bóveda en casi un mes, los demás tardan en recorrerla aproximadamente un año (el Sol, Mercurio y Venus) o más (Marte), incluso mucho más (Júpiter y Saturno). Los siete astros citados son los que aparecen en la mayoría de descripciones del cielo antiguas que nos han llegado por documentos o inscripciones e incluso, en el caso de pueblos que no han desarrollado la escritura, por transmisión oral.

No son sólo especiales el Sol y la Luna. También Mercurio y Venus lo son, por no alejarse nunca demasiado del Sol y por aparecer a veces precediéndole o a veces siguiéndole. De hecho, el identificar el lucero vespertino por excelencia con el lucero matutino por excelencia, con un solo astro que ahora llamamos Venus no es obvio. Nos lo parece ahora que conocemos su naturaleza y sabemos que orbita alrededor del Sol, y acabó siéndolo para los antiguos al observar que cuando uno es visible el otro no lo es y que su comportamiento precediendo o siguiendo el Sol es simétrico, en cuanto a duración y a separación máxima (elongación solar). Sin embargo consta que inicialmente el Venus matutino y el vespertino fueron tenidos por astros distintos. Así fue, por ejemplo, entre los primeros griegos, los de la época de Homero (c. 700 a.C.) y Hesíodo (c. 650 a.C.), que los denominaban Eosphoros y Hesperos, respectivamente. Diógenes Laercio (y lo confirma Aecio) escribe:

Y se cree que [Parménides de Elea] fue el primero en ver que la Estrella Matutina y la Estrella Vespertina son una y la misma . . . , aunque otros dicen que fue Pitágoras.

Lo mismo ocurrió en el caso de Mercurio (la estrella de Hermes, para los griegos). Su menor brillo y su cercanía al Sol dificultan su visión, por lo que aparece en pocos textos, aunque consta que era conocido por los Pitagóricos y que en su época (s. V a.C.) debió de identificarse con un solo astro. Podemos así pensar que en época anterior los griegos habrían podido identificar hasta nueve estrellas peculiares: Sol, Luna, Venus Matutino, Venus Vespertino, Mercurio Matutino, Mercurio Vespertino, Marte, Júpiter y Saturno. Posteriormente estas estrellas quedaron reducidas a las siete que, por ejemplo, cita Platón (c. 427 - 347 a.C.) en *Timeo*:

. . . el Sol, la Luna y las otras cinco estrellas que son llamadas “planetas” . . . en siete órbitas siete estrellas.

Nuestra palabra “planeta” proviene de la griega *πλανήτης* que significa errante. Es decir, los griegos se referían a las estrellas con un movimiento apreciable con respecto a las estrellas fijas como estrellas errantes. La

expresión inicial de *planetes asteres* se convirtió en *planetae* hacia el s. IV a.C.. Este es el proceso conceptual por el que los cinco astros dejaban de ser estrellas para tener una identidad propia dada por su comportamiento, aún no por su naturaleza física. Estas cinco estrellas errantes iniciaban su conversión en planetas, aunque durante algún tiempo esta denominación incluyó también el Sol y la Luna.



Figura 1: Universo geocéntrico en *Cosmographicus liber* de Apiano (Peter Bienewitz 1495 – 1552). Biblioteca del OAN.

Muchos siglos antes que los griegos, los egipcios y los mesopotámicos habían identificado y dado nombre a los cinco planetas visibles. Los egipcios los conocían como “las estrellas que no descansan”. Dieron diversas denominaciones a cada uno de ellos, pero parece claro que se referían a uno concreto en cada caso. Hay indicios de que identificaron a Mercurio en el crepúsculo matutino con el del vespertino en época bastante temprana, quizás en el segundo milenio (a.C.). En cuanto a

Mesopotamia, en la colección de presagios *Enūma Anu Enlil* (anterior al 900 a.C.) figuran diversos nombres de los cinco planetas, conocidos como “las ovejas salvajes”, y cálculos simples de sus épocas de visibilidad y de las retrogradaciones² de los planetas Marte, Júpiter y Saturno. Sin embargo el desarrollo de teorías matemáticas para la predicción precisa de sus efemérides no se produjo hasta el florecimiento de la astronomía mesopotámica en la época persa que, por lo tanto, precedió al desarrollo de la griega.

En definitiva, a pesar del interés de los mesopotámicos por los movimientos planetarios, el nombre con que denominamos a las “estrellas errantes” proviene de una palabra griega y los nombres con que los denominamos son herencia romana. Inicialmente los griegos les asociaron el nombre de un dios. En época posterior (periodo helenístico) les dieron un nombre propio descriptivo a fin de evitar la confusión causada por la asociación de varios dioses a un mismo planeta, aunque ésta parece haber sido la denominación más común. En su ordenación habitual entre los griegos, de más lento a más rápido en su movimiento, los planetas eran:

Φαινων	Phainon	Estrella de Cronos
Φαεθων	Phaëton	Estrella de Zeus
Πυροειζ	Pyroëis/Pyroeides	Estrella de Ares
Εωςφοροζ, Εσπεροζ	Eosphoros y Hesperos	Estrella de Afrodita o de Hera
Στιλβον	Stilbon	Estrella de Hermes

Los dioses de la última columna se inspiraron en algunos de los dioses que les atribuyeron los babilónicos siglos antes:

(Saturno)	Ninurta	dios de la fertilidad y la guerra.
(Júpiter)	Marduk	principal dios babilónico.
(Marte)	Nergal	dios de los infiernos, las plagas y la guerra.
(Venus)	Ištar	diosa del cielo.
(Mercurio)	Nabû	dios de las ciencias y las letras.

Los romanos adoptaron otros nombres para las *stellae errantes* o *sidera errantia* (astros errantes) asignándoles los de dioses de su propia mitología equivalentes o similares a los de los griegos:

Saturnus	dios romano identificado con el griego Cronos.
Jupiter	principal dios romano.
Mars, Hercules	
Venus, Juno, Isis	Lucifer (Lucero o Lucífero) al amanecer, Hesperus (Véspero o Vespertino) al anochecer.
Mercurius, Apollo	(Apolo se asignaba preferentemente al Sol)

²Movimientos con respecto del fondo del cielo en que el planeta parece moverse en el sentido opuesto al habitual.

Para ellos, los planetas y la Luna reflejan la luz del Sol. Plinio el Viejo (c. 24 – 79) dice de la Luna:

... como los demás astros [errantes], está gobernada por la luz del Sol, puesto que brilla con luz totalmente prestada por él...

Ello refleja otro cambio conceptual, pues se trata a la Luna y los planetas como astros sin luz propia, no como las estrellas fijas que sí la tienen. Este cambio se inició con la afirmación de Anaxágoras de Clazómene (c. 500 – c. 428 a.C.) de que la Luna no tiene luz propia sino que su luz proviene del Sol. Esto fue repetido y sostenido por muchos astrónomos posteriores y está en la base de los trabajos de Aristarco de Samos (c. 310 – c. 230 a.C.) sobre los tamaños y distancias del Sol y la Luna.

III. Systema Ægyptiorum, Vitruuij, Capellæ, Macrobij, Bedæ &c.

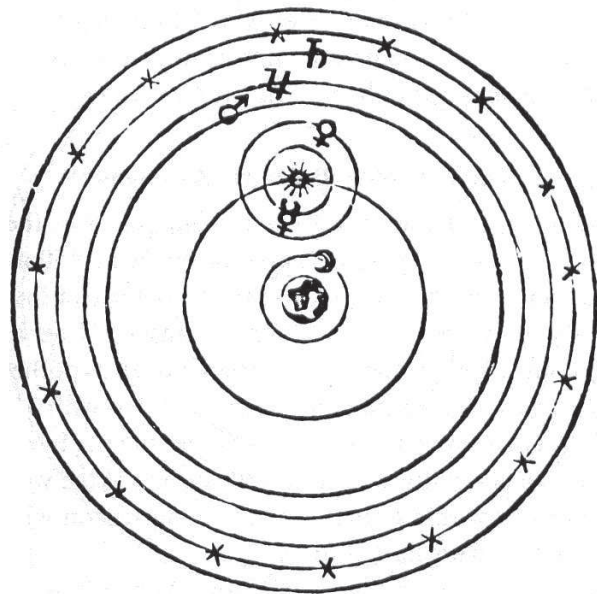


Figura 2: Sistema planetario presuntamente propuesto por Heráclito de Pontos (s. IV a.C.), en *Almagestum novum* de Giovanni Baptista Riccioli (1598 – 1671).

En lo referente a la luz de los planetas había diversidad de opiniones. En sus comentarios a los *Fenómenos* de Arato, Avieno (s. IV) dice que los planetas “son capaces de mantener fuegos incandescentes con sólo sus llamas” y, en cuanto a su aspecto, que “no sugieren ninguna figura concreta.”

El tamaño de los planetas, en cambio, fue sujeto de especulación pues no se podía medir. Abandonadas las ideas antiguas de los planetas como fuegos o aberturas en ruedas huecas llenas de fuego (Anaximandro, c. 610 – c. 546 a.C.) Tolomeo³ (s. II), utilizando el brillo de los planetas como indicador de su tamaño y comparándolos visualmente con la Luna, estimó que éstos eran similares a la Tierra, siendo Saturno, Júpiter y Marte mayores que ella, mientras que Venus y Mercurio eran menores.

Las teorías heliocéntricas

La primera propuesta de que algunos planetas giran alrededor del Sol pudo deberse a Heráclito de Pontos (c. 390 – c. 310 a.C.). Los peculiares movimientos de Mercurio y Venus, que nunca se separan mucho del Sol, le llevaron a proponer que tales planetas giran alrededor del Sol y éste gira alrededor de la Tierra. Tal explicación aparece en textos de varios autores romanos (Vitruvio, s. I; Marciano Capela, s. V), pero el único texto que cita explícitamente a Heráclito (Calcidio, s. IV) no convence a muchos estudiosos. Sí hay acuerdo en que Heráclito propuso la rotación diaria de la Tierra.

Algo mejor documentada está la teoría totalmente heliocéntrica de Aristarco. Según Arquímedes de Siracusa (c. 287 – 212 a.C.), en *El arenario*:

Aristarco de Samos ha publicado un libro ... concluye que el universo es mucho mayor ... las estrellas fijas y el Sol son inmóviles. En cuanto a la Tierra, se mueve alrededor del Sol en la circunferencia de un círculo centrado en el Sol.

Una hipótesis implícita es la rotación diaria de la Tierra. Si se entiende por planetas aquellos astros que se mueven regularmente alrededor del Sol, otra conclusión es que la Tierra es un planeta y que la Luna no lo es.

La hipótesis heliocéntrica de Aristarco sólo tuvo un valedor en la antigüedad, el matemático babilónico Seleuco de Seleucia (c. 150 a.C.), que la consideró una descripción del mundo real. Los demás autores que se refirieron a esta teoría fue para rechazarla. Incluso hubo una campaña de desprestigio popular hacia Aristarco basada en el sacrilegio que suponía para la religión griega el haber desplazado el centro del mundo de la Tierra.

³Klaudios Ptolemeios, el astrónomo más influyente de la antigüedad, autor del libro *μαθηματικὴ συντάξις*, más conocido como *Almagesto*.

al Sol, con lo que el corazón de la Casa de los Dioses, y por lo tanto la Tierra, dejaba de ser inmóvil.

La teoría heliocéntrica no reaparecería hasta que hacia 1509, dieciocho siglos después de Aristarco, empezó a circular el tratado manuscrito conocido como *Commentariolus*⁴ de Nicolás Copérnico (1473 – 1543). En él la Tierra se considera el centro de la gravedad y de la órbita lunar, mientras que es el Sol el centro de las órbitas de todos los planetas y del universo. En 1540 su discípulo Rheticus⁵ (1514 – 1576) publicó un libro comentando los nuevos trabajos de Copérnico lo que, junto con otras presiones, indujo a éste a publicarlos al completo en la obra *Los seis libros sobre las revoluciones de los orbes celestes* aparecida el año de su muerte.

Una primera consecuencia de esta teoría fue el aumento del número de planetas, pues la Tierra pasaba a ser uno de ellos. Del más lejano al más cercano al Sol los planetas eran: Saturno, Júpiter, Marte, Tierra, Venus y Mercurio, seis en total.

La teoría heliocéntrica tardó más de un siglo en consolidarse (véase el artículo de Mariano Esteban en este mismo ANUARIO) y lo hizo en gran parte gracias a los trabajos de Johannes Kepler (1571 – 1630) y de Galileo Galilei (1564 – 1642). La invención del telescopio por parte de este último en 1609 tuvo un enorme impacto en este tema. Otra importante consecuencia poco citada de tal invención fue que permitió estimar el tamaño angular de los planetas y, por lo tanto, determinar sus diámetros: con un telescopio los planetas aparecen como discos mientras que las estrellas siguen siendo puntos luminosos. Las medidas de Johannes Remus Quietanus, Pierre Gassendi (1592 – 1655) y Martinus Hortensius⁶ (1605 – 1639) dieron valores angulares algo superiores a los reales, hasta en un factor 2, pero eran mucho mejores que las estimaciones pretelescopicas. El desconocimiento de la distancia de la Tierra al Sol (la Unidad Astronómica) constituía la principal dificultad que entrañaba su conversión en distancias lineales. En 1633 Hortensius, adoptando una distancia de 1.500 radios terrestres (el valor real es de 23.455 $R_{\oplus}$), dedujo que todos los planetas salvo Saturno eran menores que la Tierra. Según sus medidas, de mayor a menor eran: Saturno, Tierra, Júpiter, Venus, Marte y Mercurio.

La observación telescópica de las fases de Venus por Galileo en 1610 demostró que los planetas, así como la Luna, no tienen luz propia sino que son cuerpos opacos que reflejan la del Sol. Con ello terminó un largo debate sobre si el Sol es la única fuente de luz, como sostuvieron griegos como Aristóteles (384 – 322 a.C.) y árabes como Averroes (Ibn Rushd, 1126 – 1198), o si los planetas son luminosos cual estrellas, como sostuvieron el

⁴*De hypothesibus motuum caelestium a se constitutis commentariolus* de Mikolaj Kopernik, forma original de su nombre.

⁵Georg Joachim von Lauchen

⁶Maarten van den Hove

romano Macrobio (s. IV) y el persa Avicena (Ibn S^īnā, 980 – 1037).

Uno de los descubrimientos más relevantes que efectuó Galileo con su telescopio fue el de los cuatro mayores satélites de Júpiter, los que conocemos apropiadamente como “galileanos”. Estos fueron los primeros astros del sistema solar descubiertos por un ser humano en tiempos históricos⁷, si exceptuamos los pasajeros cometas. Además de constituir una especie de sistema planetario en miniatura, mostraban que la Tierra no es el único planeta con satélites. Estos descubrimientos, junto con el de las fases de Venus y las variaciones de tamaño de Venus y Marte a lo largo de su órbita, constituyeron argumentos observacionales muy poderosos en favor del heliocentrismo.



Figura 3: Posiciones relativas de los cuatro satélites “mediceos” de Júpiter, descubiertos por Galileo Galilei en *Sidereus nuncius* (1610).

En la segunda mitad del s. XVII los astrónomos Christiaan Huygens (1629 – 1695) y Giovanni Domenico Cassini (1625 – 1712) descubrirían varios satélites de Saturno. En 1656 Huygens sugirió que el aspecto cambiante de Saturno se podía explicar si el planeta:

... está rodeado por un fino anillo plano que no lo toca y que está inclinado con respecto a la eclíptica⁸.

Detalles en la superficie de Marte, Saturno y Júpiter, así como el achataamiento y la gran mancha (roja) de éste, fueron puestos de manifiesto por varios astrónomos, constituyendo los primeros indicios de las diferencias y singularidades que existen entre los planetas. Francesco Fontana (1580 – 1656) publicó en 1646 *Novae coelestium terrestriumque rerum observationes*, el primer libro ilustrado con numerosos dibujos del aspecto telescópico de los planetas.

Al finalizar el siglo XVII el sistema solar se consideraba constituido por el Sol, seis planetas y diez satélites. La Unidad Astronómica (UA) se estimaba entre 20.000 y 22.000 $R_{\oplus}$ y el diámetro de los planetas se conocía con un error menor al 10 %, excepto en el caso de Mercurio que alcanzaba el 40 %. Los astrónomos eran por fin conscientes del enorme tamaño del Sol y de que Júpiter y Saturno son mucho mayores que los demás planetas.

⁷Una posible, pero incierta, observación a simple vista de un satélite de Júpiter por el astrónomo chino Gan De hacia el 364 a.C. no tuvo trascendencia.

⁸Plano de la órbita terrestre alrededor del Sol.

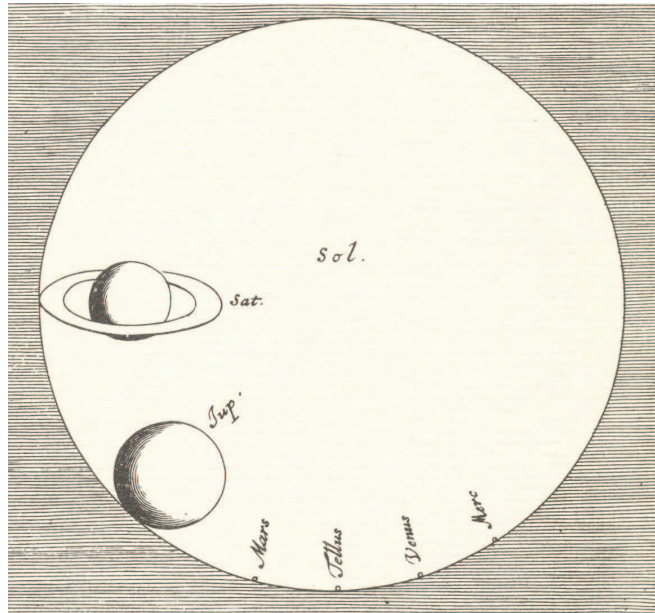


Figura 4: Tamaños relativos de los astros del sistema solar según aparecen en la obra *Kosmotheoros* de Christiaan Huygens publicada en 1698. El círculo mayor corresponde al Sol.

En esta época terminaron también las últimas especulaciones de origen medieval que consideraban los planetas como cuerpos transparentes. Ahora eran considerados sólidos, opacos y capaces de reflejar la luz del Sol. Incluso algunos divulgadores describían el nuevo sistema solar apoyándose en habitantes imaginarios que poblaban la superficie de los planetas, como hizo Bernard Le Bouyer de Fontenelle (1657 – 1757) en su obra *Entretiens sur la pluralité des mondes* (1686) reimpresa decenas de veces a lo largo de su larga vida.

Primer descubrimiento de un planeta

A lo largo del s. XVIII se sucedieron los comentarios acerca de la posible existencia de otros planetas que pudieran hallarse demasiado cerca del Sol o excesivamente lejos de él, mucho más allá de Saturno, como para ser vistos. En los libros aparecían citas a “los planetas conocidos”, dando a entender que el conocimiento del sistema planetario probablemente era incompleto. El tamaño del sistema solar se había extendido tras la predicción hecha por Edmond Halley (1656 – 1742) acerca de la existencia

de un cometa de órbita elíptica muy excéntrica alrededor del Sol con un periodo de 75 años, cometa que reapareció, tal como predijo, en 1758. Si la órbita de Saturno pudo ser considerada durante milenios como el límite del sistema solar, el afelio⁹ del cometa de Halley llegaba casi cuatro veces más lejos, un espacio en el que bien podían encontrarse otros planetas. Además, éstos podían ser muy grandes si se aceptaba la hipótesis de Huygens, negada por las medidas de diámetros, que cuanto más distante del Sol se hallaba un planeta mayor era éste.

Más comúnmente se especulaba (por ejemplo Immanuel Kant, 1724 – 1804) sobre la existencia de planetas en el gran espacio que media entre las órbitas de Marte y Júpiter, lo que se plasmaría en la famosa conjetura de Johann Daniel Titius (1729 – 1796) defendida y popularizada por Johann Elert Bode (1747 – 1826) acerca de la existencia de un planeta a una distancia de 2,8 UA del Sol, intermedia entre la de 1,5 UA de Marte y las 5,2 UA de Júpiter.

Sin embargo, el s. XVIII transcurría sin que se descubrieran nuevos planetas o satélites. En 1764 Charles Bonnet (1720 – 1793) en su obra *Contemplación de la naturaleza* hablaba de “los 17 planetas conocidos”, denominación vaga bajo la que incluía también el Sol y los satélites, el último de ellos descubierto ochenta años antes por Cassini.

El primer descubrimiento de un nuevo planeta lo debemos a un músico alemán, emigrado a Inglaterra, con una tardía afición a la astronomía. Friedrich Wilhelm Herschel (1738 – 1822), conocido después como Sir William Herschel, descubrió un nuevo astro no catalogado la noche del martes 13 de marzo de 1781, utilizando para ello un telescopio reflector de 16 cm de diámetro construido por él mismo. Creyendo haber descubierto un cometa, envió una comunicación a la Royal Society. El mucho más experimentado astrónomo real Nevil Maskelyne (1732 – 1811) sospechó que se trataba de un planeta. Efectivamente, observaciones realizadas por otros astrónomos mostraron que ni su apariencia ni su órbita correspondían a un cometa. Más aún, las órbitas calculadas por Anders Johan Lexell (1740 – 1784) y Pierre Simon marqués de Laplace (1749 – 1827) concordaban con las de un planeta en órbita ligeramente elíptica alrededor del Sol a una distancia doble que la de Saturno.

Curiosamente, este astro había sido observado al menos una veintena de veces desde que en 1690 el primer astrónomo real inglés, John Flamsteed (1646 – 1719), lo había catalogado como una estrella (34 Tau). Antes de Herschel siempre fue considerado como tal. Así pues, no fue Herschel quien lo descubrió sino quien al suponerlo un cometa, pues su posición en 1781 no concordaba con la de ninguna estrella catalogada, despertó un interés por él que llevó al hallazgo de que se trataba de un astro del sistema solar y con una órbita casi circular, como la de los planetas.

⁹La mayor distancia al Sol que alcanza un astro en su órbita alrededor de él.

La sensación que produjo este descubrimiento fue enorme, no en vano se trataba del primero de un planeta en la Historia. Ello fue determinante en la vida de Herschel, pudiendo desde entonces dedicarse por enteramente a la exploración del cielo y la construcción de nuevos telescopios, gracias a la protección y salario adjudicados por el rey Jorge III. En agradecimiento Herschel propuso llamar el planeta “Georgius sidus” (estrella de[l rey] Jorge), mientras otros proponían denominarlo Herschel. Así figura, por ejemplo, en el *Tratado de astronomía* que publicó en 1833 su hijo Sir John Frederick William Herschel (1782 – 1871), aunque en este libro también aparece con el nombre de Urano. Este nombre, sugerido por Bode, finalmente se impuso sobre aquéllos y varios más (Océano, Cibeles, Neptuno, ...). Este es el primer planeta con un nombre de dios griego en lugar de romano.

Pronto fue evidente su gran tamaño, más parecido a Saturno que a los planetas interiores. John Herschel escribe:

De Urano o Herschel todo lo que vemos es un disco pequeño, redondo y uniformemente iluminado, sin anillos, bandas, ni manchas discernibles. ... su volumen es como 80 veces el de la Tierra.

En 1787 W. Herschel descubrió los dos primeros satélites de Urano, Titania y Oberón, lo que permitió determinar la masa del planeta. Por aquel entonces recibía muchos encargos de telescopios (llegaría a construir más de medio centenar) con los que se pretendía encontrar nuevos planetas.

La búsqueda de nuevos planetas

El barón húngaro Franz Xaver Freiherr von Zach (1754 – 1832) inició en 1787 la búsqueda sistemática de un planeta entre Marte y Júpiter, animando a otros astrónomos europeos a participar en tal tarea. Uno de ellos, Giuseppe Piazzi (1749 – 1826) tuvo éxito en Palermo la primera noche del siglo XIX. A partir del 1 de enero de 1801 fue observando la trayectoria de un astro con respecto de las estrellas, concluyendo que se trataba de un cometa sin cola. En cambio von Zach, a pesar de que su órbita no había sido bien determinada, publicó inmediatamente un artículo titulado *Sobre un largamente supuesto y ahora probablemente descubierto nuevo planeta primario de nuestro sistema solar entre Marte y Júpiter*. Finalmente su órbita fue calculada por un joven Karl Friedrich Gauss (1777 – 1855), obteniendo una distancia media al Sol de 2,77 UA. Piazzi propuso llamarlo Ceres Ferdinanda, en honor a la diosa patrona de Sicilia y a Fernando IV, rey de Sicilia, mientras que von Zach propuso Hera. Actualmente se le conoce simplemente como Ceres.

La idea de que Ceres fuera un nuevo planeta del sistema solar pronto se enfrentó a grandes dificultades. La primera, su pequeño tamaño: W. Herschel calculó que su diámetro era inferior a 261 km (realmente es de unos 950 km). La segunda, su órbita inclinada más de 10° con respecto del plano de la órbita terrestre, del que los demás se separan menos de 4° , salvo Mercurio cuya inclinación es de 7° . La tercera y definitiva, el descubrimiento por Heinrich Wilhelm Matthäus Olbers (1758 – 1840) en marzo de 1802 de Palas, un segundo objeto parecido en tamaño y brillo a Ceres y con una distancia media al Sol muy similar, de 2,67 UA.



Figura 5: Telescopio de 7 pies de longitud construido por W. Herschel para el Observatorio Astronómico de Madrid. Es similar a aquel con el que descubrió Urano.

Tras el descubrimiento de Olbers se planteó la cuestión de cómo denominar estos astros. Piazzi propuso llamarlos “planetoides”, en razón de su pequeño tamaño. W. Herschel propuso “asteroides” pues, aunque no se trata de estrellas, aparecen al telescopio como tales. En el mismo año 1802 Herschel propuso una serie de criterios para decidir cuando denominar planeta o cometa a un astro. Ceres y Palas no podían ser considerados ni lo uno ni lo otro, por lo que se establecía una nueva categoría de objetos en el

sistema solar, imponiéndose la denominación de asteroides. En 1804 Karl Ludwig Harding (1765 – 1834) descubrió el tercero de ellos, Juno, y en 1807 Olbers descubrió el cuarto, Vesta. Al terminar el s. XIX se conocían 463 asteroides.

Los siete criterios que, según W. Herschel, debe satisfacer un astro del sistema solar para ser llamado planeta son: tamaño considerable, movimiento poco elíptico alrededor del Sol, órbita inclinada pocos grados respecto de la terrestre, movimiento en la misma dirección que la Tierra, distancia grande con sus planetas vecinos, puede tener satélites y anillos, y debe tener una atmósfera considerable (criterio éste que ni Mercurio ni Marte satisfacen, algo que Herschel ignoraba). Ceres no cumplía todos estos requisitos, por lo que sólo pudo ser considerado planeta durante un año.

Fracasados los intentos de descubrir un planeta principal entre Marte y Júpiter, la búsqueda debía centrarse en encontrarlo más allá de Urano. La posibilidad de que hubiera un gran planeta transurano venía apoyada por los fracasos en calcular con precisión la posición de Urano, aún teniendo en cuenta las perturbaciones gravitatorias debidas a Júpiter y Saturno. En la década de 1830, especulaciones sobre tal hipotético planeta aparecían incluso en libros de divulgación, como el famoso libro de Mary Fairfax Greig Somerville (1780 – 1872) *Sobre la conexión de las ciencias físicas*, publicado en 1834.

En la década siguiente Urbain Jean Joseph Leverrier (1811 – 1877) y, privadamente, John Couch Adams (1819 – 1892) propusieron independientemente unos elementos orbitales del desconocido planeta perturbador y en qué zona concreta del cielo los observadores debían buscarlo. El 23 de septiembre de 1846 Johann Gottlieb Galle (1812 – 1910), acompañado de su asistente Heinrich Louis d'Arrest (1822 – 1875), descubrió dicho planeta con un refractor de 23 cm de apertura. La noche siguiente determinaron su tamaño angular resultando, si se confirmaba su distancia, ser otro planeta grande aunque menor que Júpiter y Saturno. Unos días después William Lassell (1799 – 1880) descubrió su primer satélite, Tritón.

Como en el caso de Urano, el nuevo planeta había sido observado mucho antes, en 1612 por Galileo. Su lento movimiento propio y los escasos aumentos del telescopio de Galileo le impidieron darse cuenta de que no se trataba de una estrella. Asimismo, hubo desacuerdos con el nombre del nuevo planeta. Leverrier inicialmente sugirió Neptuno, nombre de un dios romano que von Zach no habría aprobado, dado que sugirió en 1801 denominar los nuevos planetas con nombres griegos a fin de distinguirlos de los conocidos en la antigüedad. En tal caso podía haberse denominado Posidón, el equivalente griego a Neptuno.

A lo largo del s. XIX poco se concluyó acerca de la naturaleza de los planetas. Se desconocía el estado en que estaba la materia acumulada a partir de una nebulosa primordial, pero se solía considerar que tenían

una superficie sólida y que, salvo Marte, estaban rodeados de una espesa atmósfera que no permitía ver detalles de su superficie (caso de Venus y, erróneamente, de Mercurio) o cuya estructura y variaciones observamos, como en el caso de Júpiter. A principios del siglo la masa y la densidad se estimaban de manera muy precaria, salvo la buena determinación para la Tierra hecha en 1797 por Henry Cavendish (1731 – 1810). En la segunda mitad de siglo se conocía que los cuatro planetas interiores tienen una densidad media unas 4 veces mayor que los planetas exteriores. Estos, además, son más grandes, están achatados por su rápida rotación y van acompañados de varios satélites. En esta época, el progresivo desarrollo del análisis espectroscópico de la luz aplicado a la astronomía permitió conocer algunos de los componentes gaseosos de las atmósferas planetarias.

El increíble planeta menguante

A lo largo del s. XIX y principios del XX hubo intentos de predecir la existencia de un planeta transneptuniano (situado más allá de Neptuno), complementados con búsquedas basadas en la nueva y eficiente técnica de la fotografía astronómica que permitía detectar astros mucho menos luminosos que las observaciones visuales. Tal búsqueda se prolongó durante décadas, hasta que el 18 de febrero de 1930 Clyde William Tombaugh (1906 – 1997) descubrió un nuevo planeta comparando fotografías tomadas con un telescopio de 33 cm de apertura los días 23 y 29 de enero. El descubrimiento fue hecho público el 13 de marzo, aniversario del de Urano. Como en los casos precedentes, se sugirieron varios nombres, imponiéndose finalmente el de Plutón, sugerido por una niña inglesa, y también había sido observado antes, pues aparece en fotografías tomadas al menos desde 1908.

La suposición inicial de que el nuevo planeta era el causante de las irregularidades en la órbita de Neptuno había llevado a suponer que su masa era unas 7 veces la de la Tierra¹⁰, a lo que correspondería un tamaño tres veces mayor que ésta (o sea, unos 40.000 km) si su densidad fuera similar a la de los demás planetas grandes. Sin embargo su baja magnitud no concordaba con un tamaño tan grande, sino más bien con un tamaño menor que el de la Tierra. En su libro *Iniciación a la astronomía*, William Herbert Steavenson (1894 – 1975) escribe en 1933:

Plutón se encuentra tan distante, que no hay telescopio con que se llegue a apreciar su disco. Sin embargo, su brillo aparente induce a creer que su tamaño es algo semejante al de Marte.

¹⁰La masa de la Tierra es: $1 \mathcal{M}_{\oplus} \approx 6$ cuatrillones (10^{24}) de kg. Su volumen es de aproximadamente 1 billón (10^{12}) de km^3 .

Suponer un tamaño mitad al de la Tierra implicaba también suponer que se trataba de un planeta denso, quizás como los cuatro primeros (densidad media de $5,0 \text{ g/cm}^3$), y no de un planeta gigante ligero (densidad media de $1,2 \text{ g/cm}^3$), como los cuatro siguientes. Además se distinguía de éstos por tener un órbita más excéntrica que ningún otro planeta y muy inclinada, de unos 17° , tanto que según los criterios de W. Herschel no debía ser considerado planeta sino asteroide.

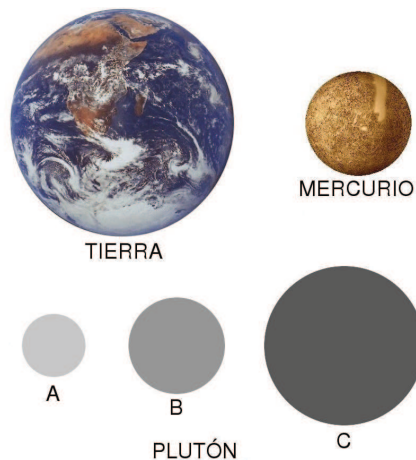


Figura 6: Dimensiones de Plutón basadas en distintos valores (A, B y C) de su reflectividad, comparadas a las de la Tierra y Mercurio. Ilustración inspirada en la del bello libro *Sur les autres mondes* (1937) del pintor y astrónomo aficionado Lucien Rudaux (1874 – 1947).

Dos dificultades añadidas eran comentadas por Josep Comas Solà (1868 – 1937) en su libro *Astronomía* (1935):

Si... se tiene en cuenta que su masa debe de ser considerable para explicar las perturbaciones observadas, hay que admitir que su densidad es muy grande. El ángulo considerable que forma el plano de la órbita de Plutón con la eclíptica y la proximidad relativa a que pasa de Neptuno, inducen a suponer que aquel planeta es un astro cosmogónicamente extraño al sistema solar y que fué capturado en tiempos remotos por Neptuno.

Es decir, recién descubierto ya se ponía en duda que Plutón se hubiese formado junto con los demás planetas. A pesar de todos estos argumentos

en contra, Plutón aparece en tales libros como el noveno planeta del sistema solar.

En 1976 la UAI recomendaba utilizar 5.000 km como diámetro de Plutón y una masa del 11% de la de la Tierra. De hecho, su masa no pudo conocerse hasta 1978, año en que se descubrió su primer satélite conocido¹¹, Caronte. El efecto inmediato de este descubrimiento fue el de disminuir el tamaño estimado para Plutón, pues su brillo debía repartirse entre ambos astros. Además, se pudo calcular por primera vez la masa de Plutón, que resultó ser cercana al 0,2% $M_{\oplus}$, 50 veces menor que lo supuesto. Este valor de la masa concordaba con un tamaño pequeño incluso si se suponía una densidad más baja que la de los planetas interiores.



Figura 7: Fotografía de Plutón y Caronte tomada en 1994 con la cámara para objetos débiles (FOC) de la Agencia Europea del Espacio instalada en el Telescopio Espacial Hubble. (Cortesía de NASA/ESA.)

En 1980 la ocultación de una estrella por Caronte mostró que el tamaño de este satélite excede los 1.200 km. Medidas radioastronómicas, infrarrojas y la serie de eclipses de Caronte por Plutón que se dieron entre 1985 y 1990 permitieron por fin determinar el tamaño de Plutón, que ha resultado estar entre 2.200 y 2.400 km, y su densidad, intermedia entre la de los planetas rocosos y la de los gaseosos.

En conclusión, se pasó de considerar que Plutón podría tener un diámetro de 40.000 km a medirse unos 2.400 km. Se trata, pues, de un planeta bastante más pequeño que todos los demás (el menor, Mercurio, tiene un diámetro de 4.879 km), menor incluso que muchos satélites (la Luna tiene 3.476 km y el mayor, Ganimedes, 5.262 km).

¹¹En 2006 se han descubierto dos nuevos satélites de Plutón, denominados Nix e Hidra.

El sistema solar a finales del siglo XX

En la década de 1990 nuestra visión del sistema solar iba a cambiar una vez más. Al inicio de la década se consideraba que el sistema solar se componía principalmente del Sol, nueve planetas (sólo dos de ellos importantes, Júpiter y Saturno, pues contienen más del 90 % de la masa planetaria), 61 satélites, unos 20.000 asteroides y unos 800 cometas.

Las órbitas de los asteroides conocidos no se limitan ya al espacio que hay entre Marte y Júpiter, conocido como “cinturón de asteroides”, sino que se les puede encontrar en órbitas que cruzan la de la Tierra como la de Atenas, el primero de tales asteroides, descubierto en 1932. Otros tienen órbitas caóticas entre Saturno y Neptuno, los denominados “centauro”, pues el primero de los cuales es Quirón, descubierto en 1977. Durante unos años a partir de 1988 Quirón aumentó su brillo, como si se tratara de un cometa, a pesar de que su tamaño de unos 200 km es un orden de magnitud mayor que el típico del núcleo de un cometa.

En 1992 se descubrió 1992 QB₁, el primero de una serie de objetos cuya órbita se encuentra más allá de la de Neptuno. En pocos años se descubrió que había muchos objetos en la región situada entre 30 y 50 UA del Sol, que pasó a denominarse “cinturón de Kuiper”¹². En esta misma década se descubrió que los asteroides más alejados del Sol parecen contener más hielos. Con todo ello parece difuminarse la diferencia clásica entre cometas y asteroides: un asteroide rico en hielos de materiales volátiles puede convertirse en un cometa si se acerca suficientemente al Sol; un cometa que ha perdido todos sus volátiles aparecerá como un asteroide de baja densidad.

Aún hubo mayores sorpresas. El ritmo de descubrimiento de objetos transneptunianos ha aumentado encontrándose algunos de gran tamaño, incluso mayores que Ceres, el mayor de los asteroides. Reciben un nombre oficial generalmente basado en divinidades de la creación o del infierno en diversas mitologías: Quaoar, Sedna, Orco, Varuna, Ixión. Las medidas realizadas con diversas técnicas sugieren que se trata de objetos similares a Plutón o a Tritón, es decir con una densidad de unos 2 g/cm³ y un elevado albedo¹³, muy superior al de los asteroides. En 2005 Michael E. Brown (1965) y sus colaboradores descubrieron un objeto inicialmente catalogado como 2003 UB₃₁₃ y actualmente nombrado Eris o Éride, siguiendo su sugerencia de usar el nombre griego de la diosa romana Discordia. Las primeras medidas revelaron que su tamaño excede ligeramente el de Plutón, por lo que no había razón para no considerarlo a su vez un planeta: se habló del décimo planeta del sistema solar.

Parece probable que el elevado ritmo de descubrimiento de este tipo de objetos pueda llevar al descubrimiento de nuevos objetos transneptunianos

¹²En honor a Gerrit Pieter Kuiper (1905 – 1973).

¹³Fracción de energía luminosa reflejada o difundida por un cuerpo no luminoso.

mayores que Plutón. Por ello, pronto se planteó el debate acerca de la idoneidad de considerar o descartar como planetas estos astros. Un grupo de trabajo establecido por la UAI con miembros de la División III (Ciencias de los sistemas planetarios) abordó el tema considerando aspectos físicos y dinámicos, pero no alcanzó un consenso en aspectos culturales como es el estatus de Plutón. Posteriormente se formó un comité más diverso, cuyos miembros presentaron en agosto de 2006 una propuesta a la Asamblea General de la UAI en Praga. Según ella, planeta es cualquier cuerpo celeste que orbita alrededor de una estrella (sin ser estrella o satélite de un planeta) y que tiene masa suficiente para que su fuerza de gravedad supere las fuerzas de sólido rígido y dicho cuerpo alcance una forma (casi) esférica de equilibrio hidrostático. Esta propuesta, que daba pie a que el número de planetas en el sistema solar pasara inmediatamente a ser de 12 y que pudiera llegar a ser de varias decenas, fue mayoritariamente rechazada.

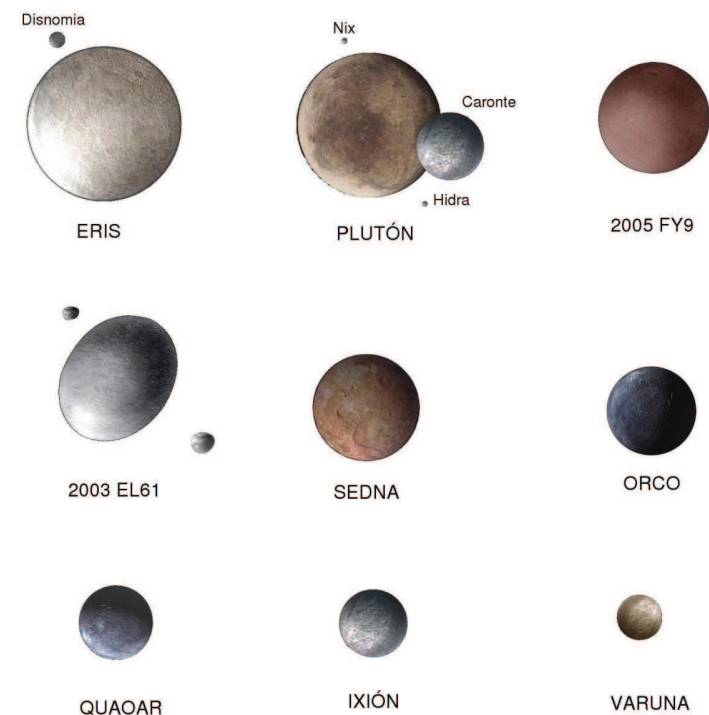


Figura 8: Dimensiones estimadas de algunos de los mayores objetos transneptunianos conocidos comparadas con las de Plutón. (Adaptada de una ilustración de NASA.)

En su lugar se aprobó una resolución (5A) mucho menos ambiciosa, pues sólo se refiere a los planetas del sistema solar. Se definen tres términos:

Planeta. Es un cuerpo celeste que orbita alrededor del Sol, que tiene masa suficiente para que su fuerza de gravedad supere las fuerzas de sólido rígido y dicho cuerpo alcance una forma (casi) esférica de equilibrio hidrostático y que ha “limpiado” la vecindad de su órbita. (En la práctica, ello es lo mismo que decir que hay 8 planetas y éstos son: Mercurio, Venus, Tierra, Marte, Júpiter, Saturno, Urano y Neptuno.)

Planeta enano. Es un cuerpo celeste que orbita alrededor del Sol, que tiene masa suficiente para que su fuerza de gravedad supere las fuerzas de sólido rígido y dicho cuerpo alcance una forma (casi) esférica de equilibrio hidrostático, que no ha “limpiado” la vecindad de su órbita y que no es un satélite. (Hasta la fecha se han incorporado a esta categoría el asteroide Ceres, y los objetos transneptunianos Plutón y Eris.)

Cuerpos menores del sistema solar. Son todos los demás objetos que orbitan alrededor del Sol, exceptuando los satélites. (Esta categoría incluye la mayoría de asteroides y de objetos transneptunianos, los cometas y otros cuerpos aún menores.)

Se aprobó, asimismo, la resolución 6 que establece que Plutón es un planeta enano, según la definición anterior, y que es el prototipo de una nueva categoría de objetos transneptunianos que aún no ha recibido un nombre oficial. Las definiciones anteriores no permiten establecer una masa o un tamaño mínimos para que un astro sea considerado planeta, pues el alcanzar una forma casi esférica depende también de su composición química, su densidad y su temperatura.

En definitiva, actualmente se considera que el sistema solar se compone del Sol, 8 planetas con 152 satélites, 3 planetas enanos con 3 satélites, unos 340.000 asteroides, más de 1.000 objetos transneptunianos y unos 2.500 cometas, además de centenares de miles de astros de un tipo u otro que presumiblemente aún quedan por descubrir.

Otros sistemas planetarios

En los últimos quince años también hemos vivido una revolución en el conocimiento de otros sistemas planetarios, cuya existencia hasta hace poco era una simple hipótesis. En 1992 se descubrió el primer planeta alrededor de otra estrella, aparentemente un planeta de tipo terrestre orbitando alrededor de una estrella de neutrones detectada como púlsar, un tipo de estrella moribunda. En 1995 se descubrió el primero de los planetas gigantes alrededor de una estrella normal. (Véase el artículo *El*

descubrimiento de planetas alrededor de otros soles de Francisco Colomer publicado en el ANUARIO de 1998.) Actualmente se conocen más de doscientos “exoplanetas”, como a veces son conocidos, la mayoría de los cuales tienen una masa superior a la de Júpiter¹⁴, alcanzando incluso 17 veces su masa. La fracción de sistemas con varios planetas es bajo (10%), aunque este porcentaje probablemente crecerá a medida que las técnicas de detección mejoren.

La existencia de estos planetas tendrá sus implicaciones en una posible futura definición más general de planeta. En particular, habrá que establecer una masa máxima por encima de la cual deba hablarse de otro tipo de astro, como es una “enana marrón”. Se trata de un objeto con una masa menor que la necesaria para mantener las reacciones de fusión nuclear que se dan en la mayoría de estrellas, pero que es suficiente para que su interior sea homogéneo, sin estratos con distinta composición química, debido al efecto de una convección global. Los límites mínimo y máximo para que un objeto pueda ser denominado enana marrón no están muy claros y son objeto de intenso estudio. Se estima que son del orden de 13 y 75 veces la masa de Júpiter, respectivamente.

En cuanto a los objetos subestelares que flotan libremente en el espacio sin estar ligados gravitatoriamente a una estrella, la UAI recomienda no llamarlos planetas. Se denominarán enanas marrones, si su masa es suficientemente grande, y subenanas marrones, si su masa es inferior a unas 13 veces la de Júpiter. Su naturaleza y propiedades físicas están aún bajo discusión.

En conclusión, la definición apresurada de planeta aprobada recientemente por la UAI es válida sólo para el sistema solar y presenta varias lagunas, como son la necesidad de decidir en comité qué objetos son considerados planetas enanos, la inadecuación de la definición de éstos (según la cual un planeta enano no es un planeta) y la falta de una denominación específica para la clase de objetos de las que Plutón es prototipo. Quizás hubiese sido más conveniente simplemente establecer convenios flexibles según los cuales denominar, en base al conocimiento de cada época, las familias de astros a fin de disponer de una nomenclatura común.

Por otra parte, la definición actual, aún con sus limitaciones, engloba objetos de distinta naturaleza y serían aún más diversos si se incluyeran los planetas encontrados alrededor de otras estrellas. En efecto, la riqueza del concepto actual de planeta es mucho mayor que lo imaginado en la antigüedad.

¹⁴La masa de Júpiter es $1M_J \simeq 318M_{\oplus}$, $13M_J \simeq 1/80 M_{\odot}$.

Referencias

Para ampliar conocimientos sobre la consideración de los planetas en distintas épocas:

T. Heath, *Aristarcus of Samos. The ancient Copernicus*, Clarendon Press, 1913 (reeditado por Dover, 1981).

A. van Helden, *Measuring the Universe. Cosmic dimensions from Aristarcus to Halley*, U. Chicago Press, 1985.

E. Grant, *Planets, stars and orbs. The medieval cosmos, 1200 – 1687*, CUP, 1996.

R. Taton y C. Wilson (eds.), *Planetary astronomy from the Renaissance to the rise of astrophysics* (2 volúmenes), CUP, 1989 y 1995.

J.H. Shirley y R.W. Fairbirdge (eds.), *Encyclopedia of planetary sciences*, Chapman & Hall, 1997

http://en.wikipedia.org/wiki/Definition_of_planet

PUESTA AL DÍA DEL TEMA

Más y más planetas

El artículo **La idea de planeta a lo largo de la historia** se publicó en el ANUARIO del año 2007. Al año siguiente la Unión Astronómica Internacional (UAI) incluyó en la lista de planetas enanos del sistema solar a Makemake y Haumea, a pesar de que todavía no se conocía con precisión su forma, pues se estimaba que su tamaño era mayor que el de Ceres, dado que en ambos casos su magnitud absoluta¹ es $H < 1$, de lo que se deduce un diámetro mayor de 836 km (que corresponde a un albedo geométrico máximo de 1). Medidas posteriores realizadas por el equipo de M.E. Brown, de Caltech, indican que Makemake es casi esférico, con un diámetro medio de unos 1430 km, y que Haumea tiene forma de elipsoide achatado ($1920 \times 1540 \times 990$ km), con un diámetro medio muy similar al anterior. A pesar de su forma alargada, Haumea se considera planeta enano pues tiene masa suficiente como para adquirir forma esférica, aunque se lo impide su rápida rotación (su período sidéreo es de 3 h 55 min).

Las dimensiones de los planetas enanos Ceres y Plutón han sido revisadas recientemente como consecuencia de las exploraciones realizadas en 2015 mediante sondas interplanetarias. Las dimensiones del mejor elipsoide que se ajusta a la forma de Ceres determinadas por la sonda *Dawn* son $965 \times 961 \times 891$ km. En cuanto a Plutón, las medidas de la sonda *New Horizons* muestran que tiene un diámetro medio de 2377 km y un achatamiento inapreciable, menor del 0,6 % (Nimmo et al 2016). Medidas de una ocultación estelar observada desde varios observatorios terrestres han permitido determinar que el diámetro de Eris es de 2326 km (Sicardy et al 2011), algo menor que el de Plutón.

Actualmente se conocen casi 2000 objetos transneptunianos (cuya distancia media al Sol es mayor que la de Neptuno, 30 ua), alguno de los cuales, tras una caracterización más detallada de sus propiedades físicas, podría pasar a engrosar la lista de planetas enanos.

Las definiciones de planeta y planeta enano acordadas en 2006 durante la 26ª Asamblea General de la UAI siguen siendo discutidas, tanto por la vaguedad de la definición de planeta enano, como por el hecho de que la definición de planeta se limitó a los objetos del sistema solar. Es decir, los planetas extrasolares, también llamados exoplanetas, no están incluidos en tal definición.

¹La magnitud absoluta, H , es la magnitud que se observaría si su ángulo de fase fuera nulo y se encontrara a una distancia de 1 ua tanto del Sol como de la Tierra.

El número de exoplanetas también se ha incrementado mucho desde el año en que se publicó el artículo original en el ANUARIO. Se conocían unos 200 y en la actualidad se conocen miles, sobretodo gracias a las observaciones realizadas con el telescopio espacial *Kepler* y con el instrumento HARPS de ESO. Concretamente, desde febrero de 1995 hasta 2017 se han descubierto más de 3500 planetas fuera del sistema solar (exoplanetas), formando parte de unos 2700 sistemas planetarios, más de 600 de los cuales cuentan con dos o más planetas. Varios de ellos cuentan con un número de planetas comparable al del sistema solar. El exoplaneta más cercano orbita alrededor de Próxima Centauri, estrella situada a 4,22 años luz del Sol, y fue descubierto en 2016 por un equipo liderado por el astrónomo español Guillem Anglada-Escudé. Las revistas *Nature* y *Science* incluyeron su descubrimiento como uno de los diez más importantes del año².

Los miles de exoplanetas conocidos muestran una gran variedad de características, tanto las físicas y orbitales del planeta, como las de la estrella o sistema estelar que orbitan. La variedad de objetos descubiertos dificulta el establecimiento de una definición suficientemente genérica de planeta (o de exoplaneta) que englobe a todos los objetos que se ajustan a lo que comúnmente consideraríamos un planeta. La imposibilidad de determinar la forma y el dominio de la órbita de los exoplanetas reduce las opciones de definirlos a considerar un límite superior a su masa, por ejemplo que ésta sea insuficiente para ser considerado una enana marrón³.

Actualmente se considera exoplaneta un objeto que orbita alrededor de un sistema estelar (puede tratarse de un sistema múltiple o incluso de una estrella difunta, si se incluyen objetos como los púlsares) y cuya masa estimada es menor que 13 veces la masa de Júpiter. El hecho de haber sido detectado y determinada su masa mediante observaciones realizadas desde la Tierra nos asegura que tiene masa suficiente como para haber adquirido una forma (casi) esférica, como en el caso de los planetas del sistema solar. El exoplaneta de menor masa detectado alrededor de una estrella hasta la fecha tiene una masa y un tamaño parecidos a los de Marte. Si se quiere imponer, como en el caso del sistema solar, la condición de que el planeta pueda limpiar o dominar dinámicamente su órbita, la masa mínima puede depender también del tamaño de su órbita y de la masa del sistema estelar, así como de la masa y la órbita de los demás planetas del sistema.

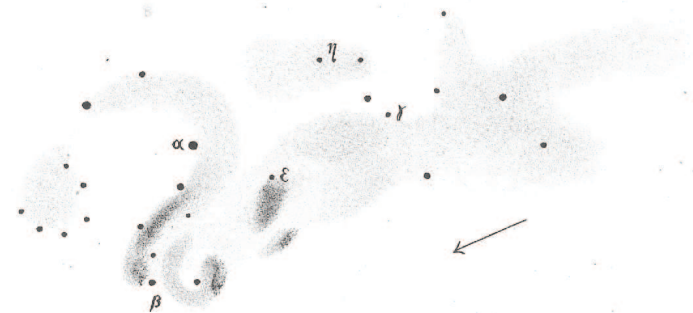
En definitiva, la idea de planeta evolucionará según vayamos caracterizando la gran variedad de objetos planetarios que se descubren.

Pere Planesas Bigas, enero 2017

²Otro, de importancia astrofísica, fue la primera detección de ondas gravitacionales.

³Objeto con masa suficiente para que se dé de manera significativa la fusión termonuclear del deuterio pero insuficiente para que se dé la del hidrógeno. Aunque la masa mínima depende de la metalicidad y otros factores, se suele adoptar el valor canónico de 13 veces la masa de Júpiter, lo que equivale a $0,012 M_{\odot}$ o a unas 4000 veces la masa de la Tierra.

EL SOL Y LAS ESTRELLAS



EL SOL: NUESTRA ESTRELLA, NUESTRA ENERGÍA

Rafael Bachiller

*Observatorio Astronómico Nacional
Instituto Geográfico Nacional - Ministerio de Fomento*

Abstract

Beyond its purely astronomical importance, the Sun is of great interest these days as it is considered to be a practically never-ending source of energy. Specifically, insolation data are particularly important for the planning of photovoltaic solar plants. This article describes, from the viewpoint of the astronomer, what the Sun is and how it works, how it was formed and how long it will last. We emphasize the origin of solar energy and the amount of it that we receive on Earth.

Introducción

La creciente demanda de energía en nuestro mundo hace que el hombre mire al Sol con nuevos ojos: como una fuente aparentemente inagotable de energía que puede cubrir todas nuestras necesidades. El Sol se nos revela ahora como una fuente de energía que puede ser explotada por la Humanidad de manera mucho más eficaz y exhaustiva de lo que lo es en la actualidad.

En efecto, directa o indirectamente, el Sol es la fuente primaria de toda la energía en la Tierra. La luz solar hace vivir a todos los organismos en nuestro planeta. Esta misma radiación solar, asimilada en la fotosíntesis, se transforma en plantas, y las plantas son el alimento básico de muchos animales. La descomposición de las plantas y de otros seres vivos hace millones de años que almacenó la energía solar en forma de combustibles fósiles (carbón, petróleo y gas natural). El Sol es también la causa de muchos

otros fenómenos energéticos en la Tierra, como los vientos atmosféricos o las corrientes oceánicas. La energía nuclear generada en la Tierra tiene su origen en los elementos pesados que se crearon bien en el centro del Sol o de alguna otra estrella. La propia Tierra, junto con los otros planetas del Sistema Solar, no es más que un producto, en cierto modo “secundario”, creado durante la formación del Sol.

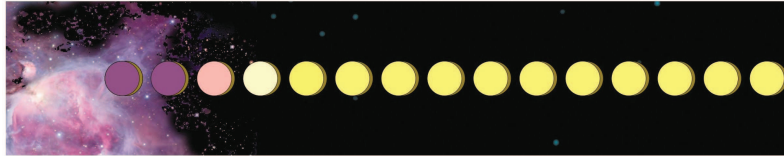


Figura 1: Representación de la evolución del Sol desde su formación en una nube interestelar, pasando por la fase de secuencia principal...

El Sol: una estrella banal

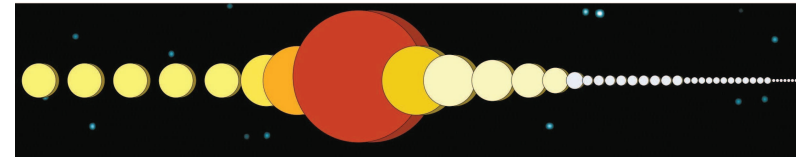
El Sol parece ser una estrella anodina situada en una galaxia anodina. En términos astronómicos, el Sol se clasifica como una estrella de tipo espectral G2 y clase de luminosidad V: una estrella “G2V”. Esto es una manera rápida de decir que es una estrella enana y amarilla, con temperatura superficial de 5 780 K¹, compuesta esencialmente por Hidrógeno (74 % en masa y 92 % del volumen) y Helio (24,5 % en masa y 7 % del volumen), a lo que se añaden pequeñas trazas de elementos pesados como Hierro, Níquel, Oxígeno, Silicio, Azufre, Magnesio, Carbono, Neón, Calcio y Cromo.

El Sol está situado en una región que no parece tener nada de especial: a unos 26 000 años luz del centro de la Galaxia (que tiene un radio de unos 60 000 años luz), en el borde interno del brazo espiral de Orión. En la Galaxia, el Sol se mueve a una velocidad de 214 km s⁻¹, lo que le permite recorrer una distancia de un año luz cada 1 400 años. Hay unos 200 mil millones de estrellas en nuestra Galaxia, la Vía Láctea, de las que más de 100 millones son de tipo G2, y aunque la mayor parte (85 %) de esas estrellas son enanas rojas, menos brillantes que el Sol, nuestro Sol es un millón veces menos luminoso que las estrellas más luminosas de la Vía Láctea. Y nuestra Galaxia no es más que una galaxia media de los miles de millones del Universo observable. Así pues, desde el punto de vista astronómico, el Sol es una pequeña estrella perdida en una galaxia banal.

Pero, para la Tierra y para nosotros sus habitantes, el Sol es una estrella única: nuestra estrella. Durante el día, el Sol hace palidecer a los otros cuerpos celestes, su brillo cegador no nos permite mirarlo directamente. El Sol

¹El Kelvin (K) es la unidad de temperatura que tiene el punto cero en -273 grados Celsius, el cero absoluto

ha sido objeto de curiosidad y estudio para todas las civilizaciones. Muchas comunidades antiguas lo consideraban un dios y le dedicaron monumentos y observatorios. En Egipto, en Grecia, en América, en el Extremo Oriente, en todo el mundo podemos encontrar innumerables ejemplos y curiosidades que ilustran la fascinación y el reconocimiento de la luz y calor proporcionados por el Sol como fuente de vida y energía.



...hasta su transformación en gigante roja y posterior ocaso como enana blanca (dibujo de F. Martín, OAN).

Aunque para el astrónomo el Sol parece una estrella anodina, como tantas otras, es muy difícil encontrar estrellas exactamente gemelas del Sol. Para algunas aplicaciones astrofísicas (calibración) tales gemelos resultan importantes pues el Sol, aunque es el patrón para ciertas medidas, está demasiado cerca y es demasiado brillante para calibrar nuestros instrumentos. Búsquedas cuidadosas de estrellas idénticas al Sol han dado resultados poco satisfactorios. Hasta ahora, las estrellas que se consideran más parecidas al Sol son 18 Scorpii (HD142633) e HIP56984 (artículo de Meléndez y Ramírez, 2007, *The Astrophysical Journal* 669, L89). 18 Sco es similar al Sol en muchos aspectos, pero difiere en otros, por ejemplo tiene una abundancia de litio que es 3 veces superior a la solar. HIP56984, sin embargo, parece idéntica al Sol dentro de las precisiones de las medidas. Esta estrella, situada a unos 200 años luz de distancia de nuestro Sol, tiene la misma edad que nuestra estrella y es por tanto un candidato excelente para la búsqueda de planetas similares a la Tierra. Si se encontrasen indicios de tales planetas, éstos serían candidatos privilegiados para emprender la búsqueda de vida extraterrestre.

Origen y evolución del Sol

El mecanismo por el que se formó el Sol fue objeto de un artículo previo en este Anuario (Mario Tafalla, *¿Cómo nació el Sol? Formación de estrellas de tipo solar*, Anuario del Observatorio Astronómico 2002). Tal y como se refiere en ese artículo, el Sol es una estrella de tercera generación que se formó hace unos 4 600 millones de años, es decir cuando el Universo tenía unos 9 000 años de edad (su edad actual es de 13 600 millones de años). Las teorías más avanzadas sobre la formación estelar

nos enseñan que el Sol se formó en una región particularmente densa de una nube interestelar. Estas nubes, que abundan en nuestra Galaxia, están formadas por gas y polvo (pequeñas partículas de material en estado sólido). La fuerza de la gravedad (el propio peso de la nube) hizo que en un determinado momento la nube comenzase a contraerse. Para alcanzar las densidades típicas observadas en las estrellas, una nube interestelar ha de comprimirse 20 órdenes de magnitud (es decir por un factor 10^{20}). Si la nube inicial fuese más o menos esférica, el radio de dicha esfera hubo de hacerse un millón de veces más pequeño en su proceso de formación del Sol.

Es la fuerza de la gravedad, pues, el motor de la formación estelar, la causa de las altas densidades alcanzadas en el centro del Sol, y el mecanismo mediante el que se desencadenarán allí las reacciones nucleares. Durante el colapso gravitatorio, la parte del material que está animada de un pequeño movimiento de rotación, no cae directamente sobre el proto-Sol sino que forma un disco que gira en torno al objeto central. De la misma manera que un patinador gira más y más aprisa según recoge sus brazos, la materia de ese disco acelera su rotación según se acerca hacia el centro. Según pasa el tiempo, el material de este disco se va aglomerando en objetos sólidos (planetesimales) que tras irse combinando en objetos mayores acaban dando lugar a un conjunto de planetas: el Sistema Solar.

En una estrella como el Sol, se estima que tras unos 30 millones de años de contracción gradual, la temperatura y la densidad centrales alcanzan los valores necesarios para que comience la fusión nuclear (la conversión de hidrógeno a helio). El inicio de la fusión marca el comienzo de la edad madura, lo que en términos de los astrónomos conocemos como la llegada a la Secuencia Principal. Esta fase será la más larga en la vida de nuestra estrella, pues el hidrógeno (el 70% de la masa total) es suficiente como para alimentar el reactor nuclear durante unos 10 000 millones de años. Pero una vez agotado este combustible, el Sol deberá sufrir unos reajustes estructurales muy dramáticos que lo llevarán a una fase de gigante roja (una estrella grande y brillante pero relativamente inestable). Tras una nueva (pero corta) fase de estabilidad en la que el helio será el combustible nuclear que contrarrestará la fuerza de la gravedad mediante su conversión a carbono, el Sol sucumbirá a la gravedad y se convertirá en una estrella enana blanca (prácticamente inerte), perdiendo en estos procesos (en forma de explosiones y eyecciones diversas) una buena proporción de su masa.

Así genera el Sol su energía

Se dice a menudo que el Sol es una esfera de gas en ebullición. En concreto, la zona más interna del Sol, es decir la que abarca desde el centro hasta unos 0,2 radios solares, se encuentra a una temperatura muy elevada (hasta 15 millones de grados) y constituye, de hecho, un inmenso reactor

nuclear. Como hemos mencionado, las tres cuartas partes de la masa del Sol están constituidas por átomos de hidrógeno. En la zona central del Sol, los átomos de hidrógeno fusionan para formar átomos de helio. El helio es un gas noble que fue detectado en el Sol antes de ser descubierto en la Tierra, de ahí proviene su nombre: de Helios, un dios solar de la antigua Grecia. Mediante una cadena de reacciones nucleares conocida como cadena p-p (protón-protón), cada cuatro átomos de hidrógeno dan lugar a uno de helio. Pero en la conversión de hidrógeno a helio hay una pequeña fracción de masa (un 0,7%) que se convierte en energía de acuerdo con la fórmula de Einstein $E = mc^2$.

Cada segundo en el interior del Sol unos $3,4 \times 10^{38}$ protones se convierten en núcleos de helio, generando unos $3,86 \times 10^{26}$ watios. Dicho de otro modo, cada segundo unos 580 millones de toneladas de gas hidrógeno se convierten en helio, consumiéndose unos 5 millones de toneladas para producir unos 90 000 millones de megatones de energía pura. Para comparación digamos que una bomba nuclear tiene típicamente varios megatones o decenas de megatones.

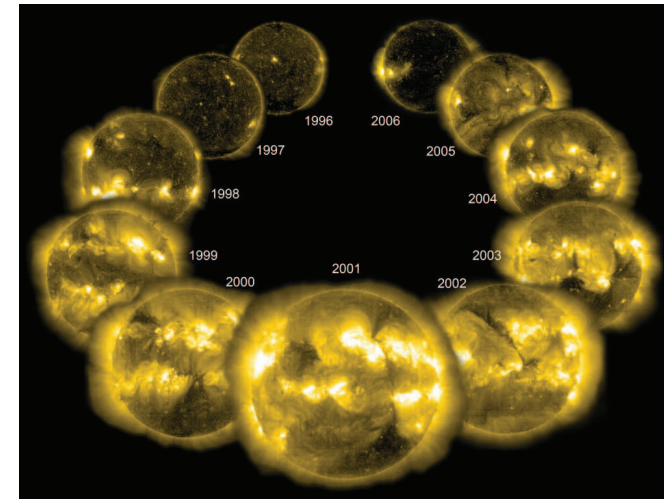


Figura 2: El telescopio espacial SOHO tomó imágenes del Sol, en el extremo ultravioleta, durante uno de sus ciclos completos de actividad de 11 años. El máximo sucedió en 2001 y los mínimos en 1996 y 2007. Cortesía: SOHO (consorcio EIT), ESA y NASA.

El Sol, según va produciendo energía nuclear, va perdiendo masa. Tal y como hemos visto, pierde 5 millones de toneladas de masa cada segundo. Pero, sin embargo, la luminosidad solar permanece muy constante. Las medidas indican que las variaciones de la luminosidad del Sol están por debajo

del 1 % (ni siquiera se ha apreciado una variación significativa con el ciclo de 11 años de las manchas solares). Aunque la potencia nuclear generada en el Sol es enorme, conviene señalar que la eficiencia en generación de energía por unidad de volumen o de masa es muy pequeña: tan sólo de unos $0,3 \mu\text{W cm}^{-3}$ (microWattios por centímetro cúbico), o unos $6 \mu\text{W kg}^{-1}$. Para orientación, esto puede ser comparado con el cuerpo humano que produce unos $1,2 \text{ W kg}^{-1}$, es decir que es varios millones de veces más eficaz. Un reactor artificial de fusión nuclear en la Tierra, para ser eficaz debe trabajar con plasma a temperaturas mucho más elevadas que las que se encuentran en el interior solar.

Como consecuencia del déficit de masa que tiene lugar en la transformación de hidrógeno en helio, las reacciones nucleares generan radiación de alta energía. Los fotones que componen esta radiación son absorbidos y reemitidos muchas veces a lo largo de la trayectoria que recorren desde el centro del Sol. Se estima que la escala de tiempo para que la radiación recorra la parte más densa del Sol (entre 0,2 y 0,7 radios solares) está entre 10 000 y 200 000 años. Esta zona se denomina *zona radiativa*, pues en ella la energía se transmite mediante absorciones y re-emisiones sucesivas de la radiación. Sin embargo, a una distancia del orden de 0,7 radios solares desde el centro, la densidad y la temperatura no son suficientemente altas como para continuar la transferencia de energía mediante radiación. En esa región comienza a operar la convección térmica: el material caliente -menos denso- emerge a la superficie donde se enfría y vuelve a caer a la base de la zona de convección. Se forman grandes células de convección que pueden ser observadas como la estructura granulada de la superficie solar. Así pues, contrariamente a lo que pensaba Anaxágoras, el Sol no es una piedra incandescente sino, como adelantábamos al principio de este apartado, el Sol es una esfera de gas cuya zona exterior se encuentra literalmente en ebullición.

La fotosfera, o superficie visible del Sol, es la capa en la que los fotones encuentran hueco libre para viajar por el espacio. Tiene un espesor de unos 100 km, una temperatura de unos 5 800 K y una densidad relativamente baja, unas 10^{23} partículas por cm^3 , tan sólo un 1 % de la densidad de la atmósfera terrestre a nivel del mar. Por encima de la fotosfera se extiende un gas tenue conocido como *atmósfera solar* que está compuesto por capas, según nos alejamos del centro: la cromosfera, la corona y la heliosfera.

Fenómenos de la superficie solar: manchas solares y fulguraciones

El material ionizado que circula por el globo solar forma una complicada red de corrientes eléctricas que son, a su vez, origen de campos magnéticos. La rotación y los movimientos ascendentes y descendentes de la convección hacen que las líneas magnéticas se compriman en ciertos puntos y

se estiren en otros. A veces las líneas magnéticas atraviesan la superficie del Sol formando bucles en los que las líneas de fuerza emergen en una zona con polaridad positiva y se sumergen en otra zona con polaridad negativa. Por los efectos de los campos magnéticos, estas regiones de la fotosfera son más frías y, por lo tanto, más oscuras, apareciendo como “manchas” en la fotosfera. Típicamente la zona más umbría de las manchas se encuentra a unos 4 200 K (a comparar con los 5 800 K a los que se encuentra la fotosfera). Las manchas suelen aparecer por pares (una positiva junto a una negativa) y por grupos que suelen permanecer visibles durante varias semanas. Naturalmente, las manchas solares son una medida fácil y directa de la actividad solar. Para estimar y tabular de manera sencilla esta actividad se utiliza el número de Wolf: una expresión que combina el número de manchas individuales y el de grupos de manchas.

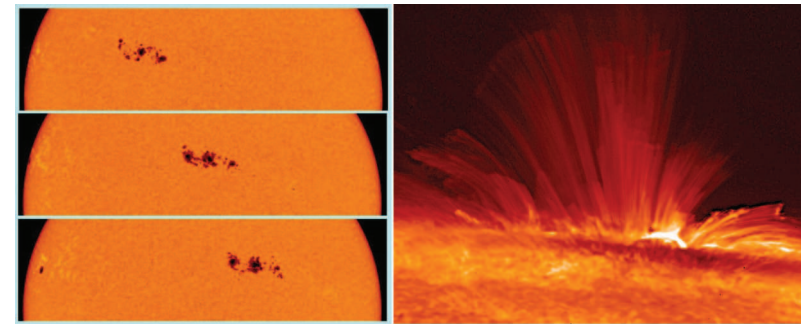


Figura 3: En la imagen izquierda el *Grupo 30* de manchas solares observadas cerca de un máximo de actividad solar por el telescopio espacial SOHO. La derecha muestra la imagen detallada de una mancha solar vista de perfil tomada por el telescopio espacial japonés HINODE. Cortesía SOHO, ESA, NASA, HINODE, JAXA.

El número de manchas solares (el número de Wolf) ha sido medido sin interrupción desde principios del siglo XVIII, pero hay estimaciones sobre cómo ha variado el número de manchas durante los últimos 10 000 años. Se observa que la actividad solar pasa por máximos y mínimos de manera periódica. La duración de este ciclo principal es de 11 años, pero a éste se superpone otro ciclo de unos 80 años. Sin embargo, el efecto de estos ciclos sobre la luminosidad total del Sol es muy pequeño. Las fluctuaciones de la intensidad apenas representan un 0,1 % de la luminosidad total del Sol, por lo que se considera que los efectos sobre la Tierra deben ser poco importantes, aunque éste es aún un tema de estudio.

Asociadas con el fenómeno de las manchas solares, se pueden observar inmensas protuberancias emergiendo de la superficie solar. Se trata de

columnas de material ionizado y a muy alta temperatura (de color rojo profundo) que pueden alcanzar alturas cercanas al millón de kilómetros. Por efecto de los campos magnéticos, frecuentemente las protuberancias se curvan para formar arcos o bucles, por encima de las grandes manchas, que pueden durar varios días.

Las inestabilidades de los campos magnéticos pueden dar origen a grandes explosiones en la fotosfera. Estas explosiones, conocidas como fulguraciones, provocan una intensa emisión de rayos X y gamma y la eyección de un viento de partículas de muy alta energía. Cuando al cabo de unas horas esta lluvia de partículas alcanza la Tierra, se produce una serie de fenómenos asombrosos: las telecomunicaciones se alteran debido a los efectos ocasionados en las capas superiores de la atmósfera y las perturbaciones en el campo magnético terrestre pueden desorientar a nuestros instrumentos de medida (brújulas y derivados) e incluso a las palomas mensajeras. Esta interacción de las eyecciones solares con el campo magnético terrestre es también la causa de las auroras boreales.

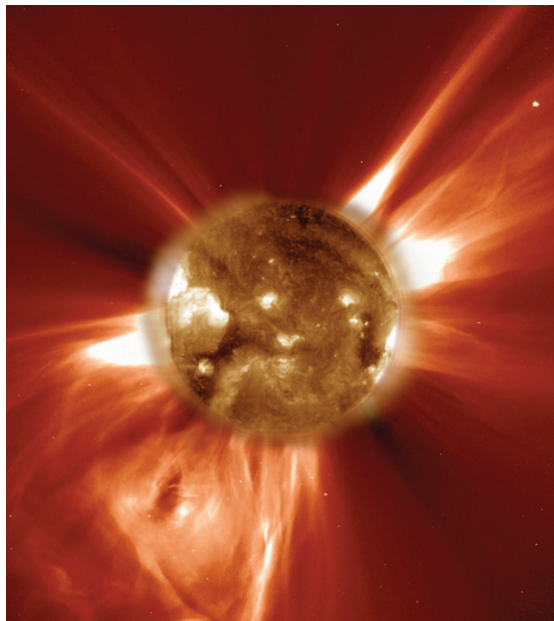


Figura 4: Eyecciones de la superficie activa del Sol observadas por SOHO en el año 2002. La luz del disco solar ha sido bloqueada y su imagen ha sido reemplazada con una imagen ultravioleta tomada simultáneamente. Cortesía SOHO, ESA, NASA.

Telescopios solares

Los astrónomos han diseñado telescopios e instrumentos específicos para la observación detallada del Sol. Desde principios del siglo XIX, los espectrógrafos han servido para analizar cuidadosamente el espectro solar, lo que permite obtener su composición química. El espectroheliógrafo es un instrumento que nos ofrece imágenes monocromáticas (filtrando un estrecho rango de frecuencias). El coronógrafo oculta de manera artificial la fotosfera (produce una especie de eclipse artificial) para estudiar la corona. La mejor batería actual de telescopios solares se encuentra en territorio nacional, concretamente en el Observatorio internacional del Teide (dependiente del Instituto de Astrofísica de Canarias). Y los telescopios montados en plataformas espaciales han podido proporcionar una información muy detallada del Sol, sin el efecto pernicioso introducido por la atmósfera terrestre en las observaciones. Entre estas misiones hay que destacar el Solar and Heliospheric Observatory (SOHO), un telescopio fruto de la colaboración entre las agencias espaciales europea (ESA) y norteamericana (NASA) que viene observado cuidadosamente el Sol desde hace unos doce años. Algunas de las imágenes que ilustran este artículo proceden de observaciones realizadas con SOHO.



Figura 5: El observatorio internacional del Instituto de Astrofísica de Canarias en Izaña (Sta. Cruz de Tenerife) alberga la mejor batería de telescopios solares del mundo. Foto: Miguel Briganti, SMM del IAC.

Energía solar en la Tierra

El Sol no sólo emite energía en el rango óptico (visible). El espectro (la distribución por frecuencias) de la radiación solar es similar al de un cuerpo negro (un cuerpo ideal que absorbe toda la radiación que llega sobre él sin reflejar nada, una fuente ideal de radiación térmica) a una temperatura de unos 5 800 K. Por tanto, hay tanta energía emitida por el Sol en el visible

como en el resto del espectro, principalmente en el infrarrojo cercano con una pequeña contribución en el ultravioleta.

Espectro de la radiación solar

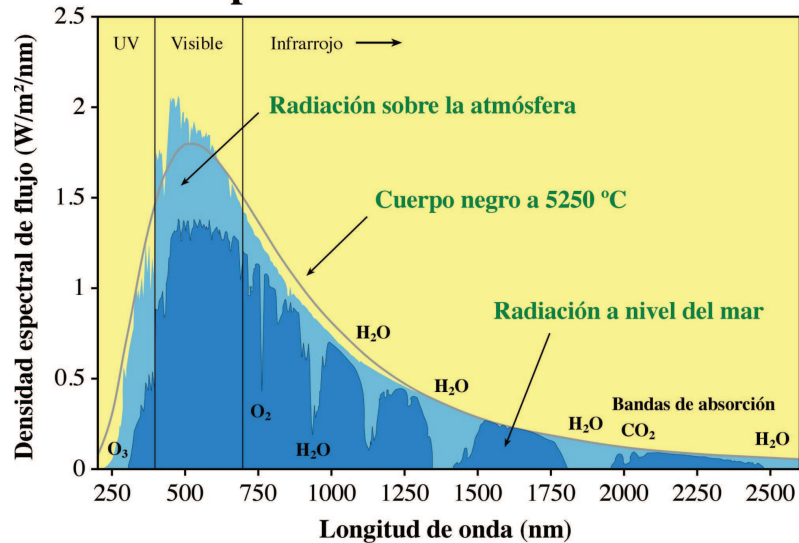


Figura 6: Espectro de radiación solar. Tras la dispersión y absorciones en la atmósfera, la radiación que alcanza el nivel del mar está muy disminuida en el ultravioleta y presenta un exceso en el azul.

Como hemos señalado anteriormente, el Sol emite al espacio (en todas las direcciones) unos $3,86 \times 10^{26}$ vatios. El ángulo sólido subtendido por la Tierra, cuando se observa desde el Sol, es de unos $1/140\,000\,000$ estereoradianes. Por lo tanto, la potencia de la radiación solar electromagnética por unidad de superficie que alcanza la atmósfera de la Tierra es de unos $1\,370 \text{ W m}^{-2}$ (vatios por metro cuadrado de la superficie perpendicular a los rayos del Sol). Esta cantidad recibe el nombre de constante solar. A veces se expresa como $1,96 \text{ Ly min}^{-1}$ (donde Ly representa 1 Langley, es decir, una caloría por cm^2). Sin embargo, este parámetro no es realmente constante, su fluctuación alcanza el 7 % entre enero y julio (según la superficie de la Tierra está más lejos o más cerca del Sol debido a su órbita elíptica). La constante solar representa, por tanto, un valor medio calculado a lo largo del año. La intensidad de la radiación solar recibida en la Tierra también varía con la latitud del lugar, además de con las estaciones. A latitudes de unos 30° , la insolación diurna (acumulada en el día) varía de unos 470 Ly en invierno a casi 1 000 Ly en verano, y la insolación anual (acumulada en el año) alcanza un valor de unos 280 000 Ly.

La atmósfera terrestre atenúa la radiación solar y sólo unos $1\,000 \text{ W m}^{-2}$ alcanzan el nivel del mar en un día claro (para una superficie perpendicular a los rayos). Si se tiene en cuenta el ángulo con que inciden los rayos, la trayectoria recorrida por la radiación en la atmósfera, la absorción y la difusión atmosféricas, y la duración del día, resulta que la insolación media de la superficie de la Tierra (ignorando nubes) es de unos 250 W m^{-2} . Esto también puede expresarse como 6 kWh m^{-2} por día. A latitudes de unos 30° , la insolación diurna atenuada varía de unos 200 Ly en el solsticio de invierno a unos 600 Ly en el de verano, y la insolación anual atenuada es de unos 150 000 Ly.

Los números descritos aquí, que tienen un origen puramente astronómico, son los que se utilizan para calcular el tamaño y rendimiento de las plantas de energía solar. Junto con el rendimiento de los paneles fotovoltaicos, los valores locales de la insolación son los que determinan el rendimiento final de una planta de energía solar.

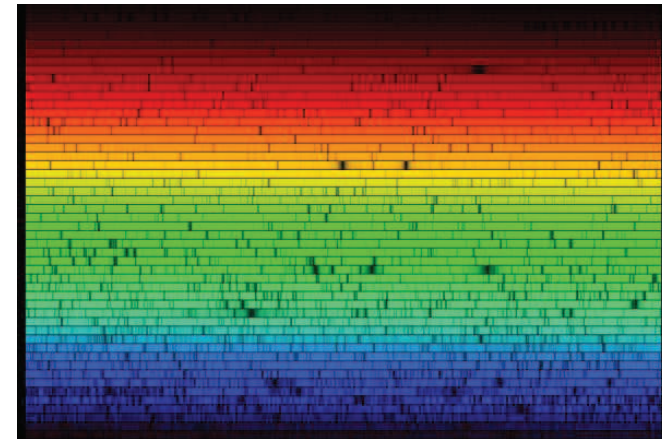


Figura 7: Espectro solar obtenido mediante el paso de luz solar a través de un dispersor similar a un prisma. El Sol emite en todos los colores del espectro visible y diferentes tipos de gas en su atmósfera absorben diferentes radiaciones creando zonas oscuras (líneas de absorción) que permiten conocer su composición química. Cortesía: Nigel Sharp (NSF), FTS, NSO, KPNO, NOAO/AURA/NSF.

Conclusión

El Sol domina de manera abrumadora la práctica totalidad de la actividad sobre nuestro planeta y, naturalmente, debe jugar un papel importante en la satisfacción de las necesidades contemporáneas y futuras

de nuestra civilización en términos de energía. El Sol es, por tanto, un ingrediente esencial en el futuro de la Humanidad.

Como hemos subrayado en este artículo, el Sol parece ser una pequeña estrella perdida en una galaxia anodina. Pero para nosotros el Sol, nuestro astro, es una estrella única e irremplazable. Aunque la Vía Láctea tiene muchas estrellas similares, cuando intentamos buscar gemelos idénticos al Sol, los astrónomos encontramos serias dificultades. Es cierto que aún no sabemos cómo es de especial el Sol, ni cuán único es nuestro Sistema Solar, ni cuán única es la Tierra, pero cuando estudiamos la Galaxia con los medios disponibles hoy día, es inevitable considerar la vida y la civilización terrestre como fenómenos sumamente especiales. El estudio del Sol y el aprovechamiento óptimo de su energía son requisitos indispensables para conservar adecuadamente la Tierra.

Referencias

- J. Audouze, G. Israël (Eds.), *Le Grand Atlas d'Astronomie*. Encyclopaedia universalis, 1988.
- J. Meléndez, I. Ramírez, *The Astrophysical Journal Letters* 669, 2007
- J. Kelly Beatty, C. Collins Petersen, A. Chaikin, *The New Solar System*, Cambridge University Press, 1999.
- K. R. Lang, *The Cambridge Encyclopedia of the Sun*, Cambridge University Press, 2001.
- M. Longair, *La evolución de nuestro universo*, Cambridge University Press, 1998.
- J. M. Pasachoff, L. Golub, *Nearest star: the surprising science of our Sun*, Harvard University Press, 2001.
- M. Tafalla, *¿Cómo nació el Sol? Formación de estrellas de tipo solar*, Anuario del Observatorio Astronómico de Madrid para el año 2002.
- J. B. Zirker, *Journey from the Center of the Sun*, Princeton University Press 2002.

http://es.wikipedia.org/wiki/Disco_solar
<http://solar-center.stanford.edu/index.html>
<http://sohowww.nascom.nasa.gov>
<http://www.solarviews.com/eng/sun.htm>
<http://www.nasa.gov/home/index.html>
<http://www.esa.int/esaCP/Spain.html>
<http://www.windows.ucar.edu/tour/link=/sun/sun.html>

¿CÓMO NACIÓ EL SOL?

Formación de estrellas tipo solar

Mario Tafalla

Observatorio Astronómico Nacional (OAN)

Instituto Geográfico Nacional (IGN)

Abstract

Solar-type stars are presently forming in our Galaxy, and by studying their properties we can attempt to reconstruct the process by which our Sun came to life 4.700 billion years ago. The Sun probably started its life as a dense core approximately 0.5 light years in size, which after a long period of stability began to collapse under the force of gravity. This collapse produced a dense, opaque object at the center (a protostar) surrounded by a disk of high angular momentum material. A powerful bipolar outflow probably helped the disk material to lose angular momentum and fall onto the protostar, in addition to disrupt the core and make the central object optically visible as a contracting T Tauri star. After about 30 million years of slow contraction, the central temperature and density reached the high values needed to sustain hydrogen burning, and the T Tauri star became the main sequence star we see every day in the sky.

El problema de la formación de estrellas

Una estrella como el Sol produce energía mediante la conversión nuclear de hidrógeno en helio en las zonas más densas y calientes de su interior. En la terminología astronómica, se dice que el Sol se encuentra en la *secuencia principal*, que corresponde a la fase más estable y duradera de la vida de una estrella (el Sol lleva unos 4.700 millones de años en esta fase y le quedan otros tantos para abandonarla). En la secuencia principal, la estructura interna de la estrella es tal que la fuerza de la gravedad es exactamente compensada por la presión interna debida al calor liberado en la combustión nuclear del hidrógeno en el centro. Entender cómo el Sol, o cualquiera de las más de cien mil millones de estrellas que le acompañan en la Vía Láctea, llegó a adquirir las condiciones físicas precisas que permiten tal equilibrio es un problema fundamental de la astronomía contemporánea.

El estudio de las estrellas en la secuencia principal ofrece pocas pistas sobre la forma en que éstas alcanzaron su estado presente. Esto difiere del

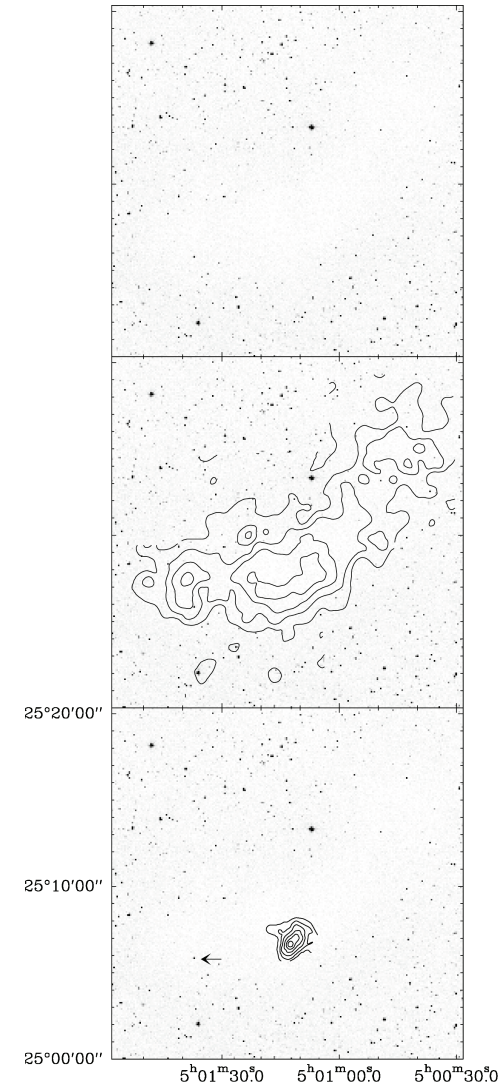
estudio de la evolución posterior (post-secuencia principal), que en parte puede inferirse mediante simulación numérica a partir de las condiciones físicas en la secuencia principal. La diferencia resulta de que algunos de los procesos que acompañan la formación de una estrella borran cualquier memoria de los eventos anteriores, y ello impide reconstruir la historia pasada a partir de la estructura presente. Afotunadamente, el proceso de formación estelar sigue activo en nuestra Galaxia, por lo que existen estrellas que están en este momento pasando por estadios que nuestro Sol recorrió hace casi cinco mil millones de años. El estudio de estas regiones de formación estelar ha aportado durante las últimas décadas una información fundamental para poder reconstruir (todavía parcialmente) los eventos que rodean el nacimiento de una estrella como nuestro Sol.

Nubes moleculares: cunas de las estrellas

Todas las regiones de formación estelar conocidas se encuentran dentro de nubes moleculares. Éstas son condensaciones más o menos amorfas (que recuerdan a las nubes del cielo) compuestas de gas molecular, en su mayoría hidrógeno, y de pequeñas partículas sólidas llamadas granos de polvo. Las nubes moleculares pueden ser extremadamente opacas dada la facilidad con que los granos de polvo absorben la luz, por lo que las regiones de formación estelar son en su mayor parte invisibles en observaciones ópticas. Esta situación cambia cuando las regiones se observan con detectores sensibles a las ondas infrarrojas y radio (milimétricas), algo técnicamente posible sólo a partir de los años 70. Para estas ondas, el material de la nube es casi o totalmente transparente, lo que permite ver con ellas el proceso de formación estelar en acción (ver Fig. 1).

Figura 1: Tres vistas de la nube molecular formadora de estrellas L1544 en Tauro. El panel de la izquierda muestra la imagen óptica roja del Digital Sky Survey (Copyright (c) 1992-8, Caltech and AURA), donde se aprecia un menor número de estrellas a lo largo de la diagonal que conecta el vértice inferior izquierdo con el superior derecho. Esta aparente falta de estrellas resulta de la presencia de una nube molecular, cuyos granos de polvo extinguen la luz de las estrellas situadas tras ella. (Las coordenadas son Ascensión Recta y Declinación época 1950). El panel central muestra los contornos de intensidad de la emisión de la molécula $C^{18}O$ (2.7 mm de longitud de onda) superpuestos a la imagen óptica. La emisión de $C^{18}O$ coincide con la zona de extinción, y permite estudiar en detalle la forma de la nube molecular. Ésta es alargada, de unos 2,5 años luz de longitud, y presenta diversos máximos de emisión. Finalmente, el panel de la derecha muestra los contornos de intensidad del ion molecular N_2H^+ , superpuestos también sobre la imagen óptica. N_2H^+ es muy sensible a la presencia de gas denso, y su emisión concentrada indica

la presencia de un núcleo denso en la zona central de la nube. Este núcleo denso parece estar a punto de colapsar para formar una estrella, o podría haber iniciado ya el proceso colapso. A su izquierda, e indicada con una flecha, se encuentra una estrella tipo T Tauri, resultado de una fase de formación anterior en la nube. (Salvo indicado, datos del autor.)



Si las estrellas se forman en nubes moleculares, el gas y el polvo de estas nubes tiene que constituir la materia prima para formar soles. Observaciones radio indican que las densidades del gas molecular en regiones de formación estelar oscilan en torno a 10^{-19} gramos por centímetro cúbico, y esto da una idea de la enorme complejidad del proceso de formación estelar: el material de una nube debe comprimirse 20 órdenes de magnitud (10^{20}) hasta llegar a las densidades típicas de una estrella. En términos de tamaño, si el material inicialmente llena una esfera, el radio de dicha esfera debe disminuir más de un millón de veces hasta llegar al tamaño de una estrella.

La variedad de procesos físicos que pueden entrar en juego en la formación de una estrella es muy grande, pues cualquier parámetro proporcional a la densidad (por ejemplo) puede magnificarse por un factor 10^{20} . Sin embargo, hay un elemento que domina a todas las escalas espaciales, y que en última instancia es el máximo responsable de la formación estelar: la inexorable fuerza de la gravedad. La vida de una estrella, desde su formación a su muerte, es en gran medida la lucha de la materia para vencer la siempre atractiva fuerza de la gravedad. A lo largo del camino existen victorias parciales, como el largo periodo de estabilidad que representa la secuencia principal, donde la gravedad es compensada por la combustión de hidrógeno. Sin embargo, uno a uno, los combustibles nucleares disponibles se agotan, y la fuerza de la gravedad vuelve siempre a recuperar su ventaja. Sólo como enana blanca o como estrella de neutrones, la materia consigue controlar la gravedad, pero pagando el alto precio de formar un objeto inerte que poco a poco se va enfriando para desaparecer en la oscuridad. La gravedad es por tanto el motor de la vida y la muerte de las estrellas. En el proceso de formación estelar, es el motor de la vida, pues convierte el oscuro material molecular en un objeto luminoso y visible. En la evolución posterior, es el implacable motor de su muerte.

Formación estelar aislada y en cúmulos

El estudio sistemático de regiones de formación estelar revela que no todas las estrellas nacen en las mismas condiciones ambientales. En la Nebulosa de Orión, por ejemplo, más de 3.500 estrellas han nacido recientemente (1 millón de años) formando una agrupación tan compacta (cúmulo) que las 100 estrellas centrales ocupan el volumen de una esfera con radio tan sólo de 0,35 años luz (al). En la región de ρ Ophiuchi, hace falta una esfera de casi 2 al de radio para contener ese mismo número de estrellas recién nacidas, y en la nube molecular de Tauro la densidad es tan baja que el radio de una tal esfera debe ser de casi 15 al. Como el volumen de una esfera crece como el radio al cubo, la densidad de estrellas jóvenes en Orión es más de 100 veces mayor que en ρ Ophiuchi, y unas 100.000 veces mayor que en Tauro. El origen de esta diferencia en densidades estelares no se conoce todavía, aunque se cree que resulta de

distintas condiciones en el gas molecular (temperatura, turbulencia, campo magnético).

Una diferencia entre las regiones de formación estelar compacta y las de formación aislada, es que sólo las primeras son capaces de formar estrellas masivas. Estrellas 10 veces más masivas que el Sol, por ejemplo, solo nacen en cúmulos densos como el de Orión, mientras que estrellas de una masa solar parecen formarse en todo tipo de entornos. Esta ubicuidad de las estrellas como el Sol plantea la pregunta de si nuestro astro nació en un medio tan denso como el cúmulo de Orión o si empezó su vida aisladamente, como lo han hecho las estrellas de la nube molecular de Tauro. Desgraciadamente todavía no somos capaces de inferir las características precisas del entorno del proto Sol, y para simplificar la presentación de este artículo, nos concentraremos a partir de ahora en la formación de estrellas de forma aislada.

Núcleos densos

Si el Sol nació de forma aislada, su historia empezó con un núcleo denso de unos 0,5 al de diámetro aproximadamente, parecido a los que actualmente se observan en Tauro o Camaleón (Figura 1). Estos núcleos densos son condensaciones discretas dentro de nubes moleculares mucho mayores, las cuales pueden extenderse por decenas de años luz y contener decenas o centenares de miles de veces la masa de nuestro Sol. Los núcleos densos son objetos extremadamente fríos (unos 10 grados por encima del cero absoluto), y la densidad en su interior suele aumentar hacia el centro, indicando que la gravedad ha empezado a jugar un papel importante. La estabilidad de un núcleo denso contra su propia gravedad procede en parte de la presión térmica del gas en su interior, que lucha por expandirse una vez comprimido (como el aire de un globo hinchado a presión). Existe además una posible contribución de un campo magnético “congelado” en su interior. El papel de los campos magnéticos en la estabilidad de los núcleos densos es un asunto de intenso debate, pues la dificultad de medirlos ha hecho que no conozcamos con precisión su intensidad o geometría. Aunque el grado de ionización del gas en una nube es pequeño, resulta suficiente para que la materia no pueda moverse sin arrastrar al campo magnético interior (de ahí el término “congelado”), y la resistencia natural de un campo magnético a comprimirse puede aportar un elemento adicional a la estabilidad del núcleo denso.

Durante un millón de años aproximadamente el núcleo denso parece mantenerse en equilibrio, pero poco a poco va perdiendo estabilidad, probablemente debido a la difusión y debilitamiento del campo magnético. Si esto es así, y el núcleo es lo suficientemente concentrado, debe llegar un momento en que la fuerza de la gravedad rompa el equilibrio y haga que el gas empiece a colapsar hacia el interior. Según modelos teóricos, el colapso gravitacional empieza en las zonas más internas del núcleo, que

caen antes, y poco a poco va afectando al gas de zonas más externas. El momento de la inestabilidad gravitacional marca el auténtico comienzo del nacimiento de una estrella.

Ha nacido una protoestrella

El gas que colapsa por acción de la gravedad lo hace hasta que llega a una densidad tal que es opaco a su propia radiación. En este momento, el gas no puede liberar fácilmente la energía cinética que gana por la aceleración gravitacional y empieza a calentarse, aumentando su presión y llegando otra vez a una situación de equilibrio. En el centro del núcleo colapsando, empieza pues a formarse un objeto denso que es el precursor de la estrella, y que va lentamente aumentando de masa gracias a la continua caída de material. Este objeto central, denominado protoestrella, emite ya luz, pero ésta no procede de reacciones nucleares en su interior (es demasiado frío para que éstas ocurran), sino de la liberación de energía cinética producida por el colapso (el ruido de un objeto que cae y choca contra el suelo también resulta de la energía ganada por acción de la gravedad). La luz de la protoestrella, sin embargo, es absorbida inmediatamente por el polvo de la condensación circundante y degradada a longitudes de onda infrarrojas o radio. La protoestrella es por tanto ópticamente invisible, pero su presencia puede ya ser detectada con instrumental sensible a estas longitudes de onda (ver Figura 2).

Desde el primer momento, la protoestrella no está aislada en el centro de la condensación colapsando, sino que está rodeada de un disco que gira a su alrededor y que resulta del material de la nube que no puede caer directamente hacia el centro por tener demasiado momento angular. Este momento angular es una medida de la rotación de la materia, y es una cantidad que se conserva durante el colapso gravitacional. Igual que un patinador que al recoger sus brazos acelera su rotación, el material de una nube que colapsa aumenta su velocidad de giro al acercarse hacia el centro. El efecto en el material que colapsa es mucho más pronunciado que en el patinador, pues éste puede variar su radio (recogiendo sus brazos) sólo un factor 3 o 4, mientras que el material de la nube colapsando disminuye de radio un factor un millón. El aumento de la rotación es tal que llega un momento en el que la fuerza centrífuga compensa la fuerza de la gravedad, y el material no sigue colapsando sino que empieza a orbitar (como un planeta) alrededor de la protoestrella. Se forma así un disco (de varios cientos de unidades astronómicas de diámetro) continuamente alimentado por material de la nube que cae pero no consigue alcanzar el centro (ver Figura 3).

El disco no es un objeto pasivo que se limite a girar alrededor de la protoestrella, sino que es un mecanismo eficaz por el cual el material de la nube es capaz de perder su momento angular y terminar alimentando a la protoestrella central. Aunque los detalles físicos no han sido todavía total-

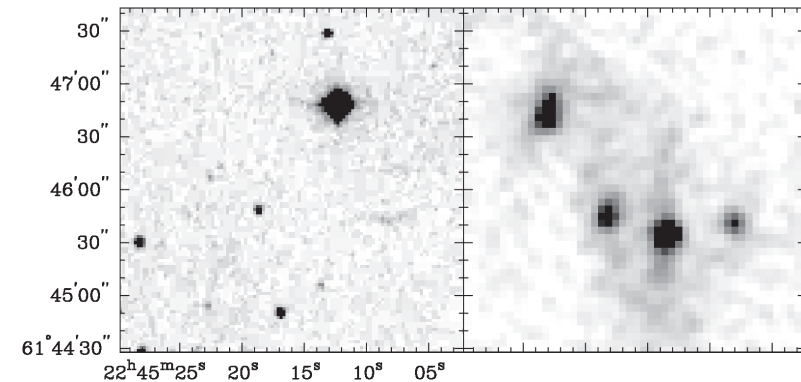


Figura 2: Estrellas y protoestrellas. El panel de la izquierda muestra la imagen óptica roja del Digital Sky Survey en dirección a la nube molecular L1211 (Copyright (c) 1994, Association of Universities for Research in Astronomy, Inc.). Las estrellas visibles en este panel no están asociadas a la nube, y simplemente se encuentran en su misma dirección, entre ella y nosotros. El panel de la derecha muestra una imagen del mismo campo a longitud de onda de 1,2 milímetros. A esta longitud de onda, las estrellas de la imagen izquierda son invisibles, mientras que aparece un grupo de cuatro protoestrellas. Estas protoestrellas se encuentran en el interior de la nube L1211 y su emisión óptica no es detectable por ser totalmente absorbida por los granos de polvo de la nube. A 1,2 mm, la nube molecular es transparente, y las protoestrellas son fácilmente detectables. (Las coordenadas son Ascensión Recta y Declinación época 1950, y la separación máxima entre las protoestrellas es aproximadamente 1,5 años luz.)

mente clarificados, de alguna manera el material en el disco es capaz de transferir momento angular del centro a la periferia, haciendo que sólo una pequeña fracción del gas acumule la mayor parte del momento, mientras que la mayor parte del material es capaz de descender lentamente en espiral para acabar cayendo a la protoestrella. Esta reorganización del momento angular, crucial para el nacimiento de una estrella, es evidente en nuestro sistema solar, donde el Sol, con un 99,9% de la masa contiene solo el 2% del momento angular (el resto está en la rotación de los planetas, especialmente Júpiter). El mecanismo responsable de la redistribución, todavía no completamente entendido, es claramente muy eficiente.



Figura 3: Dos ejemplos de estrellas jóvenes con disco en la nebulosa de Orión. Las condiciones especiales en Orión, con una fuerte emisión de fondo debida al gas ionizado, permiten ver a contraluz los discos opacos que rodean a las estrellas tipo T Tauri que se encuentra en la zona anterior de la nebulosa. En el panel de la izquierda, el disco aparece casi de frente, y la estrella central es visible como un punto brillante. En el panel de la derecha, el disco aparece casi de perfil, y es tan opaco que oculta totalmente la luz de la estrella central. El disco de la izquierda tiene un diámetro de unas 250 unidades astronómicas (0,004 años luz), aproximadamente 3 veces el diámetro de nuestro Sistema Solar. Imágenes tomadas con el telescopio espacial Hubble, cortesía de Mark McCaughrean (Max-Planck-Institute for Astronomy), C. Robert O'Dell (Rice University) y NASA

Protoestrellas binarias

Una forma alternativa de redistribuir del momento angular es acumularlo en forma de movimiento orbital de dos protoestrellas girando alrededor de su centro de masa común, es decir, formando un sistema binario. La naturaleza parece usar esta alternativa con frecuencia, pues más de dos tercios de las estrellas en la vecindad solar están formando parte de sistemas binarios o múltiples. Esta estadística se refiere a estrellas nacidas hace ya millones de años, y por tanto a sistemas binarios que han sobrevivido por largo tiempo. Estudios de frecuencia de binarias en sistemas jóvenes no han sido posibles hasta muy recientemente, y los datos preliminares parecen sugerir que la frecuencia de binarias depende del entorno en el que se forman las estrellas. Regiones de formación en cúmulo (como la Nebulosa de Orión) parecen tener una frecuencia de binarias igual a la observada en la vecindad solar, mientras que regiones de formación aislada (como Tauro)

parecen formar binarias con aún mayor frecuencia. Esta diferencia puede originarse por la ruptura de sistemas binarios en cúmulos densos (debido a la interacción gravitacional con otras estrellas), o por la dificultad de formar binarias en este tipo de regiones. Dadas las actuales estadísticas, no es posible descartar que la naturaleza sólo forme binarias, y que lo que vemos ahora como estrellas aisladas es el resultado de binarias que se han dispersado.

Aunque existe una pequeña posibilidad de que el Sol forme parte de una binaria de tan largo periodo como para que su compañera todavía no haya sido detectada, todo parece indicar que nuestro astro vive aislado. ¿Cómo explicar entonces esta anomalía, que afecta a menos de un tercio de las estrellas ya maduras? Una posibilidad es que el Sol formó parte de un sistema binario o múltiple y que éste fue destruido por alguna interacción. Otra posibilidad es que el Sol se formó en un entorno tan denso (como la Nebulosa de Orión, por ejemplo) que inhibió la formación de tal sistema. Desgraciadamente, la pregunta permanece sin respuesta, pero es posible que en el futuro, gracias a una mejor comprensión de la formación de sistemas múltiples de estrellas y de las características de sus sistemas planetarios asociados, seamos capaces de reconstruir el entorno en el que nació el Sol, casi cinco mil años después de que ésto sucediera.

Flujos bipolares

Además de la protoestrella central, el disco y el gas en proceso de colapso gravitacional, hay un elemento fundamental en la formación de una estrella que no fue descubierto hasta principios de los años 80. En esa época, el estudio de los movimientos del gas alrededor de protoestrellas reveló sorprendentemente que una gran cantidad del gas no esta contrayéndose, como debería esperarse por la acción de la gravedad, sino expandiéndose rápidamente en forma de dos corrientes opuestas. Este fenómeno, llamado flujo bipolar, muestra que la protoestrella, además de recibir material de la nube ambiente, es capaz de eyectar gas a alta velocidad que escapa de sus proximidades. Resulta paradójico que una estrella nazca de la mezcla de contracción gravitacional y flujo bipolar, pero como ahora veremos esta combinación parece ser necesaria para la formación de una estrella. La coexistencia de estos movimientos es posible porque el flujo bipolar sólo ocupa un pequeño ángulo sólido alrededor de la estrella, mientras que el gas en contracción ocupa el resto (ver Figura 4).

Aunque el mecanismo exacto que produce un flujo bipolar es todavía un motivo de discusión dentro de la comunidad astronómica, parece haber consenso en que los elementos esenciales para producirlo son la rotación del disco y la presencia de un campo magnético focalizador. Mediante la combinación de estos ingredientes, una parte del material del disco que está en proceso de caer a la protoestrella es acelerado en dirección perpendicular, parte hacia arriba y parte hacia abajo, saliendo despedido

a muy alta velocidad (alrededor de unos 100 km por segundo). Como resultado de este proceso, el material eyectado roba momento angular, permitiendo que parte del material del disco se frene lo necesario como para caer finalmente a la protoestrella central. El flujo bipolar, por tanto, es capaz de resolver el problema final que tiene el gas del disco para caer a la protoestrella: su exceso de momento angular. Según modelos recientes, esto se logra transmitiendo el momento angular sobrante (vía campo magnético) a una pequeña cantidad del gas que es eyectada (10-20%), mientras que el resto del material puede pasar a formar parte de la estrella central.

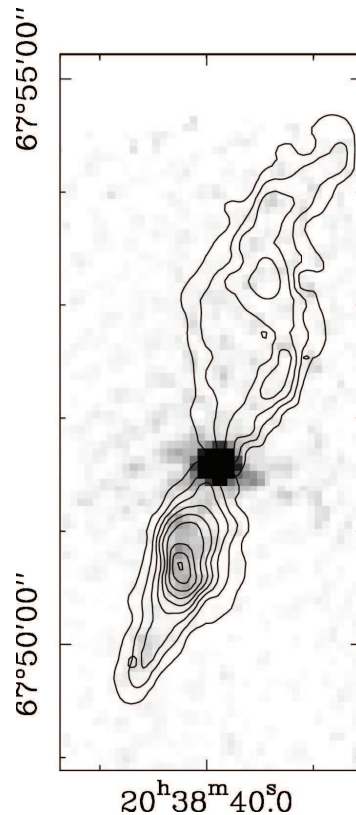


Figura 4: Protoestrella con flujo bipolar en la nube L1157. La escala de grises es una imagen a longitud de onda de 1,2 mm que muestra una protoestrella brillante hacia el centro del panel. Los contornos superpuestos muestran la emisión de la molécula ^{12}CO (1,3 mm de longitud de onda), e indican la distribución de gas acelerado por la protoestrella central. El gas forma dos lóbulos que divergen de la

protoestrella y se extienden por una distancia de unos 2 años luz. Datos del flujo bipolar cortesía de Rafael Bachiller y Miguel Pérez Gutiérrez (Observatorio Astronómico Nacional).

Además de su importante papel de redistribuidor del momento angular, el flujo bipolar parece cumplir otra misión crucial en el nacimiento de una estrella: la dispersión de su núcleo denso. Esto ocurre debido a la gran energía cinética del flujo bipolar, que es transmitida al gas ambiente mediante violentos choques supersónicos. El gas del núcleo denso va así siendo poco a poco acelerado por el flujo bipolar, y este movimiento contrarresta la atracción gravitacional de la protoestrella. Parece muy posible que de esta manera, al cabo de varios cientos de miles de años de iniciarse el colapso gravitacional, el flujo bipolar inhiba totalmente este movimiento, y dé por finalizada la fase principal de crecimiento del sistema disco-protostrella.

Estrellas pre-secuencia principal

La dispersión de parte del núcleo denso por el flujo bipolar tiene como consecuencia adicional hacer visible ópticamente a la protoestrella. Por convención, hablamos entonces de una estrella pre-secuencia principal, o en referencia al primer caso conocido, de una estrella T Tauri (Figuras 1 y 3). Este tipo de estrella no emite todavía energía por combustión de hidrógeno en helio, sino por la lenta contracción que la va aproximando al deseado estado de equilibrio. La estrella no puede realizar esta contracción de forma rápida, pues como hemos visto, para contraerse necesita liberar energía, y al ser un objeto opaco esta energía no puede escapar fácilmente. Una estrella como el Sol, necesita unos 30 millones de años para liberar la energía extra y alcanzar finalmente la secuencia principal. En este proceso, el radio de la estrella disminuye aproximadamente un factor 5, hasta alcanzar los aproximadamente 700.000 km que el Sol tiene en la actualidad.

Aunque la contracción gravitacional no es una fuente de energía muy duradera, sí es eficiente, como lo muestra el hecho de que durante la fase T Tauri una estrella como el Sol sea unas diez veces más luminosa que en la secuencia principal. Desde el punto de vista observacional, la emisión de una estrella T Tauri empieza a parecerse a la de un objeto de la secuencia principal, pero todavía incluye cierto grado de actividad atípica, como variaciones de luminosidad y emisión en líneas de hidrógeno. Estas características llamaron la atención de los astrónomos a principios de los años 40, y ahora podemos entenderlas como signos de juventud estelar. Parte de la variabilidad, por ejemplo, es debida a la presencia de enormes zonas frías (manchas) en la superficie, parecidas a las que presenta el Sol en tiempo de máximo solar pero a una escala miles de veces mayor (en una estrella T Tauri típica, las manchas pueden cubrir el 10 % de su superficie).

Estas manchas, y otros fenomenos como la emisión de rayos X, indican la presencia de intensos campos magnéticos en la superficie, que resultan de la rotación estelar y de una estructura convectiva (la convección transmite la energía del interior mediante movimientos de materia, como en un fluido en ebullición).

Tanto el disco como el flujo bipolar siguen presentes en la fase T Tauri, al menos durante los primeros millones de años. El disco se va haciendo menos masivo, pues ha perdido la fuente de material que representa el núcleo denso colapsando. Sigue sin embargo alimentando de material a la estrella, y se estima que en la fase T Tauri, ésta adquiere el 10% final de su masa total a través del disco. El flujo bipolar, por su parte, es menos energético que antes, probablemente debido a la menor actividad del disco. A medida que transcurre la fase T Tauri, el disco y el flujo van perdiendo prominencia, como los otros signos de juventud estelar representados por la variabilidad y la intensidad de las líneas de emisión de hidrógeno. La luminosidad de la estrella va también decreciendo lentamente, aproximándose al valor correspondiente a la secuencia principal.

Aproximación final a la secuencia principal

Aunque la principal fuente de luminosidad de la estrella T Tauri procede de su contracción, existen períodos donde ésta se frena temporalmente al aparecer fuentes de energía alternativas. Cuando la temperatura central de la estrella llega a un millón de grados, por ejemplo, el deuterio (hidrógeno pesado) en el interior es capaz de reaccionar, y así producir energía que compensa la fuerza gravitatoria. El deuterio, sin embargo, es un elemento minoritario en la composición estelar (un par de partes por 10.000), y no constituye por tanto una solución de largo plazo al equilibrio final. En una estrella como el Sol, la combustión de deuterio dura unos cien mil años, y una vez agotado éste, la estrella vuelve a su lenta contracción gravitacional.

A medida que la contracción progresa y la luminosidad de la estrella disminuye, llega un momento en que el interior estelar cambia de estructura, y pasa a ser radiativo. Esto significa que la energía en el interior ya no se transmite hacia la superficie mediante un movimiento de materia similar a la ebullición, sino que lo hace a través de radiación que se mueve en un medio estático. Ese cambio representa el último ajuste estructural de la estrella, que sigue contrayéndose lentamente, pero ahora con luminosidad casi constante (de hecho aumenta un poco). En una estrella como el Sol, se estima que tras unos 30 millones de años de contracción gradual, la temperatura y la densidad centrales alcanzan los aproximadamente 14 millones de grados y 100 gramos por centímetro cúbico necesarios para comenzar la combustión nuclear de hidrógeno en helio. Este momento marca la llegada de la estrella a la secuencia principal; representa el final de su juventud y el principio de su edad madura. La gran cantidad de combustible de que

dispone la estrella (un 70% de su masa es hidrógeno) permite que esta fase sea la más larga de su vida, y represente un triunfo (aunque tan solo temporal) frente a la implacable acción de la gravedad.

El ciclo de vida y muerte de las estrellas

Las nubes moleculares que están formando estrellas en la actualidad, o la que formó al Sol hace 4.700 millones de años, están compuestas mayoritariamente de hidrógeno y helio, pero también contienen elementos más pesados, que en astronomía llamamos genéricamente “metales”. Esta presencia de metales indica que las nubes no tienen una composición primordial, esto es, no tienen la composición inicial del universo tras el Big Bang (casi en su totalidad hidrógeno y helio), sino que ha sido enriquecidas con elementos más pesados. El único mecanismo eficaz de formar elementos pesados es mediante la combustión nuclear en el interior de estrellas, proceso que poco a poco va convirtiendo el hidrógeno y helio primordial en elementos pesados como el carbono, oxígeno, y nitrógeno. Ya sea mediante explosiones tipo supernova (en las estrellas más masivas) o mediante potentes vientos estelares (estrellas menos masivas), las estrellas al morir devuelven al medio interestelar parte de los elementos generados en su interior, y de esta manera, los elementos vuelven a quedar disponibles para formar nuevas estrellas. Este reciclaje a escala galáctica lleva funcionando miles de millones de años, y es el causante de la composición no primordial de todas las estrellas conocidas.

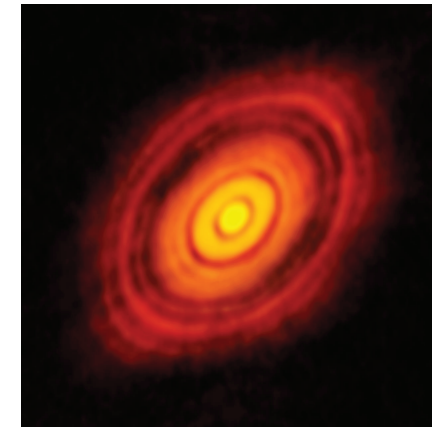
Como hemos visto en este artículo, nuestro conocimiento del proceso de formación estelar es todavía demasiado incompleto como para responder con confianza preguntas como si nuestro Sol nació aislado o formando un grupo, o qué hizo que el Sol terminase su formación con la masa que tiene ahora y no con otra mayor (o menor). La enorme actividad investigadora en este campo hace que podamos ser optimistas de que en un futuro cercano tendremos respuestas a estas preguntas, y a otras que vayan surgiendo a medida que profundizamos en nuestro entendimiento de la formación estelar. De lo que sí podemos estar seguros es de que el Sol transcurrirá su larga fase de estabilidad en la secuencia principal, para después iniciar los cambios estructurales que conlleva el agotamiento de hidrógeno en su centro. Pasada una fase de gigante roja, y tras un breve período de estabilidad quemando helio, el Sol empezará su proceso de formación de una enana blanca perdiendo por el camino gran parte del material que tomó prestado al nacer. Este material, mezclado con el procedente de otras estrellas de distintas partes de la Galaxia, volverá de nuevo a formar una nube molecular, a condensarse formando un núcleo denso, y con el tiempo, a repetir el complejo proceso que un día hizo nacer a nuestro Sol.

Quince años de progreso ininterrumpido

Han pasado quince años desde la publicación del artículo *¿Cómo nació el Sol?* En este tiempo, nuestro conocimiento del proceso de formación estelar ha mejorado notablemente gracias al trabajo de una comunidad investigadora enormemente activa que ha sabido aprovechar la puesta a punto de nuevos telescopios y el progreso en la capacidad computacional de los ordenadores. Los nuevos avances observacionales y teóricos han aclarado un gran número de incógnitas que existían al publicar el artículo en el año 2002. Aun así, las principales ideas presentadas en el texto siguen siendo válidas, pues describen los procesos más básicos que tienen lugar durante la formación de una estrella. Estos procesos no han sufrido cambios radicales desde 2002, aunque el trabajo más reciente ha permitido enriquecer su comprensión y refinar muchos de sus detalles.

Uno de los aspectos en los que más han avanzado nuestras ideas sobre la formación estelar es en lo relativo a la naturaleza y el estado dinámico de las nubes moleculares. Estas nubes representan el lugar donde nacen las estrellas, y el gas difuso que las compone constituye la materia prima para formar nuevos soles. Imágenes a gran escala de las nubes moleculares más cercanas al Sol, como las obtenidas con el telescopio espacial Observatorio Espacial Herschel, han proporcionado un nivel de detalle hasta ahora desconocido, y han revelado la presencia de una rica estructura interna en forma de complejas redes de filamentos. Estas nuevas observaciones, combinadas con los resultados de simulaciones numéricas, sugieren que las nubes moleculares no son objetos cercanos al equilibrio, como inicialmente se había propuesto, sino que representan estructuras con un alto grado de turbulencia y en un estado siempre cambiante. La formación de estrellas en estas condiciones debe ser por tanto un proceso más dinámico de lo que se pensaba, y probablemente involucra una compleja secuencia de eventos a diferentes escalas espaciales. Así, la formación inicial de los filamentos probablemente sucede mediante colisiones de grandes corrientes de gas, que generan condensaciones alargadas de materia donde la turbulencia ha sido parcialmente disipada. Algunos de estos filamentos logran acumular suficiente materia como para superar el límite de lo que la presión interna es capaz de soportar, y como resultado, se fragmentan gravitacionalmente dando lugar a múltiples núcleos densos. Estos núcleos, a su vez, son también inestables, y terminan por colapsar y formar una nueva generación de estrellas.

Otro aspecto del proceso de formación estelar con un significativo progreso reciente es el estudio de los discos protoplanetarios alrededor de estrellas jóvenes. Estos discos tienen tamaños cercanos a los de nuestro sistema solar, y son por tanto demasiado pequeños como para poder ser estudiados en detalle a las distancias típicas de las regiones de formación estelar. La reciente puesta en funcionamiento del interferómetro ALMA en Chile ha permitido superar esta limitación y realizar observaciones de discos protoplanetarios con gran resolución espacial en longitudes de onda milimétricas y submilimétricas. Uno de los primeros y más espectaculares frutos de estas observaciones es la imagen del disco que gira alrededor de la estrella HL Tau, de aproximadamente 1 millón de años de edad y una masa parecida a la de nuestro Sol. La imagen, publicada en 2015 por Brogan et al. y mostrada en la figura adjunta, revela la estructura interna de este disco, y en particular, la presencia de una serie de anillos de baja densidad que aparecen de color oscuro. Estos anillos indican la presencia de huecos en la distribución de materia del disco a distintas distancias a la estrella, y sugieren la existencia de un sistema planetario subyacente en fase de formación.



Aunque nuestro conocimiento del proceso de formación estelar es sin duda notable, todavía sufre de importantes lagunas. Por ello, en su informe de 2010 *Una visión científica para la astronomía europea*, el consorcio ASTRONET seleccionó “el origen y la evolución de las estrellas y planetas” como una de las cuatro preguntas fundamentales que la astronomía debe responder en los próximos veinte años. Esta selección ilustra por una parte la importancia de la formación estelar. Por otra, nos recuerda la complejidad del proceso, que sin duda mantendrá ocupados durante muchos años a una buena parte de la comunidad astronómica internacional.

Mario Tafalla García, febrero 2017

EL COLOQUIO DE LAS NEBULOSAS

Valentín Bujarrabal

*Observatorio Astronómico Nacional
Instituto Geográfico Nacional - Ministerio de Fomento*

Berganza.— Cipión hermano, óyote hablar y sé que te hablo, y no puedo creerlo, por parecerme que el hablar nosotros pasa de los términos de la naturaleza.

Cipión.— Así es la verdad, Berganza, y viene a ser mayor este milagro en que no solamente hablamos, sino en que hablamos con discurso, como si fuéramos capaces de razón, estando tan sin ella que la diferencia que hay del animal bruto al hombre es ser el hombre animal racional, y el bruto, irracional.

El coloquio de los perros
M. de Cervantes, natural de Alcalá de Henares

EL COLOQUIO DE LAS NEBULOSAS

Calabaza.— Pasé tanto tiempo en el oscuro vientre de aquella estrella gigante que casi no conservo recuerdo alguno de ello. Años, milenios, tenebrosos y que poco camino dejaban a la esperanza. ¿Habría de ser para siempre?

Rectángulo.— Años oscuros, sí, sin historia, sólo preparación de nuestro nacimiento (para algunos muy espectacular). El nacimiento de las nebulosas planetarias. Nuestro sol, ya viejo, pasó a ser una estrella gigantesca, que nos contenía inquietos y casi sin vida.

Sólo algunas briznas de la estrella escapaban, huyendo en una desbandada tenue y constante. Pero que poco alteraba la vida de quien quedaba en su interior y poco nos hacía presentir cual sería nuestro futuro.

Calabaza.— Mas pasaron aquellos tiempos y, repentinamente, salí de la estrella como una mariposa que sale de su capullo e inicia el primer vuelo. Desplegué mis alas de mil colores y volé rápida y alegre. No se ha visto nacimiento más espléndido ni recién nacida más hermosa, Venus misma me envidia.

El autor.— Así es, como lo cuentan nuestros amigos Calabaza y Rectángulo. Las estrellas, la mayoría de las estrellas, incluyendo nuestro Sol, cuando se hacen viejas evolucionan convirtiéndose en gigantes rojas. Las gigantes rojas son mucho más extensas que el Sol, unas mil veces mayores, aunque no contienen más masa, pues el proceso es una expansión pura. Al mismo tiempo que aumentan de tamaño, su temperatura disminuye. Llegan a tener unos dos o tres mil grados de temperatura en su superficie, mientras que una estrella como el Sol tiene unos seis mil. Por eso son rojas, porque un cuerpo con esas temperaturas emite sobre todo luz roja; nuestro Sol, más caliente, emite una luz blanquecina, tirando a amarilla. Debido a su gran tamaño (y a pesar de su menor temperatura), estas estrellas viejas son mucho más luminosas que sus predecesoras. Otra característica importante de las gigantes rojas es que su tamaño varía, o sea que pulsan, lo que hace que su luminosidad también varíe. Su diámetro puede variar de hasta un treinta por ciento con periodos de aproximadamente un año. La vida de una estrella gigante roja (en su fase extrema) no excede un millón de años, lo que es poco comparado con las fases anteriores, pues una estrella como el Sol puede permanecer en su estado, antes de empezar a dilatarse, unos diez mil millones de años.

Otra característica importante de las estrellas gigantes rojas es que pierden masa, aunque al principio a un ritmo muy lento. Esto lleva a la formación de lo que se llama una envoltura circunestelar, un material que se escapa de la estrella y que representa sólo una pequeña parte de la masa total. La razón de esta pérdida de masa es que la estrella tiene una atracción gravitatoria relativamente pequeña en su superficie, pero su alta luminosidad hace que la presión de radiación empuje el material que está en sus bordes hacia afuera. (La atracción gravitatoria es pequeña porque la gigante es muy extensa pero no contiene más masa que una estrella normal. La presión de radiación consiste en que la luz cuando choca con un objeto lo empuja; este empuje es pequeño en nuestro mundo, pero en la superficie de una estrella de gran luminosidad no es en absoluto despreciable y acelera eficazmente una parte del material estelar.) La fuerte pulsación de la estrella también ayuda, pues es como si la *diera de sí*, manteniendo una considerable cantidad de materia alejada de la superficie y lista para ser empujada por la radiación.

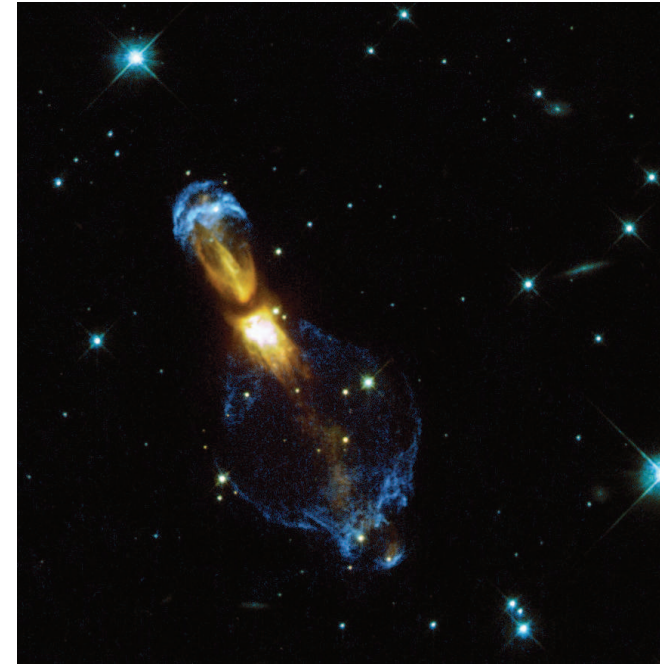


Figura 1: OH 231.8+4.2, la Calabaza, la nebulosa presumida. Esta figura es una composición de dos imágenes obtenidas con el Hubble Space Telescope (HST, cortesía de NASA/ESA/STScI). Vemos luz estelar reflejada por granos de polvo, que es la parte amarillenta; la parte azul es emisión de átomos de hidrógeno (línea H α). Sabemos por observaciones de emisión de la molécula CO que la estructura trazada por los granos contiene casi todo el material nebuloso. En total la nebulosa tiene más masa que el Sol, contiene sin duda la mayor parte de la masa que originalmente estaba en la estrella, que realmente se ha convertido en nuestra Calabaza. También sabemos por observaciones espectroscópicas que la nebulosa se expande a alta velocidad, superando el millón de kilómetros por hora. La componente masiva (la amarillenta en la figura) es fría, está a unos 250° bajo cero. Pero la componente en emisión que aquí vemos azulada ha sido calentada por choques entre diversas componentes nebulares, lo que es fácil a la vista de sus grandes velocidades, y alcanza temperaturas de unos 10 000 °C. En la región central se encuentra la estrella, que es doble (dos estrellas orbitando), no se ve directamente porque el polvo de la densa zona nebulosa en el ecuador de la nebulosa la oculta.

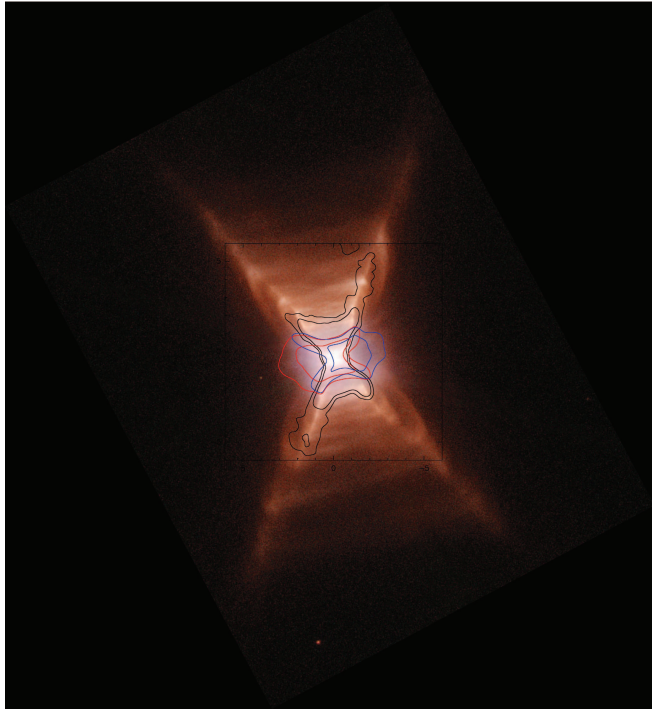


Figura 2: El Rectángulo Rojo, HD 44179, nuestra otra nebulosa. Los colores son la imagen óptica (obtenida con el HST, cortesía de NASA/ESA/STScI), que se debe principalmente a la emisión de pequeños granos de polvo excitada por la luz de la estrella central (que tampoco se ve directamente en este caso). Esta nebulosa tiene un disco gaseoso en su ecuador que está rotando, las líneas rojas y azules representan respectivamente las zonas del disco que se alejan y se acercan, medidas con gran precisión gracias a observaciones de emisión de CO con el telescopio ALMA. Las líneas grises muestran la emisión de gas que se encuentra en expansión. Éste ocupa en realidad la zona que se extiende entre el ecuador y la X que se ve en el óptico, la zona interior al doble cono (interior a la X) está casi vacía. Esta nebulosa presenta en cualquier caso velocidades mucho menores que las de la Calabaza, y su masa es mucho más pequeña que la de la estrella. El cuadrado en la imagen da la escala angular en segundos de arco, para que os hagáis una idea del pequeño tamaño angular de nuestras nebulosas; un segundo es 1/60 de un minuto, que es 1/60 de un grado.

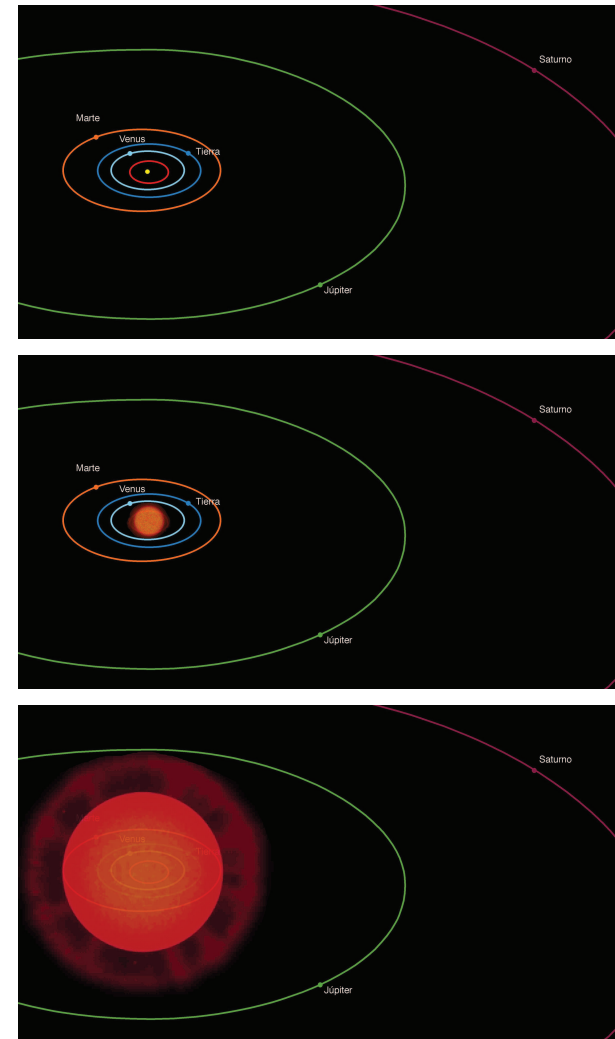


Figura 3: Se cree que cuando una estrella como el Sol envejece se extiende de forma espectacular, al tiempo que se vuelve rojiza. Llega a ser tan grande que termina engullendo algunos planetas, posiblemente éste sea el destino de la Tierra. Estas estrellas se llaman gigantes rojas, tienen un tamaño variable y límites poco nítidos.

El ritmo de pérdida de masa se acelera con el tiempo. Al final de la fase de gigante roja, este ritmo crece de manera espectacular, la estrella eyecta mucho material, casi todo el que en ella había. Posteriormente, el gas nebular es acelerado a grandes velocidades sobre todo en los polos, y forma unas nebulosas asimétricas de gran belleza. Este proceso está aún poco conocido; quizás se produzca por interacción, o sea por choques, entre las eyecciones de la gigante roja y otras ya posteriores muy rápidas y colimadas (es decir, en forma de chorro). El final del proceso es la formación de las llamadas *nebulosas planetarias*, nombre que tiene raíces puramente históricas, pues no tienen nada que ver con los planetas. Así envejecen la mayoría de las estrellas, en particular las que son similares al Sol, convirtiéndose en espectaculares nebulosas.

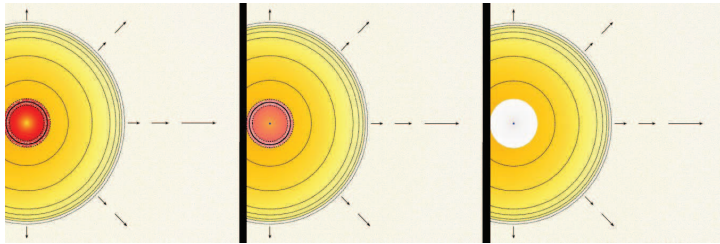


Figura 4: Las gigantes rojas pierden masa, formando las llamadas envolturas circunestelares (círculos amarillos en la figura), de gas y polvo en expansión. Al final de esta fase, el ritmo de pérdida de masa se acelera mucho, prácticamente toda la masa de la estrella se eyecta y sólo queda un remanente muy compacto y caliente, el núcleo de la gigante, que aparece en el tercer panel. A la nueva estrella se le llama estrella enana azul, pues a su alta temperatura le corresponden colores azulados.

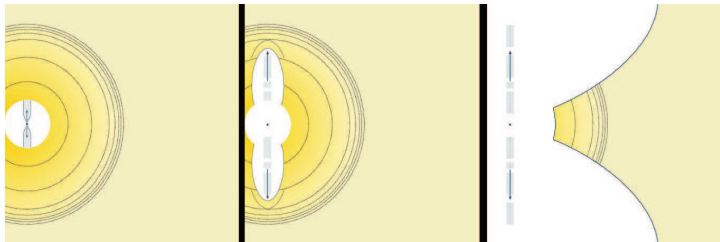


Figura 5: En la misma transición de gigante roja a enana, se produce una espectacular evolución de la envoltura. Parece ser que la nueva estrella eyecta chorros muy rápidos que chocan con la envoltura, más lenta, acelerando sus regiones polares y abriendo finalmente un hueco en ellas. Se genera así una nebulosa asimétrica, *bipolar*.

Rectángulo.— Bueno, ya sabemos que eres guapa.

Calabaza.— Si me creo bella es porque lo soy, que es lo que cuenta, y tu no seas modesto que tienes también muy buena percha.

Rectángulo.— A veces, la belleza que menos llama la atención es la más importante. Quizás en la información que se encuentra bajo un aspecto modesto nuestro autor y sus colegas puedan encontrar las claves de nuestra existencia, cuyo origen se conoce tan mal.

Calabaza.— Si, bueno, con su amigote Javier.

El autor.— Alcolea, Javier Alcolea.

Calabaza.— Y el resto, vaya pandilla, Arancha, Hans, Miguel, Carmen, ... Me los encuentro en un callejón oscuro y me dan un susto de muerte.

Rectángulo.— Pues a mí me caen simpáticos.

El autor.— La Calabaza es una nebulosa planetaria joven, que tiene por nombre de catálogo OH 231.8+4.2. No es difícil hacerla posar, podéis verla en la figura 1. Se trata de una de las nebulosas más bellas que se han observado (quizás por eso sea un poco cursi). Está formada, como las demás de su especie, de gas y polvo, sobre todo de gas. El polvo se detecta principalmente debido a la luz estelar que refleja, el gas es más bien emisor de radiación. La Calabaza, como la mayoría de las nebulosas planetarias recién formadas, presenta unas velocidades de expansión muy elevadas, particularmente en su caso, pues se expande a más de un millón de kilómetros por hora; los astrónomos preferimos medir las velocidades en kilómetros por segundo (km/s), la Calabaza alcanza velocidades superiores a los 300 km/s. El total de la masa que contiene la Calabaza es muy elevada, mayor que la masa del Sol, y en ella está sin duda la mayor parte de la masa que se encontraba en la estrella de la que nació.

Nuestro otro amigo es el Rectángulo Rojo, HD 44179, al que llamaremos simplemente Rectángulo. También es una nebulosa planetaria joven y muy bella (pero parece más modesto), os puedo enseñar una foto suya, figura 2; me han pasado fotos de otros amigos, nebulosas del mismo tipo, irán apareciendo a lo largo del artículo, leed bien los pies de las figuras. (También encontraréis en el artículo diagramas explicando los procesos evolutivos de los que hablamos.) El material que forma el Rectángulo Rojo, de nuevo gas y polvo, es sólo la centésima parte de la masa total del sistema estrella+nebulosa, es probable que la estrella aún tenga que eyectar una buena parte de sí misma. Tiene Rectángulo una característica muy interesante: en el ecuador de la nebulosa se encuentra una estructura de gas y polvo que no se expande sino que está en rotación alrededor de la estrella central. La dinámica es similar a la que tienen los planetas de nuestro sistema solar, es debida a que la atracción gravitatoria de la estrella impide que escapen. Todo esto se encuentra rodeado por gas en expansión que parece alimentarse del disco rotante, por lo que es probable que la mayor parte del mismo termine escapándose a pesar de todo. En este caso, el material

se mueve a velocidades mucho menores que las que antes mencionamos, menores que 15 km/s.

La Calabaza y el Rectángulo Rojo son nebulosas que nacieron hace poco de la eyección de masa por una estrella gigante roja y se convertirán con toda probabilidad en nebulosas planetarias plenamente desarrolladas. Pero sus características son tan diferentes, que hacen pensar que realmente esta formación no se produce siempre siguiendo el mismo proceso. Ya se sabe que los opuestos a veces se atraen y parece que hacen buenas migas, quiero decir que quizás estudiando coordinadamente objetos tan diferentes podamos encontrar respuestas a los enigmas que plantea la muerte de las estrellas.



Figura 6: NGC 6886. En esta nebulosa planetaria vemos cómo se rompe la simetría esférica de la envoltura que rodeaba a la estrella gigante roja (según se muestra en el diagrama anterior). (Cortesía de NASA/ESA/STScI.)

Rectángulo.— Si una parte de mí hubiera caído, en espiral vertiginosa, de nuevo hacia la estrella.

Calabaza.— Volver allí, ¿qué dices?, nunca.

Rectángulo.— Como la semilla se sacrifica para que en primavera nazca un espléndido y generoso campo de mies. Quizás sólo la caída de una parte de mí haya podido servir para surgir el todo con más brío, para escapar con más energía a la realización plena. Quizás así sea nuestro nacimiento y la muerte de la estrella gigante.

Calabaza.— Y si es así, ¿tengo yo también algo, o lo he tenido, que con su caída haya potenciado mi resplandor? ¿Lo tenemos todas nosotras?

El autor.— Su intuición no les engaña. De las teorías que se han propuesto para explicar estos fenómenos, la que más aceptación tiene consiste en suponer que hay una parte del material eyectado que permanece girando alrededor de la estrella, sin alejarse nunca demasiado. Ya he mencionado que sabemos que esto ocurre al menos en el Rectángulo Rojo, pues lo hemos observado. Una parte de éste material, volvería a caer sobre la estrella, adquiriendo según se acerca una velocidad cada vez mayor. La enorme energía cinética que llega a tener al acercarse a la estrella puede servir para empujar hacia afuera una porción de ese mismo gas, creándose dos chorros muy rápidos en sentidos opuestos (lo que llamamos una eyección bipolar). Estos chorros rápidos alcanzan pronto la envoltura expulsada previamente, chocan fuertemente con ella y aceleran sus regiones polares (figura 5). Se produce al mismo tiempo un calentamiento de algunas de estas componentes, que se convierten en emisoras de luz (como la parte de la Calabaza que vemos en azul en la figura 1). Al final, literalmente se abren agujeros en estas zonas polares.



Figura 7: En NGC 6302 la ruptura de la simetría es ya más amplia. (Cortesía de NASA/ESA/STScI.)

Para que el proceso funcione parece necesario que la estrella sea binaria, es decir que haya en realidad dos estrellas muy próximas en el centro de la nebulosa, rotando alrededor de su centro de gravedad. El disco puede así adquirir un movimiento de rotación por interacción con el sistema estelar; de otra manera, el material previamente expulsado por la estrella prácticamente no giraría nada (decimos nosotros que tendría un momento angular despreciable). También es necesario que exista un campo magnético bipolar (o sea con dos polos), aunque no sea muy intenso, lo que permite convertir el movimiento del material que cae en movimiento de una parte de éste hacia afuera y en forma de chorro. Este mecanismo se sabe que es eficaz, pero es poco intuitivo y no sería fácil de explicar aquí.

Sabemos que precisamente las estrellas centrales de nuestras dos nebulosas son dobles, aunque no en todas las nebulosas planetarias las hemos visto. En el caso del Rectángulo Rojo hemos detectado bien el disco en rotación. Hay indicios de que la Calabaza también lo tiene, pero es de tamaño muy reducido y su detección no es fácil. Por otro lado, podría ser que el proceso funcionase incluso si una de las estrellas es muy pequeña, o quizás sea suficiente si tenemos una estrella acompañada de un planeta grande. En este sentido las predicciones teóricas no son concluyentes.

También es posible que en algunos casos se produzca una interacción muy intensa entre ambas estrellas, debido a que la gigante engulle literalmente su compañera (como hará el Sol con algunos de los planetas, figura 3), es lo que nosotros llamamos la fase evolutiva en *common envelope*. Esto incrementaría tremendamente la pérdida de material por parte de la gigante y alteraría también la dinámica de la eyección, que se volvería asimétrica y más rápida en la dirección de los polos. La eyección múltiple implica también choques entre capas de material eyectadas antes y después. Se espera que, tras todo este proceso, la estrella principal se haya convertido en una enana muy pequeña (el núcleo de la gigante, como hemos citado ya) y, curiosamente, que la segunda estrella aparezca de nuevo, visible para nuestros telescopios y girando, como antes, como si nada hubiera pasado.

Calabaza.— Verdad es que poco se sabe con certeza de como nacemos las nebulosas. ¿Por qué una estrella, al final de su vida, lanza al exterior tan gran cantidad de material? ¿Y por qué lo hace de esta forma, que los astrónomos llaman bipolar?, que parece sólo destinada a darnos más hermosura ...

El autor.— Qué presumida.

Calabaza.— Y para que todo el mundo nos ame, por lo menos a mí.

Rectángulo.— Qué cruz.



Figura 8: NGC 6543, el Ojo de Gato, es una espectacular nebulosa planetaria, más evolucionada que las anteriores. Su estrella central ya es una enana blanca (antes fue una enana azul) y también sabemos que tiene una compañera. Se ve aún el resto de la envoltura esférica, pero también vemos como los choques avanzan rompiendo la simetría. Notemos que el eje en que se producen parece haber ido cambiando con el tiempo. En las imágenes más antiguas de esta nebulosa, sólo la región central más asimétrica se apreciaba, de ahí su nombre. El halo esférico sólo se ha visto recientemente, gracias a instrumentos más sensibles. Ojo de Gato es muy bella, pero más vieja que Calabaza. (Cortesía de NASA/ESA/STScI.)



Figura 9: Otra bonita nebulosa joven es Hen 401, en ella los choques se han extendido rápidamente y la afectan por entero. (Cortesía de NASA/ESA/STScI.)

El autor.— Añado yo: ¿por qué algunas estrellas que ya han dejado la fase de gigante roja parecen perezosas a la hora de continuar su evolución, mientras otras lo hacen de forma explosiva?, serían los casos de nuestros dos amigos. Quizás sea debido a una interacción de diferente naturaleza entre las dos estrellas entre sí y con la materia circundante, que sería mucho más intensa en el caso de Calabaza. Quizás los dos procesos que hemos citado antes, la eyección suave y continua de material para formar un disco en rotación y la interacción entre estrellas muy violenta, que aparece en cierto momento y altera fuertemente el proceso de eyección de gas y polvo, sean los responsables de estos casos que vemos tan diferentes. Pero de esto aún no podemos dar una explicación con certeza. De todos modos, ambos tipos de objetos parecen evolucionar hacia la formación de una nebulosa en expansión, la nebulosa planetaria, alrededor de un resto estelar, una estrella enana que fue el núcleo de la estrella gigante. Incluso en objetos que tienen una componente en rotación importante, como el Rectángulo Rojo, el material del disco se está escapando y pronto sólo veremos material nebuloso en expansión.



Figura 10: En NGC 2346, una nebulosa planetaria ya evolucionada, las zonas polares se han roto y las ecuatoriales son las más intensas. (Cortesía de NASA/ESA/STScI.)

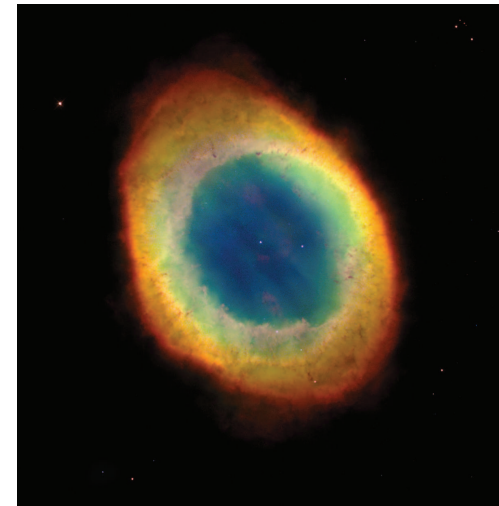


Figura 11: Al final tendremos algo como NGC 6720, el Anillo, del que sólo la zona ecuatorial es visible. En éstas dos nebulosas, como en la mayoría, el material que vemos no rota sino que se expande. (Cortesía de NASA/ESA/STScI.)

Calabaza.— Pronto no quedará atrás sino el mismo núcleo de la gigante, una minúscula estrella muy caliente, capaz, eso sí, de iluminarnos y darnos así, si cabe, aún más gloria.

Rectángulo.— Pero para mí la cosa es más lenta. Aún la estrella que dejé está lejos de hacerse tan pequeña y caliente. Pasan los cientos de años y todo va despacio.

El autor.— En casos como el de la Calabaza la estrella eyecta casi toda su masa en sólo unos cientos años. Como he mencionado ya, este material se acelera pronto a gran velocidad, de forma asimétrica. Y la estrella central se precipita en su evolución: en menos de mil años será lo que llamamos una estrella enana azul y poco después una enana blanca, una estrella muy pequeña (ahora, más pequeña que la estrella de partida, pues es sólo el núcleo de la gigante), pero muy caliente (más de nuevo que el Sol, con hasta 100 000 °C). Se ve azulada porque éste es el color de la materia así de caliente, luego se irá enfriando y se volverá más bien blanca. La radiación de esta estrella será capaz de excitar la nebulosa (ionizar, decimos nosotros en este caso, separando electrones de los átomos), dándole aún mayor resplandor y convirtiéndola en una nebulosa planetaria desarrollada.

También he mencionado que para otras nebulosas, las que presentan discos en rotación como nuestro Rectángulo, todo va más despacio. Creemos que las estrellas también se convertirán en enanas, pero la evolución parece mucho más lenta. Donde antes eran cientos de años, ahora son miles. Quizas sea tan lenta que, cuando la estrella central vuelva a ser caliente, el pobre Rectángulo ya se habrá expandido tanto, se habrá hecho tan difuso, que difícilmente pueda adornarse con la radiación de la nueva estrella.

Calabaza.— Todo llegará. Seguro que también verás, aunque sea desde lejos, la estrella excitante.

Rectángulo.— El tiempo tiene que seguir su camino. Al final, de todas las maneras, sólo habrá un rescoldo de estrella en nuestro centro. Quizás nos dé vida un tiempo, pero también ella se irá apagando.

El autor.— La fase de enana blanca es justo la fase final de la vida de la mayoría de las estrellas. A partir de ese momento, la estrella irá perdiendo energía lentamente hasta morir como un astro inerte, pequeño y frío.

La nebulosa irá, por su parte, diluyéndose, confundiéndose con el medio interestelar. Empezará a hacerse indetectable unos veinte mil años tras su nacimiento. Será a partir de estas nubes interestelares, alimentadas por los restos de estrellas anteriores, que se formarán nuevas estrellas después, como se formaron a su vez, hace mucho, las estrellas de las que surgieron las nebulosas planetarias actuales. Nacerán nuevos astros de nuestras bellas nebulosas, nacerán estrellas con planetas a su alrededor y posiblemente con vida en ellos.

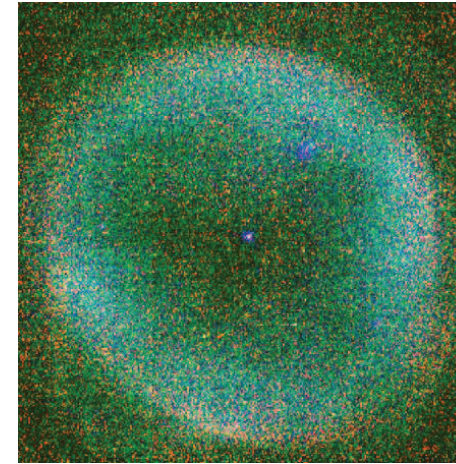


Figura 12: IC 1454 es una nebulosa planetaria muy evolucionada, muy vieja. Se ha extendido mucho debido a su dinámica expansiva y al tiempo transcurrido, su imagen es ya muy difusa. (Cortesía de NASA/ESA/STScI.)

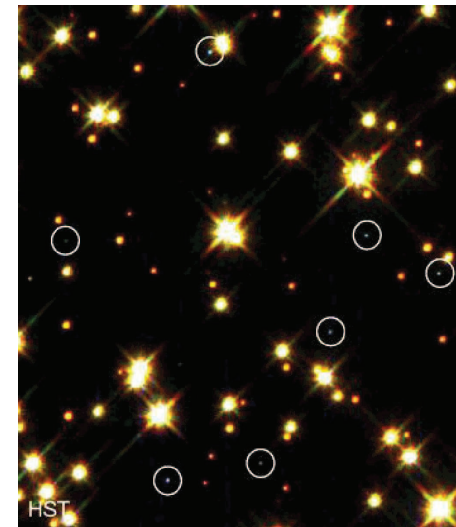


Figura 13: Imagen del cúmulo de estrellas M 4. Los círculos señalan estrellas enanas blancas muy viejas y débiles, ya no se ve rastro a su alrededor de la nebulosa. (Cortesía de NASA/ESA/STScI.)

Nuestro Sol, nuestra Tierra y nosotros mismos nos hemos formado de átomos que en su día formaban parte de nebulosas como éstas que ahora dialogan ante nosotros. Y al final, dentro de unos miles de millones de años, nuestro Sol, nuestra Tierra y nosotros mismos (o, mejor dicho, lo que quede de nuestros huesos) serán de nuevo eyectados por una gigante roja para formar parte de una calabaza o un rectángulo.

Rectángulo.— Y nosotros iremos diluyéndonos poco a poco, decayendo. Cada vez más difusos, menos conscientes de nuestro ser y nuestro pasado y ya despreocupados de nuestro destino. Hasta confundirnos en un todo inmenso y sin forma.

Calabaza.— ¡Habló la alegría de la huerta! No pienses en eso, mientras tanto somos lo más bello del Universo, y si luego hemos de servir para que vuelva a surgir la vida de nuestras carnes, tanto mejor. Soy joven y me sobra energía. Y tú, no nos engañemos, con esa cinturita y esos anchos hombros, tienes un aspecto tan varonil ...

Rectángulo.— Tú si que estás bien. Y si soy tan varonil y todo eso, digo yo que algo podríamos hacer, ¿no?

Calabaza.— (Fíate de la mosquita muerta.) Si tanto insistes.

Rectángulo.— (Tampoco ha hecho falta insistir mucho.) Se me ha ocurrido una ideíta ...

El autor.— Me parece que éste es el momento de dejarlos, y más que este artículo es de horario infantil. Quedan muchas preguntas por responder sobre este tema, y no se van a responder en un futuro próximo. O a lo mejor sí, precisamente estamos viendo que los nuevos instrumentos, como el potentísimo radiointerferómetro ALMA, están ahora mismo produciendo importantes avances en este campo. Uno de los primeros resultados de ALMA, obtenido hace sólo unos meses por nuestro grupo, se puede ver en la figura 2, la imagen del Rectángulo Rojo. Vemos en particular que algunos objetos, como éste, presentan los ingredientes necesarios para explicar la eyección de chorros que pueden crear nebulosas planetarias tal como las conocemos. Pero, precisamente en las nebulosas que presentan una evolución más rápida y profunda, como la Calabaza, estos elementos son menos evidentes. Estas nebulosas son fuertemente asimétricas, y tienen los movimientos más rápidos y las estructuras más espectaculares. ¿Estaremos ante una interacción muy violenta entre estrellas compañeras que no deja rastro en forma de disco rotante? ¿Y qué pasa si las estrellas no son dobles?, quizás sea suficiente la presencia de un planeta o quizás tendremos que admitir que estrellas solitarias (como nuestro Sol) siguen una senda evolutiva muy diferente que la que aquí hemos descrito.

No sabemos bien cómo han surgido, pero el caso es que estas nebulosas me caen simpáticas (como nosotros a ellas), hasta les toma uno cariño.



Figura 14: Gas y polvo interestelar en la nebulosa de Carina. De este material, que una vez fue expulsado por estrellas viejas, se están formando nuevos astros. Los filamentos que vemos en la imagen (por ejemplo, arriba, en el centro) señalan dónde se forman estrellas jóvenes, pues éstas, curiosamente, suelen lanzar chorros parecidos a los que vimos en las estrellas viejas y las nebulosas planetarias. (Cortesía de NASA/ESA/STScI.)

Agradecimientos

En este artículo he utilizado imágenes obtenidas con el Hubble Space Telescope (HST, ver <http://outreachoffice.stsci.edu>). La imagen de la emisión de CO la obtuvimos recientemente con el interferómetro ALMA (<http://www.almaobservatory.org>). Las figuras 3, 4 y 5 las compuso Florencio Martín, del OAN, sede de Alcalá de Henares, a quien estoy particularmente agradecido. En las referencias se citan algunos artículos científicos que detallan los estudios que he resumido tan drásticamente; como referencias generales recomiendo el libro de S. Pottasch, el artículo de B. Balick y A. Frank y, en español y en tono más divulgativo, los que respectivamente escribimos para este Anuario R. Bachiller y yo hace unos años.

Referencias

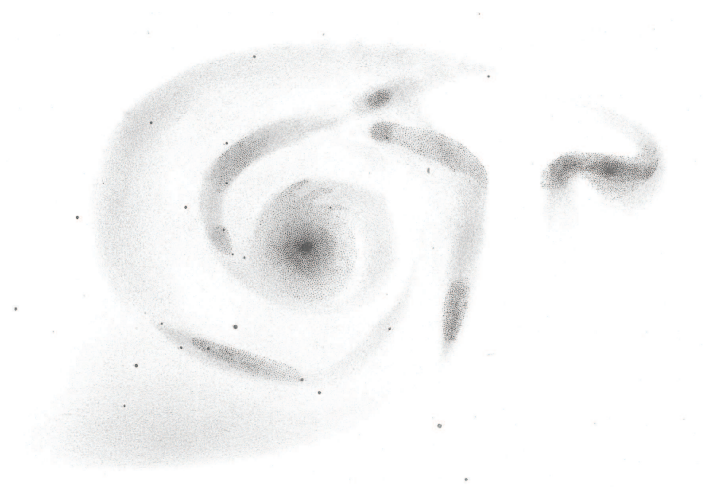
- S.R. Pottasch, 1984, “Planetary Nebulae”, Dordrecht, D. Reidel Publ. Co. (Para saber más sobre nebulosas planetarias y evolución estelar tardía.)
- B. Balick, A. Frank, 2002, ARA&A, 40, 439. (Para saber más en detalle de la formación de nebulosas planetarias.)
- R. Bachiller, “Nebulosas planetarias. Una mirada nueva sobre unos objetos viejos”, Anuario del Observatorio Astronómico 2000, pag. 307. (Referencia general en español sobre nebulosas planetarias.)
- V. Bujarrabal, “Evolución estelar tardía y nacimiento de las nebulosas planetarias”, Anuario del Observatorio Astronómico 2000, pag. 293. (Referencia general en español sobre las gigantes rojas y el nacimiento de las nebulosas planetarias.)
- V. Bujarrabal, J. Alcolea, C. Sánchez Contreras, R. Sahai, 2002, A&A, 389, 271. (Para saber más sobre las imágenes de la Calabaza.)
- J. Alcolea, V. Bujarrabal, C. Sánchez Contreras, R. Neri, J. Zweigle, 2001, A&A, 373, 932. (Para saber más sobre la emisión de CO en la Calabaza.)
- M. Cohen, H. Van Winckel, H.E. Bond, T.R. Gull, 2004, AJ, 127, 2362. (Para saber más sobre las imágenes HST del Rectángulo Rojo.)
- V. Bujarrabal, C. Castro-Carrizo, J. Alcolea, et al., 2013, A&A, en prensa, ver: <http://arxiv.org/abs/1307.5959>. (Para saber más sobre la emisión de CO en el Rectángulo Rojo y las imágenes de ALMA.)

- V. Bujarrabal, C. Castro-Carrizo, J. Alcolea, C. Sánchez Contreras, 2001, A&A 377, 868. (Para saber más sobre la dinámica de las nebulosas planetarias jóvenes.)
- A. Frank, E.G. Blackman, 2004, ApJ, 614, 737. (Para saber más sobre el mecanismo de eyección de chorros de alta velocidad en nebulosas planetarias jóvenes.)

Los artículos de divulgación aparecidos en el ANUARIO entre los años 1994 y 2013 pueden encontrarse en enlaces a los que se accede desde la página web:

<http://astronomia.ign.es/web/guest/anuario-astronomico>

*EL MEDIO
INTERESTELAR
Y LAS GALAXIAS*



EL AGUA EN EL UNIVERSO

Asunción Fuente

*Observatorio Astronómico Nacional
Instituto Geográfico Nacional - Ministerio de Fomento*

Abstract

Despite of being interesting by itself, the study of water in the Universe is important to understand the origin of water in the Earth. Thus far, our knowledge of the physics and chemistry of water outside the Solar System has been hampered by the scarce observational data. The Herschel Space Observatory has provided the opportunity of observing water vapor in all astronomical environments. In this paper we summarize the most important results.

Introducción

El agua en estado líquido es necesaria para la existencia de vida en la Tierra tal y como la conocemos. El agua es el líquido que sirve para el intercambio de alimentos y desechos entre el ser vivo y el exterior. El agua tiene además otras propiedades interesantes para los seres vivos como la capilaridad y la dilatación anómala. Gracias a la capilaridad, el agua puede subir a través de las raíces y los tallos de las plantas hasta las hojas en contra de la fuerza de la gravedad. La densidad del agua en estado sólido (hielo) es inferior a la densidad del agua en estado líquido, lo que permite al hielo flotar en el agua permitiendo que la vida continúe en las capas inferiores y evitando que los océanos se congelen. Su papel es también importante en el origen de la vida en la Tierra, pues en el océano se produjeron las reacciones químicas que dieron lugar a la compleja química orgánica que sustenta la vida.

En el Universo hay grandes cantidades de agua. El hidrógeno es el elemento más abundante en el Universo y unas veces en forma atómica (H) y otras en forma molecular (H₂) es responsable de prácticamente la totalidad de la masa visible del Universo. Ésta está dominada por las estrellas. En nuestra galaxia, la masa de las estrellas supone el ~85 % de

la masa total. El resto, $\sim 15\%$, se concentra en nubes de gas y granos de polvo que llenan el espacio entre las estrellas, es lo que se llama el medio interestelar. Los átomos pesados (aquellos con masa superior al H, principalmente los de carbono, oxígeno y nitrógeno, C, O, y N respectivamente) se forman en el interior de las estrellas. Al final de la vida de las estrellas, éstas explotan como supernovas en el caso de estrellas masivas o expulsan las capas externas de su envoltura formando nebulosas planetarias en el caso de estrellas de masa similar al Sol, de esta manera los átomos pesados se liberan al espacio incorporándose al medio interestelar. En el medio interestelar el gas y el polvo no se distribuyen uniformemente, sino que se acumulan formando nubes. En las partes más densas de las nubes interestelares, donde la radiación UV (ultravioleta) de las estrellas no puede llegar, se forman las moléculas. Actualmente se han identificado más de 160 moléculas diferentes en el medio interestelar (véase una lista actualizada en el apartado de **Tablas complementarias** de este Anuario), y el número de moléculas detectadas crece cada año debido a la mayor sensibilidad y resolución angular de los radiotelescopios.

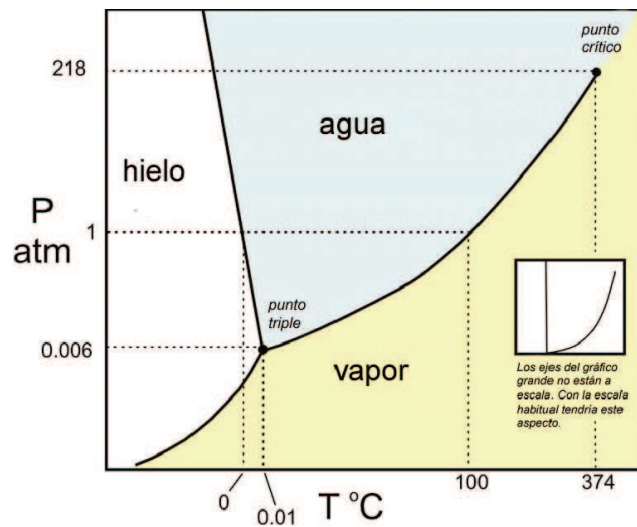


Figura 1: Diagrama de fases del agua en función de la presión y la temperatura (adaptado de *Water in the Universe*, Arnold Hanslmeier, Springer, 2010).

Después del hidrógeno molecular (H_2), el agua (H_2O) junto con el monóxido de carbono (CO) son la moléculas más abundantes. De hecho, por cada 10000 moléculas de hidrógeno, hay una molécula de agua. En astronomía, esto se suele expresar diciendo que la abundancia del agua

relativa al hidrógeno molecular¹ es 10^{-4} . El agua se encuentra en todos los entornos en el medio interestelar: regiones de formación estelar, núcleos activos de galaxias, sistemas planetarios y estrellas en las últimas fases de su vida. Se puede decir que el agua es ubicua en nuestro Universo. Sin embargo, resulta muy difícil encontrarla en estado líquido, que es el estado más habitual en la Tierra y el que tiene más interés desde el punto de vista biológico.

En el medio interestelar, el agua se encuentra generalmente en estado sólido (hielo) o gaseoso (vapor de agua). Esto se debe a las peculiares condiciones de presión y temperatura existentes. A la presión en la superficie terrestre ($P = 1$ atmósfera), el agua puede estar en estado sólido, líquido o gaseoso dependiendo de la temperatura (ver Figura 1). Para temperaturas menores que 0 grados centígrados ($= 273$ grados Kelvin) el agua se encuentra en estado sólido. Entre 0 y 100 grados centígrados (entre 273 y 373 grados Kelvin), el agua se encuentra en estado líquido. A temperaturas más altas, el agua se evapora. En el medio interestelar la presión es mucho menor que en la Tierra porque tanto la temperatura como la densidad típicas lo son: 10 K ($= -263$ grados C) y 10^4 núcleos de hidrógeno por centímetro cúbico respectivamente. Para estas presiones ($\sim 10^{-17}$ atmósferas) el agua sólo puede estar en estado sólido y gaseoso, y la temperatura de evaporación (estrictamente hablando es de sublimación pues se pasa directamente del estado sólido al gaseoso) es 100 grados Kelvin ($= -173$ grados C) (ver presiones menores que 0,006 atm en la Figura 1). En nuestra Galaxia, esta temperatura sólo se alcanza en las regiones más cercanas a las estrellas y en las zonas del medio interestelar donde se producen choques.

¿Podemos detectar agua en el espacio?

Dentro de nuestro Sistema Solar, se han enviado sondas a distintos planetas y satélites para conocer la composición de su atmósfera y su superficie. Se conoce la existencia de agua en Marte y en algunos satélites de Saturno y Júpiter. En Marte, hay dos grandes casquetes polares de hielo de agua y de dióxido de carbono. Europa, el satélite de Júpiter, está revestido de una capa de agua helada. La sonda Cassini descubrió grandes bloques de agua helada en Titán, luna de Saturno. La sonda Cassini también descubrió volcanes de hielo o géiseres que lanzan vapor de agua y otros gases al espacio en Encelado, luna de Saturno (ver Figura 2). El planeta enano Ceres y el asteroide Pallas, que orbitan en el entorno de Marte y Júpiter, también podrían contener agua. El agua es también un componente del hielo que recubre el núcleo de los cometas. Cuando el cometa transita por la cercanía del Sol, este hielo es evaporado formando la cabellera. Sin

¹O sea, el cociente entre el número de moléculas de H_2O y de H_2 es 10^{-4} .

embargo, a excepción de nuestro planeta, sobre la superficie de ningún otro cuerpo celeste del Sistema Solar se encuentra el agua en estado líquido. Los astrónomos creen que bajo la gruesa capa de hielo de los satélites Encelado y Europa existen océanos. Por esta razón, está proyectada una misión de exploración no tripulada a Europa cuya función consistiría en introducirse a través del escudo de hielo, cuyo grosor se calcula en 50 kilómetros, hasta toparse con la capa de agua líquida.



Figura 2: Representación artística de los volcanes de Encelado detectados por la sonda Cassini en el año 2009 (cortesía de NASA)

Dada la gran distancia que separa al Sol del resto de las estrellas de nuestra galaxia, la detección de agua líquida fuera del Sistema Solar no es posible. Por fortuna, podemos emplear la espectroscopía para detectar el vapor de agua incluso en las galaxias más lejanas. Los astrónomos utilizamos la espectroscopía para conocer la composición química y determinar las condiciones físicas del gas en las nubes interestelares. Dependiendo de su estructura molecular, cada compuesto químico emite radiación a unas frecuencias determinadas dando lugar a líneas espectrales, es decir tiene un espectro característico. Dicho espectro nos permite identificar una molécula de la misma manera que las huellas digitales permiten identificar a los seres humanos. La intensidad de las líneas y los cocientes entre las intensidades de unas y otras, dependen de las condiciones de presión y temperatura del gas. Por tanto, a partir del espectro de un gas se pueden identificar las moléculas que lo componen y determinar sus condiciones físicas.

El vapor de agua interestelar fue detectado por primera vez hace más de 40 años en la nube molecular de Orión a través de la emisión máser de la

línea a 22 GHz² (Cheung et al. 1969). Este tipo de emisión se produce en zonas extremadamente densas y calientes, que son la excepción en el medio interestelar. Por esta razón, el conocimiento sobre la distribución espacial del vapor de agua que se puede tener a partir de la emisión de esta línea es muy limitado. La puesta en marcha del radiotelescopio de 30m de diámetro de IRAM en Pico de Veleta permitió la detección de la emisión máser de las líneas de agua a 183 GHz y 325 GHz. La emisión máser de estas líneas procede de regiones más extensas, lo que permitió realizar por primera vez un mapa de la emisión del vapor de agua en Orión y otras regiones de formación estelar (Cernicharo et al. 1990, 1994; González-Alfonso et al. 1995). Líneas infrarrojas de agua procedentes de regiones muy calientes en la cercanía de las estrellas (a distancias menores que 10 UA) han sido observadas desde Tierra con el *Very Large Telescope* (VLT) y desde el espacio con el *Spitzer Space Telescope*. Sin embargo, la mayor parte del vapor de agua en el medio interestelar se encuentra en regiones más frías y menos densas que no emiten en estas líneas.

Para estudiar la distribución espacial del vapor de agua en el medio interestelar, necesitamos observar las transiciones de agua de más baja energía que se hallan en la región del infrarrojo lejano del espectro electromagnético. Desgraciadamente, debido a la gran cantidad de vapor de agua que hay en la atmósfera terrestre, toda la radiación proveniente del espacio a estas frecuencias es absorbida al atravesar la atmósfera, por lo que para observar estas líneas necesitamos telescopios espaciales. En los últimos años, varias misiones espaciales han sido lanzadas teniendo entre sus objetivos la observación del vapor de agua: *Submillimeter Wave Astronomy Satellite* (SWAS, Melnick et al. 2000); ODIN (e.g., Hjalmarson et al. 2003), y *Infrared Space Observatory* (ISO) (van Dishoeck 2004). Pero ha sido el *Herschel Space Observatory*, por su mayor cobertura en frecuencia, su mayor sensibilidad y su mayor resolución angular, el que ha permitido el estudio del vapor de agua en todos los entornos astrofísicos. En la Tabla 1 se listan las líneas más intensas del vapor de agua en las regiones milimétrica y del infrarrojo lejano del espectro, así como el telescopio con el que se pueden (o se han podido) observar.

El Observatorio Espacial Herschel

El observatorio espacial Herschel es el cuarto programa “cornerstone” de la Agencia Espacial Europea (ESA). Fue lanzado desde la Guayana francesa el 14 de mayo del 2009. El satélite lleva a bordo una antena Cassegrain con 3,5 metros de diámetro que es la más grande que se ha lanzado al espacio hasta ahora. La instrumentación que la acompaña permite hacer fotometría y espectroscopía en el rango de 55-671 μm

²El gigahercio es un múltiplo de la unidad de frecuencia, el hercio. 1 GHz = 10^9 Hz

Table 1: Líneas de H₂O observadas

Línea ¹	Freq.(GHz)	$\lambda(\mu\text{m})$	E_u (K) ²	Telescopio
H ₂ O 6 _{1,6} →5 _{2,3}	22.235	1.35 cm	626	desde Tierra
3 _{1,3} →2 _{2,0}	183.310	1.63 mm	199	desde Tierra
5 _{1,5} →4 _{2,1}	325.153	0.92 mm	458	desde Tierra
1 _{1,0} →1 _{0,1}	556.936	538.3 μm	61	SWAS, ODIN, Herschel
2 _{1,2} →1 _{0,1}	1669.905	179.5 μm	114.4	ISO, Herschel
1 _{1,1} →0 _{0,0}	1113.343	269.3 μm	53.4	Herschel
2 _{0,2} →1 _{1,1}	987.927	303.5 μm	100.8	Herschel
2 _{1,1} →2 _{0,2}	752.033	398.6 μm	136.9	Herschel
3 _{1,2} →3 _{0,3}	1097.365	273.2 μm	249.4	Herschel
3 _{1,2} →2 _{1,1}	1153.127	260.0 μm	249.4	Herschel
2 _{2,1} →2 _{1,2}	1661.008	180.5 μm	194.1	ISO, Herschel
2 _{1,2} →1 _{0,1}	1669.905	179.5 μm	114.4	ISO, Herschel
3 _{0,3} →2 _{1,2}	1716.770	174.6 μm	196.8	ISO, Herschel
3 _{1,3} →2 _{0,2}	2164.132	138.5 μm	204.7	ISO, Herschel
3 _{3,0} →3 _{2,1}	2196.346	136.5 μm	410.7	ISO, Herschel
4 _{0,4} →3 _{1,3}	2391.573	125.4 μm	319.5	ISO, Herschel
4 _{1,4} →3 _{0,3}	2640.474	113.5 μm	323.5	ISO, Herschel
2 _{2,1} →1 _{1,0}	2773.977	108.1 μm	194.1	ISO, Herschel
3 _{2,2} →2 _{1,1}	3331.458	90.0 μm	296.8	ISO, Herschel
7 _{0,7} →6 _{1,6}	4166.852	71.9 μm	843.5	ISO, Herschel

¹ Sólo se listan las transiciones que dan lugar a las líneas más intensas.² Energía del nivel superior de la transición expresada en grados K. Para que la emisión del vapor de agua en una transición sea intensa, el gas debe tener una temperatura mayor que $E_u/2$.

(Pilbratt et al. 2010; de Graauw et al. 2010). Varias líneas de agua caen en este rango de longitudes de onda permitiendo un estudio detallado de la distribución espacial y las condiciones físicas del vapor de agua en el medio interestelar. Para una completa información sobre el Observatorio Espacial Herschel, ver el artículo “El Observatorio Espacial Herschel” publicado por Pere Planesas y Javier Graciá Carpio en el Anuario del Observatorio Astronómico Nacional del año 2007. A continuación se detallan algunos de sus principales aportaciones en los dos años que han transcurrido desde su lanzamiento.

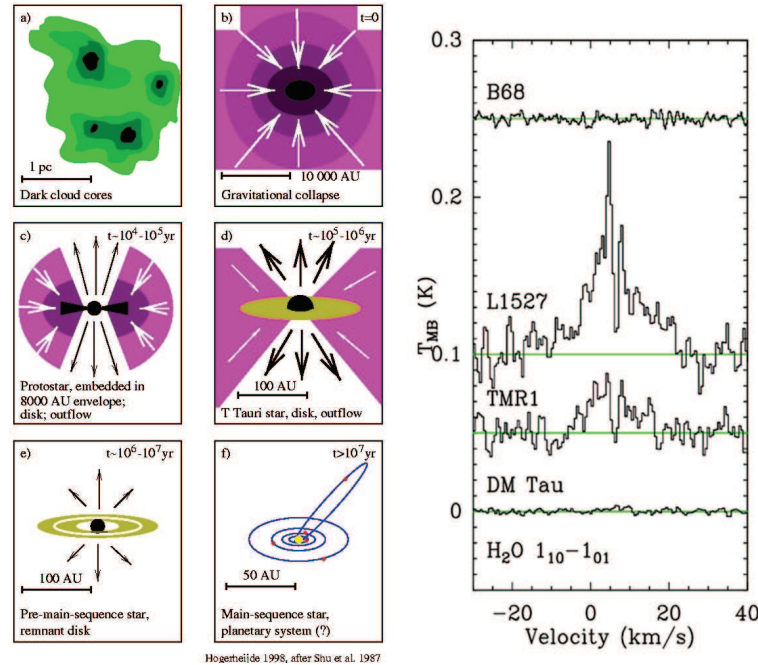


Figura 3: Representación artística del Telescopio Espacial Herschel (cortesía de la Agencia Espacial Europea, ESA).

El agua en las nubes donde se forman las estrellas

Las estrellas se forman en los núcleos fríos ($T \sim 10$ K) y densos (más de 10000 moléculas de hidrógeno por centímetro cúbico) de las nubes interestelares. A estos núcleos se les denomina condensaciones pre-estelares. En estas zonas prácticamente toda el agua se encuentra formando mantos de hielo en la superficie de los granos de polvo. El agua congelada en los granos de polvo, puede transformarse en vapor de agua si el grano se calienta hasta alcanzar una temperatura mayor que 100 K (temperatura de sublimación del agua a las presiones existentes en el medio interestelar). Cuando los granos están a una temperatura inferior, existen otros mecanismos que pueden devolver parte de las moléculas de agua a la fase gaseosa manteniendo cierta cantidad de vapor de agua en la nube. Estos mecanismos son: fotodesorción (la molécula de agua absorbe

un fotón ultravioleta y la energía adquirida le sirve para pasar a la fase gaseosa), *sputtering* (parte de la energía cinética disipada en un choque se utiliza para liberar las moléculas de agua y devolverlas al gas) y desorción por rayos cósmicos y rayos X.



Hogerthide 1998, after Shu et al. 1987

Figura 4: A la izquierda se muestra un esquema de la formación de una estrella de masa similar al Sol. Los paneles a) y b) corresponden a la fase de condensación pre-estelar. Los paneles c) y d) corresponden a la fase de protoestrella, en la que mientras la estrella adquiere la totalidad de su masa, se producen fuertes eyecciones bipolares. Los paneles e) y f) corresponden a una estrella adulta rodeada de un disco protoplanetario. A la derecha mostramos observaciones de la línea $\text{H}_2\text{O } 1_{1,0} \rightarrow 0_{0,1}$ en una condensación pre-estelar (B68), dos protoestrellas (L1527 y TMR1) y un disco protoplanetario (DM Tau). Es evidente la gran variación que se produce en la emisión del vapor de agua durante la formación y vida de una estrella (van Dishoeck et al., 2011).

Observaciones recientes con el telescopio espacial Herschel demuestran que la abundancia de vapor de agua en el interior de los núcleos densos y fríos de las nubes interestelares es menor que $1,3 \cdot 10^{-9}$ pudiendo aumentar

hasta $\sim 10^{-8}$ en la parte externa de la nube debido a la fotodesorción. Esta abundancia es mucho menor que la abundancia esperada para el agua, 10^{-4} , lo que indica que prácticamente la totalidad del agua se halla atrapada en los hielos. En la Figura 4 se muestra el espectro a la frecuencia de la transición $\text{H}_2\text{O } 1_{1,0} \rightarrow 0_{0,1}$ obtenido con Herschel en la condensación pre-estelar B68. Como vemos, el agua no ha sido detectada en este núcleo frío.

Cuando comienza el colapso, en el interior de la nube se forma un objeto protoestelar (embrión de estrella) que calienta el gas y el polvo hasta una temperatura mayor que 100 K. Esta región caliente es muy pequeña (típicamente unos pocos centenares de unidades astronómicas) y se denomina *hot core* o *hot corino* (dependiendo de la masa de la estrella). En estos núcleos calientes la temperatura de los granos excede la temperatura de evaporación del agua y prácticamente la totalidad del agua regresa a la fase gaseosa. Paralelamente, en estas primeras fases de la formación de las estrellas se producen fuertes vientos y eyecciones bipolares que crean choques en la nube materna de la misma manera que un barco crea una frente de choque cuando avanza por el océano. En estas regiones de choque, todo el agua es liberada a la fase gaseosa debido al *sputtering* y la temperatura del gas asciende a miles de grados Kelvin.

En la Figura 4 se muestran los espectros de la transición $\text{H}_2\text{O } 1_{1,0} \rightarrow 0_{0,1}$ observados con Herschel en las protoestrellas L1527 y TMR1. La emisión de la línea de vapor de agua es muy intensa en estas protoestrellas. Es también de destacar, que las líneas son muy anchas en velocidad, con anchuras mayores que 20 km s^{-1} (unos 70.000 km/h), lo que sugiere que la mayor parte de la emisión proviene de las regiones de choque producidas cuando los vientos y las eyecciones bipolares golpean a la nube. La cantidad de masa de la nube afectada por los choques es mucho mayor que la cantidad de masa que forma el *hot corino* y la temperatura del gas es también mayor. Por tanto, es la emisión del vapor de agua producido en los choques la que domina el espectro.

Esta interpretación es confirmada por los mapas de agua que se han realizado en este tipo de protoestrellas. En la Figura 5 se muestra un mapa de la emisión de agua a $179 \mu\text{m}$ en la estrella joven L1157mm. La mayor parte de la emisión no proviene de las cercanías de la estrella, sino de los choques entre las eyecciones bipolares y la nube materna. La abundancia de vapor de agua en las zonas de choque es tan alta, $\sim 10^{-4}$, que podemos decir que prácticamente todo el oxígeno se encuentra en forma de agua.

En los espectros de agua de L1527 y TMR1 mostrados en la Figura 4, se puede observar una pequeña (auto-)absorción en el centro de la línea. Esta absorción nos permite saber que la abundancia máxima del agua en el gas de la nube materna que no está afectada por los choques ni por el *hot corino* es de $\sim 10^{-8}$, similar a la que había antes de formarse la estrella.

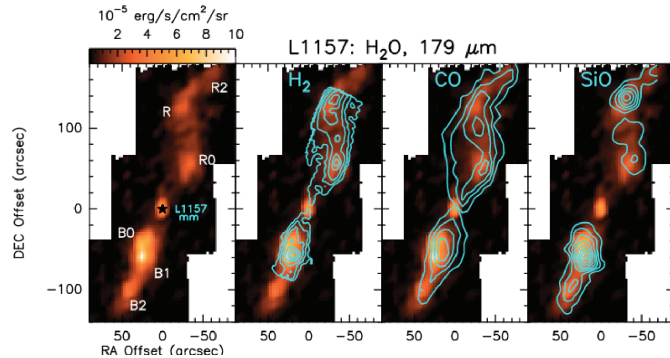


Figura 5: En el panel de la izquierda se muestra el mapa de la emisión de la línea de vapor de agua a $179 \mu\text{m}$ realizado con Herschel en la protoestrella L1157-mm. Nótese que la mayor parte de la emisión de agua procede de las zonas de choque entre las eyecciones bipolares y la nube materna. En el resto de los paneles se compara la emisión de agua con otros trazadores de choque encontrándose una buena correspondencia (Nisini et al. 2010).

El agua en discos protoplanetarios

Durante la formación de las estrellas, las eyecciones bipolares, los vientos y la radiación ultravioleta producida por la joven estrella dispersan el gas y el polvo de la nube. Cuando la estrella llega a la edad adulta, lo único que resta de la nube materna es un pequeño disco circunestelar. En este disco circunestelar se formarán los futuros planetas, por lo que también recibe el nombre de disco protoplanetario. La detección de agua en discos protoplanetarios ha sido y sigue siendo un objetivo prioritario de Herschel, debido a las implicaciones que un descubrimiento de este tipo tiene para comprender el origen del agua en la Tierra. Bergin et al. (2010) publicó una detección tentativa de la línea de $\text{H}_2\text{O } 1_{1,0} \rightarrow 0_{0,1}$ en el disco protoplanetario de DM Tau (ver Figura 4). La intensidad de esta línea es un factor 20-230 más débil de lo que predicen los modelos. La explicación más aceptada para esta falta de vapor de agua es que un 95%-99% del agua se encuentra en forma de hielo en los granos de polvo que se acumulan en el plano medio del disco. El telescopio espacial Spitzer había detectado con anterioridad vapor de agua en las partes más internas (distancias a la estrella menores que 10 UA) de algunos discos protoplanetarios con una abundancia similar a la esperada. Uniendo la información procedente de ambos telescopios llegamos a la conclusión de que la abundancia y estado del agua cambian con la distancia a la estrella en los discos protoplanetarios. Mientras que en las regiones cercanas a la estrella el agua está en estado gaseoso, en

las regiones más alejadas (~ 100 UA), el agua parece estar esencialmente en forma de hielo que se sitúa en el plano medio del disco. El límite que separa una región de otra es lo que se llama *snowline*.

El agua en las envolturas de las estrellas viejas

Durante la edad adulta, las estrellas obtienen energía mediante la fusión del hidrógeno. La fusión del hidrógeno consiste en formar un átomo de helio a partir de cuatro átomos de hidrógeno. El Sol, con una edad de 4500 millones de años, se encuentra en esta fase de su vida. Pero dentro de 4000 o 5000 millones de años, el Sol habrá gastado todo el hidrógeno de su interior y entrará en la fase de gigante roja. Durante esta fase su tamaño aumentará en más de cien veces, llegando a engullir a Mercurio, Venus, incluso quizás a la Tierra, y la temperatura de su superficie descenderá hasta unos 2000 K, lo que permitirá que se formen moléculas en su atmósfera (entre otras, el agua). Siguiendo con su evolución, en una etapa posterior, el Sol expulsará sus capas externas perdiendo una fracción significativa de su masa. Estas capas formarán primero una envoltura fría alrededor de lo que queda de estrella. Esta envoltura se expandirá hasta ser expulsada y los átomos pesados se incorporarán al medio interestelar. La última etapa se conoce con el nombre de nebulosa planetaria. Herschel nos ha permitido estudiar la abundancia de vapor de agua en las envolturas frías (unos 1000 K) y masivas de las gigantes rojas y en las envolturas más extensas y tenues de las nebulosas planetarias, planteando nuevos problemas al estudio de la química en estos objetos.

Las moléculas que se forman en la envoltura de las estrellas evolucionadas y su abundancia dependen de la razón C/O, que a su vez depende de la masa de la estrella y de su estado evolutivo. En general, podemos decir que las estrellas carbonadas (C/O mayor que 1 en su envoltura) tienen menor masa que las estrellas oxigenadas (C/O menor que 1 en su envoltura). En las estrellas oxigenadas, las moléculas que se espera que sean más abundantes son CO y H_2O , con una abundancia de vapor de agua respecto al hidrógeno molecular entre 10^{-7} y 10^{-6} , similar a otras regiones del medio interestelar. Sin embargo, en las estrellas carbonadas, las moléculas más abundante deberían ser CO y C_2H_2 con una abundancia muy pequeña de vapor de agua. Los resultados obtenidos con SWAS y Herschel contradicen estos modelos. La primera sorpresa fue la detección de vapor de agua en la estrella carbonada IRC+10216 realizada con el telescopio espacial SWAS (Melnick et al. 2001). Diversas explicaciones se propusieron para explicar esta inesperada detección de vapor de agua. Una posible explicación era que el agua procedía de las regiones de choque que se producen en la envoltura entre los vientos más recientes y el gas expulsado anteriormente. Otra, más exótica, era que el agua procedía de la evaporación de cometas u objetos rodeados de hielo que orbitarían alrededor de la estrella. La re-

cienta detección con Herschel de vapor de agua en 8 estrellas carbonadas (Neufeld et al. 2011), y la similitud entre los perfiles de las líneas de CO y H₂O contradicen esta última interpretación. La formación de agua en la envoltura de estas estrellas debe ser el resultado de choques en su envoltura, o de fotones ultravioletas interestelares que iluminan la envoltura de la estrella desde el exterior.

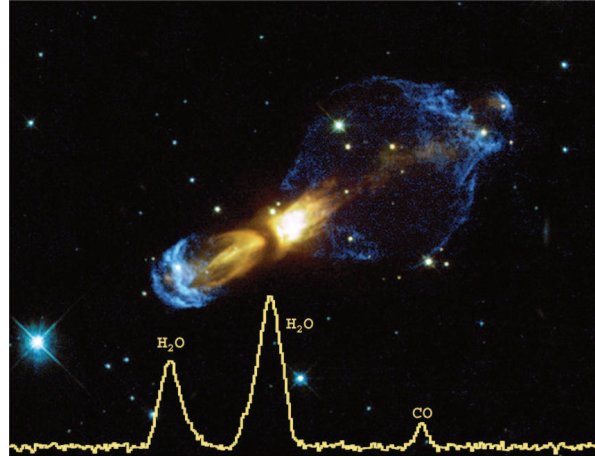


Figura 6: Ejemplo de espectro de vapor de agua y monóxido de carbono obtenido con el Observatorio Espacial Herschel (amarillo claro) en la nebulosa planetaria OH231.8+4.2, superpuesto a una imagen óptica de la nebulosa obtenida con el Telescopio Espacial Hubble (cortesía de V. Bujarrabal, investigador principal del programa clave de Herschel HIFISTARS).

El agua en otras galaxias

El vapor de agua no sólo es un ingrediente fundamental del medio interestelar en nuestra Galaxia sino también en las galaxias externas. Herschel ha permitido la observación de vapor de agua en la galaxia ultraluminosa Markarian 231 (González-Alfonso et al. 2010) y la galaxia con brotes de formación estelar M 82 (Weiß et al. 2010). El origen de la emisión es diferente en estas dos galaxias. En Markarian 231, la emisión procede del núcleo de la galaxia. Los rayos X producidos por el agujero negro que hospeda en su interior y los fuertes choques son los responsables de que gran cantidad de moléculas de agua se hallen formando vapor de agua en este núcleo activo. En M 82 la emisión de vapor de agua parece estar asociada a una componente más difusa que rodea el núcleo de la galaxia formando un halo.

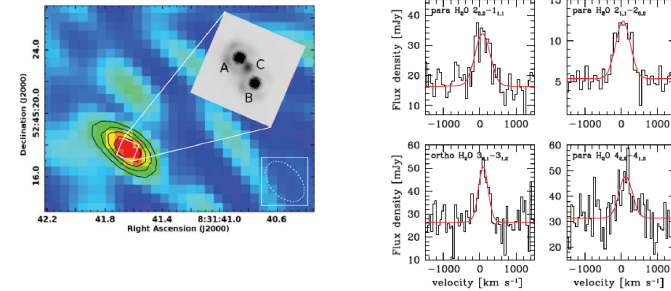


Figura 7: A la derecha se muestran los espectros de las líneas de agua observadas en APM 08279+5255 utilizando el interferómetro de IRAM del Plateau de Bure (Francia). El mapa en color corresponde a la emisión integrada de la línea de 2_{1,1}→2_{0,2} de H₂O. En grises aparece la imagen de la galaxia en infrarrojo (Van der Werf et al., 2011).

Probablemente el descubrimiento más impactante en cuanto a la detección de agua en galaxias externas ha sido la reciente detección de vapor de agua en el quásar APM 08279+5255 realizado con el interferómetro de IRAM situado en el Plateau de Bure (Francia) (Lis et al., 2011; Van der Werf et al. 2011). La galaxia APM 08279+5255 está situada a un corrimiento hacia el rojo de $z=3.91$, lo que implica que la emisión que observamos proviene del Universo temprano, cuando la edad del Universo era aproximadamente un décimo de la edad actual (la edad actual es de unos 13 700 millones de años). El análisis de las líneas de agua detectadas muestra que su emisión proviene del núcleo de la galaxia donde grandes cantidades de agua son producidas por los choques y los rayos X procedentes del agujero negro en su interior. Esta lejana galaxia parece más similar a Markarian 231 que a M 82. El estudio del agua en este caso, y su comparación con las galaxias más cercanas Markarian 231 y M 82, nos permite tener una idea de cómo eran las galaxias en el joven Universo.

¿Qué hemos aprendido con Herschel sobre el origen del agua en la Tierra?

Toda el agua que hay en nuestro planeta se formó en la más tierna infancia de la Tierra, hace unos 2500 millones de años. En la actualidad el agua puede cambiar de fase como cuando se evapora de los océanos, o de sitio con los cambios climáticos, pero la cantidad total permanece prácticamente constante. De hecho ha sido constante desde que apareció la vida en la Tierra y el agua que bebemos ahora la compartimos con todos los seres vivos que han habitado y habitarán la Tierra.

Hay diversas teorías sobre cómo se formaron los océanos, aunque ninguna de ellas da una explicación totalmente satisfactoria. En los discos protoplanetarios hay agua en estado sólido y gaseoso. El agua de los océanos podría provenir de la mezcla de rocas y gases a partir de la cual se formó la Tierra. Pero las teorías de formación de planetas predicen que la mayor parte de ese agua escaparía al espacio en los primeros millones de años de vida del planeta cuando la Tierra era menos masiva de lo que es hoy en día. La temperatura del planeta en ese momento era tan alta que todo el agua estaba en estado gaseoso y dado que la Tierra era muy poco masiva, su gravedad no era suficiente para retener las moléculas de vapor de agua en su atmósfera, que escaparían al espacio. Unos 20 o 30 millones de años después de su formación, la Tierra sufrió el impacto de un gran meteorito que le desgajó un trozo, la Luna. Algunos investigadores piensan que en ese impacto pudo formarse, además de la Luna, gran parte del agua. Otra de las teorías más aceptadas propone que el agua podría provenir del impacto de cometas recubiertos de hielo en la joven Tierra. Los choques entre cometas y planetas eran bastante frecuentes en las épocas más tempranas del Sistema Solar.

Una manera de averiguar el origen del agua de la Tierra es midiendo la cantidad relativa de agua deuterada respecto al agua normal ($\text{HDO}/\text{H}_2\text{O}$) en los cometas y compararla con la razón que se mide en los océanos. Todo el hidrógeno y deuterio (D) del Universo fue formado en el Big Bang y la razón es $\text{D}/\text{H}=0,00001$. Sin embargo, la razón $\text{HDO}/\text{H}_2\text{O}$ cambia dependiendo de las condiciones físicas del lugar en el que se forman ambas moléculas. Si el agua de la Tierra procede de un meteorito o cometa, la razón $\text{HDO}/\text{H}_2\text{O}$ en los cometas debería ser la misma que la de los océanos ($0,015\% = 0,00015$).

Hasta ahora, la razón $\text{HDO}/\text{H}_2\text{O}$ en los cometas en que se había medido era dos veces mayor que la de los océanos lo que contradecía la teoría del origen cometario del agua. Recientemente, se ha medido la razón $\text{HDO}/\text{H}_2\text{O}$ en el cometa Hartley 2 utilizando el observatorio espacial Herschel. La razón $\text{HDO}/\text{H}_2\text{O}$ es casi idéntica a la de los océanos. La diferencia entre este cometa y los cometas estudiados anteriormente es que el cometa de corto periodo Hartley 2 se originó en el cinturón de Kuiper situado más allá de Neptuno, en los confines del Sistema Solar, mientras los otros cometas estudiados, aunque se piensa que se formaron en la región de los planetas gigantes, debieron ser eyectados por estos en las primeras etapas de la formación del sistema solar hasta la nube de Oort, de donde provienen ahora. Este descubrimiento abre de nuevo la posibilidad del origen cometario del agua en la Tierra. El agua podría proceder en parte de choques de cometas formados en el cinturón de Kuiper con la Tierra.



Figura 8: Imagen del cometa 103P/Hartley 2 tomada en 2010 por la sonda EPOXI (cortesía de NASA/JPL-Caltech/UMD)

Conclusiones

El agua, un compuesto esencial para la vida en la Tierra, es una de las moléculas más abundantes del Universo. Debido a las condiciones de presión y temperatura del medio interestelar, el agua se encuentra generalmente en estado sólido o gaseoso fuera de la Tierra. El lanzamiento del Observatorio Espacial Herschel nos ha permitido conocer con detalle la distribución del vapor de agua en el Universo. Ahora sabemos que en las nubes moleculares de nuestra Galaxia el agua se encuentra esencialmente en estado sólido, siendo uno de los principales componentes de los mantos helados que rodean los granos de polvo. El vapor de agua sólo es abundante en las regiones cerca de las estrellas, donde debido a las altas temperaturas y a los choques que producen las eyecciones bipolares, los hielos se destruyen y las moléculas de agua se transfieren al estado gaseoso. En las últimas etapas de la vida de las estrellas, el agua se forma en cantidades apreciables en las envolturas de las gigantes rojas. Cuando las gigantes rojas llegan al estado de nebulosa planetaria, el gas de las envolturas va siendo expulsado poco a poco al medio interestelar donde las moléculas de agua son destruidas por la radiación ultravioleta liberando los átomos de O e H, que posteriormente servirán para formar nuevos compuestos (entre ellos, agua) en las nubes moleculares.

El agua es también una molécula muy abundante en otras galaxias. En algunas de ellas, debido a los rayos X que producen sus agujeros negros, a los choques entre nubes, o a la intensa formación estelar que hay en sus núcleos (órdenes de magnitud mayor que en el núcleo de nuestra galaxia), la abundancia del vapor de agua es tan grande que puede ser detectado con nuestros telescopios.

Respecto al origen el agua en nuestro planeta, éste sigue siendo aún un tema muy controvertido. La observación del agua en nuestro Sistema Solar

y en otros lugares del Universo, en concreto en los discos protoplanetarios, es imprescindible para desvelar nuestro propio origen.

Referencias

- A. Hansmeier, *Water in the Universe*, ed. Springer.
- E. A. Bergin, M. R. Hogerheijde, C. Brinch et al. *Astronomy & Astrophysics*, 521, L33 (2010).
- P. Caselli, E. Keto, L. Pagani et al. *Astronomy & Astrophysics*, 521, L29 (2010).
- A. C. Cheung, D. M. Rank, & C. H. Townes. *Nature*, 221, 626 (1969).
- J. Cernicharo, C. Thum, H. Hein et al. *Astronomy & Astrophysics*, 231, L15 (1990).
- J. Cernicharo, E. González-Alfonso, J. Alcolea, R. Bachiller & D. John. *Astrophysical Journal*, 432, L59 (1994).
- T. de Graauw, F. P. Helmich, T. G. Phillips et al. *Astronomy & Astrophysics*, 518, L6 (2010).
- E. González-Alfonso, J. Cernicharo, R. Bachiller & A. Fuente. *Astronomy & Astrophysics*, 293, L9 (1995).
- E. González-Alfonso, J. Fischer, K. Isaak et al. *Astronomy & Astrophysics*, 518, L43 (2010).
- Å. Hjalmarson, U. Frisk, M. Olberg et al. *Astronomy & Astrophysics*, 402, L39 (2003).
- D. C. Lis, D. A. Neufeld, T. G. Phillips, M. Gerin & R. Neri. *Astrophysical Journal*, 738, L6 (2011).
- G. J. Melnick, J. R. Stauffer, M. L. N. Ashby et al. *Astrophysical Journal*, 539, L77 (2000).
- G. J. Melnick, D. A. Neufeld, K. E. S. Ford, D. J. Hollenbach & M. L. N. Ashby. *Nature*, 412, 160 (2001).
- D. A. Neufeld, E. González-Alfonso, G. Melnick et al. *Astrophysical Journal*, 727, L29 (2011).
- B. Nisini, M. Benedettini, C. Codella et al. *Astronomy & Astrophysics*, 518, L120 (2010).
- G. L. Pilbratt, J. R. Riedinger, T. Passvogel et al. *Astronomy & Astrophysics*, 518, L1 (2010).
- P. P. van der Werf, A. Berciano Alba, M. Spaans et al. *arXiv:1106.4825* (2011).
- E. F. van Dishoeck. *Annual Review Astronomy & Astrophysics*, 42, 119 (2004).
- E. F. van Dishoeck, L. E. Kristensen, A. O. Benz et al. *PASP*, 123, 138 (2011).
- A. Weiß, M. A. Requena-Torres, R. Güsten, S. García-Burillo et al. *Astronomy & Astrophysics*, 521, L1 (2010).
- NASA, <http://www.nasa.gov/home/index.html>.
- ESA, <http://www.esa.int/esaCP/Spain.html>.
- Herschel Science Center, http://herschel.esac.esa.int/latest_news.shtml.

PUESTA AL DÍA DEL TEMA

Avances debidos a Herschel y ALMA

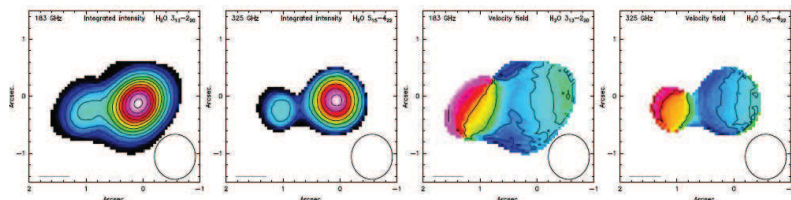
El artículo *El agua en el Universo* se publicó en el Anuario en el año 2012. La mayor parte de nuestro conocimiento sobre la distribución y abundancia del agua en el Universo proviene de las observaciones realizadas por el telescopio espacial de observación infrarroja Herschel. El Observatorio Espacial Herschel comenzó su andadura en el 2009 y la finalizó en el 2014 dejando un importante legado observacional que actualmente se ordena en un archivo público al que todos los científicos podemos acceder.

Herschel contribuyó enormemente a nuestra comprensión de la química interestelar pero dejó también importantes problemas sin resolver. Durante estos años, los astrónomos hemos continuado trabajando con estos datos y desarrollando nuevos modelos para explicar las observaciones. Por su importancia a la hora de entender el origen de la vida en nuestro planeta, yo destacaría la dificultad que aún tenemos para explicar la abundancia de vapor de agua observada en los discos protoplanetarios, es decir, en los precursores o padres de los sistemas planetarios como el nuestro. Herschel demostró que el vapor de agua en estos discos es 100 veces menos abundante de lo que predecían los modelos químicos (anteriores al 2009). Desde entonces, se ha hecho un gran esfuerzo teórico para comprender este sorprendente resultado. Actualmente, creemos que el agua se encuentra fundamentalmente en forma de hielo en estos discos. Además, estos hielos recubren granos de polvo relativamente grandes (mayores que unos cuantos micrómetros) que se acumulan en la región más fría del disco, en su plano ecuatorial. Estos granos “hidratados” están secuestrados por la fuerza gravitatoria que les impide subir a la superficie del disco donde el hielo podría evaporarse produciendo vapor de agua (van Dishoeck et al. 2014).

Otra conclusión importante de los modelos químicos posteriores a Herschel es que el agua que se encuentra en los hielos no ha tenido tiempo para formarse en el disco, sino que proviene de la nube materna de la que se formaron el disco y la estrella. Esto explicaría la similitud entre la deuteración del agua medida en los océanos terrestres y la medida en los cometas y meteoritos en términos de un origen común. Si esta teoría es correcta, concluimos que una fracción importante del agua en la Tierra y los otros cuerpos del Sistema Solar se formó antes que el Sol. Esta conclusión confirma, además, la estrecha relación existente entre la composición química de la nube materna y la de los discos protoplanetarios que se forman a partir de ella y refuerza el interés que el estudio de la química de las nubes interestelares tiene para la comprensión de la formación de nuestro propio planeta (Cleeves et al. 2014).

Tres años después de la misión Herschel, tenemos una nueva visión de cómo el agua llega a formar parte de los discos protoplanetarios y de su estado físico (sólido o líquido). Esta visión necesita ser comparada con nuevas observaciones para su confirmación y para el entendimiento más detallado de la evolución que sufre el

agua desde la nube materna hasta la formación del planeta. Estas observaciones serán posibles a partir del 2017 con la puesta en marcha de una nueva banda de frecuencias en el interferómetro milimétrico ALMA situado en la meseta de Chajnantor (Chile). Este interferómetro permitirá observar las líneas de agua a 183 GHz y 325 GHz en los discos protoplanetarios más cercanos a la Tierra con una resolución similar a la distancia de la Júpiter al Sol (5 ua), proporcionando una información clave para entender la historia de nuestro Sistema Solar.



Mapas de la intensidad integrada y del centroide de velocidad de las líneas del agua a 183 GHz y 325 GHz observadas con ALMA en la galaxia Arp 220. La extraordinaria resolución de ALMA nos permite distinguir la emisión del agua proveniente de los núcleos individuales Arp 220 Este y Arp 220 Oeste. Figura adaptada de König et al. (2017).

En julio de 2016, el interferómetro milimétrico ALMA realizó las primeras observaciones (aún en fase de prueba) de la línea del agua a 183 GHz y 325 GHz. La observación de estas líneas desde la superficie de la Tierra es complicada debido al efecto de pantalla de la atmósfera terrestre, pero las excelentes condiciones de atmosféricas (sequedad y altura) de la meseta de Chajnantor hacen posible esta observación. Uno de los objetos observados fue la galaxia Arp 220. Esta galaxia es una de las galaxias más luminosas del Universo cercano y está formada por la fusión de otras galaxias más pequeñas, Arp 220 Este y Arp 220 Oeste, cuyos núcleos pueden distinguirse fácilmente en la figura adjunta gracias a la extraordinaria resolución angular de ALMA ($<1''$). Las observaciones de agua demostraron que el choque de estas dos galaxias ha producido un brote de formación estelar. Las nuevas regiones de formación estelar son las responsables de la intensa emisión de vapor de agua observada. Observaciones con mayor resolución angular son necesarias para localizar con más precisión la situación y determinar las condiciones físicas en estos brotes.

Referencias

Cleeves, L. I., Bergin, E. A., Alexander, C. M. O. ., et al. 2014, *Science*, 345, 1590
 König, S., Martín, S., Muller, S., et al. 2016, arXiv:1612.07668
 van Dishoeck, E. F., Bergin, E. A., Lis, D. C., & Lunine, J. I. 2014, *Protostars and Planets VI*, 835

Asunción Fuente, febrero 2017

¿CUÁL ES EL DESTINO DE LA VÍA LÁCTEA?

Santiago García Burillo

*Observatorio Astronómico Nacional
 Instituto Geográfico Nacional - Ministerio de Fomento*

Abstract

The recent determination of the three-dimensional motion between Andromeda (M 31) and the Milky Way has allowed astronomers to predict with high accuracy the past history and the future evolution of the encounter between both galaxies. In about 4 billion years, Andromeda is going to start a dramatic collision with the Milky Way that will disrupt the stellar disks of both systems. The most likely scenario contemplates that the two galaxies are bound to merge and eventually form a massive giant elliptical after 10 billion years from now. During this colossal event, the supermassive black holes in our two galaxies will interact, and possibly merge. And if enough material piles up around the resulting beast, maybe our Milky Way could become a quasar.

Introducción

Nuestra Galaxia (también conocida como la *Vía Láctea*, *Milky Way* (MW) en su acepción inglesa) y la galaxia Andrómeda (M 31) son las galaxias más masivas del Grupo Local de galaxias (van den Bergh 2000). Este grupo incluye una cincuentena de objetos, en su mayoría galaxias enanas poco masivas. Como en el caso de la mayoría de los grupos de galaxias, el Grupo Local está muy probablemente desacoplado del movimiento cosmológico de expansión acelerada del universo y puede, por lo tanto, considerarse como una entidad gravitacionalmente ligada, y en cierto modo, *abandonada* a su suerte. En este contexto, conocer cuál será la evolución futura del Grupo Local, reviste un gran interés astronómico.

Situada a una distancia de unos 770 kpc (2,5 millones de años luz) de la Vía Láctea, Andrómeda es la galaxia espiral más próxima a nosotros. La componente radial de la velocidad de M 31 relativa a la Galaxia, medible fácilmente por el efecto Doppler del desplazamiento de sus líneas espectrales, se conoce con precisión desde hace mucho tiempo. Dicha velocidad radial indica claramente un *inquietante* movimiento de aproximación a la Vía Láctea ($V_{\text{rad}} \simeq -120$ km/s, medida respecto al sistema de referencia galacto-céntrico; Binney & Tremaine 1987). Este movimiento hizo sospechar a los astrónomos que una colisión violenta entre Andrómeda y la Galaxia era un escenario muy probable (Dubinski et al. 1996; Peebles et al. 2001; Cox & Loeb 2008). No obstante, la determinación de la componente transversal de la velocidad relativa de Andrómeda, esencial a la hora de caracterizar completamente su órbita en el espacio, sólo ha podido lograrse más recientemente con la precisión requerida. La razón por la que dicha medida ha tardado en ser obtenida radica en la dificultad intrínseca en medir movimientos propios (*transversales*) de objetos celestes, que por definición se encuentran muy lejanos, en escalas de tiempo *humanas* (e.g., del orden de años).

A la distancia de M 31, para medir una velocidad transversal que fuese comparable a la velocidad radial de ~ 100 km/s, se requería alcanzar una precisión en la medida de los movimientos propios del orden de 0,027 milisegundos de arco por año. Dicha precisión fué utilizada recientemente por primera vez por un equipo de investigadores para medir los movimientos propios de un conjunto de estrellas de M 31. Lograr esta hazaña técnica requirió realizar observaciones con el telescopio espacial Hubble durante un periodo de siete años en la pasada década. Los resultados de estas medidas y sus implicaciones fueron desarrollados en detalle en una serie de tres publicaciones científicas de la revista *The Astrophysical Journal*, aparecidas en el año 2012 (Sohn et al., 2012; van der Marel et al., 2012a, 2012b).

En el presente artículo se describen las consecuencias fundamentales que dichas investigaciones tienen a la hora de prever la evolución futura y el destino de nuestra Galaxia en relación a su *tormentosa* interacción con Andrómeda.

Los movimientos propios de Andrómeda

Durante el periodo referido, que va desde finales del año 2002 al 2010, el telescopio espacial Hubble registró los movimientos propios de miles de estrellas en tres zonas de M 31. Los campos observados estaban situados en el halo (cerca del eje menor de la galaxia), en el disco externo (cerca del eje mayor de la galaxia), y en una cola de interacción que se encuentra ligada al sistema. Las figuras 1 y 2 ilustran las observaciones realizadas en el campo del halo de M 31.

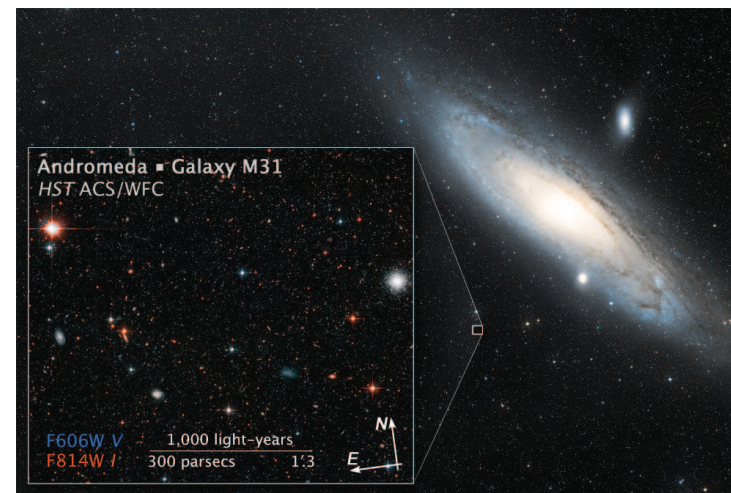


Figura 1: Una imagen óptica de la galaxia Andrómeda (M 31). Los astrónomos seleccionaron tres campos entre los que se encuentra uno situado en el halo de la galaxia, que se muestra en la figura, para que el telescopio espacial Hubble (HST) midiese los movimientos propios de sus estrellas respecto al campo profundo de galaxias, haciendo uso de la alta resolución espacial y sensibilidad del HST. Imagen cortesía de NASA, ESA, and Z. Levay (STScI).

El registro de los movimientos propios se hizo relativo a un sistema de referencia que *necesariamente* debía estar *en reposo* durante el periodo que abarcaban las medidas. Dicho sistema de referencia estuvo definido por un conjunto de centenares de galaxias compactas lejanas del fondo. Estas galaxias pudieron detectarse al mismo tiempo en cada uno de los tres campos seleccionados. Al tratarse de objetos mucho más lejanos que M 31, se puede considerar que las galaxias del fondo se encuentran *en reposo* en todas las imágenes registradas durante el periodo de tiempo en el que se midieron los movimientos propios de las estrellas de Andrómeda (ver figura 3). Aunque en principio se podían haber utilizado fuentes tipo cuásar¹ para definir el sistema de referencia en reposo, el menor número de cuásares detectables en las imágenes, en comparación con el número de galaxias del fondo, desaconsejó el uso de aquéllos en aras de lograr la mayor precisión posible en las medidas.

¹Un cuásar es una fuente astronómica de energía electromagnética, que incluye desde radiofrecuencias hasta luz visible y rayos X. En telescopios ópticos, la mayoría de los cuásares aparecen como simples puntos de luz, aunque algunos parecen ser los centros de galaxias activas. La gran cantidad de energía emitida por estas fuentes se produce durante la alimentación de los agujeros negros supermasivos que se cree se encuentran en los núcleos de estas galaxias.

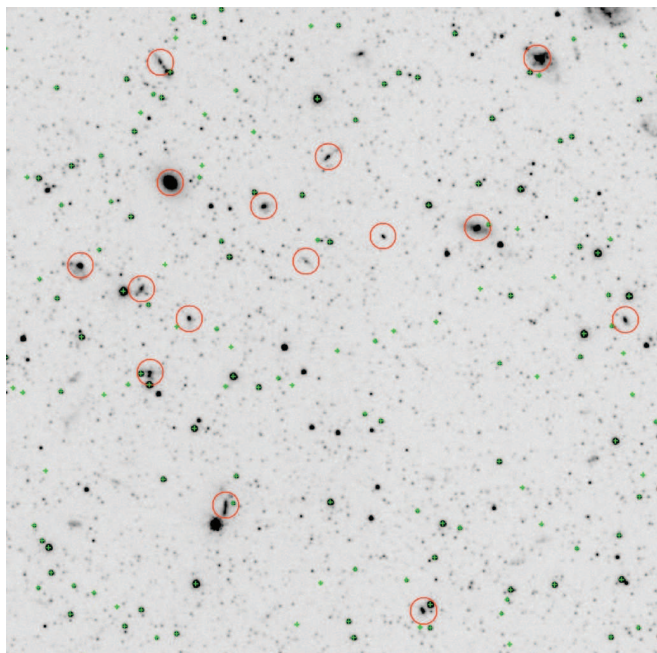


Figura 2: La figura muestra la imagen óptica de una fracción de tamaño $25'' \times 25''$ del campo observado por el telescopio espacial Hubble en el halo de la galaxia Andrómeda (M 31). Las galaxias lejanas, utilizadas como referencia para la medida de los movimientos propios de las estrellas de Andrómeda, se identifican por los círculos rojos. Las estrellas cuyos movimientos propios se registran aparecen identificadas con pequeñas cruces de color verde. Figura adaptada del artículo publicado por Sohn et al. (2012).

Como resultado de esta ambiciosa campaña de observaciones se alcanzó una precisión sin precedentes en el registro de movimientos propios: ~ 12 microsegundos de arco por año. Este record fué posible en gran medida gracias a la resolución espacial y sensibilidad del telescopio Hubble, pero también gracias a que se promediaron medidas de miles de estrellas en los tres campos seleccionados en M 31. Como prueba de la solidez del experimento, los movimientos propios que se dedujeron en los tres campos observados fueron satisfactoriamente consistentes, dentro de los límites de precisión de las medidas.

En la actualidad se está utilizando la emisión procedente de las líneas másas de agua, detectada en un número significativo de complejos en el disco de M 31 (Darling 2011), con objeto de medir los movimientos

propios de dichas regiones emisoras mediante técnicas de interferometría de muy larga base (*Very Long Base Interferometry: VLBI*). Este tipo de observaciones, que permite alcanzar precisiones de ≤ 10 microsegundos de arco por año, se estima que servirán muy pronto (sólo 4-5 años después de las primeras medidas del 2011) para corroborar los resultados obtenidos por el telescopio Hubble y, además, gracias a su mayor precisión, detectar el aumento del tamaño aparente de Andrómeda en su viaje de aproximación hacia nuestra Galaxia.

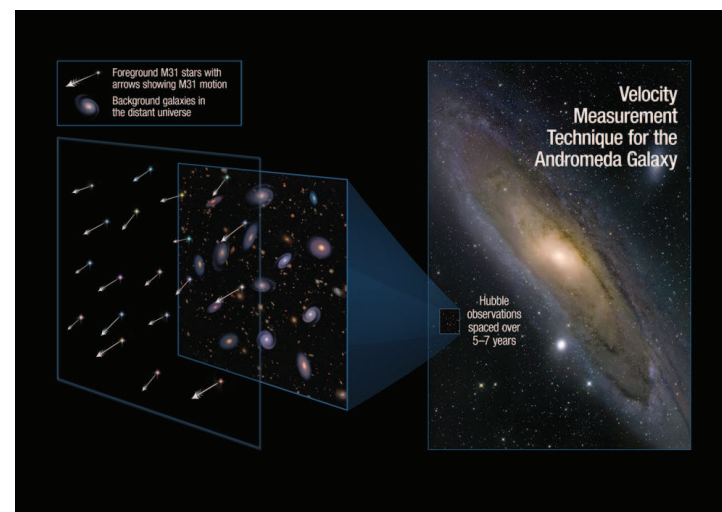


Figura 3: En esta figura se muestra una de las regiones exploradas por los astrónomos para medir los movimientos propios de las estrellas del halo de M 31 con el telescopio Hubble. Gracias a la precisión con que se registraron los movimientos transversales de las estrellas respecto a las galaxias más distantes, que permanecen fijas, durante un periodo que abarcó siete años, se pudo determinar la componente lateral del movimiento relativo global de Andrómeda respecto a la Vía Láctea. Figura cortesía de NASA, ESA, A. Feild y R. van der Marel (STScI).

Este conjunto de medidas ha permitido a los astrónomos determinar con gran fiabilidad cuál fué la historia pasada y, también, cuál será la evolución futura más probable de la interacción entre nuestra Galaxia y M 31.

La medida de la velocidad de Andrómeda

Los movimientos propios obtenidos en los tres campos de M 31 tuvieron que ser corregidos de dos efectos fundamentales con objeto de determinar

cuál es la componente transversal del movimiento propio *global* del centro de masas² de Andrómeda relativa al centro de masas de nuestra Galaxia:

- En primer lugar, hubo que corregir del movimiento de rotación interna de M 31. Gracias a la elección de los tres campos observados descritos arriba, situados a propósito en diferentes regiones de la galaxia y, por lo tanto, afectados en diferente grado por los movimientos de rotación, fué posible corregir este efecto sin añadir una excesiva incertidumbre. De hecho, se estimó que la contribución de la rotación interna era menor que la propia incertidumbre asociada a la medida de los movimientos propios efectuada por el Hubble.
- En segundo lugar, los movimientos propios en los tres campos de M 31 fueron corregidos de la distinta proyección que tiene el movimiento del sistema solar, donde se sitúa en la práctica el observador, en su rotación alrededor del centro de la Galaxia. Esta corrección se benefició del mejor conocimiento que se tenía ya en el año 2012 de la velocidad de rotación del Sol: $V_{\text{rot}} \sim 239 \pm 10$ km/s a una distancia radial $R_{\text{sol}} \sim 8,33 \pm 0,35$ kpc del centro.

Una vez efectuadas estas correcciones, se dedujo que la componente radial de la velocidad de M 31 respecto a nuestra Galaxia es $V_{\text{rad}} \sim -109$ km/s, mientras que la componente transversal era $V_{\text{tran}} \sim 17$ km/s. En otras palabras, se concluyó que la velocidad transversal de M 31 era mucho menor en valor absoluto que su velocidad radial: $|V_{\text{tran}}| \ll |V_{\text{rad}}|$. Estos valores de la velocidad indican que la órbita de M 31 respecto a nuestra Galaxia es *cuasi radial*. En la práctica, esto implica que la colisión violenta entre las dos galaxias en un futuro próximo es inevitable (ver figura 4).

Finalmente, asumiendo que las órbitas relativas son cuasi keplerianas, y basándose en la medida de las velocidades, se pudo estimar la masa total del sistema interactuante formado por M 31 y nuestra Galaxia ($M_{\text{tot}} \sim 3,2 \times 10^{12} M_{\odot}$, con una contribución similar de ambas galaxias), así como la masa de la principal galaxia satélite de Andrómeda, M 33 ($M_{\text{M33}} \sim 0,1 \times M_{\text{M31}}$).³

Como resultado final de este análisis, se pudieron estimar con la precisión requerida las posiciones iniciales, las velocidades y las masas del sistema formado por nuestra Galaxia, M 31 y M 33. Estos fueron los ingredientes esenciales que los astrónomos utilizaron para estudiar la evolución futura de la interacción entre estas tres galaxias, mediante simulaciones numéricas del encuentro.

²El centro de masas de un sistema físico es el lugar geométrico que se comporta dinámicamente como si se aplicara en él la resultante de las fuerzas externas sobre la masa total del sistema.

³M 33 se encuentra a una distancia de nuestra Galaxia similar a la de M 31, y forma junto a ésta un grupo gravitacionalmente ligado. Para M 33 ya se disponía de una medida fiable de sus movimientos propios desde el año 2005. Estos fueron estimados a través de las líneas de emisión máser del agua (Brunthaler et al. 2005).

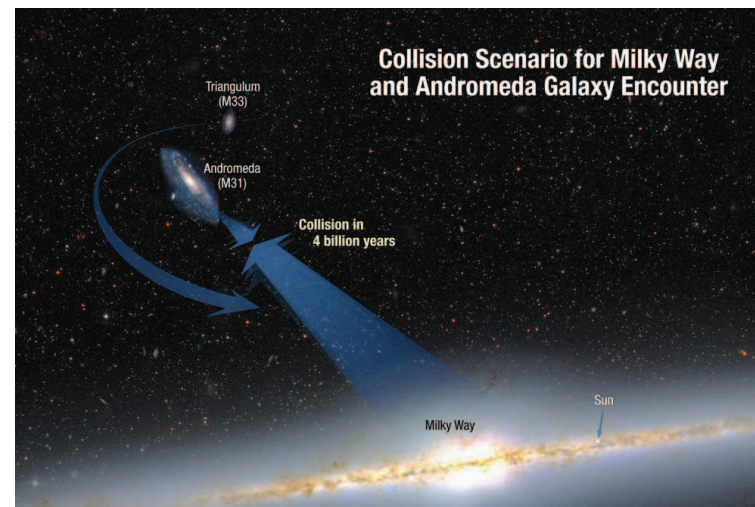


Figura 4: Esquema que muestra las tres galaxias (MW, M 31 y M 33) que se verán involucradas en un proceso de colisión en los próximos 4 gigaaños. La determinación precisa del movimiento orbital de M 31 respecto a nuestra Galaxia, ha permitido concluir que este acontecimiento ocurrirá inexorablemente. Figura cortesía de NASA, ESA, A. Feild y R. van der Marel (STScI)

La colisión entre la Vía Láctea, M 31 y M 33

Hasta que no se dispuso de una medida precisa de la componente transversal de M 31, como la que se obtuvo en el año 2012, las diferentes simulaciones numéricas publicadas hasta esa fecha, que trataban de predecir el futuro de la interacción entre Andrómeda y la Vía Láctea, eran incapaces de concluir si el encuentro podía o no resultar en la fusión completa de ambos sistemas (Dubinski et al. 1996; Peebles et al. 2001; Cox & Loeb 2008). Una fusión de dos galaxias espirales masivas, como lo son la Vía Láctea y M 31, tendría como consecuencia más probable la formación de una galaxia elíptica en escalas de tiempo de algunos gigaaños⁴.

El trabajo realizado por van der Marel et al. (2012b) permitió estudiar la evolución futura del sistema constituido por nuestra Galaxia, Andrómeda y M 33. Para ello se realizaron simulaciones numéricas que estudiaban la evolución de la componente estelar de los tres sistemas y, al mismo tiempo, se integraron, mediante una aproximación semianalítica, las órbitas de las tres galaxias durante la interacción.

⁴Un gigaaño (Ga) corresponde a mil millones de años.

En las simulaciones no se incluyó el sistema formado por la Pequeña y la Gran Nube de Magallanes debido a su baja masa ($\sim 0,1 \times M_{M31}$) y, sobre todo, debido al hecho de que su órbita indica que actualmente las dos galaxias se están alejando de la Vía Láctea. En particular, se estima que dichas galaxias no volverán a encontrarse cerca de la Vía Láctea hasta dentro de muchos miles de millones de años, lo que convierte en irrelevante su influencia en la dinámica del sistema (Besla et al. 2007; Shattow & Loeb 2009). Tampoco se tuvo en cuenta en estas simulaciones el papel de la componente gaseosa, que constituye sólo un 10 % de la masa de las galaxias interactuantes. De hecho es improbable que la inclusión del gas hubiera cambiado esencialmente la evolución futura de la interacción predicha por las simulaciones. Más bien se espera que el gas sufra *pasivamente* las consecuencias de la fusión.

La figura 5 muestra una serie de instantáneas correspondientes a la evolución temporal de la colisión entre las tres galaxias en la simulación numérica de van der Marel et al. (2015b). En dicha figura se aprecian tanto las órbitas relativas de los tres sistemas como los cambios que experimentan sus discos estelares, tal y como se contemplaría el evento desde un sistema de referencia centrado en el plano (X,Y) de la Vía Láctea. El instante inicial de las simulaciones u origen de tiempos $T = 0$ en el artículo de van der Marel et al. (2015b) corresponde a la configuración actual del sistema formado por la Galaxia (MW), M 31 y M 33. Para la realización de la simulación se tomó como escenario base para la configuración inicial el que se consideró como más probable, dentro de las incertidumbres de los parámetros clave, osea, las posiciones iniciales, las velocidades y las masas.

La secuencia mostrada en la figura 5 empieza en el instante $T = 3$ gigaaños. Hasta ese momento nuestra Galaxia se puede considerar como aislada y aún no ha sufrido los efectos mas visibles de la interacción. A medida que progresa el encuentro en instantes posteriores (i.e., $T > 3$ gigaaños), las galaxias M 31 y M 33 aparecen en escena. Los efectos de la interacción ya se hacen ostensibles tras $T = 3,97$ gigaaños, momento en el que los discos de M 31 y la Vía Láctea se superponen físicamente por primera vez en el espacio. Esta configuración corresponde al primer *pericentro* o primer evento de máxima aproximación en su órbita, como se ilustra en la figura 6. En ese instante la velocidad relativa de aproximación de M 31 es muy grande, del orden de 600 km/s, es decir casi seis veces el valor de la velocidad radial actual. Este hecho contribuye a que, a pesar de su cercanía, M 31 no produzca demasiados cambios en la morfología del disco de nuestra Galaxia. Sin embargo, sólo unos 500 millones de años después ($T = 4,47$ gigaaños), M 31 se frena y alcanza así su primer *apocentro* o instante de máxima lejanía en su órbita (figura 6). En ese estadio de la interacción, ya aparecen distorsiones en forma de colas de marea en los discos estelares tanto de nuestra Galaxia como de Andrómeda.

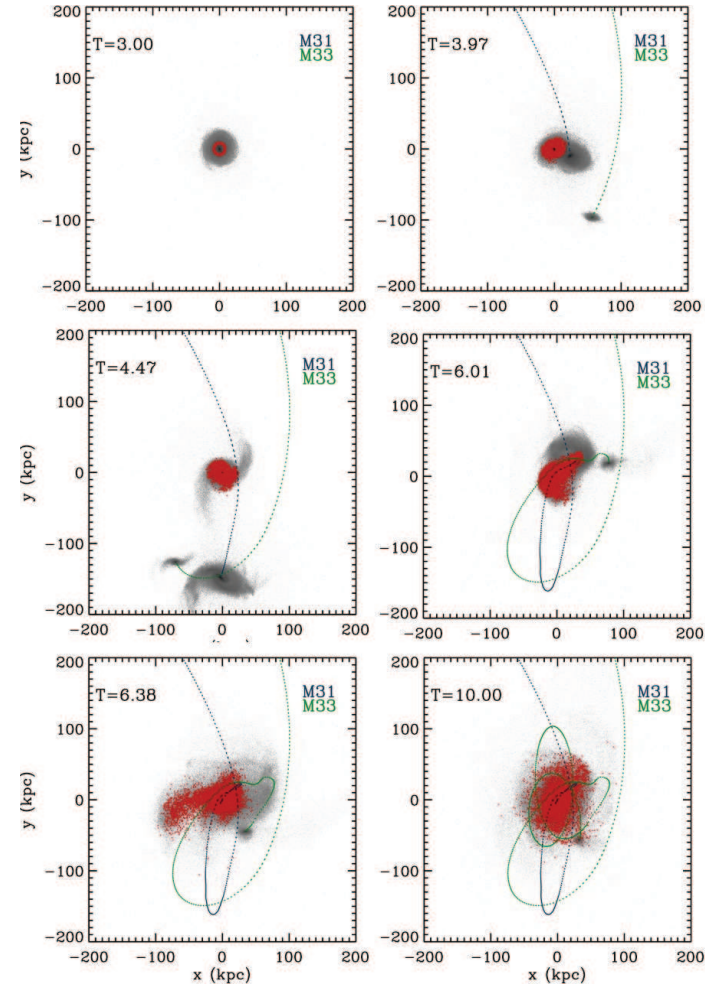


Figura 5: Instantáneas de la evolución temporal de la colisión entre la Vía Láctea (MW), Andrómeda (M 31) y M 33 vista desde la Vía Láctea. Las órbitas precedentes de M 31 (en azul) y M 33 (en verde) en cada instante ayudan a visualizar cuál es la evolución del sistema interactuante. En el panel inicial ($T = 3,00$ gigaaños) la Vía Láctea está aún aislada. En instantes posteriores ($T > 3$ gigaaños), el proceso de interacción progresa hasta cristalizar en una fusión tras un tiempo estimado de $T \sim 10$ gigaaños. Las estrellas identificadas en color rojo estarían situadas a una distancia de 8,3 kpc del centro de la Galaxia. Figura adaptada del artículo de Van der Marel et al. (2012b).

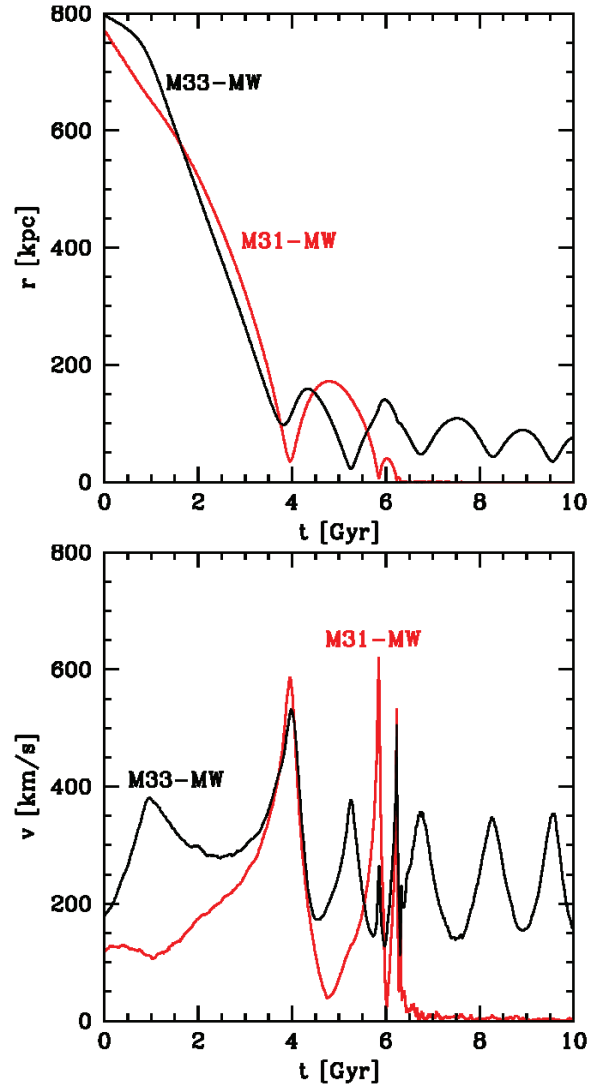


Figura 6: Las distancias (en kpc) (panel superior) y velocidades relativas (en km/s) (panel inferior) en función del tiempo (en gigaaños) entre la galaxias M 31, la Vía Láctea (MW) y M 33, tal como predice la simulación numérica de la colisión entre dichos sistemas. El origen de tiempos ($T = 0$) corresponde a la configuración actual. Figura adaptada del artículo publicado por Van der Marel et al. (2012b).

Como consecuencia de la interacción, y debido a un efecto conocido como *fricción dinámica*, la órbita de M 31 pierde progresivamente cada vez más momento angular relativo. Este hecho, común en cualquier proceso de fusión entre dos galaxias, se traduce en que los pericentros sucesivos serán cada vez más pequeños, es decir, los núcleos de ambas galaxias cada vez se acercarán más. Por la misma razón, las velocidades relativas de las dos galaxias se harán también cada vez más pequeñas (ver figura 6).

El proceso de interacción progresa hasta cristalizar en una fusión completa de M 31 con la Vía Láctea tras un tiempo estimado de $T \sim 10$ gigaaños. El resultado final más probable es la formación de una galaxia elíptica gigante que incluiría tanto a estrellas de nuestra Galaxia como de Andrómeda. Es común en la literatura referirse al sistema fusionado como *Lactómeda* (o *Milkdrómeda* en su acepción inglesa). Contrariamente a M 31, lo más probable es que M 33 siga orbitando como una galaxia independiente alrededor de *Lactómeda* durante mucho tiempo después de que la fusión se haya completado. No obstante, se espera que M 33 acabe siendo engullida por *Lactómeda* a muy largo plazo, dado que su órbita ligada está destinada a decaer con el tiempo.

Lactómeda tendrá aproximadamente el doble de masa de la Vía Láctea y será mucho más extensa que ésta, como puede fácilmente apreciarse en la figura 5. De hecho el perfil radial de la densidad superficial de estrellas en la *Lactómeda* simulada sigue una ley de potencias con el radio ($\propto R^{1/4}$). Esta ley es muy similar a la observada en otras galaxias elípticas del universo.

La figura 7 muestra cuál es la distribución final de las estrellas en el sistema resultante de la fusión entre M 31 y la Vía Láctea. Aunque las estrellas de nuestra Galaxia y de M 31 aparecen mezcladas alrededor del centro de masas común de *Lactómeda*, la galaxia fusionada no parece haber alcanzado aún tras $T \sim 10$ gigaaños el estadio final de *relajación dinámica*, en el que la distribución de las estrellas procedentes de ambos sistemas sean perfectamente indistinguibles. Por el contrario, la mayor parte de las estrellas de M 33 permanecen fuera de *Lactómeda* al final de la simulación.

Uno de los escenarios más comúnmente aceptados prevé que la fusión de dos galaxias espirales produzca brotes de formación estelar (*starbursts* en su acepción inglesa). Durante la interacción, se espera que grandes cantidades de gas pierdan momento angular y se precipiten hacia las regiones centrales del sistema, donde pueden experimentar choques y un elevado grado de compresión. El resultado probable es que se generen episodios masivos de formación de nuevas estrellas que consuman gran parte del gas. Durante esta fase, previa a la formación de una galaxia elíptica, el sistema se identificaría por su intensa emisión en el rango de longitudes del infrarrojo (de ahí su denominación en inglés *Ultra-Luminous Infrared Galaxy*, ULIRG). La evolución que probablemente experimente la componente gaseosa de nuestra Galaxia y Andrómeda durante la fusión fué

estudiada por Cox & Loeb (2008). Sus conclusiones sugieren que debido a que la componente gaseosa de ambos sistemas no es muy elevada, no se espera que la fusión de ambas galaxias dé lugar a un brote de formación estelar masivo. Antes al contrario, se espera que el proceso de formación de estrellas continúe a un ritmo sólo ligeramente superior al esperado en caso de que ambas galaxias hubieran estado totalmente aisladas.

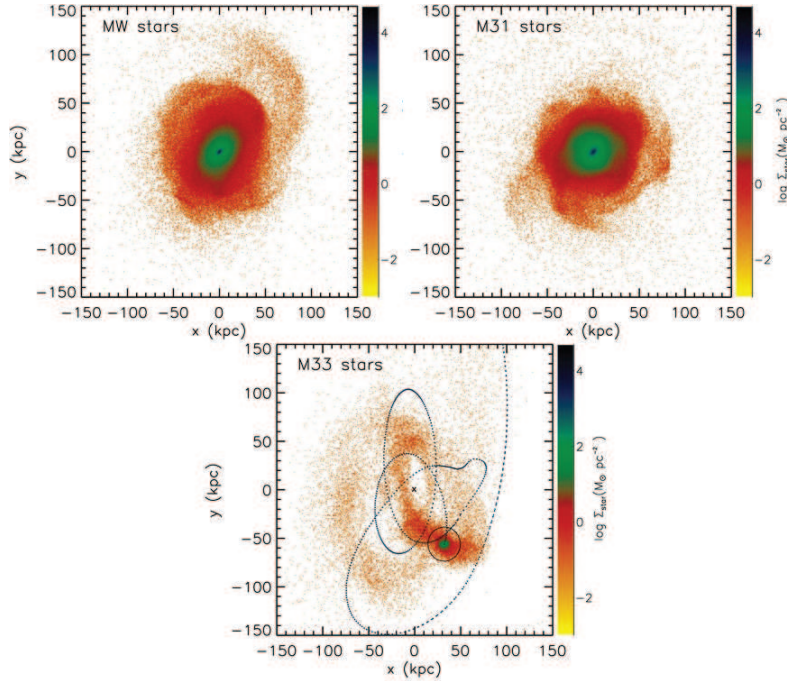


Figura 7: Distribución final de las estrellas en el sistema resultante de la fusión entre M 31 y la Vía Láctea (MW). Los tres paneles identifican cuáles son las estrellas procedentes de la Vía Láctea (MW), de M 31 y de M 33, respectivamente. Puede apreciarse que en esta simulación M 33 mantiene su propia identidad y no acaba fusionándose completamente con la Vía Láctea (MW) y M 31. Figura adaptada del artículo publicado por Van der Marel et al. (2012b).

Paralelamente al proceso de fusión de sus discos estelares, se espera que también se produzca la fusión de los agujeros negros supermasivos que se encuentran en los núcleos de nuestra Galaxia y Andrómeda. Dicho proceso producirá el nacimiento de un único agujero negro supermasivo en el núcleo de *Lactómeda*. Es más que probable que tras la fusión quede aún

suficiente gas disponible para alimentar el agujero negro central y, como indican las predicciones de los modelos teóricos (Hopkins et al. 2006), *Lactómeda* pueda convertirse en un cuásar.

¿Cuál será el destino del Sol tras la colisión?

Aunque dos galaxias colisionen violentamente y puedan acabar fusionándose en un único sistema, como es el caso de nuestra Galaxia y M 31, la probabilidad de que colisionen entre sí dos estrellas es muy pequeña. Ello se debe a la gran distancia que típicamente separa a dos estrellas, incluso en los entornos más densamente poblados de una galaxia: $\geq 1 - 2 \times 10^{11}$ km. Esa distancia promedio es siempre mucho mayor que el tamaño característico de cualquier estrella. La razón entre distancia y tamaño medios para una estrella en una galaxia típica es en cualquier circunstancia mayor que un factor $\simeq 10^5$. A la vista de estas estimaciones, es evidente que el nacimiento de *Lactómeda* no alterará en esencia la evolución interna del Sistema Solar. No obstante, sí se espera que la órbita del Sol y, por lo tanto, su posición dentro del sistema interactuante, se vean dramáticamente modificadas.

Las simulaciones de van der Marel et al. (2015b) no pudieron seguir en detalle cómo cambiaría en particular la órbita del Sol a lo largo de la interacción entre nuestra Galaxia y M 31. Sin embargo sí que permitieron predecir estadísticamente cuál será la evolución más probable de una estrella que, como el Sol, se encuentra en la actualidad a una distancia radial $R_{\text{sol}} \sim 8,33 \pm 0,35$ kpc del centro de la Galaxia y gira en torno a éste a una velocidad de rotación $V_{\text{rot}} \sim 239 \pm 10$ km/s. Dentro de esta categoría se encontraban más de 8000 estrellas en la Galaxia simulada por van der Marel et al. (2015b). Dichas órbitas, representativas de la que sigue el Sol en la práctica, fueron identificadas y monitorizadas en el transcurso de la fusión. La conclusión fundamental tras este seguimiento es que la mayor parte de las estrellas similares al Sol emigraban hacia el exterior de *Lactómeda*, es decir que su distancia radial *promedio* aumentaba hacia el final de la simulación.

Quando se habla de radios *promedio* en el caso de las órbitas típicas descritas por las estrellas de una galaxia elíptica como *Lactómeda*, es importante señalar que el radio a lo largo de dichas órbitas cambia significativamente con el tiempo. Dichos cambios se desarrollan de acuerdo con un patrón mucho menos ordenado que el que se observa en las órbitas cuasi circulares que describen las estrellas en las galaxias tipo disco, cuyo equilibrio se establece por rotación, en claro contraste con las galaxias elípticas, donde la componente asociada con movimientos caóticos domina. Probablemente el Sol, que en la actualidad gira en una órbita cuasi circular en el plano de la Vía Láctea, acabará atrapado en una de esas órbitas estelares abiertas (*loop orbits* en su acepción inglesa), como la que se muestra en la

figura 8.

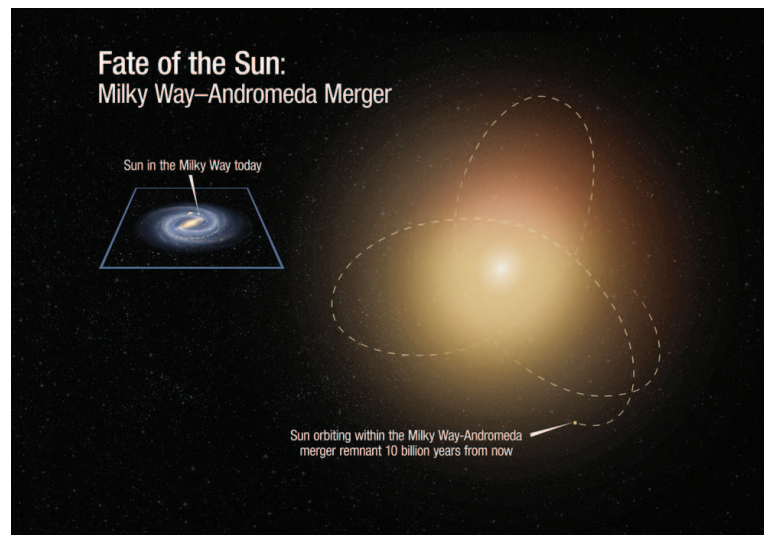


Figura 8: El destino de la Vía Láctea es convertirse, tras su fusión con Andrómeda, en una galaxia elíptica gigante una vez transcurrido un plazo de 10 gigaaños. Dentro de este sistema, el Sol acabará atrapado en una órbita estelar abierta (*loop orbit*) muy similar a la que típicamente describen las estrellas de las galaxias elípticas. Figura cortesía de NASA, ESA, A. Feild y R. van der Marel (STScI).

El resultado de estas simulaciones permite concluir, con un elevado grado de certeza, que tanto la distancia, como la velocidad y los parámetros que caracterizan el entorno inmediato del Sol (por ejemplo, la densidad estelar promedio) se modificarán debido a la interacción violenta entre nuestra Galaxia, Andrómeda y M 33. Sin embargo, eso no significa que la vida en la Tierra se vea afectada por la evolución dinámica del Grupo Local en los términos que acabamos de describir. Un evento dramático, producido por la perturbación que podría sufrir la órbita del Sol por su interacción con otra estrella que se acercara peligrosamente a nosotros, no puede formalmente excluirse. En concreto, una interacción próxima (que no colisión directa) con otra estrella podría cambiar la órbita de la Tierra al desplazar a ésta fuera de la zona de *habitabilidad*, aquella que hace posible la vida y en la que actualmente nos encontramos. Sin embargo, la probabilidad de este tipo de encuentros es, como ya hemos señalado antes, despreciable.

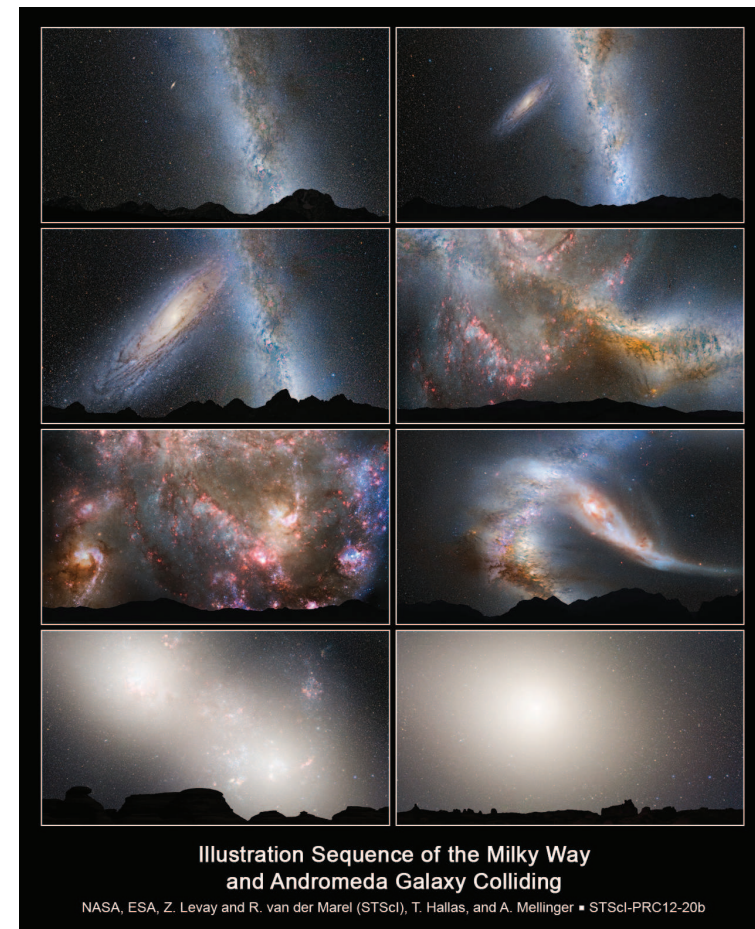


Figura 9: Imágenes ilustrativas de cómo un observador vería desde la Tierra la fusión predicha por las simulaciones numéricas entre la Vía Láctea y Andrómeda. La secuencia mostrada recorre de izquierda a derecha y de arriba a abajo el aspecto del cielo desde el instante presente, donde Andrómeda aparece al principio como una mancha difusa lejos de nuestra Galaxia, hasta la fusión final de ambos sistemas (10 gigaaños después), que da lugar a una galaxia elíptica gigante. Durante su aproximación a la Vía Láctea, Andrómeda ocupará una porción creciente del cielo, como se muestra en la secuencia de paneles, y se harán progresivamente más visibles los efectos de la interacción con la Vía Láctea. Imagen cortesía de NASA, ESA, Z. Levay, R. van der Marel (STScI), T. Hallas, y A. Mellinger.

En la práctica, nuestro porvenir en la Tierra sí se verá alterado por otro tipo de fenómenos astronómicos que nada tienen que ver con el nacimiento de *Lactómeda*. Aunque es posible imaginar cómo cambiaría durante los próximos diez mil millones de años el aspecto del cielo observable desde la Tierra debido a la colisión entre nuestra Galaxia, Andrómeda y M 33 (ver figura 9), antes de que la fusión llegue a término, el combustible necesario para que el Sol siga brillando se habrá consumido (dentro de 4,6 gigaaños; Sackmann et al 1993). Además, antes de que esto ocurra, el clima de la Tierra habrá cambiado tanto que hará imposible la vida, debido al progresivo aumento del tamaño y luminosidad del Sol, siendo ambos procesos ligados a su inexorable discurrir hacia la muerte estelar.

Siempre nos quedará el recurso de soñar que para entonces, si la civilización humana no se ha autodestruido todavía, ya sea posible la colonización de otros planetas y, por lo tanto, que ese espectáculo cósmico que refleja la figura 9 será observado por algún terrícola que haya podido escapar y sustraerse así a su *terrible* y no por ello menos cierto destino.

Referencias

- Besla, G., Kallivayalil, N., Hernquist, L., et al. 2007, “Are the Magellanic Clouds on their first passage about the Milky Way?”, *The Astrophysical Journal*, núm. 668, págs. 949-967.
- Binney, J., & Tremaine, S. 1987, “Galactic dynamics”, *Princeton, NJ, Princeton University Press*, págs. 747.
- Brunthaler, A., Reid, M. J., Falcke, H., Greenhill, L. J., & Henkel, C. 2005, “The geometric distance and proper motion of the Triangulum Galaxy (M33)”, *Science*, núm. 307, págs. 1440-1443.
- Cox, T. J. & Loeb, A., 2008, “The collision between the Milky Way and Andromeda”, *Monthly Notices of the Astronomical Society*, núm. 386, págs. 461-474.
- Dubinski, J., Mihos, J. C., & Hernquist, L. 1996, “Using tidal tails to probe dark matter halos”, *The Astrophysical Journal*, núm. 462, págs. 576-593.
- Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006, “A Unified, Merger-driven Model of the Origin of Starbursts, Quasars, the Cosmic X-Ray Background, Supermassive Black Holes, and Galaxy Spheroids”, *The Astrophysical Journal Supplement Series*, núm. 163, págs. 1-49.
- Peebles, P. J. E., Phelps, S. D., Shaya, E. J., & Tully, R. B. 2001, “Radial and transverse velocities of nearby galaxies”, *The Astrophysical Journal*, núm. 554, págs. 104-113.

- Sackmann, I.-J., Boothroyd, A. I., & Kraemer, K. E. 1993, “Our Sun. III. Present and Future”, *The Astrophysical Journal*, núm. 418, págs. 457-468.
- Shattow, G., & Loeb, A. 2009, “Implications of recent measurements of the Milky Way rotation for the orbit of the Large Magellanic Cloud”, *Monthly Notices of the Astronomical Society*, núm. 392, págs. L21-L25.
- Sohn, S. T., Anderson, J., & van der Marel, R. P. 2012, “The M31 velocity vector. I. Hubble Space Telescope proper-motion measurements”, *The Astrophysical Journal*, núm. 753, 7, págs. 15.
- van den Bergh, S. 2000, “The galaxies of the Local Group”, *Cambridge University Press, 2000 Cambridge Astrophysics Series*, núm. 35, págs. 346.
- van der Marel, R. P., Fardal, M., Besla, G., et al. 2012a, “The M31 velocity vector. II. Radial orbit toward the Milky Way and implied Local Group mass”, *The Astrophysical Journal*, núm. 753, 8, págs. 14.
- van der Marel, R. P., Besla, G., Cox, T. J., Sohn, S. T., & Anderson, J. 2012b, “The M31 velocity vector. III. Future Milky Way M31-M33 orbital evolution, merging, and fate of the Sun”, *The Astrophysical Journal*, núm. 753, 9, págs. 21.

EL MEDIO MOLECULAR EN GALAXIAS CON BROTES DE FORMACIÓN ESTELAR Y EN GALAXIAS ACTIVAS

Antonio Usero

*Observatorio Astronómico Nacional
Instituto Geográfico Nacional - Ministerio de Fomento*

Abstract

Massive star formation and nuclear activity in galaxies result from the interplay between the energy sources (stars and active galactic nuclei) and their gas reservoirs (the interstellar medium) through feeding and feedback mechanisms. These processes trigger energy and mass flows that significantly impact on galaxy evolution. We review some current attempts to characterize the interstellar medium in starburst and active galaxies by means of multispecies molecular line observations. The potential of molecular gas chemistry as a diagnostic tool for star formation and nuclear activity is emphasized.

Introducción

De acuerdo con los modelos cosmológicos de mayor aceptación la evolución del Universo está gobernada por la fuerza gravitatoria de una energía ubicua y desconocida (*energía oscura*) y la de una materia sin identificar que no emite radiación electromagnética (*materia oscura*). La materia y la radiación electromagnética como las que componen nuestro mundo más inmediato, al sumar apenas un $\sim 4\%$ de la energía total del Universo, quedan a merced de los dos componentes mayoritarios. Como resultado de esto, buena parte de la materia y de la energía ordinarias se acumula en los núcleos de grandes masas de materia oscura en forma de galaxias. Estas

son grandes concentraciones de estrellas inmersas en un medio interestelar de gas y de polvo. Gracias a la radiación electromagnética que recibimos de ellas, las galaxias actúan como faros cuyas características (tamaño, luminosidad, población estelar, etc.) y distribución en el Universo tratan de reproducir los modelos cosmológicos.

Durante los últimos años se han acumulado indicios de que, en su evolución, las galaxias no han sido meros elementos pasivos, sino que han regulado en parte su propio crecimiento. En concreto, no es fácil explicar de otro modo por qué la proporción de galaxias de gran y de pequeño tamaño que se observa es menor de lo que se preveía; tampoco por qué en las galaxias menores han seguido naciendo estrellas durante más tiempo que en las mayores pese a que estas se formaron con posterioridad a aquellas. Los mecanismos más prometedores para explicar la autorregulación son dos de los fenómenos más violentos que pueden acontecer en una galaxia: los brotes de formación estelar masiva y los núcleos activos.

Tanto las galaxias *con brotes de formación estelar masiva* (galaxias *BFE* de aquí en adelante) como las que albergan *núcleos activos* (galaxias *activas* en adelante) se cuentan entre los objetos más espectaculares del Universo. En las primeras se alcanzan tasas de formación estelar hasta mil veces mayores que en las galaxias que forman estrellas a un ritmo *normal*; en las segundas, la energía radiada desde un volumen menor que un parsec¹ cúbico llega a ser más de mil veces superior a la emitida por nuestra galaxia, la Vía Láctea. Los grandes flujos de materia y energía que tienen lugar en estos objetos no sólo afectan a las galaxias donde se originan, sino que pueden alterar el medio intergaláctico cercano.

La marcha de los brotes de formación estelar y de los episodios de actividad nuclear está regulada por la interacción entre las fuentes de energía (estrellas y núcleos activos) y el medio interestelar mediante *mecanismos de alimentación y de realimentación*: por aquellos, la materia interestelar fluye y se acumula en las regiones donde las estrellas y los núcleos activos producen energía a su costa; por estos, las estrellas y los núcleos activos depositan energía en su entorno mediante campos de radiación y flujos de materia. Los efectos de la realimentación pudieron ser de relevancia cósmica en el Universo pasado, cuando estos fenómenos eran más violentos y frecuentes.

A pesar de la importancia que se atribuye a los efectos de la formación estelar masiva y de los núcleos activos escasean las medidas precisas de hasta qué punto influyen estos en la evolución galáctica. Es muy razonable esperar que parte de la respuesta se halle en la evolución del medio molecular. El *medio molecular* es la fase del medio interestelar que

¹El parsec (pc) es una unidad de longitud empleada en Astronomía que equivale a 3,26 años luz.

predomina en los kiloparsecs centrales de las galaxias, donde surgen los brotes de formación estelar masiva y se forman los núcleos activos. De hecho, el gas molecular es la materia prima de ambos, pues las estrellas nacen en nubes moleculares densas, mientras que los núcleos activos se alimentan de los discos circunucleares de gas molecular. Por ello, es de interés evaluar cómo la radiación de alta energía y los flujos de materia que se producen en los episodios de formación estelar y en los de actividad nuclear alteran el medio molecular. Este artículo pretende ofrecer una perspectiva general de cómo el estudio de la composición del medio molecular puede ayudarnos a comprender unos fenómenos que, en última instancia, han influido en la evolución de las galaxias.

Galaxias con brotes de formación estelar masiva (BFE) y galaxias con núcleos activos (activas)

Galaxias BFE

La fuente principal de la radiación de la mayoría de las galaxias² son sus estrellas. Estas se crean a partir de densas acumulaciones de gas molecular interestelar, de cuya energía gravitatoria obtienen energía electromagnética mediante reacciones termonucleares. En galaxias *estándar*, como la Vía Láctea, la formación de nuevas estrellas es tan pausada que, de proceder a un ritmo invariable, requeriría un lapso mayor que la edad actual del Universo (unos 13 700 millones de años) para agotar todo el gas interestelar. Por contra, en las galaxias BFE las tasas de formación estelar son tan altas que todo el medio gaseoso se consumiría en unos 10 ó 100 millones de años, apenas un instante en escalas cósmicas.

Actualmente se cree que las galaxias BFE no son intrínsecamente distintas del resto, sino que se encuentran en una etapa particular de su evolución a consecuencia del nacimiento súbito de un gran número de estrellas. Durante un lapso menor de cien millones de años tras el comienzo del brote de formación estelar las estrellas de mayor masa³ ($\gtrsim 8 M_{\odot}$), cuya vida es mucho más breve que la de sus compañeras más ligeras, son más abundantes de lo habitual. Las consecuencias para el medio interestelar circundante son importantes a causa de la celeridad con la que las estrellas masivas expelen grandes cantidades de energía. Inicialmente el medio recibe grandes dosis de radiación estelar particularmente intensa en el rango del ultravioleta. Más adelante el medio se ve expuesto a los vientos estelares con los que las estrellas masivas se deshacen de parte de su masa y, finalmente, a las explosiones supernova en las que estas se destruyen.

²Ignorando por un momento el procesamiento de esa radiación por el gas y el polvo interestelares.

³Los símbolos $M_{\odot}$ y $L_{\odot}$ corresponden a la masa y la luminosidad del Sol respectivamente.

Un ejemplo de galaxia BFE es M 82, de la cual se muestra una imagen en la figura 1.

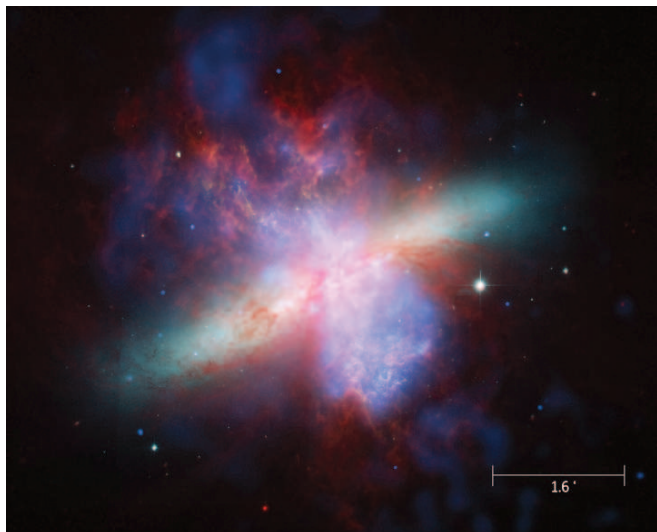


Figura 1: Mosaico multicolor de la emisión de la galaxia BFE M 82 compuesto por imágenes obtenidas con los telescopios espaciales Spitzer (continuo en el infrarrojo medio, en rojo), Hubble (continuo en la banda B del visible, en verde, y línea $H\alpha$, en naranja) y Chandra (continuo de rayos X, en azul). La figura está rotada respecto al plano del cielo para minimizar su tamaño. El mosaico muestra cómo el gas caliente del viento galáctico (rojo-azul) creado por la concurrencia de explosiones supernova se expande perpendicularmente al disco estelar de la galaxia (verde). Cortesía de NASA/CXC/JHU/D.Strickland (rayos X), NASA/ESA/STScI/AURA/Hubble Heritage Team (visible y $H\alpha$) y NASA/JPL-Caltech/Univ. of AZ/C.Engelbracht (infrarrojo).

Las galaxias BFE generan el $\sim 10\%$ de la radiación y forman el $\sim 25\%$ de las estrellas masivas del Universo en el presente. Los brotes de formación estelar masiva constituyeron la principal fuente de energía del $\sim 70-80\%$ de las galaxias *Luminosas en el Infrarrojo* (LIR), *Ultra-Luminosas en el Infrarrojo* (ULIR) e *Hiper-Luminosas en el Infrarrojo* (HLIR), que fueron relativamente abundantes en el pasado. La intensa luminosidad en el infrarrojo, L_{IR} , de estos objetos ($L_{IR} \gtrsim 10^{11} L_{\odot}$, $10^{12} L_{\odot}$ y $10^{13} L_{\odot}$ respectivamente) se debe a la emisión del polvo interestelar calentado por la radiación estelar. Es muy probable que el mismo mecanismo explique la intensa emisión ($\gtrsim 10^{13} L_{\odot}$) en el submilimétrico ($0,3 < \lambda < 1$ mm)

de las galaxias *submilimétricas*, las cuales abundaron cuando la edad del Universo era de dos o tres mil millones de años. Todos estos objetos se observan hoy día tal como eran en el pasado debido a que, al estar muy alejados de nosotros, su luz ha invertido miles de millones de años en viajar hasta la Tierra.

Galaxias activas

Una galaxia alberga un núcleo activo cuando en su región central se da una intensa producción de energía que no puede atribuirse a procesos de formación estelar. Un ejemplo de galaxia activa es NGC 1068, la cual está representada en la figura 2.

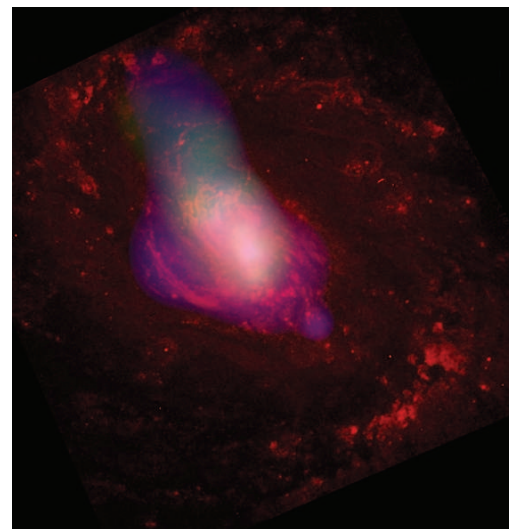


Figura 2: Mosaico multicolor de la emisión de la galaxia BFE+activa NGC 1068 compuesto por imágenes obtenidas con los telescopios espaciales Hubble (línea $H\alpha$, en rojo) y Chandra (continuo de rayos X en los rangos 0,4-1,3 keV y 1,3-8,0 keV, en azul y verde respectivamente). Los ejes horizontal y vertical corresponden a los de Ascensión Recta y Declinación, respectivamente. La figura cubre una región de 30 segundos de arco de lado. El mosaico muestra el gas ionizado del disco galáctico (rojo) y el cono de radiación que surge del núcleo activo de la galaxia (azul-verde). El gas ionizado que bordea la imagen está situado a lo largo de un anillo de intensa formación estelar. Cortesía de NASA/CXC/MIT/UCSB/P.Ogle et al. (rayos X) y NASA/STScI/A.Capetti et al. ($H\alpha$).

De acuerdo con el modelo más aceptado la energía surge de un núcleo activo por la caída de materia interestelar en un agujero negro de entre $10^6 M_{\odot}$ y $10^{10} M_{\odot}$ de masa situado en el centro galáctico. Por medio de este mecanismo la materia proporciona energía con más eficiencia que en las estrellas, lo cual explica cómo desde un volumen $\lesssim 1 \text{ pc}^3$ se llega a radiar tanta o más energía que desde una galaxia entera sin consumir una cantidad exorbitante de gas interestelar. El acrecimiento de los agujeros negros se produce probablemente a través de un disco de gas ionizado muy caliente de pequeño tamaño ($< 1 \text{ pc}$) ceñido por un denso *toro* de gas molecular y de polvo. La materia del toro entra en el disco de acrecimiento y, a medida que su energía se disipa en forma de calor y de radiación y que su movimiento se ralentiza, cae irremediamente en el pozo gravitatorio del agujero negro. A la radiación que esto genera le acompaña en ocasiones un potente chorro bidireccional de plasma que puede alejarse hasta cientos de miles de parsecs.

En el Universo local hay núcleos activos en el $\sim 43\%$ de las galaxias. La luminosidad total de todos los núcleos activos fue máxima cuando el Universo contaba con dos o tres mil millones de años de vida y ha decrecido desde entonces.

Alimentación

En general un brote de formación estelar está confinado en una región circunuclear de entre cien y mil parsecs de radio dentro de la galaxia que lo alberga. Para sostener la elevada tasa de formación estelar una buena parte del medio interestelar galáctico se ha de acumular en la región donde nacen las estrellas en el breve tiempo que dura el brote. Por su parte, los núcleos activos requieren una menor cantidad de materia para alimentarse pero, por su diminuto tamaño ($< 1 \text{ pc}$), no es fácil que el movimiento del gas interestelar se ralentice lo suficiente como para que este caiga sobre ellos. Las galaxias activas/BFE y las “estándar” se diferencian, por tanto, en la eficacia con la que ciertos *mecanismos de alimentación* conducen parte del gas interestelar hacia su interior.

En galaxias de disco con poca o nula interacción con su entorno la alimentación comienza con la aparición de irregularidades en su potencial gravitatorio. Dichas irregularidades dan lugar a la acumulación de un buen número de estrellas en forma de brazos espirales y barras de gran tamaño (varios kiloparsecs), tal como se observa en muchos objetos (véase un ejemplo en la figura 3). La fuerza gravitatoria combinada de esas estrellas induce un flujo de gas desde el disco galáctico hacia el interior de la galaxia. Por el camino algunos cúmulos de gas que siguen trayectorias secantes pueden disipar una fracción de su energía al chocar entre sí, lo cual contribuye secundariamente a su caída hacia el centro. En último extremo el gas se amontona en un anillo de unos $\sim 100 - 500 \text{ pc}$ de

radio, donde, si la densidad y la cinemática son adecuadas, surge un brote de formación estelar. El proceso de alimentación es mucho más eficaz en galaxias que colisionan o se fusionan entre sí, a causa de los violentos procesos disipativos que experimenta el medio interestelar. Son estas colisiones galácticas las que dan lugar a los brotes de formación estelar más luminosos.



Figura 3: Mosaico multicolor de la emisión de la galaxia barrada NGC 1300 compuesto por imágenes obtenidas con el telescopio espacial Hubble (bandas B y V del visible e infrarrojo). La figura está rotada respecto al plano del cielo para minimizar su tamaño. La galaxia cuenta con una barra estelar de gran tamaño (de izquierda a derecha en la figura) en cuyos extremos nacen dos brazos espirales. Cortesía de NASA, ESA y The Hubble Heritage Team.

Hacia el interior del anillo circunuclear los brazos y las barras estelares pierden su eficacia y el transporte del material del que el núcleo activo se alimenta se detiene a menos que lo retomen otros mecanismos alternativos. En primera instancia, bien por la fuerza de la gravedad de barras secundarias y pequeños brazos espirales internos, bien por la disipación de energía del gas por su propia viscosidad, el gas podría llegar a unos diez parsecs del centro. Por fin, las fuerzas de gravedad del agujero negro podrían atraer parte de esa materia hasta el disco de acrecimiento.

Realimentación

Dos vías principales por las que los brotes de formación estelar y los núcleos activos inyectan grandes cantidades de energía en el medio circundante son la *emisión de radiación* y los flujos de materia denominados *vientos galácticos*.

Radiación. Debido a sus altísimas temperaturas las estrellas masivas liberan una gran cantidad de fotones⁴ del rango del ultravioleta, cuyas energías ($h\nu \gtrsim 1$ eV) les permiten arrancar electrones (ionizar) y romper los enlaces químicos (disociar) de muchos de los átomos y de las moléculas del medio interestelar. Dado que la radiación ultravioleta interacciona muy fácilmente con el gas y el polvo, la transferencia de energía a la materia resulta muy eficiente. En torno a las zonas de formación estelar surgen regiones de gas ionizado en las que los fotones capaces de ionizar el hidrógeno atómico ($h\nu > 13,6$ eV) son absorbidos. Más allá, la radiación del *ultravioleta lejano* (UVL: $6,2 \text{ eV} < h\nu < 13,6 \text{ eV}$, $91 < \lambda < 200$ nm) restante es aún capaz de inducir la disociación de las moléculas de hidrógeno y da lugar a *regiones de fotodisociación*, también llamadas PDR (acrónimo de su nombre en inglés: *Photo-Dissociation Region*). En la figura 4 se muestra una PDR de nuestra galaxia.

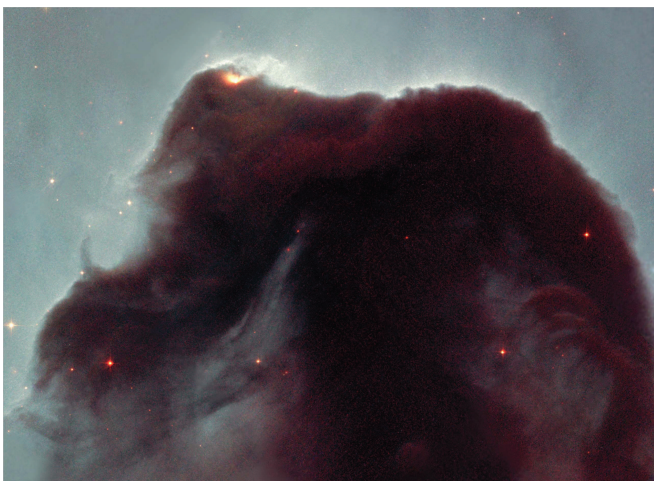


Figura 4: Imagen de la nebulosa de la Cabeza de Caballo obtenida con el telescopio espacial Hubble (bandas B, V e I del continuo y línea $H\alpha$). La imagen cubre una región de 0,7 pc de tamaño a lo largo del eje horizontal y está rotada respecto al plano celeste. La nebulosa de la Cabeza de Caballo es una acumulación de gas molecular y de polvo situada a 490 pc dentro de nuestra galaxia. En su parte superior, la radiación de las estrellas cercanas ha procesado la materia interestelar hasta crear una región de fotodisociación (PDR). Cortesía de NASA, NOAO, ESA y The Hubble Heritage Team (STScI/AURA).

⁴Los fotones son las partículas cuánticas que constituyen la radiación electromagnética (es decir, la luz). La energía de un fotón es $h\nu$, donde h es la constante de Planck y ν la frecuencia de la radiación.

A diferencia de las estrellas, las cuales radian principalmente en el visible y el ultravioleta, la radiación de alta energía de los núcleos activos abarca hasta los rayos gamma e incluye una importante componente de rayos X *duros* ($h\nu > 1$ keV). Estos son entre cien y mil veces más energéticos y penetrantes que la radiación ultravioleta, por lo cual pueden extender sus efectos a mayores distancias y afectar a mayores cantidades de materia. La radiación de un núcleo activo escapa con facilidad perpendicularmente al toro molecular que lo circunda y origina así un *cono de gas ionizado* en su camino. Por contra, en dirección al toro la gran cantidad de gas y de polvo que este contiene extingue completamente la radiación del visible al ultravioleta; cuando la cantidad de materia interpuesta es suficiente, incluso los rayos X son absorbidos por completo. No obstante, tanto si el cono de radiación llega a intersectar con el disco galáctico⁵, como si el toro no es completamente opaco, cabe que los rayos X extiendan su radio de acción más allá del entorno circunnuclear inmediato y que creen así lo que se conoce como *Regiones Dominadas por rayos X*, o XDR (acrónimo de su nombre en inglés: *X-ray Dominated Region*).

Vientos galácticos. Los *vientos galácticos* nacen en cavidades de gas caliente y sobrepresurizado que se forman en el medio interestelar cuando hay una rápida inyección de energía que el enfriamiento por radiación no llega a contrarrestar. En galaxias BFE esto ocurre por la concurrencia súbita de vientos estelares o de explosiones supernova. Otro tanto puede suceder en galaxias activas por efecto de la radiación o del chorro de plasma que brotan de los núcleos activos. La cavidad de gas sobrepresurizado da paso a una “superburbuja” expansiva que en ocasiones alcanza los límites verticales del disco galáctico y arrastra hacia el halo parte del gas y del polvo que ha barrido en su camino. La materia eyectada sólo retorna a la galaxia si su velocidad no es suficiente para superar la atracción gravitatoria del sistema.

El medio molecular

Características generales de las nubes moleculares

En las galaxias el medio molecular acumula típicamente entre el 20 % y el 80 % de la masa del medio interestelar neutro y constituye la más fría, densa y neutra de sus fases. Son esas condiciones las que posibilitan que algunas condensaciones de gas interestelar lleguen a colapsar por su propia gravedad y den así lugar a las estrellas. El medio molecular ocupa preferentemente los kiloparsecs centrales de las galaxias en forma de acumulaciones denominadas *nubes moleculares*. Estas son objetos muy inhomogéneos de un tamaño de entre 10 pc y 100 pc y una masa de

⁵ Nótese que el toro y el disco galáctico no son necesariamente coplanarios.

entre $10^2 M_{\odot}$ y $10^7 M_{\odot}$ típicamente. El $\sim 99\%$ de la masa de una nube molecular lo forma un gas heterogéneo, predominantemente neutro, cuya densidad (n) es de entre 10^2 y 10^6 partículas por centímetro cúbico y cuya temperatura cinética (T_K) está entre los 10 y los 100 K. Los componentes mayoritarios son el hidrógeno molecular (H_2), que suma hasta el 75 % de la masa, y el helio atómico (He), que aporta prácticamente el 25 %. El gas contiene además cantidades mínimas de especies atómicas y moleculares en las que se combinan el hidrógeno y los elementos minoritarios del medio interestelar (oxígeno, carbono, nitrógeno, azufre, silicio, etc.). La relevancia de estos compuestos para el estudio de las galaxias se verá más adelante.

El 1 % restante de la masa del medio molecular lo aporta un conjunto de *granos* sólidos de entre unas décimas de micra y algunos angstroms de diámetro denominado *polvo interestelar*. La composición del polvo está dominada por silicatos y por diversos compuestos de carbono, a lo que hay que añadir un manto helado de agua y de otros compuestos minoritarios que se adhiere a los granos en las regiones moleculares más densas y frías. Por último, los *Hidrocarburos Aromáticos Policíclicos* (HAP) son largas cadenas de anillo de carbono e hidrógeno que, por sus propiedades, están a caballo entre el gas y el polvo.

Composición química del gas molecular

Hasta la fecha se han detectado más de 150 moléculas (entendiendo como tales cualesquiera compuestos químicos de dos o más átomos) en el medio circunestelar e interestelar. Las particulares condiciones ambientales del espacio permiten la existencia de compuestos que, por su gran inestabilidad o reactividad, son de difícil síntesis en la Tierra.

La referencia de los estudios de la química del medio molecular son las *nubes oscuras*. Estas son nubes moleculares relativamente densas ($n \sim 10^4 \text{ cm}^{-3}$) y frías ($T_K \simeq 10 - 30 \text{ K}$) que están aisladas de la radiación ultravioleta estelar y en las que sólo penetran los rayos cósmicos⁶. Gracias a su gran energía intrínseca (entre 100 MeV y 10 GeV por partícula, típicamente) los rayos cósmicos logran ionizar una pequeña cantidad de moléculas y átomos, lo cual resulta capital para la química del gas molecular. En particular, las reacciones entre iones y moléculas neutras, al ser relativamente rápidas, determinan en gran medida la composición química del medio. De gran importancia son también las reacciones químicas que ocurren en los mantos de los granos de polvo interestelar, pues modifican indirectamente la composición del gas cuando los productos de las reacciones son liberados. Es precisamente sobre la superficie de los granos donde se forma el hidrógeno molecular que predomina en las nubes.

⁶Partículas (protones, electrones,...) que viajan por el medio interestelar a gran velocidad.

En general la composición química de una nube molecular es muy sensible a las vías por las cuales la energía se inyecta en el medio. Si bien los compuestos químicos minoritarios son escasísimos en cualquier circunstancia, sus abundancias⁷ pueden diferir de los valores medidos en nubes oscuras en varios órdenes de magnitud. Las observaciones de la emisión molecular en otras galaxias, interpretadas a la luz de los modelos teóricos químicos y comparadas con lo que se observa en regiones de la Vía Láctea bien caracterizadas, permiten comprender cuáles son los mecanismos que rigen el estado del medio molecular.

Observaciones del medio molecular en galaxias

Las líneas moleculares como trazadoras del medio interestelar

Una molécula de un compuesto cualquiera de los presentes en el medio interestelar tiene una energía interna que depende de su rotación y de la de los átomos que la componen⁸, de la vibración de dichos átomos en torno a sus posiciones de equilibrio y de la disposición de sus electrones. A esas escalas microscópicas, en las que leyes de la mecánica cuántica son patentes, cada término que contribuye a la energía puede tomar únicamente ciertos valores predeterminados que dependen de la estructura de la molécula. Se denomina niveles de energía a las posibles combinaciones de todos los términos.

El nivel de energía de una molécula aumenta (lo que se conoce como *excitación*) cuando esta absorbe un fotón de una energía determinada (igual a la diferencia de energía entre los dos niveles involucrados) y disminuye cuando lo emite (*desexcitación*). De este intercambio de fotones surgen las líneas espectrales moleculares que se suelen detectar en observaciones desde el rango de onda de radio hasta el infrarrojo. La cantidad de radiación que se recibe por medio de las líneas de cada compuesto depende del número de moléculas de dicho compuesto que se encuentran en cada uno de los niveles de energía a lo largo de la línea de mirada. A su vez, las proporciones en las que las moléculas ocupan los niveles de energía están regidas por la propia emisión y absorción de radiación y por excitaciones y desexcitaciones inducidas por colisiones con otras partículas (mayoritariamente el hidrógeno molecular). La eficacia de las transiciones por colisiones depende de la temperatura cinética del gas, que mide la agitación térmica de las partículas, y de su densidad.

⁷ En adelante, se denomina *abundancia de una especie* a la abundancia relativa de esta respecto al hidrógeno molecular, es decir, al cociente entre la cantidad de moléculas de la especie en cuestión y el número de moléculas de hidrógeno molecular.

⁸ La rotación se entiende aquí en un sentido general y puede incluir términos de espín además del término de momento angular orbital.

En definitiva, la intensidad de las líneas de una determinada especie depende del número de moléculas dispuestas a lo largo de la línea de mirada y de la densidad y la temperatura a las que estas se encuentran. La intensidad de las líneas depende también de cuál es la especie en cuestión, puesto que la disposición de los niveles y la rapidez con la que la radiación y las colisiones inducen transiciones entre ellos está determinada por la estructura interna de las moléculas. Es por todo ello por lo que las distintas líneas moleculares proporcionan una valiosa información acerca de la composición química y las condiciones físicas del medio interestelar.

Observaciones radioastronómicas

En general, en las condiciones de densidad y temperatura típicas de las nubes moleculares sólo se detectan líneas de emisión que corresponden a transiciones entre los niveles rotacionales más bajos. Estas líneas se hallan en el rango de ondas de radio de longitudes de onda milimétricas y submilimétricas, lo que corresponde a frecuencias de entre decenas y centenas de gigahertzios.

Fórmula química							
2 átomos:	OH	CO	H ₂	CH	CS	CH ⁺	CN
	SO	SiO	CO ⁺	NO	NS		
3 átomos:	H ₂ O	HCN	HCO ⁺	C ₂ H	HNC	N ₂ H ⁺	OCS
	HCO	H ₂ S	SO ₂	HOC ⁺	C ₂ S		
4 átomos:	H ₂ CO	NH ₃	HNCO	H ₂ CS	HOCO ⁺	C ₃ H	H ₃ O ⁺
5 átomos:	c-C ₃ H ₂	HC ₃ N	CH ₂ NH	NH ₂ CN			
6 átomos:	CH ₃ OH	CH ₃ CN					
7 átomos:	CH ₃ CCH						

Cuadro 1: Lista de las moléculas detectadas fuera de la Vía Láctea a enero de 2008 (Cologne Database for Molecular Spectroscopy; <http://www.astro.uni-koeln.de/site/vorhersagen/>)

Durante las últimas décadas el número de especies moleculares detectadas fuera de la Vía Láctea ha crecido de forma notable hasta alcanzar la cuarentena en la actualidad (el cuadro 1 muestra una lista de las mismas). A su vez ha aumentado el número de objetos en los que se han encontrado esas especies. Este progreso ha sido posible gracias a la puesta en marcha de diversos radiotelescopios que operan en el rango de ondas centimétricas, milimétricas y submilimétricas. En el cuadro 2 se mencionan algunos de los más relevantes de entre los que están en servicio. Dos de los instrumentos que más han contribuido en este campo por su gran sensibilidad

y resolución angular son el telescopio de Pico de Veleta (España) y el radiointerferómetro de Plateau de Bure (Francia), ambos⁹ pertenecientes al *Instituto de Radio Astronomía Milimétrica* (IRAM). El IRAM es una organización científico-técnica europea cuyos miembros actuales son el *Centre National de la Recherche Scientifique* (CNRS) francés, el *Max Planck Gesellschaft* (MPG) alemán y el *Instituto Geográfico Nacional* (IGN) español, del cual depende el Observatorio Astronómico Nacional. Mediante observaciones con los instrumentos del IRAM y con otros de los mencionados en el cuadro 2 los astrónomos del Observatorio Astronómico Nacional han contribuido al estudio del medio molecular en otras galaxias.

Telescopio	Emplazamiento	Antenas	Tamaño m	Frecuencias GHz
APEX	Chile	1	12	211 – 1384
ASTE	Chile	1	10	324 – 372
CARMA	Estados Unidos	16	6-10	85 – 270
CSO	Estados Unidos	1	10	180 – 950
Effelsberg	Alemania	1	100	<96
FCRAO	Estados Unidos	1	14	85 – 116
GBT	Estados Unidos	1	110	<50
HHSMT	Estados Unidos	1	10	210 – 720
IRAM Pico de Veleta (PdV)	España	1	30	72 – 281
IRAM Plateau de Bure (PdB)	Francia	6	15	80 – 267
JCMT	Estados Unidos	1	15	211 – 710
KPI2m	Estados Unidos	1	12	68 – 180
NMA	Japón	6	10	85 – 237
Nobeyama	Japón	1	45	20 – 116
SEST	Chile	1	15	78 – 363
SMA	Estados Unidos	8	6	180 – 702
VLA	Estados Unidos	27	25	<50

Cuadro 2: Algunos de los mayores radiotelescopios que operan en longitudes de onda milimétricas o submilimétricas. Se indica el país en el que se encuentran, el número de las antenas de las que constan, el diámetro de las mismas y el rango de frecuencias en el que operan.

Entre los instrumentos empleados en radioastronomía se distinguen dos familias: los telescopios de una única antena y las agrupaciones de varias antenas que operan coordinadamente, denominadas radiointerferómetros. Mientras que el poder de resolución angular de los primeros está limitado por el diámetro de la antena, el de los segundos está determinado por la separación entre antenas, por lo cual pueden distinguir detalles de menor tamaño. No obstante, la mayor diferencia cualitativa entre los dos tipos

⁹ En lo que sigue, *telescopio PdV* e *interferómetro PdB*, respectivamente.

de instrumentos es que las técnicas de interferometría permiten obtener radioimágenes de amplias regiones del cielo sin necesidad de apuntar el telescopio en diferentes direcciones.

El medio molecular en las galaxias

A pesar de ser el compuesto más abundante, el hidrógeno molecular es un pésimo trazador de las nubes moleculares. Al estar formado por dos átomos idénticos las transiciones radiativas entre niveles rotacionales resultan tan lentas que difícilmente se pueden detectar en el medio interestelar.

El segundo compuesto más abundante en el medio interestelar es el monóxido de carbono, CO, el cual suele aportar cerca del 0,01 % de las moléculas del medio. La estructura de la molécula de CO permite que la línea $J = 1 - 0$ (115 GHz), que corresponde a la transición entre sus dos niveles rotacionales más bajos, sea detectable cuando la densidad del gas es de varios centenares de partículas por centímetro cúbico, cosa que ocurre en prácticamente todas las nubes moleculares. La detección de transiciones entre niveles más altos, cuyas frecuencias son progresivamente mayores, es tanto más sencilla cuanto mayor es la densidad del gas. El monóxido de carbono es el trazador por excelencia del medio molecular, tanto en la Vía Láctea como en otras galaxias de nuestro entorno y del Universo primitivo. Se ha llegado a detectar en un quásar¹⁰ lejanísima emisión de CO que se originó cuando el Universo apenas contaba con 870 millones de años de vida.

En condiciones muy generales la masa total de medio molecular (incluyendo la contribución del helio), M_{mol} , se puede calcular con bastante precisión a partir de la luminosidad de la línea de CO($J = 1 - 0$), $L_{\text{CO}(1-0)}$, mediante la relación¹¹

$$M_{\text{mol}}/L_{\text{CO}(1-0)} \simeq 5 M_{\odot}/(\text{K km s}^{-1} \text{ pc}^2)$$

La masa total de gas molecular que se determina de esta manera es del orden de $10^9 M_{\odot}$ en galaxias estándar como la Vía Láctea. En algunas galaxias HLIR muy distantes la masa molecular que se infiere¹² es del orden de $10^{11} M_{\odot}$.

¹⁰ Un quásar en un tipo de galaxia activa que se caracteriza por su gran luminosidad.

¹¹ Se discute si el cociente $M_{\text{mol}}/L_{\text{CO}(1-0)}$ podría ser sensiblemente distinto en galaxias de poca metalicidad, donde el CO es escasísimo, o en los núcleos galácticos y en galaxias de tipo ULIR, donde el grueso del gas molecular tal vez no esté en forma de nubes moleculares como las de nuestra galaxia.

¹² La frecuencia de la radiación que recibimos de las galaxias más distantes es menor de lo que era en origen a consecuencia de la expansión del Universo. En algunos casos esto impide que la emisión de la línea de CO($J = 1 - 0$) se pueda detectar con la instrumentación actual. Cuando esto ocurre la masa de gas molecular se calcula a partir de observaciones de otras líneas de CO.

El gas molecular y los mecanismos de alimentación y de realimentación

Eficiencia de formación estelar

Un aspecto fundamental de las galaxias BFE es hasta qué punto el proceso de formación de estrellas en su seno difiere del que se da en galaxias estándar. Para responder a esta cuestión se ha invertido un gran esfuerzo en determinar la eficiencia con la que las estrellas nacen en el gas molecular de estos objetos. La *eficiencia de formación estelar del gas molecular*, EFE_{mol} , se define como

$$EFE_{\text{mol}} \equiv \frac{\dot{M}_{\star}}{M_{\text{mol}}}$$

donde $\dot{M}_{\star}$ es la *tasa de formación estelar*, definida como la masa de estrellas que se forma por unidad de tiempo, y M_{mol} es la masa de gas molecular calculada a partir de $L_{\text{CO}(1-0)}$. Existen diversos métodos para determinar $\dot{M}_{\star}$ en una galaxia a partir de la radiación recibida en determinadas frecuencias. Uno de los más fiables para regiones con una alta tasa de formación estelar se basa en la luminosidad de la radiación en el infrarrojo, L_{IR} , la cual se debe a la emisión del polvo calentado por las estrellas. Así, la tasa de formación estelar, $\dot{M}_{\star}$ (medida en masas solares por año), se calcula como

$$\dot{M}_{\star}/L_{\text{IR}} \simeq 2 \times 10^{-10} M_{\odot} \text{ año}^{-1}/L_{\odot}$$

Las observaciones de CO y del infrarrojo han revelado una estrecha relación entre el medio molecular y la formación estelar en galaxias. En particular, la luminosidad de la línea CO($1 - 0$), proporcional, según lo dicho, a *toda* la masa de gas molecular, y la del continuo del *infrarrojo*, proporcional a la tasa de formación estelar, se atienen a la relación

$$L_{\text{CO}(1-0)} \propto L_{\text{IR}}^{0,6}$$

De acuerdo con esta expresión, EFE_{mol} es proporcional a $L_{\text{IR}}^{0,4}$, por lo que resulta mayor cuanto mayor es la tasa de formación estelar. Podría concluirse de esta correlación que los mecanismos de la formación estelar operan de forma especial en las galaxias con mayor tasa de formación estelar, de modo que logran convertir en estrellas una fracción mayor del gas disponible. Sin embargo las observaciones de otros trazadores moleculares distintos de CO ha ofrecido un perspectiva diferente.

Eficiencia de formación estelar del gas molecular denso

Precisamente porque traza fielmente *toda* la masa molecular, independientemente de las condiciones en las que esta se encuentra, la emisión de CO(1–0) por sí sola no permite discernir las condiciones físicas del medio molecular. Por ello es difícil cuantificar a partir de la emisión de CO la fracción de gas molecular suficientemente denso como para poder colapsar y formar estrellas.

Diversos grupos de investigación han intentado determinar con fiabilidad la cantidad de gas molecular denso que alimentan los brotes de formación estelar en galaxias. Para ello se observa la emisión de especies que, por su estructura molecular, radian preferentemente desde regiones con condiciones físicas específicas. Los compuestos trazadores de gas denso que se observan con más frecuencia, tanto en La Vía Láctea como fuera de ella, son el HCN y el HCO⁺. Las abundancias de HCN y HCO⁺ son típicamente de $10^{-8} - 10^{-7}$, por lo que estos se encuentran entre las especies más profusas y de emisión más intensa después del CO. A diferencia de lo que ocurre con CO, la emisión de la transición $J = 1 - 0$ de estas dos especies (89 GHz en ambos casos) sólo es detectable cuando la densidad del gas es $\gtrsim 10^4 \text{ cm}^{-3}$ (lo que consideraremos *gas denso* de aquí en adelante). La emisión de HCN y de HCO⁺, así como de otras trazadores del gas denso, se ha estudiado en galaxias cercanas y en objetos del Universo más distante. En particular, las líneas de HCN y de HCO⁺ se han detectado en un quásar cuya emisión se originó cuando la edad del Universo era de 1600 millones de años.

En condiciones muy generales es posible calcular con bastante precisión la cantidad de gas molecular *denso*, M_{den} , a partir de la luminosidad de la línea HCN($J = 1 - 0$) (otro tanto puede hacerse a partir de la luminosidad de la línea HCO⁺($J = 1 - 0$)), $L_{\text{HCN}(1-0)}$, mediante la relación

$$M_{\text{den}}/L_{\text{HCN}(1-0)} \simeq 10 M_{\odot}/(\text{K km s}^{-1} \text{ pc}^2)$$

Los cocientes $L_{\text{HCN}(1-0)}/L_{\text{CO}(1-0)}$ en galaxias del Universo cercano difieren hasta en un factor 25, lo cual sugiere importantes diferencias en la fracción de gas molecular denso que albergan ($M_{\text{den}}/M_{\text{mol}}$). Por otra parte, al comparar las observaciones de HCN y las de infrarrojo se ha observado que $L_{\text{HCN}(1-0)}$ y L_{IR} se atienen a una estrecha relación *lineal* que abarca desde las galaxias extremadamente luminosas de tipo LIR y ULIR hasta las pequeñas condensaciones moleculares de la Vía Láctea. De dicha relación se deduce que, a diferencia de lo que ocurre con EFE_{mol} , la eficiencia de la formación estelar respecto al gas denso trazado por HCN(1–0), EFE_{den} , es independiente de la tasa de formación estelar. En concreto, se encuentra que

$$EFE_{\text{den}} \equiv \frac{\dot{M}_{\star}}{M_{\text{den}}} \simeq 2 \times 10^{-8} \text{ año}^{-1}$$

A la luz de estos resultados, que, en general, $\dot{M}_{\star}/M_{\text{mol}}$ sea mayor cuanto más luminosas son las galaxias se puede atribuir a un aumento de la proporción de gas molecular denso. La correlación HCN–IR sugiere además que la formación de estrellas está regulada por la cantidad de materia prima disponible y que, por tanto, procede de forma independiente de las características de la galaxia en la que tiene lugar.

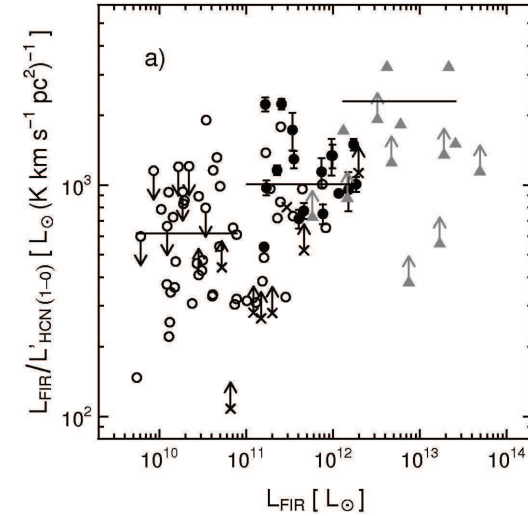


Figura 5: Cociente $L_{\text{IR}}/L_{\text{HCN}(1-0)}$ en función de L_{IR} en una muestra de galaxias LIR, ULIR y HLIR. Las flechas indican que el valor real del cociente es menor que el representado, si apuntan hacia abajo, o mayor, si apuntan hacia arriba. Se aprecia cómo el cociente tiende a ser significativamente mayor en las galaxias de mayor luminosidad infrarroja. Cortesía de J. Graciá-Carpio (adaptado de Graciá-Carpio et al. 2008, *Astronomy & Astrophysics*, 479, 703).

A pesar de que su sencillez la hace muy atractiva, la interpretación habitual de la correlación HCN–IR no está libre de incertidumbres. Una de ellas es hasta qué punto la línea HCN(1–0) (en general cualquier otra línea) traza *sin sesgos* la masa de gas molecular denso en condiciones tan anómalas como las que se dan en regiones de intensa formación estelar. Recientemente distintos estudios han concluido que, de hecho, tales sesgos existen. Por un lado, las observaciones de galaxias HLIR, aun escasas por encontrarse estas a grandes distancias, sugieren que el valor del cociente $L_{\text{IR}}/L_{\text{HCN}(1-0)}$ en estos objetos es sistemáticamente mayor de lo que la correlación HCN–IR predice (véase la figura 5). Ello sugiere que, bien la eficiencia de formación estelar real (EFE_{den}) no es proporcional al

cociente $L_{\text{IR}}/L_{\text{HCN}(1-0)}$ (este sería el caso, por ejemplo, si $L_{\text{HCN}(1-0)}$ no trazara correctamente la masa de gas denso en estas galaxias), bien las estrellas se forman más eficientemente de lo esperado. Por otra parte, al observar simultáneamente la línea $J = 1 - 0$ de HCN y HCO^+ en una amplia muestra de galaxias LIR y ULIR los astrónomos del Observatorio Astronómico Nacional han advertido que, inesperadamente, el cociente $L_{\text{HCN}(1-0)}/L_{\text{HCO}^+(1-0)}$ aumenta con la luminosidad L_{IR} . Dado que se espera que la emisión de HCN y de HCO^+ provenga de regiones con condiciones físicas similares, tal relación sugiere que la luminosidad de las líneas moleculares pudiera no reflejar fielmente la cantidad de gas denso en las galaxias más luminosas.

Las observaciones complementarias de algunas transiciones más altas ($\text{HCN}(J = 3 - 2)$, $\text{HCO}^+(J = 3 - 2)$) han aportado indicios de que, por diferencias en la composición química y en las condiciones físicas del gas, el cociente $M_{\text{den}}/L_{\text{HCN}(1-0)}$ real en galaxias ULIR es menor que el valor canónico aceptado. En tal caso la masa de gas denso de dichas galaxias sería menor de lo esperado, lo cual confirmaría que la eficiencia de formación de estrellas con respecto al gas denso aumenta en las galaxias más luminosas. Si esto es consecuencia de los efectos de la realimentación que la formación estelar ejerce en el medio, no se ha podido determinar hasta el momento. Estos recientes resultados han generado un intenso debate entre los expertos.

Diagnósticos moleculares de la formación estelar y de la actividad nuclear

Diversos estudios de la emisión molecular en galaxias han buscado el modo de diagnosticar la presencia de brotes formación estelar masiva y de núcleos activos. Subyace la idea de que la energía que las estrellas y los núcleos activos inyectan en el medio circundante podría modificar de forma apreciable la composición y las condiciones físicas del medio molecular. A diferencia de los diagnósticos más tradicionales, basados en observaciones a frecuencias altas (del infrarrojo a los rayos X), unos criterios basados en líneas moleculares no estarían limitados por la extinción de la radiación que causa el polvo interestelar, pues esta es insignificante en longitudes de onda (sub)milimétricas. De ahí que las observaciones moleculares pudieran ayudar a clarificar la naturaleza de las galaxias luminosas y ricas en gas y polvo que dominaron la formación estelar en el pasado.

Hasta la fecha varios estudios independientes han comparado la luminosidad de algunas líneas moleculares (principalmente de CO, HCN y HCO^+ , por ser las más intensas) en muestras de galaxias BFE o activas. En general se concluye que los cocientes $L_{\text{HCN}(1-0)}/L_{\text{CO}(1-0)}$ y $L_{\text{HCN}(1-0)}/L_{\text{HCO}^+(1-0)}$ son mayores en torno a los núcleos activos que en regiones con brotes de formación estelar. Las observaciones complemen-

tarias de transiciones más altas ($J = 2 - 1$, $J = 3 - 2$) sugieren además que tras esas diferencias subyacen unas condiciones físicas del gas y unas abundancias químicas distintas.

Estudios de la química del gas molecular en galaxias BFE y activas

Por qué estudiar la química del gas molecular

Las observaciones de líneas moleculares constituyen una vía muy prometedora para tasar los mecanismos de alimentación y realimentación en las galaxias BFE y activas. Sin embargo, tal como se ha visto previamente, tanto para precisar la cantidad de materia que abastece las regiones de formación estelar como para comprender los efectos de las estrellas y los núcleos activos sobre el medio es necesario determinar las abundancias químicas de las especies observadas.

Durante los últimos años, las observaciones de líneas moleculares milimétricas han demostrado que existen ciertas diferencias en la composición del medio molecular de las galaxias. No obstante, aún escasean las observaciones de moléculas cuya química sean tan particular como para revelar sin ambigüedades las causas de la composición del gas molecular en las galaxias BFE y activas. En particular, se requieren observaciones de compuestos cuyas abundancias cambien de forma sustancial en función del mecanismo que gobierne la composición del medio (choques, radiación UV, rayos X,...), lo cual no es el caso de las especies más observadas (CO, HCN, HCO^+).

A continuación, se describen algunos trabajos cuyo propósito ha sido evaluar los efectos en la química del medio molecular de los mecanismos de alimentación y de realimentación en las galaxias BFE y activas mediante observaciones de trazadores químicos específicos. Los estudios que se detallan, muchos de ellos dirigidos por astrónomos del Observatorio Astronómico Nacional, se han ocupado de determinadas galaxias BFE y activas cercanas, a fin de aprovechar la resolución angular y la sensibilidad con las que cuentan los radiotelescopios actuales.

Choques en galaxias BFE

Química dominada por choques

El medio interestelar es objeto de frecuentes perturbaciones en las que se originan *frentes de choque* que se propagan supersónicamente. Al paso de estos frentes de choque el medio interestelar es comprimido y calentado a la vez que su velocidad varía súbitamente. Estos cambios intensifican aquellas reacciones químicas que necesitan de un aporte previo de energía.

Por otra parte, a consecuencia de las colisiones de los granos con otras partículas del medio parte del material del polvo puede revertir en el gas y acrecentar durante un tiempo las abundancias de determinadas especies.

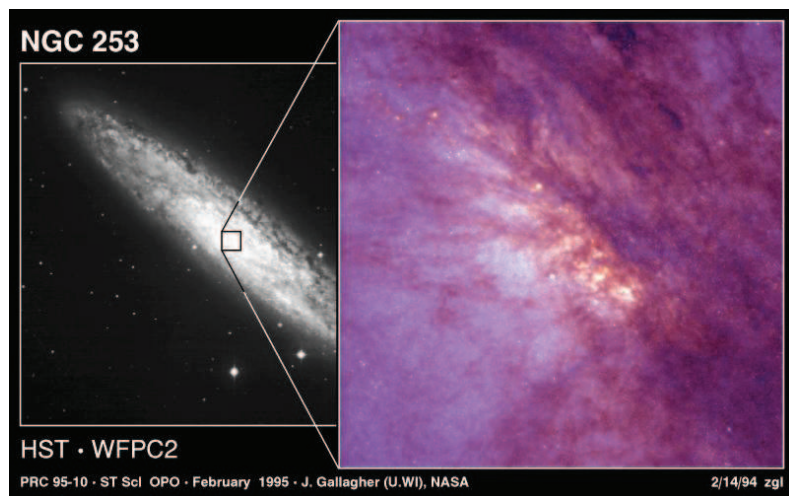


Figura 6: A la izquierda, una imagen de la emisión en el visible de la galaxia BFE NGC 253 obtenida con un telescopio terrestre; a la derecha, una imagen de los 300 pc centrales de NGC 253 obtenida por el telescopio espacial Hubble. Las observaciones de gran resolución del Hubble dan idea de cuán intensa es la formación estelar en el centro de esta galaxia. Cortesía de Jay Gallagher (University of Wisconsin-Madison), Alan Watson (Lowell Observatory, Flagstaff, AZ) y NASA (izquierda) y del Carnegie Institution of Washington (derecha).

Uno de los trazadores más versátiles y fiables de la química dominada por choques es el monóxido de silicio (SiO), cuya abundancia en fase gaseosa crece sobremedida cuando los granos (principalmente sus núcleos) son erosionados y parte del silicio es liberado. Durante el choque también pueden crecer las abundancias en fase gaseosa de ciertos compuesto sulfúreos (SO , H_2S) u orgánicos (CH_3OH , HNCO) por la destrucción parcial de los mantos de hielo de los granos. La erosión de los mantos requiere un menor aporte energético que la de los núcleos de los granos. Por ello, a diferencia de lo que ocurre con el material silíceo de los granos, los compuestos depositados en los mantos pueden ser liberados en cantidades significativas por mecanismos distintos de los choques que son relativamente habituales (sublimación, irradiación con fotones

ultravioleta). De ahí que la interpretación de la emisión del monóxido de silicio esté sujeta a menos incertidumbres¹³ que la de los otros trazadores.

Observaciones y resultados

A medida que un brote de formación estelar masiva evoluciona el medio interestelar está expuesto a diferentes perturbaciones mecánicas que pueden generar choques a lo largo de centenares de parsecs: en las fases preliminares, las barras y brazos espirales estelares que conducen el gas de la galaxia hacia el núcleo pueden inducir colisiones entre las nubes moleculares; con la evolución y la muerte de las estrellas masivas, se desencadenan superburbujas y vientos galácticos que barren la galaxia a su paso y llegan a expandirse por el halo.

Mediante observaciones con el radiointerferómetro PdB los astrónomos del Observatorio Astronómico Nacional han obtenido las primeras imágenes de la emisión de la línea $\text{SiO}(J = 2 - 1)$ en tres galaxias cercanas que albergan brotes de formación estelar de cierta intensidad: NGC 253, M 82 e IC 342. La resolución espacial de las observaciones (~ 80 pc) es similar al tamaño de las nubes moleculares gigantes. Las abundancias de SiO que se infieren en las tres galaxias a lo largo de centenares de parsecs son entre diez y mil veces mayores que las que se miden en las nubes oscuras de nuestra galaxia; ello demuestra que la química del gas molecular está condicionada por los efectos de los choques a gran escala. Los mecanismos que inducen estos choques pueden ser, sin embargo, muy dispares.

NGC 253 (figura 6) es una de las galaxias con formación estelar masiva prototípicas del Universo más próximo. Esta galaxia barrada contiene en los 300 pc centrales una región de formación estelar intensísima que probablemente se alimenta de un gran depósito de gas molecular con el que coexiste. Se cree que la concentración de gas molecular fue inducida por las fuerzas gravitatorias de la barra estelar. Las observaciones de SiO muestran que los choques ocurren en dos pseudo-anillos alineados con la barra estelar en los que aumenta la probabilidad de que las nubes moleculares, dirigidas por la gravedad de las estrellas de la barra, colisionen entre sí (véase la figura 7). Se concluye por ello que en esta galaxia la química dominada por choques acompaña al proceso de alimentación de la formación estelar.

Cabe notar que otros estudios han llegado a idénticas conclusiones a partir de observaciones con telescopios de antena única, en las cuales la mayor facilidad para observar un gran número de líneas simultáneamente suple en parte la menor resolución espacial. En particular, se ha sabido por estos estudios que abundan en el gas de NGC 253 las moléculas relacionadas con procesos químicos sobre los mantos de los granos (NH_3 , CH_3OH , CH_3CN , HNCO , SO , H_2S), lo cual da a entender que dichos mantos han sido erosionados de alguna manera.

¹³Excepto en el entorno de los núcleos activos, como se verá más adelante.

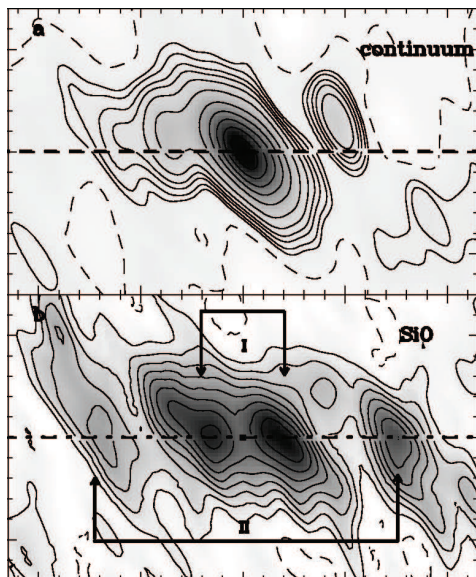


Figura 7: De arriba a abajo, mapas de la emisión del radiocontinuo a 87 GHz (caja *a*, en contornos) y de la línea $\text{SiO}(J = 2 - 1)$ (caja *b*, en contornos) en la galaxia NGC 253, ambos obtenidos con el interferómetro PdB. Los ejes horizontal y vertical corresponden a los ejes mayor y menor de la barra estelar de la galaxia. Las dimensiones de cada caja son $(46 \text{ segundos de arco}) \times (28 \text{ segundos de arco})$. De acuerdo con la emisión de SiO, los choques moleculares son particularmente intensos en dos pseudo-anillos (I y II). El grueso de la formación estelar, trazada por el radiocontinuo, está confinado en la región I. Cortesía de S. García-Burillo (adaptado de García-Burillo et al. 2000, *Astronomy & Astrophysics*, 355, 499).

IC 342 es una galaxia espiral barrada que alberga un brote de formación estelar de baja intensidad en un anillo de ~ 160 pc de diámetro. En IC 342 el gas molecular se acumula en dos brazos espirales que en su extremo interior se enroscan en torno al anillo de formación estelar. SiO es particularmente abundante en los brazos espirales, en los cuales crece la probabilidad de colisiones entre nubes moleculares inducidas por las fuerzas gravitatorias de la barra estelar (véase la figura 8). Al igual que NGC 253, IC 342 es un ejemplo de la relación entre la alimentación de la formación estelar y la química de choques a gran escala. Cabe notar que otras observaciones paralelas de la emisión de varias especies moleculares sugieren que en el anillo de formación estelar la composición del gas molecular no está tanto regulada por los choques (menos efectivos que

en el brazo espiral, según se infiere de la emisión de SiO) como por la radiación ultravioleta de las estrellas masivas. De este tipo de química se habla más adelante, a propósito del caso paradigmático de la galaxia M 82.

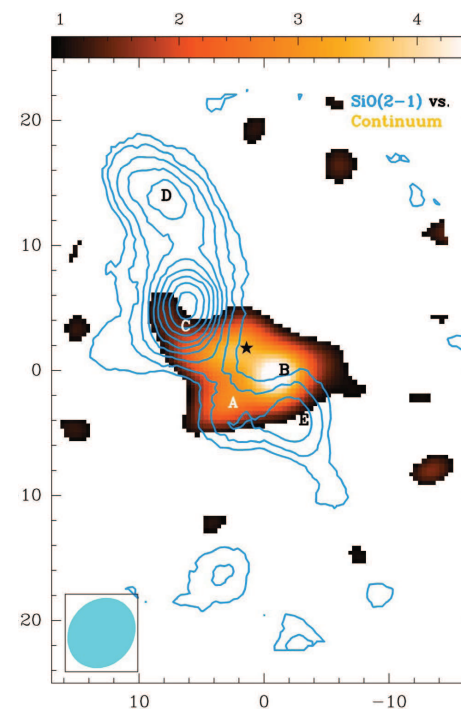


Figura 8: Mapas de la emisión de la línea $\text{SiO}(J = 2 - 1)$ (contornos azules) y del radiocontinuo a 87 GHz (color) en la galaxia IC 342, ambos obtenidos con el interferómetro PdB por el autor. Los ejes horizontal y vertical corresponden a los de Ascensión Recta y Declinación respectivamente y su escala está en segundos de arco. El grueso de la formación estelar, trazada por el radiocontinuo, está confinada en la región circunuclear (adaptado de Usero et al. 2006, *Astronomy & Astrophysics*, 448, 457).

M 82 es, junto a NGC 253, uno de los prototipos de las galaxias cercanas que albergan un brote de formación estelar masiva. En comparación con NGC 253, M 82 representa una fase más avanzada de un episodio de formación estelar. En el kiloparsec central del disco de esta galaxia irregular la alta tasa de explosiones supernova, los intensos campos UV y la producción de rayos cósmicos han alterado notablemente el medio

interestelar. La concurrencia de numerosas explosiones supernova ha originado un flujo “bicónico” de gas caliente que se expande más allá del disco (véase la figura 1). Este flujo arrastra con él grandes cantidades de gas frío y de polvo. A diferencia de lo que ocurre en NGC 253 e IC 342, en M 82 la mayor parte de la emisión de SiO no se encuentra en el disco galáctico sino en la interfaz entre este y el halo (véase la figura 9). Las dos principales estructuras que se observan son una superburbuja que se expande en torno a un cúmulo de estrella masivas y una chimenea que nace en el disco galáctico y penetra en el bulbo que rodea a este. Estas dos estructuras probablemente corresponden a dos etapas distintas del proceso por el que las explosiones supernova eyectan materia del disco galáctico al halo.

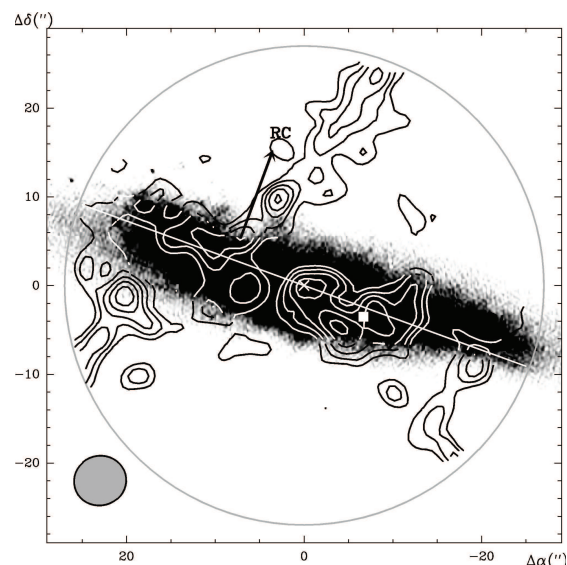


Figura 9: Mapas de la emisión en la galaxia M 82 de la línea SiO($J = 2 - 1$) (contornos), obtenido con el interferómetro PdB, y del radiocontinuo a 4,8 GHz (escala de grises saturada). Los ejes horizontal y vertical de la figura corresponden a los de Ascensión Recta y Declinación respectivamente y su escala está en segundos de arco. En la figura se observa que los choques moleculares, trazados por la emisión de SiO, se encuentran a lo largo de una chimenea perpendicular al disco galáctico y en una superburbuja que comienza a romper los límites del mismo. Cortesía de S. García-Burillo (adaptado de García-Burillo et al. 2001, *Astrophysical Journal*, 563, L27).

Radiación ultravioleta en galaxias BFE

Química dominada por rayos UV

Se conocen como *regiones de fotodisociación*, o PDR, aquellas zonas del medio interestelar cuyos balances térmico, iónico y químico están gobernados por la radiación del *ultravioleta lejano*. A lo largo de una PDR la energía de los fotones se deposita en la materia, con lo que el medio se calienta a la par que se excitan, disocian o ionizan determinadas especies. El flujo de la radiación UV se debilita principalmente por la absorción y dispersión en los granos de polvo y HAP. Los fotones UV son absorbidos también por las moléculas de hidrógeno, que resultan disociadas. El hidrógeno queda en estado atómico hasta un punto más allá del cual la cantidad de gas interpuesto protege las moléculas de hidrógeno de la disociación.

La inyección de energía en forma de radiación UV favorece las reacciones de fotodisociación. Resultan también más frecuentes las reacciones con electrones e iones a causa del mayor grado de ionización, mientras que el calentamiento y excitación de las moléculas permite que se lleven a cabo ciertas reacciones que necesitan de un aporte inicial de energía. Los fotones UV pueden procesar los mantos de hielo de los granos y favorecer así la liberación de materia de estos al gas.

Observaciones y resultados

Los astrónomos del Observatorio Astronómico Nacional han combinado observaciones con diversos radiotelescopios y radiointerferómetros (en particular los del IRAM) para caracterizar con gran detalle la composición del gas molecular en M 82. De sus estudios se concluye que, a lo largo de casi un kiloparsec, el disco de M 82 semeja una *PDR gigante* producto de la intensa radiación de las estrellas masivas en el ultravioleta. Los indicios son varios:

Alto grado de ionización. El trazador de gas denso HCO^+ cuenta con un isómero¹⁴, HOC^+ , muy reactivo e inusual. En nubes oscuras HCO^+ es unas mil veces más abundante que su isómero, pues las moléculas de HOC^+ se transforman rápidamente en HCO^+ al reaccionar con el hidrógeno molecular. En regiones moleculares con grados de ionización altos, sin embargo, esta reacción pierde importancia ya que las moléculas de ambas especies se destruyen principalmente al recombinarse con electrones. Al mismo tiempo, en estas regiones ambas especies se producen mayoritariamente mediante reacciones que generan HOC^+ en tanta o mayor proporción que HCO^+ . Como resultado de esto, aumenta la abundancia de HOC^+ , de modo que, por ejemplo, el cociente de abundancias $\text{HCO}^+/\text{HOC}^+$ llega a ser de apenas unas decenas en algunas PDR de nuestra galaxia. Uno de los resultados más destacables de las observaciones de M 82 es que el

¹⁴ Especie compuesta por los mismos átomos pero con una estructura distinta.

cociente $\text{HCO}^+/\text{HOC}^+$ de abundancias en el disco galáctico (40, en promedio) es similar al que se mide en PDR aun cuando aquel es centenares de veces mayor que estas. Mediante el interferómetro PdB se ha obtenido un mapa de la distribución de HOC^+ en M 82 (el primero tanto dentro como fuera de nuestra galaxia; véase la figura 10) que ayuda a comprender de qué manera la propagación del brote de formación estelar ha alterado el estado del medio molecular.

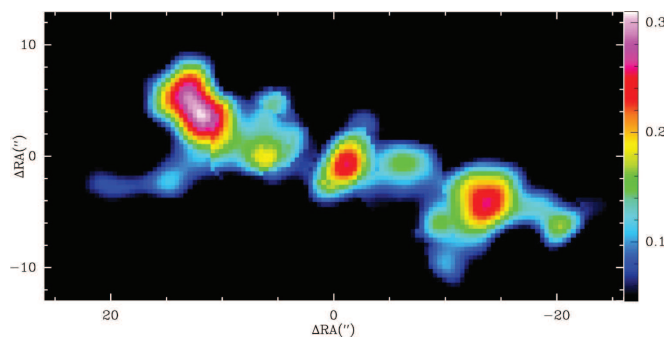


Figura 10: Mapa de la emisión de la línea $\text{HOC}^+(J=1-0)$ (color) en M 82 obtenido con el interferómetro PdB. Los ejes horizontal y vertical de la figura corresponden a los de Ascensión Recta y Declinación respectivamente y su escala está en segundos de arco. A diferencia de la emisión de SiO (véase la figura 9), el grueso de la emisión de HOC^+ , la cual se origina en las PDR inducidas por la radiación ultravioleta, está distribuida a lo largo del disco de la galaxia. Cortesía de A. Fuente.

Alto grado de disociación. El trazador de gas denso HCN libera un átomo de hidrógeno (H) con relativa facilidad al ser irradiado con fotones ultravioleta. Este y otros procesos inducidos por la radiación favorecen que el radical CN sea relativamente abundante en las PDR. Así, el cociente de abundancias CN/HCN crece al aumentar la intensidad de los rayos UV. En el disco de M 82 el cociente de abundancias CN/HCN es ~ 5 en promedio. Valores tan grandes sólo se encuentran en capas de las PDR muy expuestas a la radiación. Al igual que las grandes abundancias de HOC^+ , el cociente CN/HCN medido en M 82 da una idea de hasta qué punto los efectos del brote de formación estelar han modificado las características del medio molecular.

Destrucción de los mantos de los granos. El radical HCO es una especie relativamente infrecuente en el medio interestelar cuya abundancia en PDR llega a ser entre cien y mil veces mayor que en nubes oscuras. Esto se ha atribuido a la liberación y fotodisociación de material orgánico de los

mantos de los granos por los fotones del ultravioleta. Dentro de cada PDR la abundancia de HCO es máxima en las capas intermedias, donde el balance entre su destrucción por fotodisociación y su producción por fotoerosión de los mantos es el más favorable. En las imágenes de alta resolución espacial obtenidas con el interferómetro PdB la emisión de HCO en M 82 se extiende por el kiloparsec central de la galaxia. La abundancia media de HCO estimada en toda esa región es similar a la medida en PDR galácticas.

Fotodisociación de HAP. La facilidad con la que algunos hidrocarburos pequeños son fotodisociados no impide que las abundancias de estos en PDR puedan ser similares a las que se miden en nubes oscuras. Esto sugiere que la radiación abre también una vía adicional para su formación que probablemente sea la fotodisociación de cadenas carbónicas más largas (en particular HAP). En M 82 las abundancias de especies como $\text{c-C}_3\text{H}_2$ y $\text{CH}_3\text{C}_2\text{H}$ son altas en todo el disco central, como cabe esperar si, en efecto, los fotones UV intensifican la destrucción de parte de los HAP.

Rayos X en galaxias activas

Química dominada por rayos X

Se denomina *regiones dominadas por rayos X*, o XDR, a aquellas zonas del medio interestelar en las que la estructura química y el balance térmico están gobernados por rayos X. Su gran energía intrínseca (típicamente en el rango $1 \text{ keV} < h\nu < 100 \text{ keV}$) permite a los rayos X arrancar de la materia una primera generación de electrones muy energéticos que calientan, excitan, ionizan y disocian el gas al interactuar con otras partículas.

La química molecular en las XDR se caracteriza por un incremento de las tasas de ionización, de disociación y de reacciones con electrones. El calentamiento del gas y la excitación de algunas moléculas favorecen a su vez las reacciones que requieren un aporte inicial de energía. Los rayos X pueden además sublimar con gran eficacia los granos silíceos de pequeño tamaño y disociar los HAP más largos. Muchos de los efectos de los rayos X sobre la química del medio molecular se asemejan a los que inducen los rayos ultravioleta en PDR. No obstante, de aspectos fundamentales, como que los rayos X son más destructivos que los rayos UV o que la inyección de energía en las XDR está dominada por los electrones de primera generación y no por los fotones directamente, resultan notables diferencias químicas entre las XDR y las PDR.

Observaciones y resultados

Los núcleos activos son capaces de inyectar ingentes cantidades de energía en las galaxias que los albergan mediante intensos campos de radiación de alta frecuencia. Las regiones dominadas por rayos X (XDR) podrían predominar en las grandes concentraciones de gas molecular que se encuentran en torno a los núcleos activos en forma de *discos circunnucleares* (en

adelante, *DcN*). Hasta la fecha la mayoría de los esfuerzos por encontrar tales regiones XDR fuera de la Vía Láctea se han centrado en la galaxia NGC 1068.

NGC 1068 es una de las galaxias activas mejor estudiadas y un ejemplo paradigmático del funcionamiento de los núcleos activos. NGC 1068 es además una galaxia óptima para cuantificar los efectos de los núcleos activos sobre el medio circundante, pues cuenta con un DcN de gas molecular de ~ 200 pc de radio en el que no hay indicios de formación estelar reciente (el grueso de la formación estelar tiene lugar en un anillo alejado del centro que coincide con el anillo de gas ionizado de la figura 2). Las primeras observaciones de la emisión de líneas moleculares en el DcN de NGC 1068 indicaron que el cociente de abundancias HCN/CO era inusualmente alto. Para explicar esta anomalía se propusieron dos hipótesis que asociaban la peculiar química del DcN con los efectos del núcleo activo. La primera, una escasez inusual de oxígeno en la fase gaseosa y el consiguiente decremento de la abundancia de CO y del resto de especies oxigenadas. La segunda, el aumento relativo de la abundancia de HCN a consecuencia de la iluminación de las nubes moleculares por los rayos X del núcleo activo.

Para poder decidir cuál de los dos escenarios era el más verosímil los astrónomos del Observatorio Astronómico Nacional emprendieron un programa de observaciones del DcN de NGC 1068 con el telescopio PdV. Estas, en combinación con datos obtenidos con el interferómetro PdB, proporcionaron la primera caracterización detallada de la química del gas molecular en torno a un núcleo activo. En suma, se determinaron las abundancias de siete especies moleculares en el DcN: SiO, HCO^+ , HOC^+ , CN, CO, HCN y CS. Las altas abundancias de algunos de los compuestos oxigenados detectados demuestran que el oxígeno no puede ser tan escaso en el DcN como se había predicho. Por contra, los resultados encajan razonablemente con lo predicho por los modelos de química dominada por rayos X. De particular relevancia es la detección de las moléculas oxigenadas HOC^+ y SiO. El cociente de abundancias $\text{HCO}^+/\text{HOC}^+$ en el DcN es uno de los menores medidos en el medio interestelar y sólo es igualado por el encontrado en M 82 y algunas PDR de la Vía Láctea. Un cociente semejante es señal de un alto grado de ionización que en este caso se ha de deber a la radiación del núcleo activo. Por otra parte, la gran abundancia de SiO es señal de destrucción de granos de polvo en la región. Aun cuando la existencia de choques a gran escala no se puede descartar, cabe que los rayos X, al sublimar con gran eficiencia los granos silíceos menores ($\lesssim 10$ Å), hayan incrementado la cantidad de material silíceo en la fase gaseosa. En definitiva, la caracterización de la química molecular en NGC 1068 aporta los primeros indicios firmes de la formación de XDR circunucleares “gigantes” por efecto de la radiación del núcleo activo.

Con el propósito de mejorar la caracterización de la química inducida

por un núcleo activo en su entorno, se ha observado con el interferómetro PdB la emisión de un conjunto de líneas moleculares en el DcN de NGC 1068. Algunas de las imágenes obtenidas se representan en la figura 11. El mapa de HCN muestra cómo el gas molecular denso se distribuye en torno al núcleo activo de la galaxia. Por otra parte, de que los cocientes de líneas dentro del DcN no sean uniformes se deduce que la emisión de las distintas especies no está distribuida de igual modo, lo cual podría ser consecuencia de una composición química inhomogénea. El análisis de la relación entre los cocientes de líneas y la iluminación de los rayos X ayudará a delimitar hasta qué punto regulan estos la composición química en el DcN.

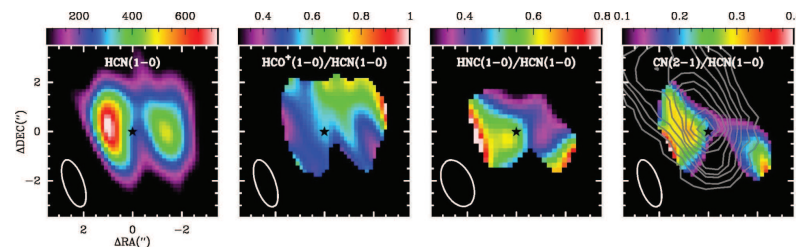


Figura 11: Mapas de la emisión de diversas líneas moleculares en el DcN de NGC 1068 obtenidos por el autor con el interferómetro PdB. En la caja de la izquierda se representa la intensidad total de la línea $\text{HCN}(J=1-0)$ y en las restantes, de izquierda a derecha, el cociente entre la intensidad total de las líneas $\text{HCO}^+(J=1-0)$, $\text{HNC}(J=1-0)$ y $\text{CN}(N=2-1)$ y la intensidad total de $\text{HCN}(J=1-0)$. Los contornos grises en la caja de la derecha indican la emisión de rayos X (imagen obtenida con el telescopio espacial Chandra de la NASA por A. S. Wilson). Una estrella señala la posición de núcleo activo. Los ejes horizontal y vertical de la figura corresponden a los de Ascensión Recta y Declinación, respectivamente, y su escala está en segundos de arco.

Balance del estudio de la química del gas molecular

El análisis de las observaciones indica que la composición química del medio molecular de estas galaxias BFE y activas es notablemente dispar a lo largo de centenares de parsecs. En ninguna de las fuentes se asemeja la composición química a la de las nubes moleculares estándar (nubes oscuras) de nuestra galaxia. De acuerdo con las abundancias moleculares estimadas las peculiaridades químicas se atribuyen al procesamiento del gas molecular por choques a gran escala, en el caso de IC 342 y de NGC 253, por intensos campos ultravioleta, en el de M 82 (también en

el del pequeño anillo circunnuclear de IC 342), y por rayos X, en el de la galaxia activa NGC 1068.

Las observaciones de IC 342, NGC 253 y M 82 aportan indicios convincentes de que, en una galaxia BFE, la composición química del medio molecular y el brote de formación estelar evolucionan a la par. Durante la fase de alimentación de las regiones de formación estelar prevalecen en el disco los choques asociados a los flujos que inducen las barras estelares. Al inicio del brote, la radiación de las estrellas masivas sucede a las barras como agente regulador de la química del medio molecular. Con el envejecimiento progresivo de la población estelar del brote surge una diferencia entre la composición química del disco galáctico, dominada por la radiación estelar, y la del halo, más afectada por los choques asociados a los vientos que inducen las explosiones supernova. La gran resolución angular de las observaciones radiointerferométricas es capital para apreciar la distribución espacial de la emisión molecular e identificar correctamente el estado evolutivo de los brotes de formación estelar.

A la vista de estos resultados se concluye que los principales mecanismos de alimentación y de realimentación que actúan en las galaxias BFE y activas dejan una impronta específica en la constitución del medio molecular a lo largo de centenares de parsecs. Como se muestra en los casos aquí analizados, mediante observaciones combinadas de la emisión de ciertas especies moleculares se puede reconocer dicha impronta.

Conclusión

La caracterización de la química del medio molecular se revela como un prometedor método para el examen y la diagnosis de la formación estelar masiva y de la actividad nuclear en galaxias. En particular, parece posible discernir y valorar los efectos de los brotes de formación estelar y de los núcleos activos mediante observaciones de líneas moleculares. Estudios como los que se han descrito pueden contribuir a la elaboración de diagnósticos certeros de la presencia de núcleos activos en galaxias. Dado que en longitudes de onda milimétricas la radiación no está afectada por la extinción del polvo, tales diagnósticos prometen ser utilísimos para averiguar qué ocultan las grandes concentraciones de materia propias de las galaxias que dominaron la formación estelar en el pasado. Tras la inminente puesta en marcha de los nuevos telescopios internacionales que han de operar en longitudes de onda milimétricas y submilimétricas, como el interferómetro Gran Conjunto de Telescopios de Atacama, ALMA¹⁵, o el telescopio espacial Herschel, en cuya construcción participa el OAN, el estudio de la química molecular puede devenir en una técnica usual y versátil de la astronomía extragaláctica.

¹⁵ Acrónimo del nombre en inglés: Atacama Large Millimeter/submillimeter Array.

Referencias

- A. Fuente, S. García-Burillo, M. Gerin, D. Teyssier, A. Usero et al, *Photon-dominated Chemistry in the Nucleus of M 82: Widespread HOC⁺ Emission in the Inner 650 Parsec Disk*. Astrophysical Journal 2005, 619, L155.
<http://arxiv.org/abs/astro-ph/0412361>
- S. García-Burillo, J. Graciá-Carpio, M. Guelin, R. Neri, P. Cox et al, *A New Probe of Dense Gas at High Redshift: Detection of HCO⁺ (5-4) Line Emission in APM 08279+5255*. Astrophysical Journal, 2006, 645, L17.
<http://arxiv.org/abs/astro-ph/0605656>
- J. Graciá-Carpio, S. García-Burillo, P. Planesas, A. Fuente, & A. Usero, *Evidence of enhanced star formation efficiency in luminous and ultraluminous infrared galaxies*. Astronomy & Astrophysics, 2007, 479, 703.
<http://arxiv.org/abs/0712.0582>
- S. Joglee, *The Fueling and Evolution of AGN: Internal and External Triggers*. Physics of Active Galactic Nuclei at all Scales, editado por Danielle Alloin, Rachel Johnson y Paulina Lira, Lecture Notes in Physics, 2006, 693, 143.
<http://arxiv.org/abs/astro-ph/0408383>
- J. H. Krolik, *Active galactic nuclei: from the central black hole to the galactic environment*. Princeton University Press, 1999.
- S. Martín, R. Mauersberger, J. Martín-Pintado, C. Henkel & S. García-Burillo, *A 2 Millimeter Spectral Line Survey of the Starburst Galaxy NGC 253*. Astrophysical Journal Supplement Series, 2006, 164, 450.
- A. Usero, S. García-Burillo, J. Martín-Pintado, A. Fuente, & R. Neri, *Large-scale molecular shocks in galaxies: the SiO interferometer map of IC 342*. Astronomy & Astrophysics, 2006, 448, 457.
<http://arxiv.org/abs/astro-ph/0510290>
- A. Usero, S. García-Burillo, A. Fuente, J. Martín-Pintado, & N. J. Rodríguez-Fernández, *Molecular gas chemistry in AGN. I. The IRAM 30 m survey of NGC 1068*. Astronomy & Astrophysics, 2004, 419, 897.
<http://arxiv.org/abs/astro-ph/0402556>

La evolución de galaxias con brotes y activas

El artículo **El medio molecular en galaxias con brotes de formación estelar y en galaxias activas** se publicó en el Anuario del año 2009. En los años posteriores, el estudio del gas molecular en el Espacio se ha beneficiado del progreso de la instrumentación astronómica que opera en longitudes de onda (sub)milimétricas. Un hito fundamental fue la puesta en marcha en 2011 del radiointerferómetro internacional ALMA, uno de los telescopios más revolucionarios de este siglo. Gracias a su emplazamiento a 5000 metros de altitud en Chile, a las 66 antenas que lo componen y a la distancia máxima entre ellas, ALMA alcanza una sensibilidad y una resolución angular sin parangón. Paralelamente, el desarrollo técnico ha permitido que algunos de los instrumentos ya existentes en 2009 sigan siendo competitivos en la era de ALMA. Dos claros ejemplos son los telescopios del Instituto franco-hispano-alemán de Radio-Astronomía Milimétrica (IRAM), en el cual participa el Instituto Geográfico Nacional: el radiointerferómetro de Plateau de Bure (cuya ampliación se denomina NOEMA) y el radiotelescopio de 30 metros, en Sierra Nevada, que ha incorporado receptores cada vez más sofisticados y ha octuplicado el ancho de banda instantáneo disponible en 2009. Esto facilita la observación simultánea de distintas líneas moleculares. Cabe destacar también que el Observatorio Espacial Herschel, operativo de 2009 a 2013, exploró el rango de ondas submilimétricas y del infrarrojo lejano, cuya observación desde tierra es difícil o casi imposible por efectos atmosféricos.

En los últimos años, nuestra imagen de la evolución de las galaxias a lo largo de la historia de Universo se ha vuelto más nítida. Se sabe que, desde hace al menos diez mil millones de años¹, el $\sim 90\%$ de las nuevas estrellas han surgido en una *secuencia principal* de galaxias cuya tasa de formación estelar específica² es muy similar en cada época. Se cree que el crecimiento de estas galaxias se debe a la caída continua de gas intergaláctico y a la fusión con galaxias más pequeñas. El resto de estrellas se origina fundamentalmente en un conjunto menor de galaxias (galaxias *starburst*) cuyas tasas de formación estelar específica son significativamente mayores. Probablemente, la mayoría de estos objetos con formación estelar intensa correspondan a interacciones violentas entre galaxias. Las observaciones de líneas moleculares en objetos de diversas épocas son imprescindibles

¹Desplazamiento al rojo (z) menor que 2.

²Cociente entre la tasa de formación de estrellas y la masa estelar total.

para comprender las diferencias entre estas dos poblaciones de galaxias, puesto que las estrellas han nacido en nubes de gas molecular durante la mayor parte de la historia del Universo. Los astrónomos han detectado ya líneas moleculares en cientos de galaxias del Universo *lejano*³. Los últimos resultados indican que, en cada época, el gas molecular de las galaxias de la secuencia principal forma estrellas con una eficiencia muy similar, mientras que la eficiencia es sustancialmente mayor en las galaxias *starburst*. Aunque las líneas de monóxido de carbono (CO) siguen siendo el principal instrumento para conocer el medio molecular en todas las épocas, las líneas de otras especies (HCN o agua, por ejemplo), se detectan cada vez con más frecuencia, incluso en los objetos más distantes⁴. Esto redonda en la medida de las propiedades físicas y químicas del gas.

Como sugiere el párrafo anterior, la eficiencia de formación estelar del gas molecular es un parámetro fundamental de la evolución galáctica. Actualmente, diversos grupos de astrónomos observadores y teóricos tratan de determinar hasta qué punto dicha eficiencia depende de las propiedades de las nubes de gas y de las del entorno en el que residen. Los radiotelescopios actuales facilitan la obtención de imágenes de gran resolución espacial, con las que estudiar nubes individuales, y/o de gran campo, con las que estudiar entornos distintos en una misma galaxia. Gracias a ellas, se cree que la relación entre eficiencia y densidad del gas es más sensible a otros parámetros (turbulencia, presión, etc.) de lo que se pensaba hace apenas unos años.

Gracias a ALMA, los astrónomos están al fin estudiando el último paso de otro proceso trascendental de la evolución galáctica: el crecimiento, a partir del gas circundante, de los agujeros negros supermasivos que residen en los centros galácticos. Dicho proceso da pie al fenómeno de la *actividad nuclear*, por el cual el entorno de un agujero negro llega a inyectar grandes cantidades de energía en la galaxia que lo alberga. Recientemente, ALMA ha proporcionado las primeras imágenes del gas molecular del toro que circunda el agujero negro de una galaxia *activa*. Los toros circunnucleares son un elemento clave de los modelos teóricos de los núcleos activos. Pronto será posible estudiar cómo fluye el gas desde ellos hacia su interior.

Por último, un resultado reciente y de gran impacto es el descubrimiento de vientos de gas molecular en un número creciente de galaxias *starburst* y *activas*. Estos vientos, que son un ingrediente capital de los modelos teóricos de evolución galáctica, surgen por la eyección de energía y materia en los brotes de formación estelar y desde los núcleos activos. Si bien los vientos de gas ionizado y atómico ya eran conocidos, el estudio de su fase molecular es clave para medir su impacto en las galaxias.

Antonio Usero Villanueva, mayo 2017

³ $z > 1$, lo cual equivale a los primeros seis mil millones de años del Universo.

⁴Se han detectado líneas moleculares hasta $z \sim 6 - 7$ (primeros mil millones de años).

LA EXPANSIÓN ACELERADA DEL UNIVERSO

Mario Tafalla

*Observatorio Astronómico Nacional
Instituto Geográfico Nacional - Ministerio de Fomento*

Abstract

The 2011 Nobel prize in physics recognized the discovery of the accelerating expansion of the Universe. Here we review at an introductory level how this discovery was possible, and follow the study of the Hubble diagram from the 1920s to the latest supernova results.

Introducción

“Por el descubrimiento de la expansión acelerada del universo a partir de observaciones de supernovas lejanas”¹. Con estas palabras, la Real Academia Sueca de Ciencias anunciaba la concesión del premio Nobel de física de 2011 a tres astrónomos de origen estadounidense: Saul Perlmutter (medio premio), Brian P. Schmidt y Adam G. Riess (conjuntamente el otro medio). Los tres astrónomos, junto con numerosos colaboradores en universidades y centros de investigación de América, Europa y Australia, son responsables de uno de los descubrimientos más revolucionarios de las últimas décadas: la expansión acelerada del universo.

La idea de que el universo se está expandiendo no es nueva. Fue establecida observacionalmente durante los años veinte del siglo pasado con el descubrimiento de una correlación entre el desplazamiento al rojo de la luz de las galaxias lejanas y la distancia a que éstas se encuentran. Esta correlación indica que las galaxias del universo se separan de

¹http://www.nobelprize.org/nobel_prizes/physics/laureates/2011/

nosotros sistemáticamente siguiendo un patrón simple: cuanto más lejana es una galaxia, mayor es su velocidad de separación. Como resultado, el cociente entre la velocidad de alejamiento y la distancia de una galaxia es aproximadamente el mismo para todas las galaxias del universo. El valor de esta constante constituye un parámetro fundamental del universo, y determinar con precisión su valor fue uno de los objetivos principales de la cosmología observacional del siglo XX. El trabajo de los grupos de Perlmutter por un lado y de Schmidt y Riess por otro, representa la culminación de esta línea de trabajo, con la particularidad del uso de observaciones de supernovas para extender la medida de distancias a los valores más altos, y por tanto, a los tiempos más remotos.

El resultado de Perlmutter, Schmidt, Riess y colaboradores no pudo ser más sorprendente. En vez de observar que la expansión del universo se ha ido decelerando con el tiempo, como era esperado por el efecto gradual de la atracción gravitacional entre los distintos elementos materiales del universo, estos investigadores observaron una débil pero significativa tendencia a la aceleración (Riess et al. 1998, Perlmutter et al. 1999). Este sorprendente resultado ha sido confirmado y refinado por numerosos investigadores que han participado en un sistemático esfuerzo observacional durante la última década y media. La concesión del premio Nobel a los principales responsables del descubrimiento muestra la confianza de la comunidad científica en un resultado observacional revolucionario y todavía misterioso. Los datos nos indican que el universo está acelerando su expansión. La naturaleza del agente causante de esa aceleración, que es tan poderoso como para dominar la evolución del universo, constituye uno de los misterios más fascinantes de la física del siglo XXI.

La escalera de distancias cósmicas

Desde un punto de vista observacional, la evidencia de que el universo se está expandiendo viene de la existencia de una relación de proporcionalidad entre la velocidad a la que las galaxias se alejan de nosotros y la distancia a la que se encuentran. Medir esta relación requiere determinar con precisión dos parámetros físicos muy distintos, la velocidad y la distancia de una galaxia.

Medir con precisión la velocidad a la que las galaxias se alejan de nosotros es relativamente sencillo. La luz que éstas emiten cambia ligeramente de frecuencia dependiendo de su movimiento respecto a nosotros. Este cambio es a veces atribuido al llamado “efecto Doppler”, bien conocido en el laboratorio y denominado así en honor al físico austríaco Christian Doppler (1803-1853). En realidad, el cambio en frecuencia de la luz de las galaxias lejanas es un efecto relativista asociado

a la expansión del espacio que nos separa de ellas. En una sección posterior discutiremos este efecto con más detalle. Por ahora, lo importante es que el efecto es medible con gran precisión comparando la frecuencia a la que se observan ciertas líneas espectrales de elementos químicos presentes en las galaxias con la frecuencia de las mismas líneas medida en el laboratorio. Típicamente, esta es una medida en la que se pueden alcanzar precisiones de varias cifras significativas, salvo en los objetos más lejanos, cuya luz está al límite del nivel de detección de la generación actual de telescopios.

Medir la distancia a la que están las galaxias es mucho más complejo. Requiere un método indirecto, basado en una cadena de suposiciones y correcciones, cada una de ellas sujeta a distinto grado de incertidumbre. Los astrónomos denominan este método la “escalera de distancias cósmicas”, pues se asemeja a una secuencia de escalones de altura creciente. Los escalones más bajos son las determinaciones de distancias a objetos cercanos, que pueden hacerse con métodos geométricos de gran precisión, como la paralaje. Estos métodos, solo pueden aplicarse a objetos dentro de nuestra galaxia, y por tanto no sirven para hacer medidas cosmológicas. Usándolos, sin embargo, es posible medir distancias a tipos especiales de estrellas, cuyo brillo intrínseco puede ser así calibrado y usado posteriormente como una “candela estándar” para medir distancias mayores.

Entre las diferentes candelas estándar, destacan las estrellas variables tipo RR Lyrae, y sobre todo, las variables cefeidas, que son fáciles de identificar por sus cambios regulares de brillo a lo largo de días o semanas. Las variables cefeidas son estrellas supergigantes, y su gran brillo las hace observables a distancias extragalácticas, por lo que tradicionalmente han sido usadas para estudiar la expansión del universo. El uso de cefeidas como indicadores de distancia está basado en la existencia de una relación casi universal entre su periodo, luminosidad y color (PLC). Esta relación, inicialmente sólo entre el periodo y la luminosidad, fue descubierta por la astrónoma estadounidense Henrietta Swan Leavitt (1868-1921) (Leavitt 1908), y permite usar dos medidas observacionales de una estrella (periodo y color) para determinar su luminosidad intrínseca. Conocida la luminosidad intrínseca de la estrella, es posible inferir su distancia dado el brillo aparente con que se observa.

A pesar de su simplicidad, la medida de distancias usando cefeidas no está exenta de dificultades. Por una parte, es necesario primero calibrar la relación PLC, y esto requiere disponer de un número elevado de cefeidas de distancia bien determinada. Al ser las cefeidas estrellas supergigantes, son relativamente raras, y el número de ellas a distancias suficientemente pequeñas como para poder aplicar técnicas geométricas de paralaje es muy limitado. Además, las cefeidas son estrellas relativamente jóvenes (unos cien millones de años), y por tanto suelen estar todavía rodeadas de gas y polvo que absorbe su luz y las hace parecer más débiles y

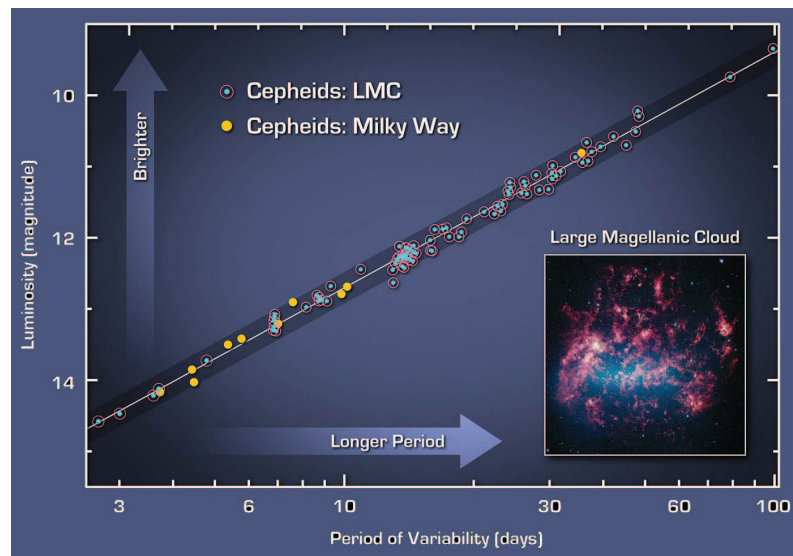


Figura 1: Relación periodo-luminosidad de las variables cefeidas. El eje horizontal indica el periodo en días y el vertical la luminosidad en magnitudes. La figura representa tanto cefeidas en la Vía Láctea (“Milky Way”) como en la Gran Nebulosa de Magallanes (LMC). Midiendo el periodo de una cefeida y usando esta relación es posible determinar su luminosidad absoluta. Comparando ésta con el brillo observado es posible determinar la distancia. Imagen cortesía NASA, JPL-Caltech y W. Freedman.

rojas. Este efecto necesita ser corregido si la relación PLC quiere aplicarse correctamente. Finalmente, diferencias en la metalicidad de las cefeidas de distintas galaxias pueden introducir desviaciones sistemáticas en la relación PLC (ver Freedman & Madore 2010 para una discusión detallada).

Un método alternativo, más simple pero de menos precisión, es el uso de las estrellas más brillantes de una galaxia. La idea detrás de este método es que aunque las estrellas de una galaxia presentan un rango muy amplio de luminosidades, el escoger la estrella más brillante, o el grupo de cierto número N de estrellas más brillantes, selecciona una población similar en todas las galaxias. Estas estrellas pueden por tanto usarse como candelas estándar. Algunas de las primeras determinaciones de distancias hechas para inferir la expansión del universo usaban este método, que desgraciadamente es muy sensible a errores sistemáticos. Una versión más precisa usa un criterio de color para seleccionar sólo estrellas en la rama de las gigantes rojas.

A distancias mayores de unos 20 Mpc (20 megaparsecs, o unos 65 millones de años luz²), no es posible usar el método de las cefeidas o de la estrella más brillante debido a la dificultad de resolver estrellas individuales y eliminar la contribución a su luminosidad de estrellas vecinas. Por ello, la escalera de distancias cósmicas requiere un peldaño adicional. Este peldaño lo proporcionan ciertas relaciones empíricas, como la de Tully-Fisher, que liga la luminosidad total de una galaxia espiral con la velocidad de rotación de sus estrellas, o la relación de Faber-Jackson, que relaciona de forma análoga la luminosidad total de una galaxia elíptica con la velocidad de dispersión de sus estrellas. Con estas relaciones, es posible de nuevo estimar distancias comparando el brillo observado de una galaxia con la estimación del brillo intrínseco basada en una medida de la cinemática de sus estrellas. De esta manera, es posible estimar distancias de galaxias hasta unos 100 Mpc (unos 325 millones de años luz).

El peldaño final de la escalera de distancias cósmicas lo aporta la observación de supernovas, en especial las de tipo Ia. El desarrollo de este método es lo que permitió a los grupos de Perlmutter, Schmidt y Riess extender el estudio de la expansión del universo hasta distancias en las que la aceleración es detectable. Dado su interés en el proceso del descubrimiento de la expansión acelerada del universo, estudiaremos este método con más detalle en una sección posterior.

¿Quién descubrió la expansión del universo?

Frecuentemente se atribuye al astrónomo estadounidense Edwin Hubble (1889-1953) el descubrimiento de la expansión del universo, dada la importancia y calidad de sus estudios de la relación velocidad-distancia de galaxias. Por ello, la relación de proporcionalidad entre la velocidad y la distancia de las galaxias se denomina comúnmente “ley de Hubble”, y el valor de la constante que las relaciona “constante de Hubble”. Sin embargo, la historia del descubrimiento de la expansión del universo es más compleja que la aportación genial de un individuo, ya que fueron varios los astrónomos que contribuyeron tanto o más que Hubble a lo que sin duda es uno de los resultados más importantes de la ciencia contemporánea. Por ello, recientemente han surgido voces que intentan recuperar la memoria de los distintos autores de esas contribuciones, y que pretenden distribuir el crédito por este descubrimiento de una forma más equitativa (Nussbaumer et al. 2009).

²La unidad de distancia en astronomía es el parsec (abreviado pc), que equivale a 3,26 años luz o unos $3,1 \cdot 10^{18}$ cm, y corresponde aproximadamente a la distancia a las estrellas más cercanas. Las distancias entre galaxias se suelen medir en megaparsecs (Mpc), es decir, millones de parsecs.

Uno de los astrónomos que con sus medidas contribuyó de manera fundamental al descubrimiento de la expansión del universo fue el estadounidense Vesto M. Slipher (1875-1969), del Observatorio Lowell en Arizona. Vesto Slipher fue pionero en observaciones espectroscópicas de “nebulosas espirales”, que ahora sabemos corresponden a galaxias como la Vía Láctea, pero cuya naturaleza era todavía discutida en 1920 (año del famoso debate Curtis-Shapley). A finales de 1912, Slipher realizó la primera medida de la velocidad radial de una nebulosa espiral, la de Andrómeda, la más cercana de las galaxias parecidas a la Vía Láctea, que se encuentra a unos 0,8 Mpc. De esta medida, Slipher concluyó que esta nebulosa se está acercando a nosotros con una velocidad de 300 km/s, valor que está de acuerdo con estimaciones modernas. Durante los siguientes años, Slipher midió un buen número de velocidades radiales de nebulosas espirales, y en 1917 presentó una compilación de sus resultados (Slipher 1917). De un total de 25 nebulosas, Slipher encontró que 21 se alejan de nosotros, mientras que tan sólo 4 (una de ellas Andrómeda) se nos acerca. Esta preponderancia de velocidades de recesión (velocidades positivas en notación astronómica) era notable. Desgraciadamente, sin una idea clara de la naturaleza de las nebulosas, debido al desconocimiento de su distancia, la interpretación de esta tendencia observacional no resultaba todavía posible.

A la vez que Slipher realizaba sus observaciones espectroscópicas de nebulosas espirales en Norteamérica, el físico alemán Albert Einstein (1879-1955) desarrollaba su teoría de la Relatividad General en Europa. Esta teoría resulta de aplicar los conceptos relativistas desarrollados por el propio Einstein en su teoría de la Relatividad Especial (Einstein 1905) a la fuerza de la gravedad, hasta entonces interpretada con la teoría de Newton. La nueva teoría de Einstein (Einstein 1916) describe la atracción mutua entre los objetos del universo como el resultado de la curvatura del espacio-tiempo, que es producida por la mera presencia de estos objetos. El espacio y el tiempo para Einstein, son por tanto elementos susceptibles de ser alterados por el contenido material y energético, pues materia y energía son equivalentes.

En una de sus primeras aplicaciones de la teoría de la Relatividad General, Einstein considera las implicaciones cosmológicas de sus nuevas ecuaciones (Einstein 1917). Al hacerlo, se encuentra con el problema de que la única fuerza a gran escala en estas ecuaciones es la fuerza de la gravedad, y sólo con esta fuerza atractiva es imposible conseguir una solución de universo estática, que es el tipo de solución que se consideraba natural en 1917. Por ello, Einstein modifica sus ecuaciones, e introduce un nuevo término en ellas, la denominada *constante cosmológica*. Este nuevo término no tiene analogía en la mecánica newtoniana, pues representa una fuerza a gran escala que se opone a la gravedad. Sin embargo, el término está presente en la forma más general de las ecuaciones. Además, permite la solución estática que Einstein buscaba.



Figura 2: Algunos de los principales protagonistas del estudio de la expansión del universo. Fila superior: Vesto M. Slipher, Albert Einstein y Georges Lemaître. Fila del medio: Edwin Hubble, Allan R. Sandage y Gérard de Vaucouleurs. Fila inferior: Saul Perlmutter, Brian P. Schmidt y Adam G. Riess. Fotos cortesía Wikimedia y Fundación Nobel (U. Montan).

Como se descubrió pocos años después, la solución estática de Einstein presenta varios problemas. Uno de ellos es que es inestable (Eddington 1930). Por ejemplo, si se produce una ligera expansión en cierto punto del universo, esta expansión disminuirá la densidad de materia, y esto a su vez disminuirá la contribución de la gravedad frente a la constante cosmológica. Como resultado, la expansión inicial aumentará, y la densidad de materia del punto disminuirá aún más, amplificando un ciclo que acabará por desbocarse. Lo mismo sucedería con una pequeña contracción. Por otra parte, las observaciones iniciadas por Slipher y continuadas por Hubble y otros harán inevitable la conclusión de que el universo se está expandiendo. Con ello desaparece la motivación inicial de introducir la constante cosmológica en las ecuaciones de la Relatividad General.

Uno podría pensar que la inconsistencia de la solución estática de Einstein y el descubrimiento de la expansión del universo marcan el fin de la constante cosmológica. Como veremos, la historia es más compleja. Desde un punto de vista lógico, la constante cosmológica es parte natural de las ecuaciones de la Relatividad General, y sólo una medida observacional de alta precisión podría demostrar que su valor es nulo. Por otra parte, el descubrimiento de que la expansión del universo se está acelerando ha renovado el interés en la constante cosmológica. La aceleración del universo requiere la presencia de una fuerza adicional de expansión, y la constante cosmológica representa una posible solución.

Un tercer trabajo con importantes implicaciones cosmológicas aparece también en 1917. Pocos meses después de que Einstein introdujera el concepto de constante cosmológica, el astrónomo holandés Willem de Sitter (1872-1934) propone un modelo cosmológico alternativo, en el que esta constante juega un papel fundamental. La solución de de Sitter (1917) no contiene materia, y por tanto representa un universo dominado por la constante cosmológica. Como el de Einstein, este universo es considerado como estático, pero tiene la peculiaridad de que si en él se introducen partículas de prueba, éstas se separan aceleradamente por efecto de la constante cosmológica. Una predicción de este modelo es la preponderancia de movimientos de recesión de los objetos más lejanos que, como de Sitter reconoce en su trabajo, en el momento de su publicación se limita a sólo dos casos entre tres medidas de nebulosas espirales. Este “efecto de Sitter” gozará de cierta popularidad durante los años veinte, cuando las nuevas observaciones empiecen a mostrar una clara mayoría de galaxias alejándose de nosotros.

Una familia alternativa de soluciones de las ecuaciones de la Relatividad General fue propuesta pocos años después por el matemático ruso Alexander Friedmann (1888-1925). Las soluciones de esta familia no son estáticas, como las de Einstein y de Sitter, sino que representan un universo que varía con el tiempo, e incluyen la posibilidad de un estado global de

expansión (Friedmann 1922). Desgraciadamente, este trabajo pionero no parece haber sido reconocido como tal en su tiempo. Einstein mismo afirmó primero que era incorrecto matemáticamente, y tardó varios meses en darse cuenta de su error al criticarlo. La muerte prematura de Friedman tres años después contribuyó a su relativo olvido.

En el campo observacional, el problema de la naturaleza y distancia de las nebulosas espirales queda finalmente resuelto en 1925, cuando Hubble anuncia el descubrimiento de cefeidas en la nebulosa de Andrómeda. Aunque ya había indicaciones previas, basadas en observaciones de novae, de que Andrómeda estaba a tal distancia que tenía que ser un objeto externo a la Vía Láctea, el descubrimiento de cefeidas en Andrómeda (Hubble 1925) zanja la disputa definitivamente y abre el camino a la medida de distancias extragalácticas usando cefeidas. Desgraciadamente, ahora sabemos que hay dos tipos diferentes de variables cefeidas, y que tienen una luminosidad muy distinta. Las llamadas cefeidas de tipo I (o cefeidas clásicas) son estrellas relativamente jóvenes, de unos 100 millones de años, que se observan normalmente en los brazos espirales de galaxias. Las cefeidas de tipo II (también llamadas variables W Virginis) son unas estrellas 100 veces más viejas y 4 veces más débiles que las cefeidas de tipo I, y se observan típicamente en cúmulos globulares. Las cefeidas descubiertas por Hubble en Andrómeda y otras galaxias eran de tipo I (las más brillantes), pero la relación periodo-luminosidad usada para determinar la distancia estaba basada en una mezcla de cefeidas de ambos tipos, y por tanto mal calibrada. Como resultado de este error, la distancia a Andrómeda se subestima significativamente (0,3 Mpc en vez de 0,8 Mpc), problema que se repite con la medida a otras galaxias durante los primeros años de la aplicación del método. La diferencia entre los dos tipos de cefeidas no será entendida, y el error de las distancias corregido, hasta varias décadas después, gracias al trabajo del astrónomo alemán Walter Baade (1893-1960) (Baade 1956).

La primera persona que parece haber conectado correctamente las medidas de las nebulosas espirales con el marco de la Relatividad General fue el astrónomo y sacerdote belga Georges Lemaître (1894-1966). En 1927, Lemaître publica, en francés y en una revista relativamente oscura, un artículo titulado “Un universo homogéneo de masa constante y radio creciente que explica las velocidades radiales de las nebulosas extragalácticas” (Lemaître 1927). En este artículo, Lemaître deduce una solución de las ecuaciones de la Relatividad General que representa un universo alternativo a los de Einstein y de Sitter, que considera incompletos, y que es equivalente a la solución propuesta por Friedmann en 1922 (que Lemaître parece desconocer en ese momento). Además de presentar esta solución de universo, Lemaître utiliza valores de las velocidades de galaxias espirales de la literatura (en gran medida debidas a Slipher) y estimaciones de la distancia a partir de trabajo previo de Hubble.

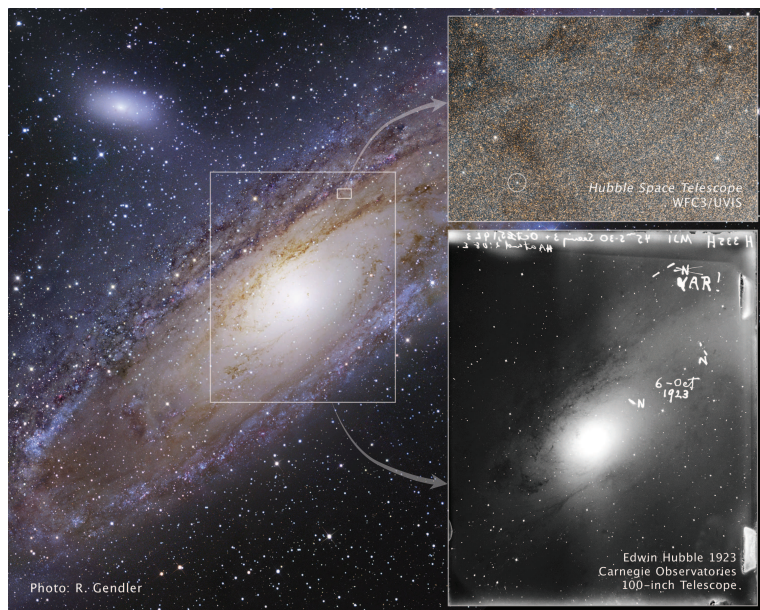


Figura 3: Galaxia de Andr meda y su primera variable cefeida. El panel de la izquierda es una imagen a gran escala de la galaxia de Andr meda. El panel inferior derecho muestra la imagen en blanco y negro tomada por Edwin Hubble el 6 de octubre de 1923. Hubble inicialmente identific  tres novae, y las marc  con una letra N. M s tarde se di  cuenta de que una de las “novae” era en realidad una variable cefeida, la primera descubierta en Andr meda. Por ello corrigi  su anotaci n de “N” a “VAR!”. Este descubrimiento, publicado en 1925, demostr  que Andr meda es una galaxia externa a la V a L ctea. El panel superior derecho muestra una imagen moderna del telescopio espacial Hubble con la estrella cefeida dentro de un c rculo. Imagen cortes a NASA/ESA y Hubble Heritage Team.

Con estos valores Lema tre deriva, entre otros par metros, una tasa de expansi n del universo (que ahora llamamos “constante de Hubble”) de unos 600 (km/s)/Mpc.

Dada la importancia de estos resultados, hoy resulta sorprendente que el trabajo de Lema tre fuera en gran medida ignorado. Probablemente contribuy  a ello el hecho de su publicaci n en una revista de menor importancia que estaba escrita en franc s en vez del ingl s o alem n dominantes en el trabajo cient fico de la  poca. Otro factor importante fue la poca aceptaci n que en aquel momento ten a la idea de un universo

en expansi n. El mismo Einstein, en conversaci n privada con Lema tre, calific  tal tipo de universo como de “abominable” (Luminet 2013).

Tres a os despu s de la publicaci n del trabajo de Lema tre, en una reuni n de la Royal Astronomical Society, el astr nomo brit nico Arthur Eddington (1882-1944) sugiere p blicamente la posibilidad de usar un modelo de universo en expansi n para explicar el desplazamiento al rojo sistem tico de la luz de las nebulosas espirales (Luminet 2013). Lema tre parece haber leido el informe de la reuni n y escribe a Eddington, que hab a sido profesor suyo en Cambridge, para recordarle los resultados de su art culo de 1927. Eddington, que parece haber olvidado el trabajo de Lema tre, se da cuenta que  ste ha resuelto el problema que  l hab a planteado, y que el trabajo de 1927 ha sido pr cticamente ignorado. Por ello, invita a Lema tre a preparar una versi n revisada de su trabajo original para publicarla en la revista brit nica *Monthly Notices of the Royal Astronomical Society*, de mucha mayor difusi n e impacto. El nuevo art culo se publica en 1931 (Lema tre 1931), y es una traducci n al ingl s con muy pocos cambios del art culo original en franc s de 1927.

El  nico cambio significativo en el art culo de 1931 es una ausencia de la discusi n original sobre la proporcionalidad entre la velocidad de expansi n de las galaxias y su distancia, lo que hoy llamamos ley de Hubble. Esta ausencia, ahora tan notable, ha sido calificada de sospechosa por algunos autores, que han llegado a sugerir la existencia de una conspiraci n para despojar a Lema tre del cr dito que su trabajo merec a (van den Bergh 2011). Sin embargo, la correspondencia original entre Lema tre y la revista, descubierta recientemente por Livio (2011), muestra que fue el propio Lema tre el responsable de la traducci n al ingl s del art culo y de la eliminaci n del p rrafo sobre la relaci n velocidad-distancia. Aunque Lema tre no lo dice expl citamente en su correspondencia, parece que en 1931  l ya pensaba que los datos de su discusi n de 1927 hab an sido mejorados por el trabajo de Hubble de 1929 (ver m s abajo), y que por tanto, la discusi n original de la relaci n velocidad-distancia no era necesaria en la nueva versi n en ingl s. Esta falta de pretensi n de Lema tre por recibir cr dito de lo que ahora consideramos un descubrimiento fundamental de la cosmolog a puede verse en algunas de sus declaraciones y publicaciones posteriores.

Si el trabajo de Lema tre pas  desapercibido, el de Hubble no hizo lo mismo. En 1929, Hubble combina medidas de velocidad de galaxias (de nuevo en su mayor a debidas a Slipher pero tomadas de una compilaci n de Eddington), con estimaciones propias de distancia para demostrar gr ficamente la existencia de una relaci n de proporcionalidad entre velocidad y distancia (Hubble 1929). Aunque observadores anteriores hab an buscado evidencia de relaciones velocidad-distancia, en gran medida motivados por la b squeda del llamado “efecto de Sitter” (Wirtz 1924, Lundmark 1925), el trabajo Hubble de 1929 es el primero que

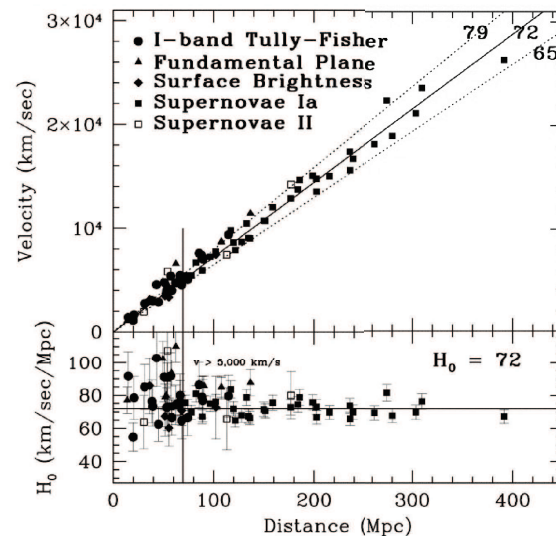
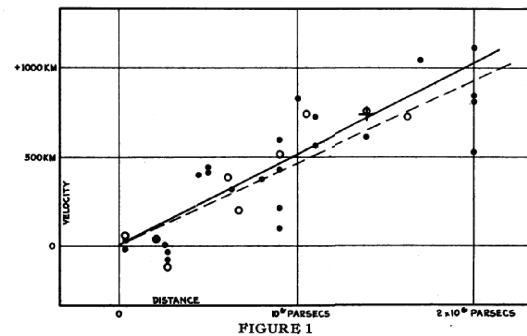


Figura 4: Dos versiones del diagrama velocidad-posición separadas más de 70 años. *Arriba*: diagrama original presentado por Hubble (1929). *Abajo*: Versión moderna obtenida por Freedman et al. (2001) usando datos del telescopio espacial Hubble. Nótese como la versión moderna extiende el rango original de distancias por dos órdenes de magnitud. Todos los puntos de Hubble (1929), si bien calibrados, caerían dentro de la primera subdivisión de eje de distancias en el segundo diagrama. Cortesía de The Astrophysical Journal.

muestra una tendencia significativa en los datos. Con sus valores, Hubble estima una constante de proporcionalidad de 500 (km/s)/Mpc, no muy distinta a la deducida por Lemaître unos años antes. A diferencia de Lemaître, Hubble no interpreta la relación en términos de un universo en expansión. Dadas todas las incertidumbres en las estimaciones, Hubble es muy cauto con los resultados, y considera “premature discutir en detalle las obvias consecuencias de los resultados presentes”. Aun así, apunta el interés de la relación velocidad-distancia en términos del “efecto de Sitter”, en lo que entonces se consideraba un modelo de universo estático.

Aunque Hubble no apreció todas las implicaciones cosmológicas de la relación velocidad-distancia, sí entendió la importancia del resultado. Por ello, continuó trabajando sistemáticamente para mejorar su determinación observacional. Pocos años más tarde, él y Milton L. Humason (1891-1972) publican un nuevo análisis de la relación velocidad-distancia, esta vez usando velocidades y distancias determinadas por ellos mismos (Hubble & Humason 1931). Con estos nuevos datos, que extienden el estudio de 1929 a distancias 18 veces mayores, Hubble y Humason estiman una constante de proporcionalidad de 538 (km/s)/Mpc, y sugieren un nivel de incertidumbre de tan solo un 10 %.

Como veremos enseguida, la estimación del nivel de incertidumbre de Hubble y Humason era extraordinariamente optimista. Ahora sabemos que su constante de proporcionalidad estaba sobreestimada por un factor 7 debido a un error sistemático en la determinación de distancias extragalácticas. Este error es enorme incluso para los estándares habitualmente generosos de las medidas astronómicas. A pesar de ello, el trabajo de Hubble y Humason (1931) puede considerarse la obra definitiva de la época pionera en la determinación de la relación velocidad-distancia. A partir de este momento, la existencia de una relación velocidad-distancia se considera firmemente establecida, y la determinación exacta de esta constante pasa a ser un importante objetivo observacional. Alcanzar la precisión de 10 % mencionada por Hubble y Humason, sin embargo, requerirá décadas de trabajo y avances en muy diversas ramas de la astronomía.

Del astrónomo Hubble al telescopio espacial Hubble

Aunque en la actualidad consideramos la existencia de una relación velocidad-distancia como la prueba definitiva de la expansión del universo, ese no era el caso en los años treinta del siglo XX. Por entonces, la relación se consideraba como un hecho observacional probado, pero su interpretación en términos cosmológicos era para algunos incierta. El propio Hubble mantuvo hasta su muerte una posición ambigua sobre su

significado, y todavía se discute si de verdad creía que el universo está en un estado de expansión (Kragh & Smith 2003).

Uno de los problemas que planteaba la interpretación de la relación velocidad-distancia en términos de un universo en expansión era que, de acuerdo con la teoría de la Relatividad General, el inverso de la constante de Hubble representa la edad del universo si la constante cosmológica es nula (como ya aceptaban los principales teóricos: Einstein & de Sitter 1932). Al valor de 500 (km/s)/Mpc derivado por Hubble y otros le corresponde una edad del universo de unos 2 mil millones de años, que ya en los años treinta se consideraba demasiado pequeña. Por entonces, medidas a partir de cocientes isotópicos sugerían que la edad de la Tierra era de varios miles de millones de años. Una Tierra más vieja que el resto del universo representa una clara contradicción.

El problema de que la constante de Hubble implique una edad del universo inferior a la edad de objetos astronómicos como la Tierra o los cúmulos globulares más antiguos (con edades de unos 10 mil millones de años) ha acompañado gran parte de la historia de su determinación durante el siglo XX. Un universo más viejo requiere una constante de Hubble más pequeña, y afortunadamente, la medida de la constante de Hubble ha sufrido con el tiempo una serie de correcciones a la baja.

La primera gran corrección importante de la constante de Hubble se debe a la ya mencionada recalibración de la escala de distancias hecha por Baade en los años cincuenta al reconocer la existencia de dos tipos de estrellas cefeidas. Esta recalibración aumenta automáticamente la distancia a las galaxias en aproximadamente un factor 3, disminuyendo el valor de la constante de Hubble por el mismo valor. Este resultado coincide aproximadamente con la muerte de Hubble en 1953, por lo que el trabajo de recalibrar la constante con la nueva escala de distancias lo llevan a cabo sus colaboradores y discípulos. En 1956, Humason, Mayal y Sandage publican los resultados de un programa observacional muy ambicioso comenzado por Hubble, quien, como dicen los autores, “de haber vivido, hubiera participado como autor senior en el análisis y la discusión”. Este monumental trabajo muestra que los objetos que Hubble y otros habían considerado como estrellas individuales en galaxias, y que habían sido la base para estimar distancias usando el método de las estrellas más brillantes, eran en realidad cúmulos densos o regiones HII (gas ionizado por estrellas masivas). Combinando la corrección de la escala de distancias con la determinación de velocidades de más de 800 galaxias, Humason, Mayal y Sandage (1956) estiman un nuevo valor para la constante de Hubble de 180 (km/s)/Mpc.

Dos años después, Allan R. Sandage (1926-2010), antiguo estudiante de doctorado de Baade y colaborador de Hubble, vuelve a corregir la estimación de la constante de expansión del universo. A partir de un reanálisis de la escala de distancias usando cefeidas y del problema de



Figura 5: Telescopio espacial Hubble. Cortesía NASA/ESA.

la confusión de regiones HII, Sandage (1958) sugiere que el valor de la constante de Hubble se encuentra en el intervalo 50-100 (km/s)/Mpc.

El resultado de Sandage fue premonitorio, pues el nivel de incertidumbre en la medida de la constante de Hubble se mantuvo en un factor dos durante varias décadas. Sandage mismo dedicó una gran parte de su carrera científica a refinar la escala de distancias cósmicas y a mejorar la determinación de la constante de Hubble. En una serie de trabajos durante los años setenta y ochenta, él y sus colaboradores usan una variedad de métodos para medir distancias a galaxias, obteniendo sistemáticamente valores de la constante de Hubble en la zona inferior del intervalo, es decir, entre 50 y 60 (km/s)/Mpc. Por otra parte, el astrónomo francés de la Universidad de Texas Gérard de Vaucouleurs (1918-1995), junto con otro grupo de colaboradores, estima valores en la zona superior del intervalo, típicamente en torno a 100 (km/s)/Mpc. A lo largo de casi dos décadas, cada grupo publica trabajos que apoyan sus resultados pero son incompatibles con las medidas del otro. La polémica sobre el valor exacto de la constante de Hubble se enquistó durante los años setenta y ochenta.

El lanzamiento del telescopio espacial Hubble en 1990 marca un hito en la historia de la determinación de la constante del mismo nombre. Medir la constante con una precisión del 10% es una de las motivaciones para construir este telescopio, y llevar a cabo esa medida representa una importante inversión de tiempo de observación. Sin la degradación causada

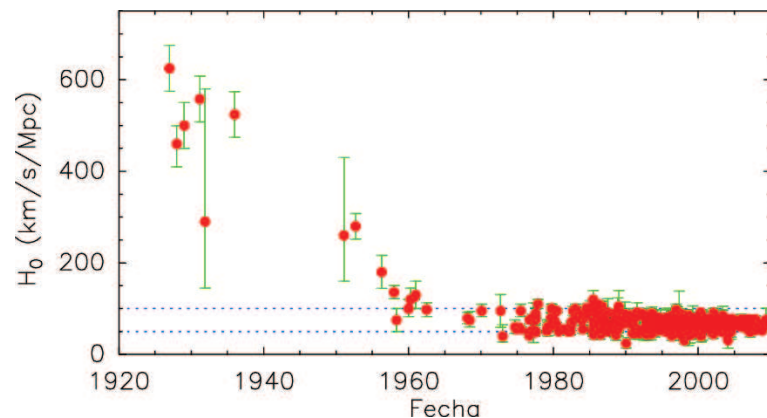


Figura 6: Evolución de la estimación de la constante de Hubble (H_0). Hasta finales de los años cincuenta, las medidas sobrestiman H_0 sistemáticamente debido a un error en la escala de distancias extragalácticas. A partir de los años setenta, los valores se distribuyen de forma casi uniforme dentro del intervalo 50-100 (km/s)/Mpc, y a partir del año 2000, tienden a concentrarse en la parte baja de dicho intervalo. Figura adaptada y basada en la compilación de John Huchra, disponible en <https://www.cfa.harvard.edu/~dfabricant/huchra/hubble.plot.dat>

por la atmósfera terrestre, las imágenes del telescopio espacial (2,4 m de diámetro) son capaces de resolver cefeidas individuales en galaxias hasta el cúmulo de Virgo (unos 20 Mpc). Esto permite extender el método de las cefeidas a distancias 10 veces mayores de lo que es posible hacer desde telescopios terrestres, aumentando así el volumen espacial cubierto por este método en un factor 1000. Observando dentro de este mayor volumen calibradores secundarios, como explosiones supernova, se puede mejorar significativamente la precisión de estos calibradores, y extender así la precisión de la constante de Hubble a distancias verdaderamente cosmológicas.

A pesar de la gran calidad de los datos obtenidos con el telescopio espacial, su uso para determinar la constante de Hubble con precisión requiere un enorme cuidado. Son todavía necesarias numerosas correcciones de difícil calibración, y el resultado final está sujeto a un nivel de incertidumbre que es necesario evaluar con detalle. Tras casi una década de trabajo, el grupo de astrónomos del llamado “Proyecto clave para medir la constante de Hubble con el Telescopio Espacial”, liderado por la astrónoma Wendy Freedman, presenta en 2001 su mejor estimación de la constante de Hubble. Según este grupo, $H_0 = 72 \pm 8$ (km/s)/Mpc (Freedman et al. 2001), que es un valor casi equidistante de las estimaciones de 50 y 100 (km/s)/Mpc que

dominaron la polémica Sandage-de Vaucouleurs durante las tres décadas anteriores. Con setenta años de retraso, por fin parece haberse cumplido el sueño original de Hubble y Humason de alcanzar una precisión del 10% en la determinación de H_0 .

Aunque una estimación de la constante de Hubble con un 10% de incertidumbre empieza a poder ser caracterizada como “cosmología de precisión”, todavía existe un margen considerable para disminuir el nivel de error. Mejoras en la instrumentación y el lanzamiento de los telescopios espaciales de próxima generación GAIA y JWST prometen afinar la determinación de H_0 usando la relación velocidad-distancia.

Métodos alternativos, como el estudio del fondo cósmico de radiación o la distribución espacial de galaxias, proporcionan determinaciones de la constante de Hubble independientes de la clásica relación velocidad-distancia, y con el tiempo han alcanzado una gran precisión. Entre las medidas más recientes, destaca la determinación con el satélite Planck, que a partir de observaciones del fondo cósmico de radiación sugiere un valor de $H_0 = 67 \pm 1$ (km/s)/Mpc (Planck Collaboration 2013). La consistencia mutua entre las medidas obtenidas por distintos métodos proporciona la confianza de que finalmente conocemos con bastante precisión el valor de este parámetro fundamental del universo.

¿Qué mide la ley de Hubble?

Desde un punto de vista observacional, la ley de Hubble es una simple correlación lineal entre el desplazamiento al rojo de la radiación de una galaxia y la distancia a la que se encuentra. Para entender su origen físico es necesario usar los conceptos de la Relatividad General, que es la teoría que mejor describe el universo a gran escala. Como demostraron inicialmente Friedmann y Lemaître en los años veinte, y formalizaron con más detalle Howard Robertson y Arthur Walker una década más tarde, existe una familia de soluciones de las ecuaciones de la Relatividad General en las que un universo homogéneo e isótropo³ se encuentra en estado de expansión (o contracción) global. La descripción matemática de estas soluciones supera el nivel de este artículo, por lo que en esta sección sólo veremos los conceptos más básicos. Una descripción más completa pueden encontrarse en monografías especializadas como la de Peebles (1993).

La familia de modelos de un universo en expansión con principio cosmológico se suele llamar de Friedmann-Lemaître-Robertson-Walker, o alguna otra combinación del nombre de estos autores. Una propiedad característica de estos modelos es que la velocidad de separación entre dos

³El universo es homogéneo e isótropo si sus propiedades físicas no dependen de la posición o la dirección. La hipótesis de que el universo a gran escala tiene estas características se denomina “principio cosmológico”.

galaxias cualesquiera es proporcional a la distancia (propia) que las separa, lo que puede escribirse como

$$v = Hr,$$

donde v es la velocidad de separación, r es la distancia propia entre las galaxias, y H es un parámetro independiente de r (“constante de Hubble”). En el marco de la Relatividad General, el movimiento de separación entre galaxias produce un desplazamiento al rojo de la radiación de una cuando es observada desde la otra, y este efecto es la causa de lo que observacionalmente reconocemos como la ley de Hubble.

A diferencia de lo que supusieron Hubble y otros inicialmente, el desplazamiento al rojo de la radiación observado no se debe al conocido efecto Doppler, que es el causante de los cambios de frecuencia medidos en los objetos que se mueven dentro de la galaxia. En las observaciones extragalácticas que dan lugar a la ley de Hubble, el efecto dominante no son los movimientos propios de los objetos. Cada galaxia permanece casi estacionaria respecto a su entorno (salvo pequeños movimientos propios) y es el espacio entre las galaxias el que se expande. En esta situación, el desplazamiento al rojo de la radiación se debe a que durante su tiempo de viaje de una galaxia a la otra, el espacio se expande, y la longitud de onda de la radiación que viaja por él hace lo mismo. A velocidades bajas, el desplazamiento al rojo causado por la expansión del espacio sigue una ley equivalente a la del efecto Doppler correspondiente a la velocidad de recesión de una galaxia, por lo que las velocidades inferidas usando las dos interpretaciones son casi iguales. A velocidades altas, la ley del desplazamiento al rojo cosmológico es distinta al efecto Doppler, y las velocidades estimadas por los dos métodos difieren.

Por otra parte, el concepto de velocidad de recesión de una galaxia debe ser refinado, pues no es necesariamente lo mismo la velocidad de recesión en el momento en que la galaxia emitió la radiación que en el momento en que esa radiación nos llega, dado que la velocidad de expansión del universo puede variar con el tiempo. Debido a ello, estimar la velocidad de recesión presente de una galaxia requiere conocer cómo ha evolucionado la velocidad de expansión del universo entre el momento de emisión de la radiación y el presente, lo que no es posible sin usar un modelo cosmológico. Para evitar esta complicación, los astrónomos prefieren usar el parámetro adimensional de desplazamiento al rojo cosmológico z , que es directamente medible y viene dado por

$$z = \frac{\lambda_0}{\lambda_e} - 1,$$

donde λ_0 y λ_e son las longitudes de onda de la radiación medida por el observador y por el emisor (i.e., su valor en un laboratorio en reposo).

El parámetro z tiene un valor muy bajo, cercano a cero, para las galaxias más cercanas, y crece a medida que las galaxias se encuentran más lejos y su velocidad de recesión aumenta. Multiplicando z por la velocidad de la luz c , se obtiene una cantidad con unidades de velocidad, que a velocidades bajas es igual a la velocidad de recesión de la galaxia con ese z , pero no es necesariamente igual a velocidades altas. Tanto z como cz son normalmente usados para construir las modernas versiones del diagrama de Hubble, aunque hay una ligera diferencia entre usar velocidad y z en el diagrama. Un diagrama velocidad-distancia es estrictamente lineal en el caso de un universo homogéneo (pero requiere un modelo de universo para ser determinado). Un diagrama z o cz -distancia, por otra parte, puede obtenerse directamente de medidas observacionales, aunque presenta una pequeña curvatura para valores elevados de z (Harrison 1993).

Una ventaja adicional del parámetro observacional z es que proporciona una medida directa del factor de expansión del universo entre el momento en que una galaxia emitió la radiación observada y el presente. Ello es debido a que la longitud de onda de la radiación aumenta con la expansión del universo, y por tanto, lleva impresa la marca de tal expansión, que es igual a $z + 1$. Así, es inmediato determinar que el universo se ha expandido un factor 2 desde el momento en que una galaxia con $z = 1$ emitió su luz y el momento actual cuando esa luz nos llega.

Una diferencia importante entre la interpretación cosmológica del desplazamiento al rojo de la radiación y el efecto Doppler es que la velocidad de recesión de las galaxias puede superar la velocidad de la luz, lo que sería imposible si el desplazamiento al rojo se debiera al efecto Doppler. Esta posibilidad no constituye una violación del principio de la relatividad, que prohíbe explícitamente que cualquier objeto viaje por el espacio más deprisa que c . Como hemos visto, la velocidad de recesión no representa un movimiento de objetos dentro del espacio sino que es el resultado de la expansión del propio espacio entre los objetos, y esta expansión no tiene límite de velocidad. De hecho, galaxias con z mayor que aproximadamente 1,5 se están separando de nosotros con velocidades mayores que c , y estas galaxias son observadas rutinariamente por los astrónomos. Más detalles sobre ésta y otras paradojas aparentes de la expansión del universo pueden verse en el artículo de divulgación por Lineweaver y Davis (2005).

Es posible entender de forma intuitiva los modelos de Friedmann-Lemaître-Robertson-Walker usando una analogía simple de la mecánica newtoniana. En esta analogía, la expansión del universo se compara al movimiento de un objeto lanzado hacia arriba desde la superficie de nuestro planeta. Si la velocidad de lanzamiento es pequeña, la energía cinética del objeto no es capaz de vencer la atracción gravitacional de la Tierra. En este caso, la velocidad de ascensión inicial irá disminuyendo hasta llegar a cero y después se invertirá cuando el objeto empiece a caer. Si por el

contrario la velocidad de lanzamiento del objeto es lo suficientemente alta como para superar la atracción gravitacional de la Tierra, la velocidad de ascensión del objeto disminuirá gradualmente, pero nunca llegará a ser cero y el objeto acabará escapando de la gravedad terrestre. En términos generales, el parámetro que regula la evolución del objeto es el cociente entre la velocidad de lanzamiento y la llamada velocidad de escape, que depende de la masa y radio de la Tierra. Si este parámetro es mayor que la unidad, el objeto escapa, y si es menor, el objeto vuelve a caer.

La solución de las ecuaciones de la Relatividad General para un universo en expansión sin constante cosmológica tiene un comportamiento dual como el del objeto lanzado hacia arriba en la analogía newtoniana. La expansión del universo puede invertirse o continuar indefinidamente, igual que el objeto lanzado puede volver a caer o escapar para siempre. El parámetro que regula este comportamiento se denomina Ω , y es de nuevo igual al cociente entre dos cantidades. En este caso el cociente resulta de dividir la densidad material del universo por la llamada densidad crítica. Si la densidad del universo supera la densidad crítica, la expansión inicial se invierte, y el universo acaba contrayéndose. Si la densidad es inferior a la densidad crítica, el universo se expande indefinidamente. El caso en que Ω es exactamente igual a uno corresponde a una expansión infinita con una velocidad que se hace arbitrariamente baja con el tiempo, y es equivalente en la analogía newtoniana al caso de un lanzamiento con velocidad igual a la de escape.

Si la evolución del universo depende de que su densidad supere o no la densidad crítica, es importante conocer el valor de ambas cantidades. Las ecuaciones de la Relatividad General muestran que la densidad crítica del universo depende de la constante de Hubble y de la constante gravitacional, y que es aproximadamente igual a cinco átomos de hidrógeno por metro cúbico. Esta densidad es muy baja para estándares terrestres o incluso galácticos, pero a escalas intergalácticas es relativamente alta. De hecho, estimaciones de la densidad promedio del universo basadas en la observación de estrellas y de gas caliente y frío quedan sistemáticamente por debajo del valor crítico por más de un orden de magnitud, sugiriendo que el universo no tiene suficiente materia para contrarrestar su expansión inicial. Desgraciadamente, la estimación de la densidad del universo está sujeta a un alto grado de incertidumbre, pues existen posibles componentes invisibles que podríamos estar ignorando. Una de estas componentes es la llamada “materia oscura”, que como luego veremos, parece dominar la atracción gravitacional de galaxias y cúmulos de galaxias.

Otro elemento que puede afectar a la expansión del universo es la constante cosmológica. Durante la mayor parte del siglo XX, la constante cosmológica fue ignorada tanto por razones teóricas como observacionales. Sin embargo, los resultados de la observación de supernovas que veremos más abajo han recuperado el interés por este término adicional de las

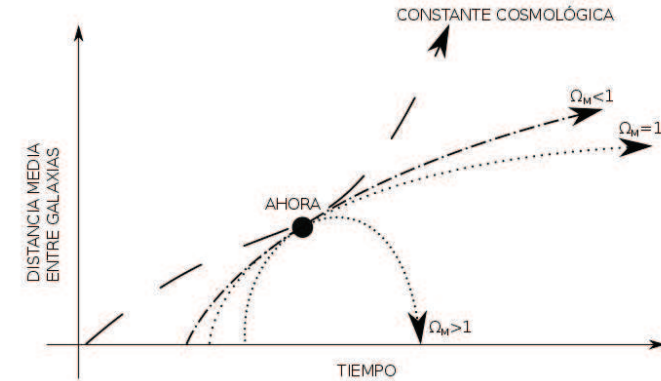


Figura 7: Evolución de la expansión del universo dependiendo de su contenido material y energético. Si la constante cosmológica es nula, la expansión original se convertirá en contracción si Ω_M es mayor que 1. La expansión continuará indefinidamente si Ω_M es menor o igual a 1. Si el universo está dominado por la constante cosmológica, como parece ser el caso, la expansión se acelerará con el tiempo tras una primera época de deceleración. Figura adaptada de Wikimedia.

ecuaciones. Si la constante cosmológica no es nula, el comportamiento de las soluciones de universo es parecido al caso sin constante, pero definiendo ahora Ω como la suma de dos componentes. Una componente representa la contribución de la materia, tanto visible como invisible, y que denominamos Ω_M . La otra componente representa la contribución de la constante cosmológica, y es denominada Ω_Λ porque en las ecuaciones de la Relatividad General la constante cosmológica se representa por Λ . En términos de esas ecuaciones, el valor total de Ω determina la geometría del espacio-tiempo. Un universo con Ω mayor que uno es un universo esférico y un universo con Ω menor que uno es un universo hiperbólico. El caso límite con $\Omega = 1$ es un universo plano como la geometría euclídea a la que estamos acostumbrados en la vida diaria.

Dado que las contribuciones de la materia oscura y la constante cosmológica no son predecibles teóricamente, la única manera de determinar la evolución de la expansión del universo es mediante observaciones. De este modo, el problema de entender la naturaleza del universo se invierte, y es usando el estudio de su evolución temporal como esperamos determinar su composición invisible. Para ello es necesario cubrir con las observaciones la mayor cantidad posible de la historia de la expansión del universo, de manera que pequeños cambios en la tasa de expansión hayan acumulado un efecto observable. Esto requiere mirar hacia atrás en la historia de la expansión del universo, remontándose al

pasado más remoto y midiendo allí la velocidad de expansión. Para ello aprovechamos que la velocidad de la luz es finita, y que podemos ver directamente este pasado observando objetos lejanos. La radiación que ahora nos llega de estos objetos trae la imagen fósil del momento en que se emitió. La gran frontera en cosmología es por tanto la observación de los objetos más lejanos, cuya luz presenta un desplazamiento al rojo z más alto. Por su enorme distancia, estos objetos son extremadamente débiles, y necesitan ser observados usando los telescopios más grandes a nuestra disposición. Determinar su distancia ha supuesto tener que crear un nuevo peldaño en la escalera de distancias cósmicas. Ello ha sido posible gracias a la observación de algunos de los objetos más brillantes conocidos, las explosiones supernova.

Supernovas como candelas estándar

El término supernova engloba varios tipos de explosiones estelares de gran energía que suceden en las últimas fases de la vida de una estrella masiva o de un sistema binario. Las explosiones supernova han sido observadas de forma casual a lo largo de la historia, aunque su origen no fue entendido en su momento. Entre otras supernovas históricas, destacan la “estrella invitada” de 1054, observada por astrónomos chinos y árabes en la constelación de Tauro, y las supernovas de Tycho Brahe (1572) y Johannes Kepler (1604) estudiadas, entre otros, por los dos conocidos astrónomos. Hubo que esperar al siglo XX para conocer los procesos que regulan la estructura y la evolución estelar, y así entender el origen físico de estas explosiones que hacen que una estrella pueda brillar brevemente más que el resto de la galaxia donde se encuentra.

Desde un punto de vista observacional, las supernovas se dividen en tipos I y II dependiendo respectivamente de la ausencia o presencia de líneas de hidrógeno en el espectro. Cada uno de estos tipos se subdivide a su vez en subtipos según características adicionales del espectro, que se denotan con una letra. Para la determinación de distancias cosmológicas, las supernovas de mayor interés son las de tipo Ia, donde la letra “a” indica que durante las primeras fases de la explosión es posible observar en el espectro líneas de silicio ionizado una vez.

Desde un punto de vista físico, las supernovas tipo Ia (abreviadas SNe Ia) corresponden a explosiones termonucleares en un sistema binario en el que una de las componentes es una enana blanca. Este tipo de sistemas binarios resulta de la evolución de dos estrellas de distinta masa ligadas gravitacionalmente y orbitando una alrededor de la otra. La estrella más masiva agota su combustible nuclear primero, pues su producción de energía es desproporcionadamente más rápida. Si la masa inicial de esta estrella es menor que unas ocho veces la masa de nuestro Sol, la estrella



Figura 8: Supernova SN 1994D en la galaxia NGC 4526. Se trata de una supernova de tipo Ia en una galaxia relativamente próxima (17 Mpc) que, como su nombre indica, fue detectada en 1994. La supernova es el objeto puntual brillante en la zona inferior izquierda. Imagen tomada con el telescopio espacial Hubble. Cortesía NASA/ESA, The Hubble Key Project Team y The High-Z Supernova Search Team.

perderá suficiente materia durante su evolución para al final de su vida convertirse en un objeto del tamaño del planeta Tierra pero con una masa como la del Sol. Este objeto denso, compuesto principalmente de carbono y oxígeno, que prácticamente no produce ya energía por reacciones nucleares y que se enfría lentamente, es lo que denominamos una enana blanca.

La densidad de una enana blanca es más de un millón de veces superior a la del agua o a la densidad media del Sol. En estas condiciones extremas, la materia sólo puede estabilizarse frente a su autogravedad mediante el fenómeno cuántico de presión de degeneración de los electrones. Esta fuerza de presión resulta de la imposibilidad de que dos electrones (técnicamente partículas fermiónicas) compartan el mismo estado cuántico, y ofrece una resistencia a la contracción muy superior a la presión térmica de un plasma o a la presión de la radiación, que son los mecanismos que estabilizan a las estrellas en sus fases iniciales. En 1931, el físico indio Subrahmanyam Chandrasekhar (1910-1995) demostró que

existe un valor máximo para la masa de una enana blanca que puede estabilizarse mediante el mecanismo de la presión de degeneración de sus electrones (Chandrasekhar 1931). Este valor, denominado ahora “límite de Chandrasekhar”, corresponde a aproximadamente 1,4 veces la masa del Sol, y es un parámetro crítico en la evolución estelar.

En las primeras fases de la evolución de un sistema binario que va a producir una supernova tipo Ia, la estrella enana blanca tiene una masa inferior al límite de Chandrasekhar, y esta situación de estabilidad permite a la estrella compañera evolucionar con normalidad. Cuando esta segunda estrella empieza a agotar el combustible nuclear de su centro, su estructura interna cambia, y se convierte en lo que llamamos una gigante roja. En este tipo de estrellas, las capas más externas se expanden, se enfrían y quedan débilmente ligadas al núcleo denso central. Se produce entonces un fenómeno de transferencia de materia de la gigante roja a la enana blanca, la cual empieza a aumentar gradualmente de masa. Con el tiempo, la masa de la enana blanca se acerca al valor límite de Chandrasekhar.

La física de lo que sucede cuando la enana blanca alcanza el límite de Chandrasekhar es compleja, y su solución requiere modelos numéricos al límite de las capacidades computacionales. El consenso actual es que la enana blanca sufre una reacción nuclear desbocada y global que convierte el carbono y el oxígeno que la componen principalmente en níquel, y que libera suficiente energía como para destruir completamente la estrella. Esta explosión nuclear a escala estelar es lo que observamos como una supernova, y su brillo visual absoluto puede superar el del Sol por un factor de varios miles de millones. Este enorme brillo resulta de una combinación de la energía de la propia explosión y, en las semanas posteriores a ella, de la energía de la desintegración de níquel y cobalto radiactivos producidos en la explosión, que calientan el material eyectado y matienen un brillo que disminuye gradualmente.

Como candelas estándar, las supernovas de tipo Ia presentan múltiples ventajas. Su enorme brillo permite observarlas en galaxias tan lejanas que en ellas no es posible distinguir estrellas cefeidas individuales. De esta manera, las supernovas tipo Ia permiten extender la relación velocidad-distancia a valores más altos. Por otra parte, las explosiones supernova tipo Ia forman un grupo muy homogéneo. Aunque su brillo máximo no es siempre el mismo, existe una correlación estrecha entre el brillo máximo y la tasa a la que este brillo disminuye con el tiempo, lo que permite estimar con mucha precisión su brillo intrínseco a partir de observaciones.

El uso de supernovas como candelas estándar fue propuesto por Walter Baade y Fritz Zwicky (1898-1974) a finales de los años treinta, pero su uso sistemático en la determinación de la constante de Hubble no fue posible hasta mucho más tarde. Por una parte, fue necesario entender las diferencias entre los distintos tipos de supernovas y refinar la clasificación de éstas en subgrupos según las propiedades de su espectro.

Además, fue necesario determinar con precisión el brillo intrínseco de una muestra suficientemente completa de supernovas, y de esta manera disponer de un patrón de calibración fiable. Medidas precisas del brillo intrínseco de supernovas empezaron a ser posibles en gran medida gracias a observaciones con el telescopio espacial Hubble a partir de 1994. Estas observaciones permitieron encontrar estrellas cefeidas en galaxias donde había ocurrido una supernova tipo Ia, y de esta manera determinar la distancia a la galaxia e inferir de ella el brillo absoluto de la supernova. Gracias a estos avances teóricos y observacionales, a principios de los años noventa la escalera de distancias cósmicas tenía un nuevo peldaño. Usando ese nuevo peldaño, dos grupos de astrónomos van a intentar extender la ley de Hubble a escalas espaciales hasta entonces inalcanzables.

“Supernovas de encargo”

Aunque el uso de supernovas para medir distancias cósmicas presenta multitud de ventajas, también tiene inconvenientes de difícil superación. Un elemento incontrolable es la naturaleza aleatoria del fenómeno supernova. Se estima que, en promedio, una galaxia como la Vía Láctea tiene una explosión supernova tipo Ia cada 500 años aproximadamente. Esta explosión sucede sin aviso previo que indique dónde y cuándo mirar, por lo que no es posible preparar la observación de una supernova con antelación. Para encontrar supernovas, es necesario observar un gran número de galaxias, y reconocer entre ellas las pocas que casualmente han tenido una supernova recientemente. Dado lo infrecuente del caso, estas galaxias con supernova representan una muy pequeña fracción del total.

Para observar un gran número de galaxias e identificar aquellas en las que acaba de ocurrir una supernova es necesario usar un procedimiento sistemático. Este procedimiento consiste en tomar dos imágenes separadas varios días o semanas, y usar la primera imagen como comparación de la segunda, sustrayéndola digitalmente. En la imagen diferencia, un candidato a supernova aparece como un objeto puntual y brillante en alguna de las galaxias contenidas en la imagen. Para confirmar este candidato como una supernova de tipo Ia, es necesario observar el espectro del objeto, lo que sirve además para determinar su desplazamiento al rojo z . Esto se realiza comparando el espectro observado con una serie de espectros de supernovas cercanas, y encontrando el desplazamiento en frecuencia necesario para hacer coincidir las líneas espectrales de los dos objetos.

El grupo que primero desarrolló una técnica sistemática para buscar supernovas de alto z fue el Supernova Cosmology Project (SCP, o Proyecto de cosmología con supernovas), creado en 1988 y liderado por Saul Perlmutter en Berkeley (Estados Unidos). Este grupo ideó una metodología denominada “supernovas de encargo” (*supernovae on demand*

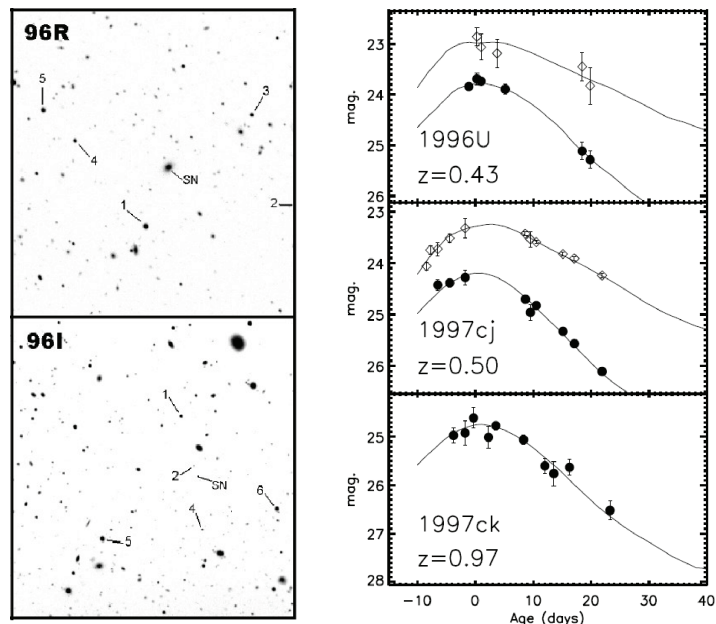


Figura 9: Supernovas tipo Ia usadas por el grupo HZT para determinar distancias (Riess et al. 1998). *Izquierda:* regiones del cielo donde se encontraron las supernovas (marcadas SN). Los números indican estrellas usadas como referencia. *Derecha:* curvas de luminosidad de varias supernovas. Los círculos vacíos indican el brillo en el filtro visible V y los rellenos en el filtro azul B.

en inglés), que permitía asegurar que para cierta fecha, el grupo tendría aproximadamente dos docenas de supernovas en fase cercana al máximo en las que observar un espectro y determinar el tipo de supernova y el valor de z . Esta estrategia era necesaria porque la observación de espectros de supernovas lejanas requiere el uso de telescopios grandes, como los Keck en Hawái (10 m de diámetro), y las observaciones con este tipo de telescopios se conceden de forma competitiva y se programan con antelación. Tener supernovas recientes en la fecha asignada para la observación espectroscópica requiere haberlas encontrado en las semanas anteriores. Para ello, el grupo de Perlmutter realizaba una campaña previa usando el telescopio de 4 m del Cerro Tololo Inter-American Observatory (CTIO) en Chile. En estas observaciones se tomaban, separadas por tres semanas, imágenes de entre 50 y 100 regiones del cielo, cada una con casi mil galaxias. De la comparación de las imágenes en las dos fechas, se identificaban los candidatos a supernova, que por la secuencia de observación se encontraban necesariamente cerca del máximo de brillo

(una supernova tipo Ia tarda unas dos semanas en alcanzar su máximo brillo). Estas estrellas se observaban casi inmediatamente con uno de los telescopios Keck para obtener un espectro y determinar el tipo de supernova y el valor de z . Una vez identificado un objeto como supernova tipo Ia y con parámetro z determinado, éste se reobservaba durante meses para seguir su curva de luminosidad. Para ello, el grupo contaba con una red de colaboradores en diferentes telescopios.

Una técnica parecida a la del SCP fue la usada por un grupo competidor, el denominado High- z Supernova Search Team (HZT, Equipo de búsqueda de supernovas de alto z), liderado por Brian Schmidt, y del que formaba parte Adam Riess. Este grupo contaba con numerosos expertos en supernovas de Harvard y Berkeley, entre otras instituciones, y empezó su búsqueda de supernovas en 1994, algo más tarde que el SCP. Durante varios años, los dos grupos fueron encontrando y midiendo suficientes supernovas tipo Ia como para remontarse en el pasado aproximadamente la mitad de la historia de la expansión del universo ($z \approx 0,8$).

Un resultado inesperado

Tanto el SCP como el HZT usaron para reconstruir la historia de la expansión del universo una versión ligeramente modificada del diagrama de Hubble. Como se muestra en la figura 10, en vez de representar velocidad frente a distancia, este diagrama modificado representa los dos principales parámetros observacionales del estudio de las supernovas, el brillo máximo, medido en magnitudes, y el desplazamiento al rojo z . En este diagrama, los distintos modelos de universo predicen diferentes relaciones entre la magnitud y z , y es el objetivo de las observaciones distinguir entre ellos. Para pequeños desplazamientos z , los diferentes modelos son prácticamente iguales entre sí, pues las diferencias entre las magnitudes de las supernovas que predicen son menores que el nivel de error de las medidas. A medida que aumenta z , las diferencias entre los modelos de universo aumentan, y para valores de z superiores a aproximadamente 0,3, las diferencias empiezan a ser medibles con el nivel de precisión alcanzado por las observaciones.

En 1998, los dos grupos competidores habían acumulado suficientes observaciones de calidad como para que los resultados empezaran a poder distinguir entre los diferentes modelos de universo. En una auténtica carrera contra reloj, las publicaciones se siguieron en el plazo de unos pocos meses. El grupo del HZT lo hizo primero, con un trabajo que presentaba observaciones de 10 supernovas de alto z (Riess et al. 1998). Pocos meses después, el SCP presentaba los resultados del análisis de 42 supernovas en un artículo aceptado por la revista en 1998 pero publicado al año siguiente (Perlmutter et al. 1999).

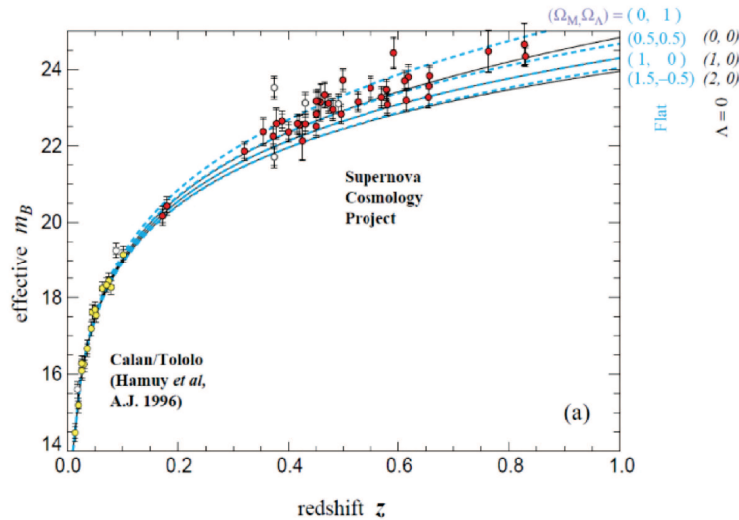


Figura 10: Diagrama magnitud-desplazamiento z usado por el grupo SCP para estudiar la expansión del universo (Perlmutter et al. 1999). Los puntos representan medidas con supernovas y las líneas son posibles modelos de universo. Sólo modelos con aceleración, representada por la constante cosmológica Λ , son capaces de explicar las magnitudes altas (brillos bajos) de las supernovas con z más alto.

Antes de publicar sus resultados los dos grupos tuvieron que vencer muchas reticencias. Ambos equipos esperaban observar que la expansión del universo se ha ido decelerando a lo largo de la historia, como estaba previsto debido a la atracción gravitacional entre las distintas componentes materiales. Los resultados, sin embargo, contradecían esa expectativa inicial. Desde un punto de vista observacional, los datos de ambos grupos mostraban que supernovas más distantes tenían sistemáticamente una magnitud mayor de lo esperado, es decir, eran más débiles de lo predicho si el universo estuviera decelerando su expansión (debido a una convención antigua de unidades, la magnitud de un objeto aumenta a medida que su brillo disminuye). Si las supernovas más lejanas eran más débiles de lo esperado, es porque estaban más lejos de los previsto. La única manera de explicar esta discrepancia era concluir que la expansión del universo se ha ido acelerando con el tiempo, lo contrario de lo inicialmente esperado.

Una manera de entender lo sorprendente del resultado obtenido con las supernovas es usando de nuevo la analogía con el lanzamiento de un objeto hacia arriba. Independiente de que la velocidad de lanzamiento supere la velocidad de escape, el objeto se irá frenando a medida que sube. El universo observado por los grupos SCP y HZT usando supernovas se

comporta de una manera totalmente distinta. Es como si al tirar el objeto hacia arriba, éste no sólo subiera, sino que acelerara su ascensión a medida que gana altura.

La aceleración del universo requiere la presencia de una fuerza a gran escala opuesta a la gravedad. Como vimos antes, la teoría de la Relatividad General proporciona un mecanismo natural con esas características, la constante cosmológica introducida por Einstein en 1917. Por ello, la constante cosmológica fue el mecanismo elegido tanto por el HZT como el SCP para interpretar sus observaciones de supernovas. Esto puede verse en la figura 10, que muestra el diagrama de Hubble creado por el equipo SCP en su trabajo de presentación de los primeros datos de supernovas de alto z . Estas supernovas (42 en total) están representadas por círculos rojos, y complementan los valores obtenidos por otro grupo observando supernovas más próximas (círculos amarillos). Las líneas azules y negras muestran las predicciones teóricas de modelos de universo de distinta densidad de materia (parametrizada con Ω_M) y constante cosmológica (parametrizada con Ω_Λ). Como puede verse, las supernovas observadas son aproximadamente un 15 % más débiles que el caso más extremo sin constante cosmológica (un universo vacío con $\Omega_M = \Omega_\Lambda = 0$). Sólomente añadiendo una constante cosmológica no nula es posible explicar las observaciones obtenidas.

Aunque los resultados de las supernovas eran sorprendentes, el hecho de que fuesen obtenidos por dos grupos independientes y competidores entre sí les dió gran credibilidad desde el principio. Aún así, un resultado tan sorprendente necesita ser validado con observaciones independientes por una comunidad diversa con intereses muchas veces contrapuestos. Por ello, a las publicaciones originales de 1998-99 siguió una avalancha de observaciones para encontrar más supernovas con las que confirmar o refutar los resultados iniciales. Tras más de una década de intenso trabajo, las nuevas medidas han confirmado el resultado inicial. Las supernovas más lejanas son sistemáticamente más débiles, y esto indica que la expansión del universo se ha acelerado con el tiempo. Un ejemplo de los últimos resultados es la figura 11, que presenta una compilación de más de 500 supernovas obtenidas por diferentes grupos a lo largo de una década de trabajo. La enorme cantidad de puntos observacionales, y su bajo nivel de dispersión, permite observar un comportamiento extraordinariamente sistemático que puede ajustarse satisfactoriamente con un modelo de universo en expansión acelerada.

La observación de supernovas adicionales y la extensión del diagrama de Hubble a valores de z cada vez más altos tiene como objetivo ir más allá de verificar la existencia de una aceleración en la expansión del universo. Medir supernovas de z creciente significa remontarse hacia atrás en el tiempo y, de esta manera, poder explorar la historia de la expansión en detalle. Los resultados de supernovas obtenidos en la última década han

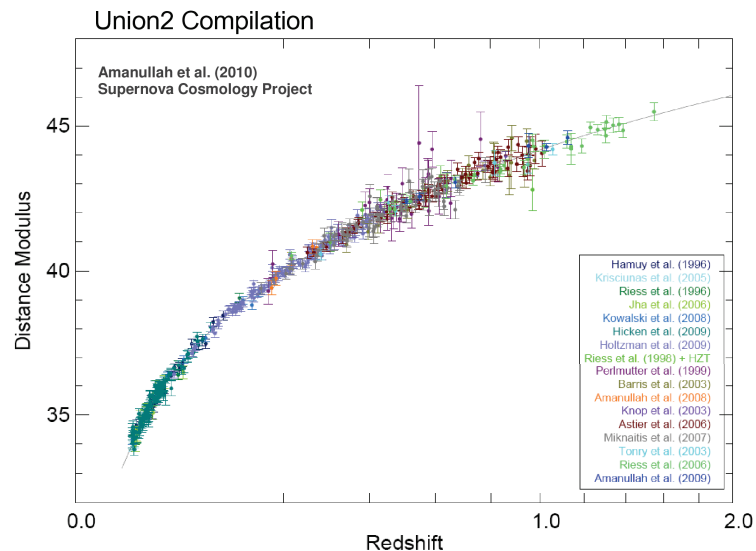


Figura 11: Compilación reciente de supernovas por Amanullah et al. (2010). La combinación de datos de 557 supernovas tipo Ia no sólo confirma las conclusiones originales de los primeros trabajos, sino que permite reconstruir con gran precisión la historia de la expansión acelerada del universo, modelizada en la figura por la línea gris.

permitido reconstruir esa historia de la expansión midiendo los cambios que ésta ha sufrido a lo largo de los últimos casi 10 mil millones de años, lo que representa una parte considerable de la historia total del universo, que tiene 14 mil millones de años. Según estas observaciones, la aceleración del universo es un efecto relativamente reciente, que sólo ha sucedido en los últimos cinco o seis mil millones de años. En épocas anteriores, el universo se deceleraba de la manera esperada por el efecto de la gravedad.

Una combinación de deceleración inicial y aceleración posterior es exactamente lo que predice un modelo de universo con constante cosmológica. Como hemos visto, esta constante juega un papel de fuerza opuesta a la gravedad, y su contribución en cada punto es proporcional al volumen de espacio considerado. Por ello, el aumento de volumen asociado a la expansión del universo hace crecer el efecto de la constante cosmológica de forma sistemática. En una primera fase de su evolución, el universo era relativamente compacto y estaba dominado por la materia. En esta fase, la expansión del espacio estaba limitada por la atracción gravitacional entre las componentes materiales, y el universo se iba decelerando lentamente. A medida que la expansión del espacio progresó,

la contribución de la constante cosmológica fue aumentando gradualmente de forma proporcional al incremento de volumen, y hace cinco o seis mil millones de años esta contribución superó al efecto opuesto de la gravedad. A partir de ese momento, el balance de fuerzas se invirtió, y la constante cosmológica comenzó a dominar la expansión del universo. El resultado es el universo acelerado que observamos actualmente.

Más allá del diagrama de Hubble

Hasta ahora hemos estudiado la expansión del universo centrándonos en el diagrama de Hubble. Históricamente, este diagrama proporcionó las primeras pruebas tanto de la expansión del universo como de su aceleración, pues representa la manera más directa de medir el estado cinemático del universo. Sin embargo, no es la única forma posible. Como sería de esperar, la expansión del universo ha tenido un efecto global y por ello ha dejado su marca en muchos otros fenómenos observables. Estudiando estos fenómenos es posible poner a prueba las conclusiones obtenidas analizando el diagrama de Hubble, y en algunos casos, añadir nueva información sobre la historia de la expansión del universo. Aunque hay un gran número de observables, en esta sección sólo veremos brevemente los dos más importantes.

Una fuente de información única sobre la historia del universo y sus de propiedades tempranas es el fondo cósmico de radiación. Este fondo cósmico es una radiación difusa remanente de la época densa y caliente en la que la materia y la radiación del universo dejaron de convivir en equilibrio. Ello sucedió unos 400 000 años después del principio de la expansión, lo que llamamos el Big Bang. En la actualidad, el fondo de radiación es detectable sobre todo a frecuencias radio, y su descubrimiento accidental en 1965 fue galardonado con un premio Nobel de física a Arno A. Penzias y Robert W. Wilson en 1987.

Una propiedad importante del fondo de radiación es su extraordinaria homogeneidad espectral y espacial. La distribución de la radiación en frecuencias es, dentro de las incertidumbres de la medida, la de un cuerpo negro perfecto de aproximadamente 2,7 K de temperatura. En cuanto a su distribución espacial, la radiación es homogénea (salvo la presencia de una componente dipolar) mejor que una parte en 10 000. Sólo con medidas de alta precisión, como las del satélite Cosmic Background Explorer (COBE), ha sido posible medir las inhomogeneidades espaciales en la distribución del fondo de radiación. Estas inhomogeneidades reflejan las fluctuaciones de densidad y temperatura del universo en el momento en que se generó el fondo de radiación, y son las semillas de la estructura que dio lugar a las galaxias y los cúmulos de galaxias actuales. El descubrimiento en 1992 de estas inhomogeneidades con observaciones del satélite COBE fue

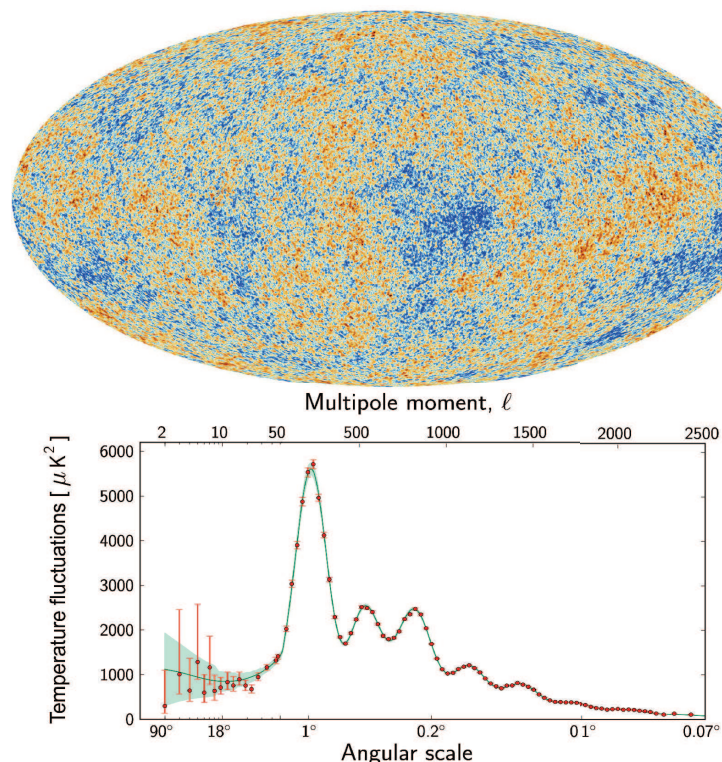


Figura 12: Resultados del análisis de las inhomogeneidades del fondo cósmico observadas con el satélite Planck. Arriba: mapa de la distribución espacial de las inhomogeneidades (aumentadas un factor 5.400 para hacerlas visibles). Abajo: descomposición de las inhomogeneidades en función de la escala angular, desde 90° a $4'$. La línea verde es un modelo de universo con materia y energía oscura que ajusta los datos dentro del nivel de ruido. Cortesía ESA y Planck Collaboration.

galardonado también con el premio Nobel de física a John C. Mather y George F. Smoot en 2006.

El estudio de las inhomogeneidades del fondo cósmico se realiza observando la intensidad de la radiación en todas direcciones, y midiendo el nivel de los cambios de brillo a cada escala angular, desde decenas de grados hasta segundos de arco. A las medidas originales del satélite COBE se han añadido nuevas observaciones obtenidas desde tierra, globos sonda y nuevas generaciones de satélites, como el estadounidense Wilkinson Microwave Anisotropy Probe (WMAP) y el europeo Planck. Estas medidas

revelan el estado del plasma de bariones y fotones que componía el universo cuando la radiación se originó, y corresponden a una medida del universo con un desplazamiento al rojo z de aproximadamente 1.100. Desde este momento de emisión, la radiación del fondo cósmico ha viajado por el espacio y, por ello, las inhomogeneidades originales llevan superpuestas los efectos de la expansión del universo. Observando esas inhomogeneidades, podemos remontarnos a una época en la que todavía no había ni galaxias ni supernovas, y que es por tanto inaccesible usando la técnica del diagrama de Hubble.

Otra fuente de información sobre la expansión del universo son las llamadas oscilaciones acústicas de bariones (abreviadas BAO en inglés). Estas oscilaciones son el resultado de la evolución temporal de las fluctuaciones originales de densidad y temperatura que dieron lugar a las inhomogeneidades del fondo cósmico. Con el tiempo, estas inhomogeneidades dejaron su huella en la materia bariónica que dio lugar a las galaxias y, por ello, su presencia se manifiesta como una ligera tendencia de las galaxias a presentar un exceso de separaciones a cierta distancia característica. Observaciones de decenas de miles de galaxias, como las realizadas por el proyecto Sloan Digital Sky Survey, han encontrado evidencia de este exceso de galaxias a separaciones de unos 150 Mpc, confirmando las expectativas teóricas y proporcionando una importante medida de una escala característica de la estructura del universo.

La combinación de los resultados de supernovas, anisotropías del fondo cósmico de radiación y oscilaciones acústicas de bariones proporciona una prueba de autoconsistencia muy fuerte al modelo de expansión acelerada del universo. De forma independiente, cada observación sugiere un universo en expansión acelerada. Juntas, las observaciones pueden usarse para determinar con gran precisión la composición material y energética del universo. Esto puede verse en la figura 13, que muestra las regiones del plano (Ω_M , Ω_Λ) favorecidas por cada uno de los tres métodos: diagrama de Hubble con supernovas (SNe), fondo cósmico de radiación (CMB), y oscilaciones acústicas de bariones (BOA). Debido a la existencia de errores experimentales y a cierta ambigüedad en la determinación de parámetros, ninguno de los tres métodos es capaz de medir con precisión los valores de Ω_M y Ω_Λ del universo. Cada método favorece una banda de probabilidad, y por tanto una posible combinación de los dos parámetros. La intersección de las tres bandas, sin embargo, restringe enormemente el nivel de ambigüedad, y determina con mucha precisión el valor más probable de Ω_M y Ω_Λ . Como puede verse, los resultados actuales sugieren que la densidad total de materia en el universo Ω_M es aproximadamente 0,3 veces la densidad crítica, y que la contribución de la constante cosmológica Ω_Λ es del orden de 0,7.

Un resultado adicional de la determinación de Ω_M y Ω_Λ mostrada en la figura 13 es que la suma de las dos cantidades es muy próxima a la unidad.

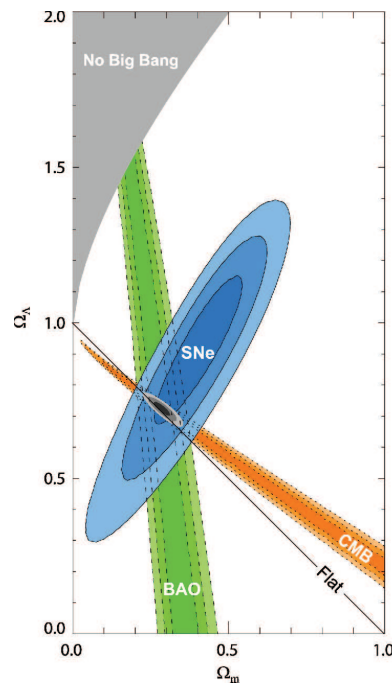


Figura 13: Determinación de los parámetros cosmológicos Ω_M y Ω_Λ combinando datos de supernovas (SNe), fondo cósmico de radiación (CMB) y oscilaciones acústicas de bariones (BAO). Cada uno de los métodos favorece una amplia región del plano $(\Omega_M, \Omega_\Lambda)$, mientras que la intersección restringe los valores a aproximadamente 0,3 y 0,7. (Kowalski et al. 2008).

Como vimos, la suma de las dos contribuciones proporciona una medida de la geometría global del universo, y en particular de su curvatura. Un valor igual a uno equivale al espacio plano al que estamos acostumbrados en la geometría euclidiana, en el que las rectas paralelas no se cruzan y los ángulos de los triángulos suman 180 grados. Tal tipo de espacio es precisamente el predicho por el modelo inflacionario, según el cual, inmediatamente después del Big Bang, existió una época en la que el universo se expandió de manera desbocada.

¿De qué está hecho el universo?

La consistencia entre los resultados de los distintos métodos ilustrados en la figura 13 sugiere que la determinación de parámetros cosmológicos

como Ω_M y Ω_Λ ha alcanzado un alto grado de fiabilidad. Aunque el acuerdo entre los distintos métodos no es perfecto, como ilustran algunas diferencias existentes entre la determinación de la constante de Hubble a partir de supernovas y a partir del fondo cósmico, las diferencias son del orden del 5%, lo que está muy lejos del factor dos que existía en un pasado no muy lejano. Una lectura optimista de la situación es que hemos alcanzado una determinación cuantitativa precisa de los parámetros fundamentales del universo.

En contraste con lo anterior, la interpretación de los resultados cosmológicos en términos físicos dista mucho de ser satisfactoria. Por ejemplo, hemos visto que la densidad de materia en el universo Ω_M es aproximadamente 0,3 veces la densidad crítica. Sin embargo, un gran número de observaciones independientes muestra que la densidad bariónica del universo no llega a 0,05 veces la densidad crítica, lo que representa sólo la sexta parte del total. Esto significa que la mayor parte de la materia del universo pertenece a una categoría no bariónica cuya naturaleza desconocemos, y que es distinta de la materia (bariónica) de que están hechos los planetas, las estrellas, y el plasma caliente que se observa en el interior de los cúmulos de galaxias.

La discrepancia entre la cantidad materia que observamos y la que parece contribuir a las fuerzas gravitacionales en el universo no es nueva. En los años treinta Fritz Zwicky notó que la masa necesaria para equilibrar el cúmulo de galaxias de Coma es mayor que la observada, y que la estabilidad de ese sistema requiere una importante cantidad de material invisible (“oscuro”) (Zwicky 1937). En 1980, observaciones hechas por Vera Rubin y colaboradores mostraron que la rotación de numerosas galaxias espirales requiere también la presencia de grandes cantidades de materia invisible en las zonas más externas (Rubin et al. 1980). Estos resultados han sido confirmados por un gran número de observaciones adicionales, tanto de galaxias como de cúmulos de galaxias y lentes gravitacionales. Todas estas observaciones sugieren que el universo, además de la materia que emite luz y otros tipos de radiación, contiene una cantidad aún mayor de “materia oscura”. Aunque algunos modelos han tratado de explicar este tipo de observaciones como resultado de una desviación de la fuerza de la gravedad de su valor teórico y no como consecuencia de la presencia de materia oscura, el hecho de que Ω_M supere por un factor 6 la estimación de la densidad de materia bariónica hace prácticamente inevitable concluir que la mayor parte de la materia en el universo es de una naturaleza no bariónica que todavía desconocemos.

Si el término material Ω_M encierra una contribución mayoritaria de materia exótica, el término Ω_Λ representa una componente aún más misteriosa. Como vimos, esta componente parece representar una fuerza del vacío que se opone a la gravedad y no tiene análogo en la física newtoniana. Un candidato natural para ella es la constante cosmológica,

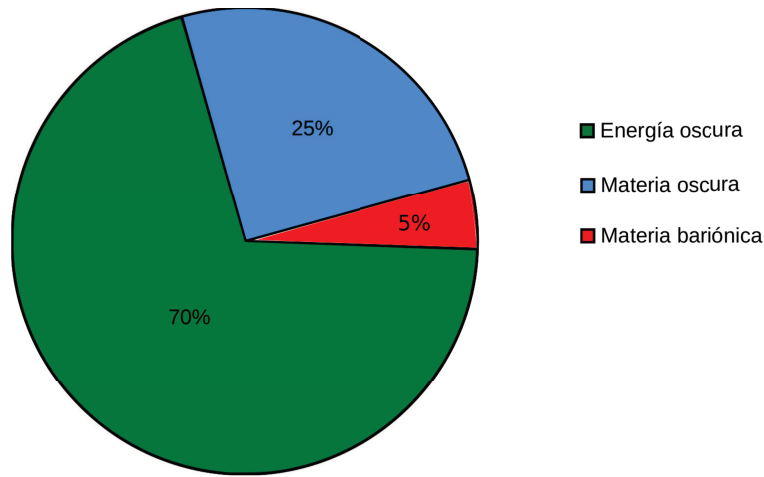


Figura 14: Composición aproximada del universo. Sólo un 5% del universo está hecho de la materia bariónica que conocemos (principalmente neutrones, protones y electrones). El resto son las componentes “oscuras”, cuya naturaleza todavía se nos escapa.

y los primeros resultados de supernovas fueron interpretados en términos de esta constante. Este punto de vista sigue siendo mayoritario en la comunidad astronómica, y está reforzado por el hecho de que la constante cosmológica explica también la observación de que la aceleración actual fue precedida de una deceleración inicial.

Aunque esté bien descrita por las ecuaciones de la Relatividad General, la constante cosmológica es una fuerza misteriosa. Como reconoció el propio Georges Lemaître (Lemaître 1934), la constante cosmológica puede entenderse en términos de una energía propia del vacío y, en el contexto de la mecánica cuántica, este tipo de energía surge de manera natural debido a las fluctuaciones cuánticas que resultan de la creación de pares de partículas virtuales. Desgraciadamente, no disponemos todavía de una teoría satisfactoria que combine los principios de la mecánica cuántica y la Relatividad General, por lo que la equivalencia entre la constante cosmológica y la energía del vacío cuántico es incompleta. Además, una estimación numérica de esta energía cuántica predice un valor incorrecto para la constante cosmológica por 120 órdenes de magnitud, lo que supone un nivel de inconsistencia alarmante. Por ello, es común referirse al agente causante de la aceleración del universo con la expresión “energía oscura”, que combina en un mismo término la idea de una componente energética del vacío con la idea de que su origen es tan desconocido como la naturaleza de la materia oscura.

Aunque consistente con las observaciones, la constante cosmológica no es la única forma posible de energía oscura. Un gran número de mecanismos alternativos han sido propuestos para explicar la expansión acelerada del universo, y el estudio de sus implicaciones teóricas y observacionales es un campo de gran actividad. Una alternativa popular es la llamada quintaesencia, que representa una componente energética que, a diferencia de la constante cosmológica, puede variar con el tiempo (Caldwell et al. 1998). Otra línea de investigación activa es la posibilidad de que la aceleración del universo se explique en términos de correcciones a las ecuaciones de Einstein, y no represente una componente energética adicional (Carroll et al. 2004). Esta riqueza y variedad de alternativas propuestas (en promedio una al día en palabras de Saul Permuter) refleja tanto la creatividad de los científicos como la importancia del problema.

Por supuesto, la última palabra sobre origen de la expansión acelerada la tiene el propio universo. A falta de un laboratorio terrestre donde explorar las distintas alternativas, las observaciones astronómicas son la única herramienta disponible para entender el origen de esta aceleración. Por ello, vivimos un momento de extrema actividad observacional, con nuevos proyectos que empiezan a observar, como el Dark Energy Survey en Chile, y otros que empiezan a ponerse en marcha, como el satélite europeo Euclid. La carrera para entender la misteriosa naturaleza de la energía oscura acaba de comenzar, y nadie duda de que hay un premio Nobel en la meta esperando al equipo ganador.

Referencias

(La mayoría de las siguientes referencias pueden consultarse libremente a través del servicio Astrophysics Data System (ADS) de SAO/NASA, http://adsabs.harvard.edu/abstract_service.html).

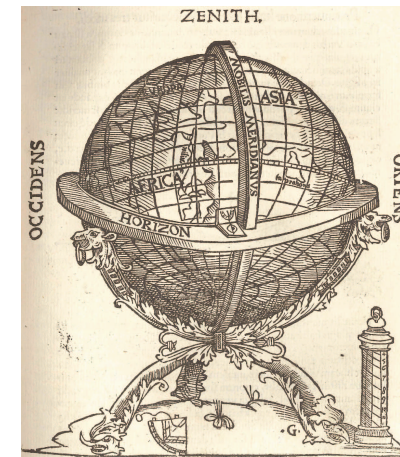
- Amanullah, R. et al. 2010, “Spectra and Hubble Space Telescope Light Curves of Six Type Ia Supernovae at $0.511 < z < 1.12$ and the Union2 Compilation”, *Ap.J.*, 716, 712
- Baade, W. 1956, “The Period-Luminosity Relation of the Cepheids”, *P.A.S.P.*, 68, 5
- Caldwell, R. R., Dave, R., & Steinhardt, P. J. 1998, “Cosmological Imprint of an Energy Component with General Equation of State”, *Physical Review Letters*, 80, 1582
- Carroll, S. M., Duvvuri, V., Trodden, M., & Turner, M. S. 2004, “Is Cosmic Speed-up Due to New Gravitational Physics?”, *Physical Review D*, 70, 043528

- Chandrasekhar, S. 1931, "The Maximum Mass of Ideal White Dwarfs", *Ap.J.*, 74, 81
- de Sitter, W. 1917, "Einstein's Theory of Gravitation and its Astronomical Consequences. Third paper", *M.N.R.A.S.*, 78, 3
- Eddington, A.S. 1930, "On the Instability of Einstein's Spherical World", *M.N.R.A.S.*, 90, 668
- Einstein, A. 1905, "Zur Elektrodynamik bewegter Körper", *Annalen der Physik*, 322, 891
- Einstein, A. 1916, "Die Grundlage der allgemeinen Relativitätstheorie", *Annalen der Physik*, 354, 769
- Einstein, A. 1917, "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie", *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin)*, 142
- Einstein, A., de Sitter, W. 1932, "On the Relation between the Expansion and the Mean Density of the Universe", *P.N.A.S.*, 18, 213
- Freedman, W.L. Madore, B.F. 2010, "The Hubble Constant", *Annual Reviews of Astronomy and Astrophysics*, 48, 273.
- Freedman, W. L., et al. 2001, "Final Results from the Hubble Space Telescope Key Project to Measure the Hubble Constant", *Ap.J.*, 553, 47
- Friedmann, A. 1922, "Über die Krümmung des Raumes", *Zeitschrift für Physik*, 10, 377
- Harrison, E. 1993, "The Redshift-Distance and Velocity-Distance Laws", *Ap.J.*, 403, 28
- Hubble, E. P. 1925, "Cepheids in Spiral Nebulae", *Popular Astronomy*, 33, 252
- Hubble, E. 1929, "A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae", *P.N.A.S.*, 15, 168
- Hubble, E., & Humason, M. L. 1931, "The Velocity-Distance Relation among Extra-Galactic Nebulae", *Ap.J.*, 74, 43
- Humason, M. L., Mayall, N. U., & Sandage, A. R. 1956, "Redshifts and Magnitudes of Extragalactic Nebulae", *A.J.*, 61, 97
- Kowalski, M. et al. 2008, "Improved Cosmological Constraints from New, Old, and Combined Supernova Data Sets", *Ap.J.*, 686, 749

- Kragh, H., & Smith, R. W. 2003, "Who Discovered the Expanding Universe?", *History of Science*, 41, 141
- Leavitt, H. S. 1908, "1777 Variables in the Magellanic Clouds", *Annals of Harvard College Observatory*, 60, 87
- Lemaître, G. 1927, "Un Univers Homogène de Masse Constante et de Rayon Croissant Rendant Compte de la Vitesse Radiale des Nébuleuses Extra-galactiques", *Annales de la Societe Scitifique de Bruxelles*, 47, 49
- Lemaître, G. 1931, "Expansion of the Universe, A Homogeneous Universe of Constant Mass and Increasing Radius Accounting for the Radial Velocity of Extra-galactic Nebulae", *M.N.R.A.S.*, 91, 483
- Lemaître, G. 1934, "Evolution of the Expanding Universe", *P.N.A.S.*, 20, 12
- Lineweaver, C. H., & Davis, T. M. 2005, "Misconceptions about the Big Bang", *Scientific American*, 292, 36
- Livio, M. 2011, "Lost in Translation: Mystery of the Missing Text Solved", *Nature*, 479, 171
- Luminet, J.-P. 2013, "Editorial Note to: Georges Lemaître, A Homogeneous Universe of Constant Mass and Increasing Radius Accounting for the Radial Velocity of Extra-galactic Nebulae", *General Relativity and Gravitation*, 45, 1619
- Lundmark, K. 1925, "The Motions and the Distances of Spiral Nebulae", *M.N.R.A.S.*, 85, 865
- Nussbaumer, H., Bieri, L., & Sandage, A. 2009, "Discovering the Expanding Universe", Cambridge, UK: Cambridge University Press.
- Peebles, P. J. E. 1993, *Principles of Physical Cosmology*, Princeton University Press
- Perlmutter, S. et al., 1999, "Measurement of Ω and Λ from 42 High-Redshift Supernovae", *Ap.J.*, 517, 565
- Planck Collaboration, 2013, "Planck 2013 results. XVI. Cosmological Parameters", *arXiv:1303.5076*
- Riess, A.G. et al., 1998, "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant", *A.J.*, 116, 1009

- Rubin, V. C., Ford, W. K. J., & Thonnard, N. 1980, "Rotational Properties of 21 SC Galaxies with a Large Range of Luminosities and Radii, from NGC 4605 ($R = 4\text{kpc}$) to UGC 2885 ($R = 122\text{kpc}$)", *Ap.J.*, 238, 471
- Sandage, A. 1958, "Current Problems in the Extragalactic Distance Scale", *Ap.J.*, 127, 513
- Slipher, V.M. 1917, "Nebulae", *Proceedings of the American Philosophical Society*, 56, 403
- van den Bergh, S. 2011, "The Curious Case of Lemaître's Equation No. 24", *J.R.A.S.C.*, 105, 151
- Wirtz, C. 1924, "De Sitters Kosmologie und die Radialbewegungen der Spiralnebel", *Astronomische Nachrichten*, 222, 21
- Zwicky, F. 1937, "On the Masses of Nebulae and of Clusters of Nebulae", *Ap.J.*, 86, 217

EL TIEMPO Y SU MEDIDA



LA MEDIDA DEL TIEMPO

Pablo de Vicente Abad

*Observatorio Astronómico Nacional
Instituto Geográfico Nacional - Ministerio de Fomento*

Abstract

The measure of time is a quest that has accompanied mankind since its origin thousands of years ago. Although the nature of time is still unknown, humans have made huge progress in the techniques to measure it. Such is the precision reached in the measurement of time that some of our basic measurement units, like the meter, depend currently on measurements of time. This article makes a historical trip from ancient calendars to clocks, making some stops on scientific and technical details that have played an important role on the development of the current time system.

Introducción

Escribir sobre el tiempo resulta mucho más difícil que hacerlo sobre la medida del tiempo. El tiempo es omnipresente y no ocupa espacio, se puede medir pero es inasible para los sentidos, todo el mundo comprende lo que es pero nadie sabe definirlo y parece que sólo crece en una dirección. El ahora es tan efímero que deja de serlo inmediatamente. Algunos filósofos y científicos, como Newton o Einstein, han dedicado muchos años a la búsqueda de una definición apropiada, de modo que parece bastante claro que escribir sobre el tiempo es más complicado que hacerlo sobre su medida.

El tiempo es una de las magnitudes fundamentales en Física y su unidad en el Sistema Internacional es el segundo. A primera vista resulta evidente que es diferente del resto de las magnitudes fundamentales, como

la masa o la longitud. Tradicionalmente la Oficina Internacional de Pesos y Medidas ha conservado patrones de las magnitudes fundamentales, como por ejemplo el patrón de longitud, una barra de platino e iridio, y el de masa, un cilindro del mismo material. En la actualidad la definición de la primera depende del tiempo y no está asociada a un elemento físico.

El tiempo se mide empleando eventos periódicos y midiendo cuantos de estos eventos periódicos transcurren mientras se desarrolla lo que se quiere medir. Parece claro que se deben elegir cuidadosamente los eventos periódicos en función de lo que se quiere medir.

Tradicionalmente los eventos periódicos más sencillos de usar han sido los asociados a los fenómenos astronómicos porque están siempre presentes en la vida de los humanos. El evento periódico por excelencia es la duración del día. Otro evento periódico de mayor duración son las fases de la Luna y uno mayor, la duración del año medida observando los ciclos de la naturaleza, es decir las estaciones, o la posición del Sol en el cielo en el mismo momento del día.

Por tanto podemos dividir los eventos periódicos en fenómenos de larga duración, que se emplean en la elaboración de los calendarios, y en eventos de corta duración basados en relojes que se emplean para la determinación de la hora. En este artículo se hablará por tanto sobre calendarios y relojes.

Definiciones básicas: el día, el mes y el año

El día es un fenómeno astronómico producido por la rotación terrestre y se corresponde aproximadamente con el tiempo que tarda la Tierra en completar dicha vuelta. Para medir el tiempo transcurrido es necesario utilizar un cuerpo celeste visible desde la Tierra. El más sencillo es el Sol.

En principio uno podría intentar definir el día solar como *el tiempo transcurrido entre dos posiciones «iguales» del Sol en el cielo*. Sin embargo si uno registra las trayectorias del Sol en una lámina a lo largo de un mismo año, a una cierta latitud, observará que el Sol describe trayectorias diferentes cada día. La técnica de registro de estas trayectorias se denomina solarigrafía y se pueden encontrar algunas de ellas en Internet. Las «fotografías» que se obtienen, producen líneas curvas que se corresponden con el camino seguido por el sol en el cielo a lo largo de los días. Cada día el Sol aparece por el horizonte, si nos encontramos fuera de los círculos polares, alcanza su máxima altura al Sur (si nos encontramos en el hemisferio norte) o al Norte (si nos encontramos en el hemisferio sur) y se oculta por el horizonte. A medida que nos desplazamos del invierno al verano la trayectoria registrada en el cielo se abomba hacia arriba y se estira por los lados, y en el intervalo de tiempo del solsticio de verano al de invierno la trayectoria se encoge hacia abajo y la trayectoria se hace más corta. En el hemisferio Sur el comportamiento es idéntico pero las

estaciones están cambiadas. Por tanto no existe una posición «idéntica» del Sol de un día para otro que nos permita decir que ha transcurrido un día.

Sin embargo es posible definir el día solar como algo más preciso advirtiendo que el Sol siempre culmina al Sur o al Norte, es decir alcanza su mayor altura a su paso por el meridiano local. Por tanto un día solar medio se puede definir como *el tiempo transcurrido entre dos tránsitos del Sol por el meridiano*. Los humanos han empleado diferentes momentos para marcar el comienzo del día: el paso por el meridiano (occidente en la antigüedad), la salida del sol (egipcios), la puesta del sol (sumerios y babilonios) o el paso por el meridiano opuesto o medianoche (en la actualidad).

Al igual que el Sol, el resto de los astros también dibujan una trayectoria en el cielo y alcanzan su máxima altura al pasar por el meridiano local. Se puede definir un día sidéreo medio como *el tiempo transcurrido entre dos tránsitos de una estrella por el meridiano*. Un día sidéreo dura menos que un día solar, porque la Tierra está girando en torno al Sol y tras realizar una rotación completa en torno a su eje (día sidéreo) debe recorrer un pequeño arco más ($\approx 1^\circ/365$) para que el Sol vuelva a estar en el meridiano del lugar. Es decir, el día solar es $1/365 \approx 0,027$ veces más largo que el día sidéreo. La figura de la página siguiente muestra un esquema de la posición del Sol, la Tierra y una estrella para explicar la diferencia entre día sidéreo medio y día solar medio.

Un mes lunar sinódico se puede definir como el intervalo transcurrido entre dos fases idénticas de la Luna, o dos lunaciones. Dado que las fases lunares dependen de la orientación entre la Tierra, la Luna y el Sol, el mes lunar se repite cada vez que, por ejemplo, se produce una luna nueva, es decir, cada vez que la Luna se sitúa entre la Tierra y el Sol. También existe un mes sidéreo, que es más corto que el mes sinódico ya que se corresponde con una rotación de la Luna en torno a sí misma, utilizando como referencia el fondo de estrellas. El mes sinódico dura 29,53 días y el mes sidéreo 27,32 días.

Finalmente el otro gran evento periódico es el año. El año trópico es el tiempo transcurrido entre dos equinoccios iguales (los de otoño o los de primavera) y por tanto está asociado al ciclo de las estaciones. Otro modo de expresar el año trópico es el tiempo transcurrido para que la Tierra se mueva entre dos pasos sucesivos por el primer punto de Aries. El primer punto de Aries es el lugar donde se corta la eclíptica con el ecuador terrestre en la dirección de la constelación de Aries. La eclíptica es el plano en el que orbita la Tierra en torno al Sol y forma aproximadamente 23 grados con el ecuador terrestre. Ambos planos se cortan produciendo una línea recta llamada línea de los equinoccios. Esta línea a su vez corta la órbita terrestre en dos puntos. El punto que se encuentra en la dirección de la constelación de Aries vista desde la Tierra se llama primer punto de Aries. Este punto no permanece fijo en el tiempo porque la Tierra tiene un movimiento de precesión y de nutación que hace que el punto de Aries se desplace a través

de la trayectoria de la órbita, realizando una vuelta completa al cabo de 26 000 años.

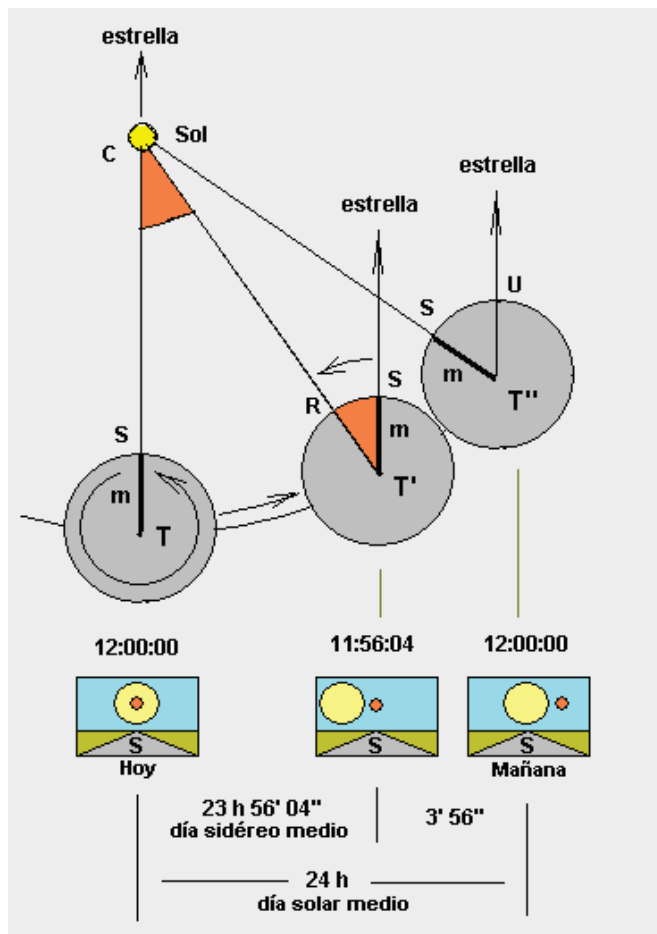


Figura 1: Día solar medio y día sidéreo medio. Se muestran la posición relativa del Sol, la Tierra y una estrella en tres momentos diferentes y desde dos perspectivas distintas: desde fuera de la Tierra y por encima del plano de la eclíptica, y el aspecto del Sol y de la estrella de referencia en el cielo desde la superficie de la Tierra. La escala no es correcta y está exagerada para poder apreciar la geometría fácilmente. La imagen es cortesía de Francisco Javier Blanco González (Wikimedia Commons).

El año sidéreo es el tiempo transcurrido para que la Tierra complete una vuelta al Sol tomando como referencia el fondo de estrellas y su duración es de 365,256363 días solares medios, diferente de la del año trópico. El año sidéreo está directamente relacionado con los signos del zodiaco. El signo del zodiaco indica la constelación que se ve tras el Sol durante una cierta época del año. Estas constelaciones se pueden distinguir al amanecer y al atardecer, cuando el brillo del Sol molesta menos a los observadores. Las constelaciones del zodiaco están situadas aproximadamente en el plano de la eclíptica y en diferentes direcciones. Debido a la precesión de los equinoccios el punto de corte entre la eclíptica y el ecuador varía lentamente respecto de esas constelaciones a lo largo de miles de años y por tanto el comienzo de cada estación sucede en posiciones diferentes de esas constelaciones. Sin embargo estos cambios son muy lentos y sólo se perciben a simple vista al cabo de miles de años. En la actualidad, el 21 de marzo el Sol aparece en la constelación de Piscis, y hace 4 000 años estaba en la constelación de Aries.

El problema de los calendarios

El objetivo de los calendarios es dividir el tiempo en unidades y subunidades que resulten prácticas para la sociedad. Idealmente las unidades son múltiplos de las subunidades pero esto no es posible si empleamos el año tropical como unidad y los meses lunares y los días solares medios como subunidades. El principal problema es que un año no dura un número entero de meses lunares y por tanto no se puede asociar un cierto mes lunar a una determinada época del año, algo que resulta muy útil para la organización de las fechas. Un año lunar, compuesto de 12 meses lunares, dura 354 días.

Las antiguas civilizaciones resolvieron este problema insertando meses o días de modo arbitrario o siguiendo una regla. Cada mes lunar recibía un nombre y estaba normalmente asociado a la constelación del zodiaco en la que estaba el Sol. Sin embargo como el año no está compuesto de un número entero de meses lunares era necesario duplicar algún mes cada cierto tiempo o añadir días extras.

A continuación se describen aquellos calendarios que han tenido importancia en el mundo occidental para comprender su evolución hacia el calendario civil que se utiliza en la actualidad.

Breve historia del calendario occidental

Existen tres tipos básicos de calendarios: los calendarios lunares, los lunisolares y los solares. Los primeros se rigen por las lunaciones y los meses se van desplazando a lo largo del año. Los calendarios lunisolares

intentan integrar las fases lunares con la duración del año, para conseguir que los meses lunares se produzcan siempre en la misma época del año. Los calendarios solares subdividen el año en partes iguales y están totalmente desconectados de las fases de la Luna.

Los calendarios lunares se han utilizado desde el paleolítico, porque son los más sencillos. Existen hallazgos arqueológicos, como los del hueso de Blanchard encontrado en una región de Francia, que demuestran que se utilizaban desde hace 27 000 años aproximadamente. El hueso de Blanchard muestra sobre su superficie las fases de la Luna durante tres meses. Este tipo de calendario resultaba adecuado para las sociedades cazadoras-recolectoras previas a la aparición de la agricultura.

Con la aparición de la agricultura, las estaciones del año se convierten en eventos muy importantes para las sociedades agrícolas, ya que son fundamentales los momentos de siembra y recolección. De hecho se comienzan a asociar las festividades religiosas con los actividades agrícolas más importantes. En el neolítico la sociedad se vuelve más compleja y aparece una diversificación de oficios y de estratos. La mayor parte de la población depende de la agricultura y de la ganadería. Aparece una clase dirigente y un ejército que mantiene el poder de la primera y que justifica su existencia en la protección que ofrece al resto de la sociedad, fundamentalmente campesinos, de los enemigos externos. Surgen los impuestos con la finalidad de sufragar al ejército, a la clase religiosa y a la clase dirigente. El calendario de los impuestos tiene que asociarse con la cosecha que es cuando los campesinos disponen de mayor riqueza. Por otra parte la cosecha se hace en el momento más propicio para la recolección, y cuando lo indica la clase religiosa. Este conocimiento reside en los sacerdotes que observan el cielo y determinan el momento del año, es decir de los solsticios y equinoccios, a partir de la posición del Sol respecto de las constelaciones del zodiaco.

En este tipo de sociedad surge la necesidad de establecer la subdivisión del año en partes aproximadamente iguales, asociadas a las estaciones del año y por tanto a los ciclos de la naturaleza, pero ligadas a las fases lunares que permitan una contabilidad más sencilla. Sin embargo los meses lunares no se pueden ajustar fácilmente con la duración del año, por lo que se buscan soluciones en las que se añade un mes lunar cada cierto tiempo.

La agricultura surgió en el creciente fértil y con ella florecieron civilizaciones como Sumeria y Babilonia. Los sumerios y los babilonios utilizaban un calendario lunisolar. El año comenzaba después de la cosecha y estaba dividido en meses de 29 o 30 días. Cada 3 o 4 años el rey, aconsejado por los sacerdotes que a su vez se regían por las observaciones, añadía un mes. Cada mes comenzaba después de una luna nueva, y cada día se dividía en 6 partes diurnas y 6 nocturnas, cuya duración era variable dependiendo de la época del año. Cada mes lunar recibía un nombre. Habitualmente el mes que se insertaba era Adaru, que se duplicaba y estaba asociado a la conste-

lación de Piscis (febrero-marzo en el calendario gregoriano) y en algunas ocasiones se duplicaba Ululu. Los nombres de los meses de este calendario se emplearon después en el calendario hebreo y algunos de ellos en el asirio.

En el año 499 a. de C., a partir de Dario I, los meses se comenzaron a regular de acuerdo con un ciclo de 19 años compuesto por 235 meses lunares. Este ciclo debe su nombre al griego Metón, que lo aprendió de los babilonios y lo progagó por Occidente. Los meses se insertaban de acuerdo con la siguiente regla: en los años 3, 6, 8, 11, 14 y 19 de un ciclo se añadía un segundo mes de Adaru; en el año 17 se añadía un segundo mes de Ululu. El ciclo de Metón indica el número de años en los que un mes lunar comienza el mismo día del año. Es decir, cada 19 años se produce luna nueva el mismo día del año, por ejemplo el 26 de junio. El ciclo de Metón se obtiene de calcular el mínimo común múltiplo de 365 y 29,5 días, las duraciones aproximadas de un año trópico y un mes lunar.

El primer calendario solar de la historia que se conoce se utilizó en Egipto. Egipto vivía gracias al Nilo y los desbordamientos anuales del río marcaban el ritmo de la sociedad egipcia, que vivía de la agricultura en sus orillas. Los años duraban 365 días y se dividían en 12 meses de 30 días más 5 días adicionales conocidos, en griego, como epágómenos. El año comenzaba con el orto helíaco de Sirio, que se asociaba con la crecida del Nilo en Memphis, capital del antiguo reino, y con el solsticio de verano. Ese día, Sirio emergía por el horizonte unos instantes antes y por el mismo lugar que el Sol. A esta coincidencia se le atribuyó un origen divino. Esta relación se desfasaba debido a la errónea duración del año que empleaban pero sabían que se repetía cada 1460 años, en lo que se conocía como ciclo sótico (de Sotis). Es decir, se producía una diferencia de 365 días en 1460 años que es equivalente a añadir 1 día cada 4 años. Este hecho se tuvo en cuenta mucho tiempo después para reformar el calendario egipcio.

En el año 238 a. de C. los sacerdotes y miembros de la clase religiosa se reunieron en Cánope y establecieron que el año trópico duraba 365 días y 6 horas. Resolvieron añadir un día a los epágómenos cada 4 años. La resolución tomada se encuentra escrita en tres lenguas en la Piedra de Cánope, en lo que se conoce como el decreto de Canopus. La reforma del calendario Ptolomeico no se llevó a cabo, posiblemente por diferencias entre los miembros de la clase religiosa, pero fue retomada doscientos años después por los romanos.

En el año 46 a. de C., Julio Cesar encargó a Sosígenes, un astrónomo alejandrino, que le asesorara para crear un nuevo calendario. Antes de la reforma el calendario romano era lunisolar. La inserción de meses y el cómputo de años era función de los sacerdotes que mantenían esta actividad en secreto y en ocasiones la ejecutaban de manera arbitraria según sus propios intereses.

Sosígenes conocía la fallida reforma de Cánope y propuso a Julio Cesar abandonar el calendario lunisolar y adoptar un calendario solar. La implantación del nuevo calendario se hizo en el año 45 a. de C.. Entonces se estableció que el año duraba 365 días y que cada 4 años se añadiría un día extra. Con el tiempo, estos años recibirían el nombre de bisiestos. Se dispuso que hubiera 12 meses de duraciones de 29, 30 y 31 días y se desplazó el comienzo del año de marzo a enero. Este año fue conocido como el año de la confusión porque se añadieron 85 días para poder corregir el desfase del calendario anterior, y por tanto el año tuvo 445 días. Hasta el año 8 DC el calendario juliano sufrió algunas pequeñas modificaciones por errores que se habían producido en el cómputo de los días.

El desplazamiento del comienzo del año de marzo a enero, se puede observar en el nombre de los meses: septiembre se refiere al mes siete, cuando en realidad es el nueve, y lo mismo ocurre con octubre, mes ocho, noviembre, mes nueve y diciembre, mes diez. El mes de Julio recibió su nombre para honrar a Julio Cesar y el de Agosto para hacer lo mismo con Augusto, que no quería ser menos que su tío abuelo Julio Cesar. De hecho para que Julio y Agosto tuvieran la misma duración de 31 días hubo que quitar un día a febrero que quedó con 28 días. Otros emperadores posteriores pretendieron cambiar los nombres de los meses, pero estas modificaciones no perduraron.

La semana de 7 días fue instaurada en el calendario juliano por Constantino el Grande, copiandola del calendario lunisolar judío. Decretó que el domingo fuese el día del descanso para adorar a Dios, ya que Jesucristo resucitó en domingo. Los nombres de la semana están asociados a los planetas, nombre en griego que significa errantes. Los errantes eran los astros más luminosos que se movían respecto del fondo de estrellas y se corresponden con los 5 planetas del Sistema Solar más próximos a la Tierra más el Sol y la Luna. En castellano, el domingo es el día del Sol, «sunday» en inglés y el sábado, nombre heredado de los judíos, es el día de Saturno, en inglés «saturday». La asociación del resto de los días de la semana con los planetas no requiere explicación porque es muy sencilla y se deja como ejercicio para el lector.

El calendario juliano comete un error de un día cada 125 años porque la duración del año trópico no es 365,25 sino 365,242189 días. En el año 1582, el papa Gregorio XIII promovió la reforma del calendario juliano porque la fiesta de Pascua se iba adelantando a lo largo del año. En el año 325, el concilio de Nicea estableció que el domingo de Pascua se debía celebrar el domingo siguiente a la primera luna llena después del equinocio de primavera. Todas las demás fiestas religiosas cristianas estarían referidas a esta y debido a la diferencia de duración del año entre el calendario juliano y el año trópico esta fiesta se adelantó 10 días desde aquel concilio. Para resolver el problema se estableció una nueva regla para los bisiestos: Un año es bisiesto si es divisible por 4 excepto si es divisible por 100 y

no lo es por 400. Con este ajuste la duración del año queda en 365,2425 días. Además ese año se eliminaron los 10 días del calendario que se había desplazado la fiesta de Pascua desde el concilio de Nicea; al día 4 de octubre de 1582 del calendario siguió el día 15 de octubre de 1582.

Esta reforma fue adoptada inmediatamente por los países de religión católica romana. Otros países de religión protestante, anglicana y ortodoxa lo fueron adoptando años e incluso siglos después. Algunos países ortodoxos, como por ejemplo Rusia, Bulgaria o Grecia lo implementaron en el siglo XX. Algunas iglesias ortodoxas todavía siguen utilizando el calendario juliano para la determinación de la fiesta de pascua y por eso la navidad se celebra en dichas iglesias en los primeros días de enero.

Día juliano y día juliano modificado

En 1583 J. Scaliger, un religioso holandés, propuso la contabilización del paso del tiempo empleando una cuenta continua de días y fracciones. A la unidad empleada le llamó “día juliano” y propuso como comienzo del día 0 el mediodía en Alejandría del 1 de enero del año 4713 a. de C.. Esta elección la hizo empleando algunos argumentos un tanto abstrusos para los contemporáneos. A finales del siglo XIX los astrónomos adoptaron la fecha juliana, o contabilización del paso del tiempo en días julianos, utilizando el meridiano de Greenwich en lugar del meridiano en Alejandría. Los días julianos comienzan a mediodía siguiendo el criterio de Ptolomeo, que contabilizaba los periodos astronómicos contando los días desde el mediodía. La elección se hizo porque el paso del Sol por el meridiano local ocurre siempre a la misma hora aparente, independientemente de la época del año, al contrario de lo que sucede con el amanecer o el atardecer, momentos, que como hemos dicho antes, empleaban algunas civilizaciones para determinar el comienzo del día.

La fecha juliana se utiliza en la actualidad para el cálculo de las efemérides astronómicas. Existen algoritmos sencillos que permiten pasar de una fecha de calendario gregoriano a día juliano y viceversa. El 1 de enero de 2000 a mediodía corresponde al día juliano 2 451 545,0. La fecha juliana es muy útil para calcular intervalos temporales porque estos se obtienen por simple substracción.

El “día juliano modificado” fue introducido por el Smithsonian Astrophysical Observatory en el año 1957 y se define como el día juliano menos 2 400 000,5. Por tanto los días julianos modificados comienzan a medianoche, como en el calendario civil. Este número también se emplea con frecuencia en astronomía para datar eventos.

La división del día

Para poder medir eventos cuya duración es menor que la de un día solar medio es necesario dividir este en fracciones. En la antigüedad los sumerios y babilonios dividían el día en doce partes: seis durante las horas de luz y seis durante la noche. Los egipcios dividían el día en doce partes diurnas y doce nocturnas. La subdivisión del día egipcio es por tanto el origen de las actuales horas. La división del día en doce o veinticuatro horas está relacionado con la utilización del sistema sexagesimal en el medio oriente. Se cree que procede de la costumbre de contar empleando el pulgar sobre las 12 falanges del resto de los dedos de esa mano. Además, el 60 es divisible por 1, 2, 3, 4, 5, 6, 10, 12, 15, 20 y 30 lo que resultaba muy útil y facilitaba el cálculo de fracciones. Este es el origen de la subdivisión de las horas, y los grados, en 60 minutos y estos en 60 segundos.

La duración de las horas diurnas y nocturnas para los antiguos era diferente y además dependía de la época del año. En invierno las horas nocturnas eran más largas que las diurnas y en verano ocurría lo contrario. De hecho la duración del día, entendido en este caso como la fracción diurna del día solar medio, y de la noche dependen de la latitud del lugar y de la época del año. Esta variabilidad en la duración de la parte diurna y nocturna del día se debe a la inclinación del eje de rotación de la Tierra respecto de la eclíptica. La inclinación, de 23 grados, también es el origen de las estaciones.

Este tipo de horas, llamadas temporarias o desiguales, se utilizaron hasta el siglo XIV, a pesar de que en el año 130 a. de C. Hiparcos propuso el uso de horas de igual duración. La desigual duración de las horas no tenía mayor importancia para las sociedades agrícolas porque el ciclo de vida de sus habitantes estaba regido por el amanecer y el atardecer.

Relojes antiguos

Los primeros relojes de los que se tiene noticia son los relojes de agua, también conocidas como clepsidras, y los relojes de sol.

Es posible que los relojes de agua fueran empleados por primera vez en China hace 4000 años. El primer reloj de agua del que se tiene constancia física es una vasija egipcia datada entre los años 1415 a 1380 a. de C. perteneciente al reinado de Ahmenotep III encontrada en uno de los templos de Karnak. Los relojes de agua eran vasijas cerámicas, habitualmente con forma de cono truncado, con un orificio en el fondo y marcas en su interior. La vasija se llenaba una vez y el agua, al ir descendiendo, descubría las marcas que indicaban el paso de las horas. No es correcto suponer que el flujo del agua fuera uniforme porque el peso de la columna de agua varía a medida que la vasija se vacía. Más adelante los griegos inventaron mecanismos para corregir este efecto.

Los relojes de agua se utilizaban de noche en Egipto, pero en la sociedad griega y romana se empleaban para fijar el tiempo de los oradores en los juicios y en las guardias nocturnas del ejercito romano. Se siguió utilizando durante centenares de años después como un método eficiente para medir el tiempo. Ctesibio, un inventor y matemático griego, concibió un reloj de agua que marcaba las horas empleando un puntero y un dial. El reloj también empleaba un mecanismo que garantizaba un flujo de agua uniforme. La figura 2 muestra un grabado del siglo XIX.

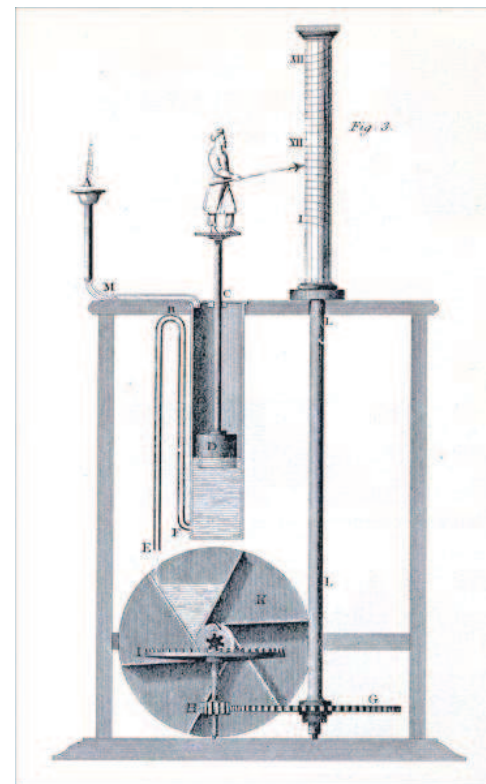


Figura 2: Diagrama de una clepsidra de Ctesibio según un grabado del siglo XIX. El indicador de las horas asciende a medida que lo hace el agua en el tubo. Cuando esta llega a su punto más alto, el tubo se vacía llenando el cangilón de abajo y se hace rotar una rueda dentada que, a su vez, mueve la lámina sobre la que apunta el puntero. La lámina contiene líneas que tienen en cuenta la desigual duración de las horas.

Los primeros relojes de sol de los que se tiene constancia física son los obeliscos, algunos de los cuales datan del año 3500 a. de C.. Los relojes de sol están hechos de un estilete llamado gnomón que proyecta su sombra sobre una superficie plana habitualmente colocada verticalmente u horizontalmente. A medida que el sol se desplaza por el cielo la sombra del gnomón recorre la superficie, que está grabada con marcas, que se corresponden con las horas del día. Si el reloj de sol es vertical el gnomón debe estar en el meridiano local y formar un ángulo con la superficie igual a la latitud del lugar. El mayor reloj de sol de la antigüedad fue el reloj de Augusto y fue descrito por Plinio el Viejo. La figura 3 muestra una reconstrucción hipotética del siglo XIX. El gnomón era un obelisco que en la actualidad está en la plaza de Montecitorio en Roma.

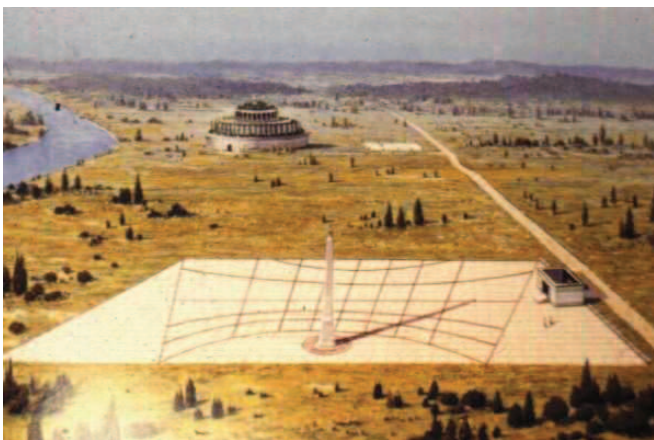


Figura 3: Reconstrucción del reloj de sol de Augusto según un grabado del siglo XIX. Estaba integrada con el Ara Pacis a lo largo de la Vía Flaminia.

Los relojes de sol marcan el tiempo solar aparente. El tiempo solar medio se obtiene a partir del día solar medio que es un promedio de los días solares aparentes a lo largo de un año trópico. La diferencia entre el tiempo solar medio y el tiempo solar aparente se conoce como ecuación del tiempo. El tiempo solar aparente, es decir el movimiento diario aparente del sol, no es uniforme debido a dos causas: la oblicuidad de la eclíptica que forma 23 grados aproximadamente con el ecuador terrestre como ya se mencionó anteriormente, y la excentricidad de la órbita elíptica de la Tierra en torno al Sol. La última causa es fácil de entender. La Tierra se mueve en torno al Sol siguiendo una órbita elíptica. De acuerdo con las ecuaciones de Kepler el movimiento de la Tierra en torno al Sol es más rápido en sus

proximidades de acuerdo con la ley de áreas iguales barridas en tiempos iguales. Cuando la Tierra está próxima al Sol los días son más largos que cuando está más alejada.

Un reloj de sol puede adelantarse hasta 16 minutos y retrasarse 14 minutos a lo largo del año. La figura 4 muestra la ecuación del tiempo en función del día del año. En el capítulo «Tiempos» de este anuario se incluye la ecuación del tiempo que permite corregir la medida de un reloj de sol y obtener el tiempo solar medio.



Figura 4: Representación gráfica de la ecuación del tiempo a lo largo del año. Los relojes de sol se adelantan hasta 16 minutos en Noviembre y se retrasan 14 minutos en febrero. Fuente: Wikimedia.

Relojes mecánicos

En la Edad Media comenzaron a aparecer relojes en Europa que usaban pesos y muelles. Los relojes medievales utilizaban un mecanismo de escape de paleta y balancín, también conocido como Verge y Foliot. Un peso unido a una cuerda enrollada en un cilindro movía una rueda de corona. La rueda de corona disponía de unos dientes en su circunferencia. La rueda giraba por el peso pero era detenida por unas paletas solidarias con un eje vertical que en su parte superior disponía de un balancín con dos pesos. El balancín oscilaba en un sentido y en otro y el eje solidario con el, permitía el paso de un diente de la rueda de corona. Este mecanismo se llama «de escape», ya que cada oscilación detiene la rueda y luego la libera brevemente, permitiéndola girar un poco. La frecuencia del reloj se ajustaba separando o acercando los pesos. La oscilación del balancín

producía un sonido rítmico característico de los relojes. La figura 5 muestra un diagrama del mecanismo de escape.

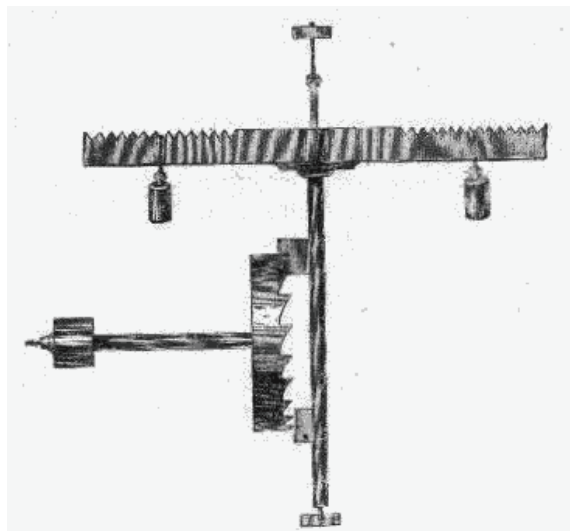


Figura 5: Mecanismo de paleta y balancín de Henri de Vick de 1379 (Paris). La imagen procede del tratado de 1849, «Historia de la Relojería» de Pierre Dubois. Fuente: Wikimedia.

A comienzos del siglo XIV el herrero alemán Peter Henlein inventó un reloj mecánico de muelle que utilizaba un mecanismo similar de rueda de escape. Estos relojes son los precursores de los relojes cronómetros que mencionaremos más adelante. Los relojes medievales mantenían el tiempo con un error de 15 minutos por día y por tanto no disponían de minuterio.

A comienzos del siglo XVII Galileo descubrió el isocronismo de los péndulos y propuso su uso para la medida del tiempo. Galileo advirtió que el período de oscilación de un péndulo no depende de la amplitud de la oscilación ni de la masa suspendida en su extremo, tan sólo de su longitud y de la gravedad. Los péndulos largos oscilan más lentamente que los cortos. Este periodo de oscilación «natural» es una propiedad que permitía la fabricación sencilla y repetitiva de relojes con un funcionamiento similar y por tanto poco error. Más tarde se conoció que el periodo de oscilación depende ligeramente de la amplitud de esta, pero cuanto menor es la amplitud menor es el error cometido.

En 1656 Cristian Huygens, astrónomo, físico y matemático holandés, fabricó el primer reloj de péndulo empleando un mecanismo de escape de Verge (rueda de corona) similar a los de los relojes medievales. Este reloj

mantenía el tiempo con un error de 10 segundos por día lo que supuso un avance considerable respecto de aquellos. Este momento supone un cambio drástico en la medida del tiempo. La figura 6 muestra un dibujo del propio Huygens que corresponde al segundo reloj de péndulo que fabricó en 1673.

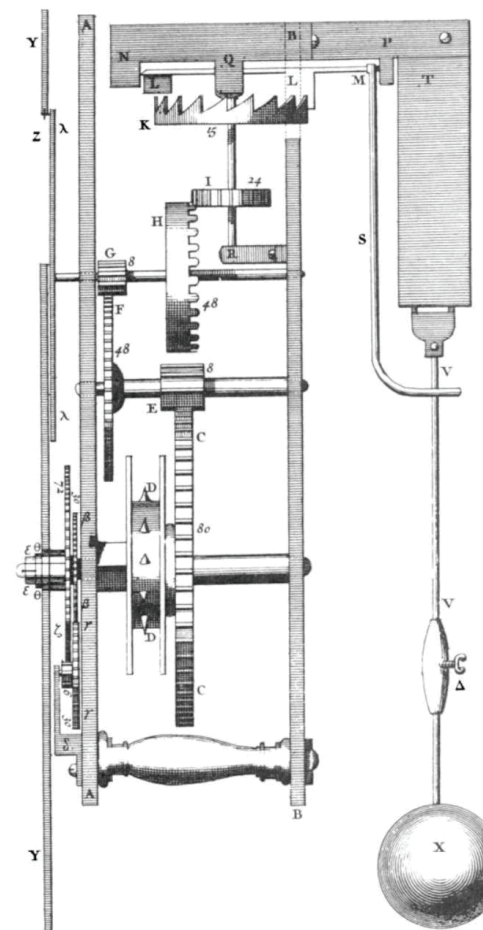


Figura 6: Diagrama de C. Huygens correspondiente al segundo reloj de péndulo que fabricó en 1673. El mecanismo de escape con la rueda de corona y las paletas están marcadas con las letras K y L en la parte superior del dibujo. Fuente: Wikimedia.

Los relojes de péndulo se emplearon hasta 1927. A lo largo de los siglos fueron mejorando constantemente. Se trabajó fundamentalmente en reducir la fricción, en minimizar las variaciones de longitud producidas por cambios de temperatura y en aportar energía de modo óptimo para mantener su oscilación. El péndulo más preciso que se fabricó fue el péndulo de Shortt que mantenía el tiempo en 2 ms por día. Se trataba de una combinación de dos péndulos: uno esclavo que suministraba energía y otro libre que oscilaba en el vacío y que actuaba de maestro.

El problema de la longitud

Desde el siglo XVI y sobre todo en el siglo XVII, Europa comenzó una expansión más allá del continente que dio lugar a un comercio marítimo muy importante. La navegación se convirtió en uno de los principales asuntos de estado. La dificultad de determinar la posición de los barcos en el mar era un problema muy grave que en ocasiones originaba el naufragio de navíos y grandes retrasos tras sufrir tormentas que modificaban la ruta de navegación de modo imprevisible. Todo ello generaba un considerable quebranto para los propietarios de los navíos, para la actividad comercial y para las naciones cuyos navíos de guerra sufrían los mismos problemas.

La posición sobre la superficie de la Tierra se determina estableciendo la latitud y la longitud del lugar. La latitud se obtiene, durante el día, a partir de la altura del sol sobre el horizonte a mediodía y por la noche, midiendo la altura de la estrella polar sobre el horizonte (en el hemisferio norte). La longitud se obtiene calculando la diferencia horaria entre la hora local y la de un lugar de referencia, por ejemplo la del Observatorio de Greenwich en Inglaterra. La hora local se puede determinar observando el sol y su paso por el meridiano. Sin embargo para conocer la hora de Greenwich es necesario disponer de un reloj que no se retrase ni adelante en exceso y que se pone en hora cuando el barco abandona el puerto. El péndulo no funcionaba bien en los barcos debido al balanceo de estos.

Hasta tal punto la determinación de la latitud era crucial que la corona inglesa estableció en 1713 un premio de 20 000 libras para quien construyese un reloj fiable que funcionara en el mar. John Harrison, un relojero inglés, empleó 40 años en la fabricación de cinco relojes, llamados cronómetros marinos. Tras una historia muy azarosa, descrita magníficamente por Dava Sobel en su novela «Longitud», pudo cobrar el premio completo. El H4, penúltimo reloj de la serie que fabricó, mantuvo el tiempo con un error de 54 segundos en una travesía de 5 meses de duración. El reloj era extraordinariamente complejo y empleaba un muelle como fuente de energía y un mecanismo de escape. Estaba preparado para minimizar la dilatación y contracción de su piezas por los cambios de temperatura. Harrison empleó rubíes y diamantes para minimizar el rozamiento. El H4 se conserva actualmente en el Museo Marítimo Nacional de Greenwich.

Los cronómetros marinos de Harrison son los precursores de los actuales relojes de pulsera de cuerda.

La calidad de un reloj

Un reloj es un dispositivo que produce un fenómeno periódico, contabiliza el número de ciclos transcurridos y los muestra en un visor. El dispositivo que produce un fenómeno periódico se llama resonador. Un resonador perfecto funcionaría eternamente, sin embargo, en la práctica la fricción hace que se detenga al cabo de un tiempo. Un resonador al que se aplica energía para que mantenga su oscilación se llama oscilador. Algunos resonadores son mejores que otros. El modo de medirlo es calcular el número de oscilaciones que realiza hasta que su energía disminuye una cierta cantidad respecto de la energía inicial que se le suministró. Este número se llama factor de calidad o «Q». Un resonador de Q alta tiende sólo a oscilar si lo hace en torno a su frecuencia natural.

La calidad de un reloj se mide fundamentalmente con dos parámetros: la precisión o estabilidad y la exactitud. Se denomina precisión a la capacidad de un oscilador para producir ciclos de duración similar. Se llama exactitud a la capacidad de un oscilador para producir ciclos de la duración deseada. Para comprender la diferencia podemos pensar en un reloj preciso, es decir cuyos ciclos son todos de duración similar, pero inexacto porque duran, por ejemplo, menos de lo esperado y por tanto se adelanta. Un reloj exacto pero impreciso sería aquel cuyos ciclos en promedio son de la duración correcta pero unas veces son más largos que otras y por tanto no tiene un comportamiento estable.

El factor de calidad de un reloj está muy ligado a la precisión y exactitud de un reloj. Un reloj que sólo oscila en su frecuencia natural y que por tanto mantiene su oscilación durante mucho tiempo es exacto y estable. Por tanto una Q alta implica mayor exactitud y precisión.

El valor de la Q oscila desde 1 000 para los relojes de pulsera económicos hasta 10^{10} para los fuentes de cesio.

Relojes de cuarzo

En 1880 Jacques y Pierre Curie descubrieron la piezoelectricidad. Los materiales piezoeléctricos se deforman al aplicar sobre ellos una diferencia de potencial eléctrico. Del mismo modo, un material piezoeléctrico con electrodos en dos caras opuestas, al sufrir una deformación vertical a la superficie de sus caras, genera una corriente eléctrica. Esta propiedad que más tarde, se emplearía en el desarrollo del sónar, es la base de la fabricación de los relojes de cuarzo. Entre los años 1918 y 1927, Nicholson, Cady, Harrison y Die, estadounidenses los tres primeros y británico el

último, jugaron un papel fundamental en el desarrollo de los relojes de cristal de cuarzo.

El reloj de cuarzo está formado por un cristal cortado de un modo especial que vibra al serle aplicado un voltaje de corriente alterna. La fricción interna del cuarzo es muy pequeña y la Q vale desde 10^5 hasta $2 \cdot 10^6$. La frecuencia de oscilación del cuarzo depende del modo en que está cortado, de su tamaño y de la frecuencia resonante con la que se excite. Las oscilaciones del cristal pueden ser desde varios miles de ciclos hasta varios millones de ciclos por segundo. El resonador de cuarzo se conecta a un sistema de realimentación de modo que la frecuencia de oscilación se mantiene muy cercana a la frecuencia natural de resonancia. Los mejores cristales de cuarzo mantienen el tiempo con un error máximo de 0,1 ms por día. Algunos de ellos funcionan en pequeños hornos (de tamaño ligeramente superior al del cristal) para mantener su temperatura estable e impedir la contaminación por el entorno exterior.

Los relojes de cuarzo se comenzaron a utilizar en Bell Telephone en 1932 y ese mismo año la Oficina de Patrones de EEUU (NIST) comenzó a utilizar un reloj de cuarzo como patrón para mantener la hora. La precisión de estos relojes permitió medir por primera vez variaciones estacionales en la rotación terrestre.

Estados Unidos, Suiza y Japón trabajaron en la utilización práctica de los cristales de cuarzo y contribuyeron a la aparición de los primeros prototipos de relojes de pulsera en la década de 1960, momento en el que comienza su popularización en la sociedad. Suiza llegó a emplear hasta 90 000 personas en la década de 1970 en la fabricación de los relojes de cuarzo. En la actualidad los relojes de cuarzo son omnipresentes, ya que no sólo están en las muñecas de muchas personas; también se utilizan en una gran cantidad de aparatos electrónicos como ordenadores, cámaras fotográficas y teléfonos móviles.

Relojes de atómicos

Los relojes atómicos son la última etapa en el desarrollo de la medida del tiempo. Se han otorgado 13 premios Nobel entre los años 1943 y 1997 por trabajos relacionados con los relojes atómicos, lo que nos da una idea de la magnitud del estudio y trabajo de científicos e ingenieros en este área de la Física.

Un reloj atómico está basado en la frecuencia natural de las partículas atómicas. Los átomos están compuestos de un núcleo y una nube de electrones a su alrededor que ocupan diferentes orbitales. El estado de energía de un átomo depende del estado de los protones, neutrones y electrones que lo componen. Por ejemplo, en un estado excitado un electrón puede estar en un orbital de mayor energía, habitualmente más lejano

del núcleo. Como dicho estado no es estable, el electrón pasa a otro orbital de menor energía, emitiendo, en dicho paso, un fotón de frecuencia equivalente a la diferencia de energía entre ambos estados. La emisión de ese fotón se podría considerar una «oscilación» natural del átomo. Un de los ejemplos más sencillos y comunes es la transición hiperfina del átomo de hidrógeno que sirve para ilustrar el funcionamiento de casi todos los relojes atómicos actuales.

El átomo de hidrógeno está compuesto por un protón y un electrón que orbita en torno a él. El protón y el electrón tienen una propiedad física denominada espín, asociada con su momento angular intrínseco, que puede tomar dos valores: $1/2$ o $-1/2$ en el caso del electrón y 1 , 0 o -1 en el caso del protón. El espín se descubrió porque al hacer pasar partículas atómicas por un campo magnético se advertía que las partículas se desviaban en varios haces. Cuando las partículas eran electrones estos se dividían en dos haces, dependiendo del valor del espín. Cuando en un átomo de hidrógeno bajo un campo magnético externo, el espín del protón y del electrón son paralelos (ambos tienen el mismo signo) el átomo está en un estado de energía ligeramente superior a cuando son opuestos. El átomo de hidrógeno tiende a pasar espontáneamente del estado de mayor energía al de menor, es decir el electrón cambia su espín, y en esa transición emite un fotón de longitud de onda de aproximadamente 1420 MHz de frecuencia. Los relojes atómicos hacen uso de esta propiedad y utilizan la radiación emitida para para controlar la señal de un reloj de cuarzo.

El reloj atómico más paradigmático es el de cesio. También se basa en una transición hiperfina causada por un cambio de espín en el electrón más exterior. El mecanismo utilizado para generar una señal útil es como sigue. Se hace pasar un flujo de átomos de cesio a través de un imán que divide los átomos en dos grupos. Aquellos en los que el electrón de la capa exterior tiene un espín paralelo al núcleo y aquellos cuyo espín es opuesto. Uno de los haces se inyecta dentro de un tubo, que en realidad es una cavidad resonante, sobre la que se aplica una radiación electromagnética de frecuencia similar a la del fotón que se emite con el paso de un espín del electrón al otro. La mayor parte de los átomos que atraviesan la cavidad, al recibir la radiación, cambian de estado. A este proceso se le llama absorción estimulada. A su salida del tubo otro imán los vuelve a separar en dos haces. Aquellos cuyo estado final era el que se deseaba conseguir se dirigen a un detector y los demás se descartan. El detector controla un cristal de cuarzo que a su vez controla la radiación electromagnética en el tubo de tal modo que cuando la señal en el detector es máxima, la señal del cristal de cuarzo no cambia y si desciende se modifica hasta conseguir aumentar el número de detecciones. Se trata pues de un mecanismo de realimentación para conseguir la señal que maximiza el cambio de estado en el cesio y que se corresponde con una frecuencia determinada de gran pureza.

El primer reloj atómico que se construyó fue en realidad un reloj mo-

lecular de amoniaco en 1949 en Estados Unidos. Este reloj, que utilizaba una transición molecular, presentó una Q menor de lo esperado y su desarrollo se abandonó. Los esfuerzos se dirigieron a la fabricación de un reloj atómico de cesio. El primer prototipo fue ideado por Polykarp Kusch (premio Nobel en 1955) y contruido por Harold Lyons en 1952 en Washington DC. Sin embargo presentó algunos problemas de funcionamiento y no sería hasta tres años después, en 1955, cuando Louis Essen y Jack Parry, a partir de un método nuevo de Norman Ramsey (también premio Nobel en 1989), finalizaron el primer reloj atómico operativo en Gran Bretaña. Durante un año fue el único reloj atómico funcional en todo el mundo.

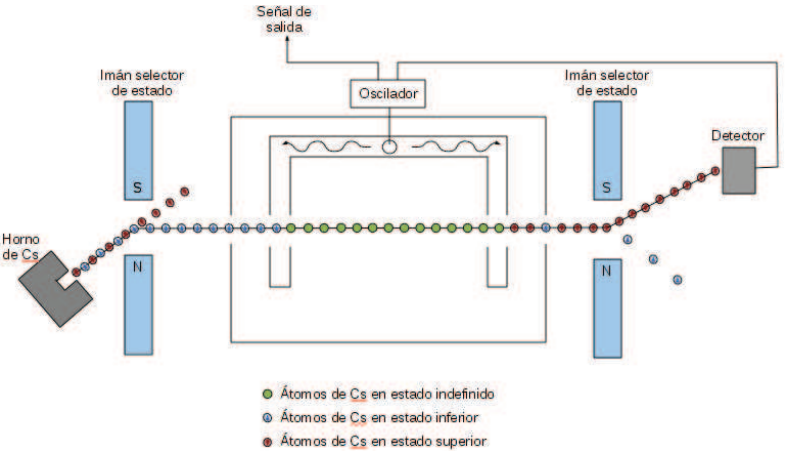


Figura 7: Esquema de un reloj de cesio. Un horno produce átomos de cesio en los dos estados descritos en el texto. Un imán selecciona los átomos dejando pasar hacia la cavidad resonante sólo aquellos que están en el estado inferior. La radiación estimula el paso del estado inferior al superior. Este proceso es tanto más eficiente cuanto más próxima es la frecuencia de la radiación a la de la transición hiperfina. A la salida se vuelven a separar los átomos entre aquellos que están en el estado inferior, y que no han cambiado a su paso por la cavidad, y los que estén en el estado superior. Estos últimos llegan a un detector que controla la frecuencia del oscilador que genera la radiación electromagnética en la cavidad. El sistema es realimentado de modo que la frecuencia del oscilador se modifica para maximizar el recuento del número de átomos en el estado superior. La frecuencia del oscilador es la que se emplea para generar una señal exterior.

En la actualidad los relojes atómicos comerciales tienen una calidad superior a la del cesio de Essen y Parry. Más tarde se han fabricado

relojes de cesio bombeados ópticamente, o fuentes de cesio que emplean mecanismos para reducir el ensanchamiento de la línea. El error que cometen es de aproximadamente 1 nanosegundo (10^{-9} segundos) por día. La incertidumbre en frecuencia más pequeña conseguida hasta la fecha es de 10^{-17} con fuentes de cesio. Otros relojes atómicos en uso son los de rubidio y los máseres de hidrógeno. Estos últimos se emplean habitualmente en observatorios astronómicos, como el Centro Astronómico de Yebes en España, que realizan observaciones de VLBI.

Los últimos desarrollos científicos y técnicos están dirigidos a conseguir relojes compuestos por un sólo átomo o ión de modo que la línea producida por la transición atómica sea extraordinariamente estrecha y produzca una línea muy pura, es decir muy precisa y exacta. También se están realizando trabajos para conseguir relojes atómicos de bolsillo.

Cuadro resumen con las fechas y la mejor precisión y factor de calidad de algunos relojes a lo largo de la historia

Reloj	Fecha	Precisión	Q
Paleta y Balancín	1200 – 1650	1000 s/día	
Péndulo	1650 – 1927	2 ms/día	10^5
Cronómetro	1761 –	0,2 s/día	10^3
Cuarzo	1927 –	0,1 ms/día	10^6
Reloj atómico de cesio	1955 –	10 ns/día	10^8
Máser de hidrógeno	1960 –	400 ps/día	10^9
Fuente de cesio	1999 –	100 ps/día	10^{10}

El cuadro anterior contiene un breve resumen sobre las fechas aproximadas de uso de diferentes tipos de relojes y su mejor grado de precisión o estabilidad y factor de calidad. Los péndulos en el aire tiene un factor de calidad, Q, de 10 000, mientras que en el vacío es de aproximadamente 100 000. En el cuadro figura el mejor valor (el más alto).

¿Cuánto dura un segundo?

En los siglos XVIII y XIX se habían ido acumulando evidencias, basadas sobre todo en la posición de la Luna, de que la rotación de la Tierra no es uniforme. El uso de los relojes de cuarzo en la década de 1920 permitió confirmar esa hipótesis ya que se observaron variaciones estacionales. Con el advenimiento de los relojes atómicos fue posible medir con mucho más detalle las irregularidades de la rotación de la Tierra.

En este momento se sabe que existen tres tipos de irregularidades:

- La velocidad de rotación de la Tierra se hace cada vez menor debido a los efectos de marea producidos fundamentalmente por la Luna.

Se cree que un día terrestre duraba 21 horas hace 600 millones de años. Como consecuencia la Luna se aleja de la Tierra.

- Los polos (asociados a la rotación terrestre) realizan un movimiento aproximadamente circular de varios metros de amplitud en un año respecto del polo asociado a la figura geométrica de la Tierra. Su movimiento se representa como un trazo sobre la superficie de la Tierra en el polo. Estos movimientos están probablemente causados por efectos estacionales y reorganización de las capas internas de la Tierra.
- Las fluctuaciones regulares están asociadas con las variaciones estacionales sobre la superficie de la Tierra. Estas variaciones tienen que ver con la conservación del momento angular y los cambios producidos por la evaporación del agua en los mares y su acumulación en hielos y nieve en las montañas y en su proceso inverso.

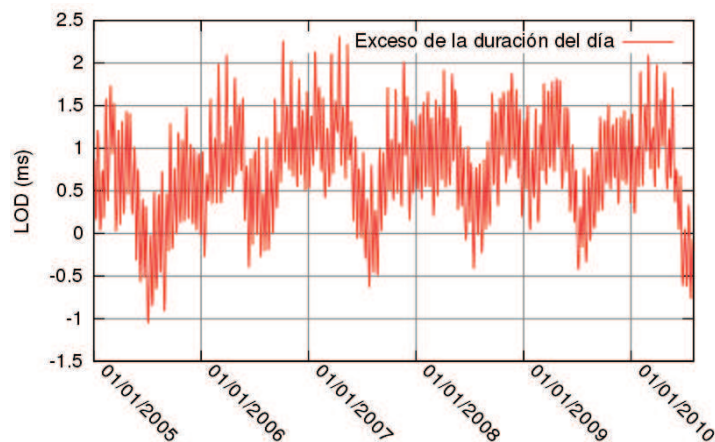


Figura 8: Variaciones en la longitud del día en los últimos 5 años. Las variaciones rápidas son debidas a efectos de marea. Las oscilaciones anuales que se observan corresponden a variaciones estacionales. Imagen generada a partir de los datos del IERS.

Todos estos efectos ocasionan que la escala de tiempo basada en la rotación terrestre, y a la que se llama Tiempo Universal o UT («Universal Time» en inglés), no sea adecuada por su falta de uniformidad y han originado los cambios de definición del segundo como unidad de tiempo. La escala UT es importante a efectos prácticos, porque la humanidad y todas sus actividades están ligada a la Tierra y su rotación.

Existen tres tipos de escalas de UT: UT0, que es la escala asociada a la rotación terrestre con todas sus irregularidades. UT1 que es la escala anterior corregida por el movimiento del polo. Esta escala es la que se emplea para realizar observaciones astronómicas y para el cálculo de las efemérides tanto de satélites artificiales como astros naturales. Finalmente la escala UT2 es un intento de hacer la escala más uniforme y de eliminar las fluctuaciones estacionales de la escala UT1. Apenas se utiliza.

Antes de 1960 el segundo, o segundo solar, se definía como la 86 400 parte de un día solar medio. Esto implicaba que se estaba utilizando una escala variable (UT) que requería un ajuste de vez en cuando. Entre 1960 y 1967 la unidad de tiempo era el segundo de efemérides que se definió como la 31 556 925,9747 parte del año trópico de 1900, y la escala de tiempo se denominó Tiempo de Efemérides (ET). Al contrario que en el caso anterior esta escala es uniforme pero requería muchas observaciones a lo largo de los años para compararla con la rotación terrestre, y poder determinar cuánto se alejaba de la escala UT.

En 1960 un conjunto de observatorios y laboratorios decidió establecer la escala UTC (Tiempo Universal Coordinado en inglés) como escala de uso civil. El objetivo era la utilización de una escala uniforme similar a la del tiempo atómico pero que se mantuviera sincronizada aproximadamente con la rotación terrestre. Se decidió que los años tuvieran el mismo número de segundos, pero dada la deceleración de la rotación terrestre esto significaba que cada año los segundos tenían una duración diferente. El ajuste se hacía cada vez que los segundos UT diferían más de una décima de los segundos ET. La escala UTC utilizaba por tanto una unidad de medida variable que se ajustaba a posteriori a partir de las observaciones. Esta estrategia resultó ser muy engorrosa y fuente continua de problemas y se abolió en 1971.

A partir de 1967 la IAU estableció como unidad de tiempo el segundo de Tiempo Atómico Internacional (TAI) que se define como la duración de 9 192 631,770 oscilaciones de la transición entre dos niveles hiperfinos del estado fundamental del cesio 133 a 0 K de temperatura. La escala de tiempos TAI está totalmente desacoplada de la rotación terrestre y de cualquier fenómeno astronómico, aunque el segundo se eligió para que coincidiese con el segundo de la escala ET, tras un programa de observación de tres años de duración que finalizó en 1958. La primera escala de tiempo atómico comenzó en junio de 1955.

Para resolver la falta de sincronismo entre el TAI y el UT, en 1971 se acordó que el Tiempo Universal Coordinado (UTC), como escala de uso civil, tuviese la misma unidad que el Tiempo Atómico Internacional y que se mantuviese en sincronismo con la rotación terrestre, intercalando o sustrayendo un segundo a final de año según fuera necesario. La regla actual dice que el UTC debe permanecer siempre a una distancia máxima de 0,9 segundos respecto de la escala UT1. En enero de 1972 la diferencia

entre el TAI y el UTC se estableció en 10 segundos y desde entonces ha ido creciendo hasta 34 segundos.

Uno de los problemas que el TAI heredó de la escala ET es que esta última se eligió de modo que un segundo sea igual a un segundo solar medio en 1900. La elección de esta fecha, aparentemente arbitraria, se debe a que fueron necesarias numerosas observaciones astronómicas hechas en los siglos XVIII y XIX para determinar su duración. Desde entonces hasta que se estableció la escala de Tiempo Atómico Internacional la deceleración de la rotación terrestre continuó y esto implica que la escala ET ya no reflejaba en la década de 1950, y aún menos en la actual, la duración del segundo UT2. Por tanto el segundo TAI, y el ET, son más cortos que el UT2 y esta es la principal razón por la que es necesario el uso del segundo intercalar.

La diseminación del tiempo

Actualmente la escala de tiempos TAI se realiza a partir de un grupo de 260 relojes atómicos situados en todo el mundo, calibrados por unos pocos patrones primarios de frecuencia. Los patrones primarios son el núcleo del sistema mundial de tiempos. Lo forman unos 10 relojes de cesio, fuentes de cesio, y Cesios bombeados ópticamente en Alemania, Estados Unidos, Francia y Japón. Los relojes comerciales sirven para generar el UTC local en cada país. La realización del TAI es función de la Oficina Internacional de Pesos y Medidas (BIPM) que emplea observaciones con satélites GPS para comparar los diferentes relojes.

El Servicio Internacional de la Rotación Terrestre (IERS) realiza medidas de la rotación terrestre y produce el tiempo universal UT1. Este tiempo se pasa a la BIPM que decide si añade o sustrae un segundo intercalar al TAI. El UTC se distribuye en una circular en la que se indica a 69 instituciones diferentes, las correcciones que deben hacer a las escalas de tiempos UTC locales. Estas instituciones retransmiten el UTC local corregido a la sociedad mediante diversas técnicas. En España el encargado del mantenimiento y diseminación del UTC local es el Real Observatorio de la Armada en San Fernando.

Existen tres esquemas básicos para la diseminación del tiempo desde un reloj maestro y la sincronización de otros relojes con él. La diseminación en una sola dirección es el método más usado. Desde el reloj maestro se emite una señal con el tiempo y el receptor sincroniza su reloj cuando la señal llega hasta él. El principal problema es el retraso desconocido introducido por la transmisión de la señal. La diseminación del tiempo en dos direcciones pretende resolver este problema. El emisor envía una señal de prueba con la hora del reloj maestro y el receptor envía de vuelta el eco. Entonces el emisor modifica el momento de su emisión para compensar por el retraso introducido en la propagación de la señal. El tercer

esquema utiliza el método de visión común. Un tercer reloj, que no debe ser necesariamente exacto, envía señales que el emisor y el receptor reciben. Entonces el emisor envía la señal del reloj maestro junto con una marca temporal correspondiente al momento de la emisión, según el reloj común. El receptor recibe la señal y la marca también con la hora del reloj común, realiza la diferencia entre ambas marcas, que indica el retraso producido por la transferencia, y corrige su reloj. Es importante que el receptor y el emisor conozcan su distancia al reloj común para que la corrección sea buena. Puede haber variaciones sobre estos tres esquemas básicos.

El método de una sola dirección se emplea para difundir el tiempo a la sociedad a través de campanadas, señales acústicas de TV y radio y ondas de radio especialmente dedicadas para ello. En el capítulo de Tiempos de este anuario hay un cuadro resumen con las emisoras de señales horarias de emisión continua. Se dividen en dos, de onda corta (5 a 15 MHz) y onda larga (60 a 78 kHz).

La sincronización a través de Internet se realiza a través de NTP. NTP (Network Time Protocol) es el protocolo de sincronización de Tiempo de Red, y puede emplear tres sistemas de diseminación que se corresponden básicamente con los esquemas descritos anteriormente. En uno de ellos el emisor de señales, y el receptor, habitualmente ambos son dispositivos con microprocesadores, intercambian paquetes marcados temporalmente con sus respectivos relojes para medir el retraso de la señal a través de Internet o de una red local. El cliente en el receptor corrige la señal recibida y pone el reloj del ordenador cliente en hora. NTP utiliza un sistema de estratificación en el que se clasifican los servidores de acuerdo con la calidad de sus señales. Los servidores «stratum 0» son relojes atómicos, receptores GPS o radios. Los servidores estrato 1, obtienen su hora de los estrato 0, y transmiten su hora a los de estrato 2.

Otro de los métodos de recepción de la hora son los receptores GPS. El GPS es un sistema de navegación global por satélites. El segmento espacial está formado por una constelación de 24 satélites distribuidos en 6 órbitas en torno a la Tierra, equipados con relojes atómicos. El segmento de tierra sirve para la operación y gestión de los satélites y para sincronizar los relojes en órbita. Los receptores GPS comerciales reciben la señal de al menos 4 satélites diferentes, permitiéndoles determinar la posición del receptor, y el tiempo UTC. El receptor procesa las señales de los diferentes satélites y, utilizando un algoritmo, corrige el tiempo de transmisión de la señal. La hora se puede conocer con una precisión de 1 nanosegundo.

Otra constelación de satélites actualmente en uso es el sistema ruso GLONASS. En el futuro cercano, estará operativo el sistema europeo Galileo.

El tiempo relativista

La precisión alcanzada en la medida del tiempo, hace necesario tener en cuenta los efectos relativistas sobre los relojes. Este hecho es especialmente relevante en los sistemas de navegación global con satélites como el GPS, GLONASS y Galileo, donde los relojes están en órbita. Finalizaremos este artículo del anuario con un breve repaso sobre la influencia de la relatividad especial y general sobre la determinación del tiempo.

La relatividad especial fue formulada por Albert Einstein en 1905. De acuerdo con ella no existe un marco de referencia absoluto, las leyes de la física son iguales en dos sistemas que se mueven con una velocidad de traslación uniforme entre ellos y la velocidad de la luz se emite a la misma velocidad independientemente de la velocidad del emisor de la luz. Este último principio es extraordinariamente importante para la sincronización de dos relojes en movimiento entre sí. De acuerdo con la relatividad especial un reloj que sufre una aceleración respecto de un reloj en reposo, marcha más lentamente que el que está en reposo. Este principio, conocido popularmente por la paradoja de los gemelos, se ha observado en relojes que se han trasladado en avión desde su emplazamiento original a otro y luego han regresado a su lugar de origen.

En 1915 Albert Einstein publicó la teoría de la relatividad general. Según ella la gravedad es una propiedad geométrica del espacio tiempo y la luz se desplaza siguiendo sus curvas geodésicas. Esto significa que en la práctica la radiación electromagnética, la luz, es atraída por un campo gravitatorio. Este efecto se ha visto confirmado por observaciones astronómicas. Muy relacionado con la propiedad anterior está el del enlentecimiento del tiempo en las proximidades de un cuerpo masivo. Este efecto también se ha confirmado midiendo la frecuencia de la señal de los relojes situados en valles y montañas. Un reloj situado primero en una montaña, donde la gravedad es menor, marcha más rápido que cuando se le sitúa después en un valle.

Las consecuencias de la relatividad especial y general sobre los relojes son especialmente importantes cuando están montados sobre satélites en órbita. Estos efectos se tienen en cuenta en su operación y sincronización con los relojes maestros en tierra. De hecho resulta curioso advertir que los relojes en órbita marchan más rápido que los terrestres por estar en un entorno de gravedad menor pero también marchan más lentos por encontrarse en rotación en torno a la Tierra. Se estima que un satélite GPS se atrasa 7 microsegundos por día por su movimiento relativo respecto de un observador en tierra, y se adelanta 45 microsegundos por día por la curvatura del espaciotiempo debida la masa de la Tierra. El efecto neto es un adelanto de 38 microsegundos por día que es necesario corregir. Otros efectos de orden menor también se tienen en cuenta.

Conclusión

El tiempo es probablemente la cantidad que los humanos han conseguido medir con mayor precisión, y como resultado, otras unidades básicas se pueden obtener a partir de él. Este es el caso del metro, que actualmente se define a partir del segundo, asumiendo que la velocidad de la luz es constante e igual para todos los observadores independientemente de su velocidad relativa. De hecho en astronomía las distancias se miden en tiempo. Por ejemplo, el Sol se encuentra a 8 minutos luz, y las estrellas más cercanas en Alfa Centauri están a 4,4 años luz de la Tierra. Existen intentos para definir la carga eléctrica a partir del tiempo empleando el efecto Josephson.

Todavía seguimos sin poder definir el tiempo pero la búsqueda de métodos de medida cada vez más precisos ha ensanchado nuestro conocimiento sobre él e impulsado el desarrollo científico y técnico de la humanidad.

Referencias

J. Jespersen, J. Fitz-Randolph, *From Sundials to Atomic Clocks*. Superintendent of Documents, U.S. Government Printing Office 1999.

T. Jones, *Splitting the Second. The Story of Atomic Time*. IOP Publishing 2000.

Marrison, *The Evolution of the Quartz Crystal Clock*. The Bell System Technical Journal, Vol. XXVII, pp. 510-588, 1948.

A. Pannekoek, *A History of Astronomy*. Dover Publications 1989.

Servicio Internacional de Rotación Terrestre y Sistemas de Referencia.
<http://www.iers.org>.

Oficina Internacional de Pesos y Medidas.
<http://www.bipm.org>.

Avances en la medida del tiempo

El artículo **La medida del tiempo** se publicó en el año 2011 y prácticamente todo su contenido es actual. Sin embargo se han producido cambios en dos asuntos que por su relevancia se mencionan a continuación.

Desde hace más de una década se viene discutiendo la utilidad del segundo intercalar en la escala de tiempos UTC y su posible abolición. Esta propuesta de la Unión Internacional de Telecomunicaciones (ITU) no ha reunido el suficiente consenso entre diversas instituciones y en noviembre de 2015 la ITU, en su conferencia mundial de radiocomunicación (WCR), decidió retrasar la decisión sobre los segundos intercalares hasta 2023.

La Unión Astronómica Internacional IAU, a instancias de la ITU-R, creó un grupo de trabajo para estudiar este asunto y publicó sus conclusiones (http://hpiers.obspm.fr/eop-pc/earthor/utc/report_WG.UTC_2014.pdf) en 2014. En las conclusiones no se emite ninguna recomendación sobre su abolición o permanencia pero sí se menciona que, en caso de eliminar los segundos intercalares, el adjetivo «Universal» de UTC debería desaparecer de esta nueva escala pues este está asociado a la rotación terrestre. Además se recomendaba la elección de un nuevo nombre para la nueva escala de tiempos y un periodo de 5 años para realizar la transición. Finalmente la declaración termina con la disolución del grupo de trabajo.

En la práctica, si el segundo intercalar desapareciese, la determinación de UT1 se haría exactamente igual que ahora excepto que en lugar de añadir la corrección DUT1 se añadiría un término creciente en el tiempo con la diferencia entre el TAI y el UT1. El Servicio de Orientación de la Tierra (IERS) seguirá informando sobre la duración del día y la determinación de la diferencia entre UT1 y el TAI estará disponible en sus boletines electrónicos y en papel. Sin embargo numerosos equipos que actualmente transmiten UT1 dejarían de hacerlo y requerirían una modificación que implementara estos cambios.

La principal consecuencia de la abolición de los segundos intercalares es que la duración del día ya no estaría asociada a la rotación de la Tierra quedando la definición de día como la duración de 794 243 384 928 000 oscilaciones hiperfinas del cesio 133. Realmente el impacto social se haría visible con el paso de los años pero no de unos pocos, sino de centenares o miles de ellos. Téngase en cuenta que la diferencia entre el TAI y el UTC es en este momento de 37 segundos, correspondiendo 10 segundos a una diferencia que se introdujo en 1972 y el resto a 27 segundos intercalares

desde entonces. Parece claro que en un futuro no muy lejano la humanidad utilizará una escala de tiempos atómica que tendrá pleno sentido cuando colonicemos otros mundos. De cualquier modo, el hecho de que tras 15 años de discusiones no se haya llegado a una decisión quiere decir que el paso de la definición del día asociada al Sol a una asociada a un reloj atómico todavía no genera un consenso suficientemente amplio.

Mientras tanto, los desarrollos tecnológicos siguen adelante y la medida del tiempo va mejorando e incluso puede que antes de que se tome una decisión sobre el segundo intercalar la definición de segundo de tiempo haya cambiado de nuevo. De hecho existe una cierta probabilidad de que ésta se produzca a finales de esta década.

La utilización de relojes basados en fuentes de cesio ha sido sustituida por relojes de entramado óptico. Los relojes atómicos tradicionales han sido relojes de «microondas», es decir, basados en transiciones cuánticas en esa región del espectro, en el caso del cesio unos 9,2 GHz. En los últimos años la emisión por transiciones cuánticas se desplaza al rango visible, cuya frecuencia es de unos 400 THz, cinco órdenes de magnitud superior. Se trata de una tecnología que proporciona una mayor estabilidad y exactitud.

Los avances se han producido en dos frentes: los relojes de iones individuales atrapados en una trampa electromagnética y los de miles de átomos neutros capturados en un entramado óptico generado por un láser. Estos últimos generan más potencia. En la actualidad se han conseguido estabilidades de una parte en 10^{15} en unos pocos segundos, similar a la conseguida por los relojes atómicos tradicionales de microondas pero después de promediar las medidas hasta 5 días. Si los tiempos de promediado aumentan se pueden llegar a conseguir estabilidades de algunas partes en 10^{18} con relojes de ión iterbio o de estroncio. Hasta la fecha la mejor estabilidad se ha conseguido con dos relojes de ión iterbio en el National Institute of Standards and Technology (NIST). Los relojes de microondas han sido hasta hace muy poco los más precisos pero no los más estables lo que llevaba al uso de otros relojes más estables, como los máseres de hidrógeno, que se calibraban periódicamente con los primeros. Los relojes «ópticos» son más estables y más precisos y esto cambiará la práctica de su uso y el modo en el que se distribuye el tiempo.

Las aplicaciones prácticas son numerosas y algunas de ellas desconocidas todavía. Estos relojes podrán ser utilizados en experimentos astronómicos de interferometría óptica, en la detección de ondas gravitacionales, en geodesia de precisión por su capacidad de medir variaciones locales del campo gravitatorio, así como en medidas de precisión de campos eléctricos, magnéticos, movimiento, fuerza o temperatura. Los relojes «ópticos» incluso tienen implicaciones en cosmología: sirven para acotar potenciales variaciones minúsculas en las constantes fundamentales de la naturaleza.

LA HORA OFICIAL EN ESPAÑA Y SUS CAMBIOS

Pere Planesas

*Observatorio Astronómico Nacional
Instituto Geográfico Nacional - Ministerio de Fomento*

Abstract

In this article we collect all the legal time changes –including daylight saving time– occurred in Spain since 1901, listing the date and time of the hour shifts, the Decree or Order identification, and the publication date. The future of the legal time in Spain is briefly discussed.

Establecimiento de una hora universal

En este artículo se recogen todos los cambios de la hora oficial habidos en España desde su implantación, documentando la fecha y la hora de su aplicación e identificando el Decreto u Orden correspondiente así como el lugar y fecha de su publicación. Se discuten brevemente varias opciones para el futuro de la hora oficial en España.

La hora oficial es puramente convencional, no estando ligada estrictamente a la hora solar local (que depende de cada lugar), sino que pretende unificar la hora en un territorio dado. En una primera instancia, pudo corresponder a la hora solar local media de la capital de un país (en ocasiones, la de su observatorio astronómico) o de la ciudad principal de un territorio. A lo largo del siglo XIX fue haciéndose cada vez más patente la necesidad de coordinar a nivel internacional el establecimiento de las horas nacionales, referidas a una hora universal de referencia. Hubo estudios, progresos a nivel nacional y reuniones de trabajo internacionales, siendo las más influyentes la de Roma, en 1883, y la de Washington al año siguiente.

En la Conferencia Internacional sobre el Meridiano que tuvo lugar en Washington en octubre de 1884 se aprobaron siete resoluciones, entre ellas: (I) la de establecer un meridiano único de referencia como origen de la longitud geográfica, (II) siendo este el del observatorio astronómico de Greenwich, y (IV) la adopción de un día universal, (V) siendo este el día solar medio de Greenwich, iniciándose a medianoche y contando las horas de 0 a 24. España votó afirmativamente a las resoluciones citadas, excepto la (V). Japón fue el único de los 25 países asistentes en adoptar legalmente estas resoluciones de manera casi inmediata, el 1 de enero de 1888. La falta de un procedimiento establecido para la adopción de tales resoluciones y las reticencias de algunos países en adoptarlas se tradujeron en un retraso y descoordinación entre los países del mundo en cuanto a la implantación de las distintas resoluciones. El proceso fue gradual y se extendió durante casi un siglo, en el que incluso el tiempo de referencia cambiaría.

Paralelamente se fue implantando un sistema de husos horarios mundial. Su origen se encuentra en la estandarización de los horarios de trenes, iniciada en 1840 y completada en 1847 en Inglaterra y Escocia (*Railway Time*), seguida de Holanda, Francia y, en 1883, en el extenso territorio de los EEUU (*Standard Railway Time*). En menos de diez años, otros países europeos habían unificado sus horarios de trenes, influyendo con ello en los horarios civiles, y la mayoría lo hizo estableciendo un horario que difiriera de la hora solar media del meridiano de Greenwich (GMT) en un número entero de horas. Curiosamente, territorios del extenso imperio británico (por ejemplo, India o Australia), que tanto podía beneficiarse del establecimiento de horas civiles adaptadas a las 24 zonas horarias (o husos horarios), tardaron más en adoptarlas.

Impulsos adicionales fueron dados en la década de 1910, con el inicio de la emisión de señales horarias estandarizadas basadas en la hora solar media del meridiano de Greenwich, la adopción de dicho meridiano para los mapas marinos de las naciones europeas con costa y la aceptación por parte de EEUU y el Reino Unido de la propuesta francesa de extender los husos horarios a mares y océanos, que daría lugar a la publicación, en Francia, del primer planisferio de husos horarios en 1917 (en España véase Fig. 1). Progresivamente, todos los países, uno a uno, han adoptado los husos horarios como base para la hora legal en sus territorios. En general, esta difiere de la hora universal en un número entero de horas, aunque en algunos casos difiere en un número entero de medias horas (India) e, incluso, de cuartos de hora (Nepal). La expresión “Tiempo Universal” (UT), aplicada al tiempo de referencia, fue adoptada por la Unión Astronómica Internacional en 1928 para su uso en cuestiones científicas, dando lugar a varias versiones (UT0, UT1, UT2) a medida que ha mejorado el conocimiento de la rotación de la Tierra. El 1 de enero de 1972 el Tiempo Universal Coordinado (UTC), basado en el Tiempo Atómico Internacional corregido de segundos intercalares, se convirtió en la base del tiempo civil en el mundo.

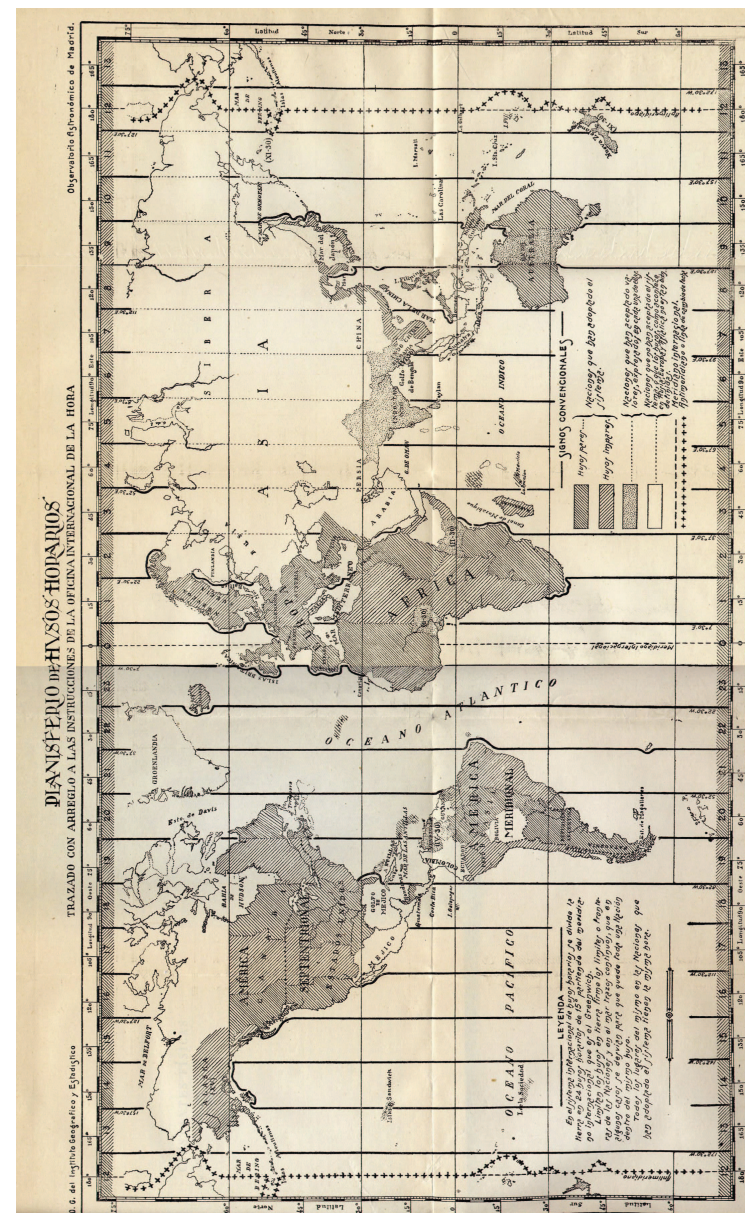


Figura 1: Planisferio de husos horarios terrestres y marítimos publicado en 1921 en el *Anuario del Observatorio de Madrid para 1922*.

La implantación de la hora oficial en España

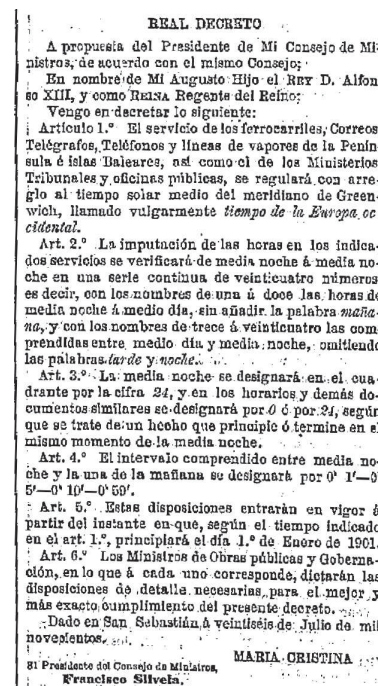
La adopción del meridiano de Greenwich como referencia para la longitud geográfica y para la hora oficial se produjo de manera muy escalonada. En efecto, nada más iniciarse el siglo XX España adoptó el tiempo solar medio en el meridiano de Greenwich como referencia para su hora legal. En cambio, la adopción de dicho meridiano como referencia para las longitudes en el mar o en tierra fue más tardía. España adoptó como primer meridiano el de Greenwich para los usos de la marina mediante el Real Decreto de 3 de abril de 1907 (Fig. 2b), publicado en la *Gazeta de Madrid* del día 6. En particular, de 1910 en adelante las efemérides del *Almanaque Náutico* se calcularon con referencia a dicho meridiano. En cuanto a la cartografía oficial, no fue hasta la publicación del Decreto 2303/1970, de 16 de julio, publicado en el BOE de 24 de agosto, cuando se sustituyó el meridiano de Madrid (que pasa por el Observatorio Astronómico) por el meridiano de Greenwich para el Mapa Nacional, base de toda la cartografía civil y militar del país.

La adopción del tiempo referido al meridiano de Greenwich fue más temprana. En un Real Decreto (RD) de 26 de julio de 1900 (Fig. 2a), publicado en la *Gazeta de Madrid* de 28 de julio y firmado por la reina regente María Cristina de Habsburgo-Lorena, se estableció que los servicios de ferrocarriles, correos, telégrafos, teléfonos y líneas de vapores de la Península e islas Baleares, así como los servicios de los ministerios, tribunales y oficinas públicas, se regularan a partir del día 1 de enero de 1901 con arreglo al tiempo solar medio del meridiano de Greenwich, conocido entonces vulgarmente como *tiempo de la Europa Occidental*. Además se estableció que las horas se contaran en una serie continua de 24 iniciándose a medianoche, instante que se designaba por la hora 0 o por 24 según se tratara de un hecho que se iniciaba o terminaba en el mismo momento de la medianoche, respectivamente.

Unos meses después, en una Real Orden (RO) de 16 de noviembre de 1900 (*Gazeta de Madrid* de 20 de noviembre), del “Director general de Obras públicas del Ministerio de Agricultura, Industria, Comercio y Obras públicas”, se detallaron instrucciones para la aplicación práctica del RD en el caso particular del servicio de ferrocarriles, que entonces se regía por la hora de Madrid. En particular, tuvieron que retrasar su salida en un cuarto de hora a fin de seguir enlazando con los de Francia y Portugal, países que no realizaron el cambio de hora de referencia hasta 1911. Además, a las 23^h 45^{min} de la noche del 31 de diciembre hubo que adelantar todos los relojes de las estaciones en quince minutos. También hubo que añadir a los relojes de muchas estaciones una numeración adicional escribiendo las horas de las 13 a las 24, en tinta roja.

Las pretensiones del gobierno eran, como se detalla en la exposición previa al RD, eliminar las diversas horas locales o regionales todavía en

uso en las distintas provincias españolas, unificando el tiempo en toda la península y las Baleares en una hora nacional, y participar en el establecimiento de unas horas internacionales con meridiano inicial único, a fin de “realizar armonías internacionales exigidas por la vida de relación, cada vez más frecuente, hasta llegar á la intimidad por todos deseada.”



Gaceta de Madrid.—Num. 96

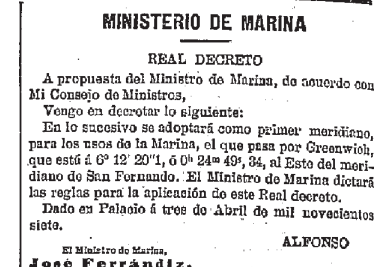


Figura 2: (a) Real Decreto de 26 de julio de 1900, tal como se publicó en la *Gaceta de Madrid*, núm. 209, pág. 384. (b) Real Decreto de 3 de abril de 1907, *Gaceta de Madrid*, núm 96, pág. 78.

Hora oficial, hora legal, hora civil

En la normativa referida al tiempo se encuentran las expresiones “hora oficial” y “hora legal”, siendo mucho más frecuente esta última. El *Diccionario de la Lengua Española* (22^a edición) de la RAE define la primera como la hora establecida en un territorio por decisión de la autoridad competente con adelanto o retraso con respecto a la solar. Fundéu (Fundación del español urgente) difiere: “en el uso técnico, la hora legal es la adoptada mediante disposiciones legales. En cambio, la hora oficial es la hora civil

del meridiano central del huso correspondiente. Es decir, la oficial es la del huso horario, mientras que la legal modifica la oficial, principalmente para adaptarse a los límites o las fronteras.” En esta explicación se cita la “hora civil” en su sentido astronómico tradicional, es decir, se trata de la hora solar media referida a la medianoche. En la vida cotidiana se denomina “hora civil” al tiempo reglamentario o legal establecido por las autoridades civiles (un gobierno o una administración pública), o sea, el que indican los relojes.



Figura 3: En la Unión Europea se utilizan oficialmente 3 husos horarios (0, +1 y +2), a pesar de que sus territorios abarcan también el -1 y aun otros, si se incluyera Groenlandia. (Adaptado de Wikimedia Commons.)

En el *Vocabulario Científico y Técnico*, publicado en 1983 por la Real Academia de Ciencias Exactas, Físicas y Naturales (RAC), el tiempo oficial se define como la “hora civil del meridiano central del huso correspondiente”, mientras que el tiempo legal se define como el “tiempo oficial [sic] adoptado por un determinado país mediante las oportunas disposiciones legales.” En algún texto de astronomía se pueden encontrar las definiciones opuestas: “Hora legal es la que resulta de la división del globo terrestre en 24 husos horarios. Hora oficial es la hora del huso rectificada según las conveniencias de cada país” (*Apuntes de Astronomía*, J.J. de Orús y M.A. Catalá).

El uso de tal variedad de expresiones se puede encontrar en las siguientes órdenes o decretos que, tras el de 1900, han modificado algún aspecto del horario legal, bien sea para cambiar su diferencia con el tiempo de referencia o para establecer uno nuevo. Veremos que las referencias a la hora oficial no se ajustan a lo definido por Fundeu o RAC, ni las referencias a la hora legal a lo definido por Orús y Catalá.

La Orden de 7 de marzo de 1940, publicada en el BOE del día siguiente, dispuso que el día 16 la *hora legal* se adelantara en una hora, a fin de que el horario nacional marchara de acuerdo con los de otros países europeos, en particular con el de Francia, que había adelantado su horario legal el 25 de febrero anterior. Este cambio, que pretendía ser transitorio, resultó ser definitivo, manteniéndose aún hoy en día, de forma que desde entonces la hora legal en España va ligada a la “hora de Europa central” (CET), que corresponde al huso horario número +1. Como puede verse en la figura 3, la mayor parte del territorio español se encuentra en el huso número 0 (el que está centrado en el meridiano de Greenwich), mientras que Galicia y las islas Canarias se encuentran en el huso número -1.

El Real Decreto 2781/1976, de 30 de octubre (BOE de 8 de diciembre), estableció que la escala de Tiempo Universal Coordinado (UTC) se considerara la base de la *hora legal* en España. Se cita que la XV Conferencia General de Pesas y Medidas, en su tercera sesión celebrada el 30 de mayo de 1975, adoptó por unanimidad la Resolución 5 en la que considerando que el UTC es ampliamente usado y constatando que constituye la base del tiempo civil, cuyo uso es legal en la mayor parte de los países, se estimaba que su uso es perfectamente recomendable.

En el Artículo 149 del Capítulo tercero del Título VIII de la Constitución española de 1978 se menciona que el Estado tiene competencia exclusiva sobre la determinación de la *hora oficial*. (Esta es una de las pocas menciones a la hora oficial.)

En el RD 1308/1992 de 23 de octubre (BOE de 24 de noviembre), que desarrolla el aspecto del tiempo de la Ley 3/1985 de Metrología, se establece que el Laboratorio del Real Instituto y Observatorio de la Armada (ROA) tiene como misión el mantenimiento de la unidad básica de tiempo (o sea, el segundo), declarado a efectos legales como Patrón Nacional de dicha unidad, así como el mantenimiento y difusión oficial de la escala de Tiempo Universal Coordinado (UTC(ROA)), considerada a todos los efectos como la base de la *hora legal* en todo el territorio nacional.

Menciones a la *hora oficial* y a la *hora oficial a nivel europeo* aparecen en el RD 4/2010 de 8 de enero (BOE de 29 de enero), regulador de la interoperabilidad en el ámbito de la administración electrónica. Así ocurre también en las páginas web del ROA, que se refieren a la *hora oficial*.

En definitiva, las expresiones hora legal y hora oficial se usan indistintamente en el BOE para referirse a la hora civil usada en la vida cotidiana.

La hora oficial en las islas Canarias

En el RD de 1900, que estableció la hora legal en España, no se hacía mención de las islas Canarias ni de los territorios españoles en el continente africano. Este olvido se subsanó en parte con el RD de 11 de febrero de 1922 (GM de 12 de febrero) en el que se estableció que, a partir del 1 de marzo, el servicio de los gobiernos civil y militar, tribunales, correos, telégrafos, teléfonos, líneas de vapores y demás transportes y oficinas públicas se regularan con arreglo al tiempo solar medio correspondiente a su huso horario, o sea, con una hora de retraso en relación con la península.

Con ello se subsanaba una situación confusa, por ser distintas las horas observadas en distintas islas e incluso distintas las horas de distintos servicios, alguno de los cuales se regía por el horario peninsular. En una misma ciudad se podían estar usando simultáneamente distintas horas (hora de Greenwich, hora del meridiano de la ciudad, hora del reloj de la catedral). La búsqueda de una solución fue propiciada por reiteradas preguntas del Almirantazgo británico realizadas en 1921 sobre si se había establecido en las islas Canarias una hora oficial de acuerdo con el huso horario correspondiente (Fig. 3).

En la sucinta Orden de 7 de marzo de 1940 antes citada, por la que se adelantó transitoriamente la hora legal en 60 minutos, no se hacía mención explícita a la península o a las islas Canarias, por lo que se sobreentendía que era de aplicación a todo el territorio español. Desde entonces, la hora normal en Canarias está adelantada con respecto a su huso horario, correspondiendo en su lugar al huso 0.

La hora oficial en Galicia

Por la misma razón, desde 1940 la hora de referencia en la península no es la del huso horario que le corresponde (el 0) sino que, como en el caso de Francia, Bélgica, Holanda y Luxemburgo, se ha adoptado la del huso +1. Como consecuencia de ello, en la mitad más occidental de Galicia, que se encuentra en el huso -1, hay un gran desacuerdo, el máximo en Europa, entre la hora solar local y la hora oficial. En verano, con la implantación del horario adelantado del que se hablará a continuación, se da la circunstancia de que en algún municipio la diferencia entre el mediodía solar verdadero y las 12^h oficiales llega a exceder los 160 minutos. Sólo en el oeste de China se supera tal desfase.

En 2006 un grupo político discutió la conveniencia o no de que Galicia adoptara la hora oficial vigente en Portugal, que es también la del Reino Unido, Irlanda, Islandia y las islas Canarias. Aún cuando hay muchos países con diferentes horas legales en distintas partes de su territorio (Australia, Brasil, Canadá, EEUU, Indonesia, México, Rusia, etc), la tendencia en la Unión Europea (UE) ha sido la de que la mayoría de países

compartan una misma hora. Incluso entre 1968 y 1970 el Reino Unido cambió su hora oficial por la de Europa central, siendo suspendida por la fuerte oposición generada en muchos ámbitos. Además, con la ampliación de la UE a los países de Europa oriental, resulta inevitable el mantener horas oficiales distintas (Fig. 3). Dada la situación actual, cabría incluso considerar la posibilidad de que países como España y Francia adoptaran la hora del huso que les corresponde, lo cual aliviaría la cuestión del gran desajuste entre la hora oficial y la hora solar en Galicia.

Como veremos a continuación, la tendencia actual es la contraria, pues es la de aumentar las horas de sol por la tarde y reducir las de la mañana.

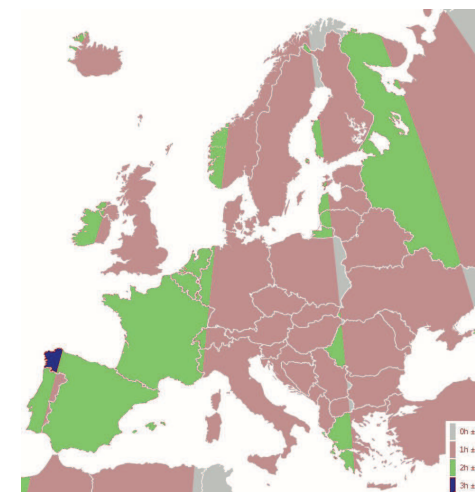


Figura 4: Diferencias entre la hora oficial en verano (2011) y la hora solar local en Europa. En la península y las islas Baleares tal diferencia se encuentra en el rango que va de 1^h 30^{min} a 2^h 30^{min} (color verde). El único territorio en Europa donde la diferencia excede tal valor (color azul marino) es en la mitad occidental de Galicia. (Adaptado de Naive cynic/Wikimedia Commons.)

La hora de verano en España

Se entiende por “periodo de hora de verano” el periodo del año durante el cual se adelanta la hora en sesenta minutos con respecto a la hora del resto del año. Realmente se trata de un periodo de primavera-verano.

Se aplicó por primera vez durante el transcurso de la primera guerra mundial, a fin de mantener las fábricas abiertas durante una hora más desde la primavera hasta septiembre u octubre. Esta medida excepcional

que tomaron los gobiernos alemán y austro-húngaro en abril de 1916 fue adoptada en ese mismo año por Holanda, Reino Unido (mayo), Francia y Portugal (junio). Otros países involucrados en la contienda lo harían después. Algunos lo mantuvieron hasta terminar la segunda guerra mundial.

Primera etapa de la hora de verano en España (1918-1931)

En España la hora de verano se adoptó por primera vez en 1918, por el RD de 3 de abril (GM de 4 de abril) “disponiendo que el día 15 del mes actual, y a las veintitrés horas, se adelante la hora legal en sesenta minutos, y que el día 6 de Octubre del corriente año se restablezca la hora normal.” La razón aducida fue la escasez de carbón provocada por la primera guerra mundial, que obligó a intensificar la producción y a reducir el consumo y, al mismo tiempo, armonizar el horario con el de los países vecinos.

En los días siguientes se publicaron una serie de reglas para su aplicación. Quizás las más relevantes fueron la de la aplicación del cambio de hora legal a los horarios del servicio de ferrocarriles (GM de 6 de abril de 1918) y la RO de 11 de abril (GM de 12 de abril) del Ministerio de Gracia y Justicia al objeto de evitar perturbaciones “en el importante servicio de la Administración de justicia y en aquellos otros que auxilian y completan el funcionamiento de la misma”. También se publicó una nota para que los gobiernos civiles (GM de 12 de abril) informaran convenientemente a ayuntamientos, industria y otras autoridades “para evitar dudas y confusiones que pudieran suscitarse en lo referente á los plazos y jornadas de trabajo” y una RO de 11 de abril (GM de 14 de abril) del Ministerio de Instrucción Pública y Bellas Artes para que se informara a los jefes de todos los centros dependientes de tal ministerio.

Unos días antes de la finalización del periodo de hora de verano se publicaron la RO de 19 de setiembre (GM de 23 de setiembre) recordando a gobernadores civiles y otras autoridades la finalización de tal periodo y la RO de 30 de setiembre (GM de 1 de octubre) relativa a la aplicación del cambio de hora a los horarios de los ferrocarriles.

En el año 1919, terminada la guerra, también se introdujo la hora de verano, atendiendo el Ministerio de Abastecimientos a la petición razonada “de las Corporaciones y entidades representantes de los grandes intereses económicos nacionales”. No se adoptó en los años siguientes, reintroduciéndose durante la dictadura del general Miguel Primo de Rivera, en los años 1924 a 1929 exceptuando el año 1925. La justificación dada en 1924 fue el mantenimiento de tal medida por parte de “otras naciones ligadas a nosotros por una relación constante de intereses y vecindad”, tras la propuesta de la Sociedad de las Naciones de acordar un periodo común mundial. En años sucesivos la justificación, dada por la Presidencia del Consejo de Ministros, desapareció o se redujo a comentar la necesidad de realizar los cambios simultáneamente con los países vecinos.

Cambios de hora oficial en España, 1901-1938 (*)

Fecha y hora oficial del cambio (día de la semana: lmxjvsd)	Nueva hora of.	TO-TU (h)	Decreto u Orden	Fecha de publicación	
Regencia de María Cristina (1885-1902)					
1901, 1 de enero (m)		0	RD 26/jul/1900	GM 28/jul/1900	
Reinado de Alfonso XIII (1902-1931), Primera Guerra Mundial (1914-1918)					
1918, 15 de abril (l)	23:00	24:00	1	RD 3 de abril	GM 4 de abril
1918, 6 de octubre (d)	25:00	24:00	0	RO 19 de sept.	GM 23 de sept.
1919, 6 de abril (d)	23:00	24:00	1	RD 28 de marzo	GM 29 de marzo
1919, 6 de octubre (l)	25:00	24:00	0	RO 22 de sept.	GM 23 de sept.
Dictadura de Miguel Primo de Rivera (1923-1930)					
1924, 16 de abril (x)	23:00	24:00	1	RD 7 de abril	GM 8 de abril
1924, 4 de octubre (s)	25:00	24:00	0	RO 25 de sept.	GM 25 de sept.
1926, 17 de abril (s)	23:00	24:00	1	RD 9 de abril	GM 10 de abril
1926, 2 de octubre (s)	25:00	24:00	0	"	"
1927, 9 de abril (s)	23:00	24:00	1	RD 23 de marzo	GM 24 de marzo
1927, 1 de octubre (s)	25:00	24:00	0	RO 28 de sept.	GM 29 de sept.
1928, 15 de abril (**) (d)	00:00	01:00	1	RO 31 de marzo	GM 1 de abril
1928, 6 de octubre (s)	25:00	24:00	0	"	"
1929, 20 de abril (s)	23:00	24:00	1	RO 16 de abril	GM 17 de abril
1929, 5 de octubre (s)	25:00	24:00	0	"	"
Dictablanda de Dámaso Berenguer (enero de 1930 - febrero de 1931)					
No hay cambio de hora en 1930		0	RO 27 de marzo	GM 28 de marzo	
Último gobierno del rey Alfonso XIII (hasta el 14 de abril de 1931)					
1931, 18 de abril (s)	23:00	24:00	1	RO 9 de marzo	GM 11 de marzo
Segunda República Española (1931-1936)					
se anula la RO anterior antes de su aplicación		0	O de 15 de abril	GMR 16 de abril	
Guerra civil española (1936-1939)					
Gazeta de la República:					
1937, 16 de junio (x)	23:00	24:00	1	D. 6 de junio	GR 8 de junio
1937, 2 de octubre (s)	25:00	24:00	0	D. 24 de sept.	GR 25 de sept.
1938, 2 de abril (s)	23:00	24:00	1	D. 26 de marzo	GR 27 de marzo
1938, 30 de abril (s)	23:00	24:00	2	O. 27 de abril	GR 28 de abril
1938, 2 de octubre (d)	24:00	23:00	1	O. 29 de sept.	GR 1 de oct.
Boletín Oficial del Estado:					
1937, 22 de mayo (s)	23:00	24:00	1	O. 17 de mayo	BOE 18 de mayo
1937, 2 de octubre (s)	25:00	24:00	0	"	"
1938, 26 de marzo (s)	23:00	24:00	1	D. 19 de marzo	BOE 21 de marzo
1938, 1 de octubre (s)	25:00	24:00	0	"	"

TO: Tiempo oficial, TU: tiempo universal.

(*) Los cambios indicados corresponden a la hora oficial en España exceptuando las islas Canarias, cuya hora oficial difiere de la anterior en una hora desde el 1 de marzo de 1922, por Real Decreto de 11 de febrero de 1922 publicado en la Gazeta de Madrid de 12 de febrero.

(**) Se consideró que el adelanto se aplicaba oficialmente a las 24:00 horas del día anterior.

Abreviaturas: **BOE** Boletín Oficial del Estado (1936-61, 1986-); Boletín Oficial del Estado: Gaceta de Madrid (1961-86). **D.** Decreto. **GM** Gazeta de Madrid (1697-1934). **GMR** Gazeta de Madrid: Diario Oficial de la República (1934-36). **GR** Gazeta de la República (1936-39). **O.** Orden; Orden Circular. **RD** Real Decreto. **RO** Real Orden; Real Orden Circular.

Con la instauración del gobierno interino del general Dámaso Berenguer a principios de 1930, en lo que se calificó entonces de “dictablanda”, se publicó la RO de 27 de marzo (GM de 28 de marzo) en que se disponía la no modificación de la hora legal al estimar el Gobierno “que las actuales circunstancias del país aconsejan no introducir variación alguna en la hora normal durante este año.” Posiblemente este comentario se refiriera a la inestabilidad política, las revueltas sociales y la crisis económica consecuencia de la fuerte caída de la bolsa estadounidense en el año anterior. Tras la dimisión de Berenguer y su gobierno el 14 de febrero de 1931, el almirante Juan Bautista Aznar ocupó la presidencia del gobierno, durante la cual se publicó la RO de 9 de marzo (GM de 11 de marzo) por la que se disponía que el 18 de abril se adelantara la hora. En la GM de 11 de abril se publicaron las consabidas instrucciones sobre el cambio de hora en el servicio de ferrocarriles. El resultado de las elecciones municipales celebradas el día siguiente supuso la caída del último gobierno del rey Alfonso XIII y la proclamación, el día 14, de la Segunda República Española. Al día siguiente, el gobierno provisional, “accediendo a numerosas peticiones formuladas por distintos sectores de la opinión pública” (GM de 16 de abril) anuló el cambio de hora previsto para el día 18.

Vaivenes en la hora oficial durante la Guerra Civil

En los años 1937 y 1938 los dos bandos contendientes en la guerra civil española establecieron horarios de verano, aunque en distintos periodos. La presidencia de la República inicialmente estipuló (Gaceta de la República, GR, de 8 de junio de 1937) que la hora de verano se extendería desde el 16 de junio al 6 de octubre, fecha que posteriormente (GR de 25 de septiembre) fue cambiada al 2 de octubre. El bando sublevado o nacional publicó su orden en el BOE de 18 de mayo en un texto más extenso que citaba explícitamente las reglas establecidas en 1918 para ferrocarriles, administración de justicia, industria y trabajo, y organismos oficiales.

En 1938 el gobierno de la República incluso llegó a adelantar el horario de invierno en una hora, la primera vez que ello se hacía en España, con lo que aumentó la confusión de horarios en la península (Fig. 5). La introducción de una hora adicional tuvo lugar a finales de abril y fue comunicada dos días antes (GR de 28 de abril), cuando ya estaba en vigor la hora de verano (desde el 2 de abril, GR de 27 de marzo). Como consecuencia la hora oficial se adelantó en dos horas al TU. En la GR de 1 de octubre se recordó la orden de retrasar la hora legal en una hora al día siguiente. Hasta el fin de la guerra hubo dos horas oficiales en la península.

En el periodo 1918-1938 los avisos de cambio de la hora oficial se realizaron con muy escasa antelación, de unas pocas semanas e, incluso, de días (por ejemplo, en 1929). Además, al realizar los cambios de modo que la nueva hora coincidiera con la medianoche, el día de otoño en que se

recuperaba la hora normal se extendía oficialmente hasta las 25^h00^{min}.



Figura 5: Cambios en la hora legal habidos en los territorios controlados por los dos bandos contendientes durante la guerra civil.

La hora oficial en la década de 1940

Tras la finalización de la guerra, la hora de verano se implantó entre los años 1939 y 1949, excepto en 1941, 1947 y 1948. Después, como en otros países europeos, dejó de adoptarse. Francia dejó de hacerlo antes, en 1946. El cambio de hora adoptado en marzo de 1940 (O. de 7 de marzo, BOE de 8 de marzo) merece especial consideración. Está redactado como un típico adelanto de la hora oficial de verano, señalando que “oportunamente se señalará la fecha en que haya que restablecerse la hora normal.” Tal restablecimiento no tuvo lugar en otoño, como era habitual, ni ha tenido lugar desde entonces hasta la actualidad. En la práctica, supuso un cambio permanente de la hora oficial de España adoptando, sin decirlo y, posiblemente, sin planearlo, la del huso horario +1 para la península y las islas Baleares y la del huso horario 0 para las islas Canarias. En 1941 no hubo ningún cambio de hora y en 1942 (O. de 1 de mayo, BOE de 2 de mayo) se reanudó la implantación de la hora de verano, argumentando “las ventajas que en los momentos actuales reporta, desde el punto de vista de una economía conveniente, que la duración de la jornada de trabajo se adapte lo más posible a la jornada solar”. Este fue el único argumento empleado en los años sucesivos, hasta 1946.

Tras dos años sin adoptar la hora de verano, en 1949 se hizo de nuevo argumentando “la situación creada por la continuada sequía que, de manera tan sensible, afecta a los suministros de energía eléctrica” (BOE de 26 de abril). Sin embargo, la hora de verano no volvería a adoptarse en España hasta la llegada de otra crisis mundial de carácter energético.

Cambios de hora oficial en España, 1939-1985 (*)

Fecha y hora oficial del cambio (día de la semana: lmxjvsd)	Nueva hora of.	TO-TU (h)	Decreto u Orden	Fecha de publicación
Dictadura de Francisco Franco (1939-1975)				
1939, 15 de abril (s)	23:00 24:00	1	O. 2 de abril	BOE 4 de abril
1939, 7 de octubre (s)	25:00 24:00	0	O. 2/4 y 3/10	BOE 4/4 y 5/10
1940, 16 de marzo (s)	23:00 24:00	1	O. 7 de marzo	BOE 8 de marzo
1942, 2 de mayo (s)	23:00 24:00	2	O. 1 de mayo	BOE 2 de mayo
1942, 1 de sept. (***) (m)	01:00 00:00	1	O. 28 de agosto	BOE 29 de agosto
1943, 17 de abril (s)	23:00 24:00	2	O. 24 de marzo	BOE 25 de marzo
1943, 3 de octubre (***) (d)	01:00 00:00	1	O. 8 de sept.	BOE 10 de sept.
1944, 16 de abril (**) (d)	00:00 01:00	2	O. 20 de marzo	BOE 21 de marzo
1944, 15 de abril (s)	23:00 24:00	2	Rectificación	BOE 22 de marzo
1944, 1 de octubre (***) (d)	01:00 00:00	1	O. 13 de sept.	BOE 14 de sept.
1945, 14 de abril (s)	23:00 24:00	2	O. 22 de marzo	BOE 24 de marzo
1945, 30 de sept. (***) (d)	01:00 00:00	1	O. 7 de sept.	BOE de 12 de sept.
1946, 13 de abril (s)	23:00 24:00	2	O. 23 de marzo	BOE 25 de marzo
1946, 29 de sept. (***) (d)	01:00 00:00	1	O. 4 de sept.	BOE 7 de sept.
1949, 30 de abril (s)	23:00 24:00	2	O. 25 de abril	BOE 26 de abril
1949, 2 de octubre (***) (d)	01:00 00:00	1	O. 18 de sept.	BOE 20 de sept.
1974, 13 de abril (s)	23:00 24:00	2	O. 5 de abril	BOE 6 de abril
1974, 6 de octubre (***) (d)	01:00 00:00	1	O. 18 de sept.	BOE 20 de sept.
1975, 12 de abril (s)	23:00 24:00	2	O. 3 de marzo	BOE 11 de marzo
1975, 5 de octubre (***) (d)	01:00 00:00	1	"	"
Reinado de Juan Carlos I (1975-...)				
1976, 27 de marzo (s)	23:00 24:00	2	O. 22 de marzo	BOE 24 de marzo
1976, 26 de sept. (***) (d)	01:00 00:00	1	"	"
1977, 2 de abril (s)	23:00 24:00	2	O. 11 de marzo	BOE 12 de marzo
1977, 25 de sept. (***) (d)	01:00 00:00	1	"	"
1978, 2 de abril (d)	02:00 03:00	2	O. 13 de marzo	BOE 22 de marzo
1978, 1 de octubre (d)	03:00 02:00	1	"	"
1979, 1 de abril (d)	02:00 03:00	2	O. 20 de febr.	BOE 5 de marzo
1979, 30 de sept. (d)	03:00 02:00	1	"	"
1980, 6 de abril (d)	02:00 03:00	2	O. 25 de enero	BOE 22 de febrero
1980, 28 de sept. (d)	03:00 02:00	1	"	"
1981, 29 de marzo (d)	02:00 03:00	2	O. 26 de enero	BOE 6 de febrero
1981, 27 de sept. (d)	03:00 02:00	1	"	"
1982, 28 de marzo (d)	02:00 03:00	2	O. 26 de febrero	BOE 27 de febrero
1982, 26 de sept. (d)	03:00 02:00	1	"	"
1983, 27 de marzo (d)	02:00 03:00	2	O. 23 de febrero	BOE 25 de febrero
1983, 25 de sept. (d)	03:00 02:00	1	"	"
1984, 25 de marzo (d)	02:00 03:00	2	O. 22 de febrero	BOE 28 de febrero
1984, 30 de sept. (d)	03:00 02:00	1	"	"
1985, 31 de marzo (d)	02:00 03:00	1	O. 26 dic. 84	BOE 7 de enero
1985, 29 de sept. (d)	03:00 02:00	1	"	"

(*) Los cambios indicados corresponden a la hora oficial en España exceptuando las islas Canarias, cuya hora oficial difiere de la anterior en una hora desde el 1 de marzo de 1922, por Real Decreto de 11 de febrero de 1922 publicado en la Gazeta de Madrid de 12 de febrero.

(**) Se consideró que el adelanto se aplicaba oficialmente a las 24:00 horas del día anterior.

(***) Se consideró que el retraso se aplicaba oficialmente a las 25:00 horas del día anterior.

La hora de verano tras la crisis energética (1974-1985)

Tras el periodo 1918-1949, en que se aplicó de manera discontinua y con escasa consistencia en las fechas, la hora de verano dejó de aplicarse en España en el periodo 1950-1973. A partir de 1974 se reinstauró en muchos países europeos como consecuencia de la crisis energética derivada del gran incremento de los precios del petróleo causado por el embargo iniciado en octubre de 1973 por algunos de los principales países productores. España reinstauró la hora de verano de forma urgente en 1974, Francia lo hizo en 1976 y Portugal en 1977.

Desde 1974 hasta 1977, una orden anual de la Presidencia del Gobierno establecía el inicio del periodo de verano a finales de marzo o principios de abril, pasando las 23^h a ser las 24^h, y su conclusión a finales de septiembre o principios de octubre, día en que las 25^h pasaban a ser las 0^h del día siguiente. La urgencia con que se reimplantó en abril de 1974 se puede observar en la ausencia de una larga exposición de motivos y la de una fecha de finalización, siguiendo el modelo de breves órdenes dadas en el periodo 1939-1949. La Orden de 5 de abril (BOE de 6 de abril) expuso simplemente que: “Teniendo en cuenta las repercusiones que se derivan para la economía nacional del encarecimiento de los productos energéticos, se considera necesario aplicar todas aquellas medidas que puedan contribuir al ahorro de energía y, entre ellas, la consistente en el adelantamiento de la hora legal en relación con la solar.” Como hemos visto antes, la hora legal ya estaba adelantada en una hora con respecto a la solar.

La orden de 3 de marzo de 1975 (BOE de 11 de marzo) sobre la “modificación de la hora legal” incluyó la fecha de inicio y también la de finalización del periodo de horario adelantado, iniciando una costumbre que se mantendría hasta 1989. En la Orden de 1976 se empleó aún otra expresión, la de “regulación de la hora oficial”, y se comentó más extensamente la necesidad de coordinación de la hora de verano con la de otros países europeos situados en el mismo contexto geográfico (“área geográfica coherente”), a fin de evitar problemas (que en el BOE se calificaban de “graves”) en el ámbito de las relaciones internacionales y especialmente en el sector de transportes. Las fechas y horas de cambio coincidieron con los de Francia, que reinstauró la hora de verano en esa ocasión.

En 1978, tras exponer que “la experiencia de cuatro años acredita como satisfactorios los resultados”, se inició la costumbre de realizar los cambios siempre a la 1^h 0^{min} UTC de un domingo. Es decir, en primavera a las 2 de la madrugada peninsulares (1:00 UTC) la hora oficial se adelanta en sesenta minutos y en otoño a las 3 de la madrugada (1:00 UTC) la hora oficial se retrasa en sesenta minutos. En la actualidad sigue haciéndose así. Al realizar el cambio dentro de un mismo día (no hay cambio de sábado a domingo) y en un momento de mínima actividad se evitan algunos de los problemas que se planteaban al hacer los cambios a media noche.

Por otra parte, en 1980 se publicó la primera directiva del Consejo de la Comunidad Económica Europea (CEE), la directiva 80/737/CEE de 22 de julio de 1980, concerniendo a las disposiciones relativas a la hora de verano. Considerando que las diferencias entre el inicio y el fin del periodo de verano entre los distintos Estados miembros (España aún no lo era) complicaban y aumentaban el coste de los transportes y telecomunicaciones, iniciaba un proceso tendente a unificar estas fechas. Este largo proceso culminaría en 1996. En 1980 se estableció la aplicación de la hora de verano a la 1:00 UTC y las fechas de inicio para los años 1981 y 1982 para los Estados miembros, exceptuando Groenlandia y las territorios de ultramar. No se especificaron las fechas de finalización, que se dejaron a discreción de cada país.

La segunda directiva, la 82/399/CEE de 10 de junio de 1982, estableció la fecha de aplicación de la hora de verano (el último domingo de marzo) para los años 1983, 1984 y 1985. Se establecieron dos fechas de finalización, el último domingo de septiembre para todos los Estados miembros excepto los del huso horario de Greenwich (Irlanda y Reino Unido) para los cuales finalizaría el cuarto domingo de octubre.

Los cambios de hora habidos en España entre los años 1980 y 1985 se adaptaron a los establecidos en las directivas de la CEE, coincidiendo en todo caso con Francia y con los demás países que dependían del huso horario +1. La tercera directiva, la 84/634/CEE de 12 de diciembre de 1984, modificada parcialmente el 20 de diciembre de 1985 por la adhesión de España y de Portugal a la Comunidad Económica Europea, estableció la fecha y hora de aplicación y finalización del horario de verano en todos los países miembros para el trienio 1986-1988 en términos parecidos a los de la segunda directiva salvo porque Portugal y las Canarias, teniendo la hora del meridiano 0, realizarían los cambios como los países del meridiano +1.

La orden dada en España en 1985 (BOE de 7 de enero) es la última que cita la RO de 1918 relativa a la Administración de Justicia.

La hora de verano tras la incorporación de España a la CEE

Desde 1980 a 2001 la hora de verano se ha aplicado en la CEE a través de ocho directivas, tres de ellas ya citadas. Las directivas cuarta a sexta no consiguieron una mayor armonización en los horarios y se limitaron a establecer las fechas y horas de aplicación y finalización para un número reducido de años. La cuarta directiva del Consejo, la 88/14/CEE de 22 de diciembre de 1987, estableció las fechas y horas para 1989, la quinta directiva, la 89/47/CEE de 21 de diciembre de 1988, para el trienio 1990-1992, y la sexta directiva, la 92/20/CEE de 26 de marzo de 1992, para los años 1993 y 1994.

La armonización de horarios dentro de la Unión Europea progresó con las dos siguientes directivas. La séptima directiva, la 94/21/CE del

Parlamento Europeo y del Consejo, de 30 de mayo de 1994, estableció la fecha y hora de inicio y finalización del periodo de la hora de verano para el trienio 1995-1997. Como novedad, se estableció que a partir de 1996 la finalización del periodo fuera el último domingo de octubre para todos los Estados miembros, pues se consideraba importante para el funcionamiento del mercado interior fijar una fecha y hora comunes para el principio y el fin del periodo de la hora de verano, válidas en todo el espacio comunitario. Como excepción, se permitió que en 1995 siguiera siendo el cuarto domingo de octubre para Reino Unido e Irlanda. La octava directiva, la 97/44/CE de 22 de julio de 1997, consolidó lo anterior y estableció las fechas para el cuatrienio 1998-2001.

Todas las directivas anteriores dieron lugar a las correspondientes órdenes ministeriales publicadas en el BOE (véase la tabla siguiente).

La directiva 2000/84/CE, del Parlamento Europeo y del Consejo, de 19 de enero de 2001, la última relativa a las disposiciones sobre la hora de verano, estableció disposiciones con carácter indefinido: del año 2002 en adelante el cambio de hora sería el mismo para todos los Estados miembros y se realizaría a la 1^h UTC del último domingo de marzo y del último domingo de octubre. Se publicaron las fechas para el quinquenio 2002-2006. Desde entonces, cada 5 años la Comisión comunica las fechas para sucesivos quinquenios sin necesidad de publicar nuevas directivas, salvo que fuera necesario para cambiar tales disposiciones. En 2006 la Comisión comunicó las fechas de inicio y fin del periodo de hora de verano para el quinquenio 2007-2011 y en 2011, para el quinquenio 2012-2016.

En España, el Real Decreto 236/2002 de 1 de marzo, publicado en el BOE del 2 de marzo, incorporó al ordenamiento jurídico español la directiva 2000/84/CE. Este decreto incluyó una larga introducción en la que se repasaba la historia del establecimiento de la hora de verano en España y se resaltaba el hecho de que en España existen dos horas oficiales, una para la península y archipiélago balear (como es habitual, no se citan explícitamente Ceuta y Melilla) y otra para el archipiélago canario. También se incluía una disposición final para que se determinara la conveniencia de establecer una forma de identificación de cada una de las dos horas que, con igual denominación, corren consecutivamente entre las dos y las tres horas de la madrugada (la una y las dos horas en Canarias) del último domingo del mes de octubre de cada año. No parece que se haya establecido.

En 2007 se publicó una comunicación de la Comisión de la Comunidades Europeas al Consejo, al Parlamento Europeo y al Comité Económico y Social Europeo, sobre la Directiva 2000/84/CE. La Comisión consideraba adecuado mantener el régimen actual de la hora de verano, resaltando que ningún estado miembro había manifestado su deseo de abandonar la hora de verano o de modificar las disposiciones de la directiva vigente en la actualidad.

Cambios de hora oficial en España, 1986-2009 (*)

Fecha y hora oficial del cambio (día de la semana: lmxjvsd)	Nueva hora of.	TO-TU (h)	Decreto u Orden	Fecha de publicación
Pertenencia a la Comunidad Económica Europea / Unión Europea (1986-...)				
1986, 30 de marzo (d)	02:00	03:00	2	O. 27 dic. 85
1986, 28 de sept. (d)	03:00	02:00	1	"
1987, 29 de marzo (d)	02:00	03:00	2	O. 9 de enero
1987, 27 de sept. (d)	03:00	02:00	1	"
1988, 27 de marzo (d)	02:00	03:00	2	O. 29 de enero
1988, 25 de sept. (d)	03:00	02:00	1	"
1989, 26 de marzo (d)	02:00	03:00	2	O. 6 de marzo
1989, 24 de sept. (d)	03:00	02:00	1	"
1990, 25 de marzo (d)	02:00	03:00	2	O. 15 de febrero
1990, 30 de sept. (d)	03:00	02:00	1	"
1991, 31 de marzo (d)	02:00	03:00	2	"
1991, 29 de sept. (d)	03:00	02:00	1	"
1992, 29 de marzo (d)	02:00	03:00	2	"
1992, 27 de sept. (d)	03:00	02:00	1	"
1993, 28 de marzo (d)	02:00	03:00	2	O. 30 nov. 92
1993, 26 de sept. (d)	03:00	02:00	1	"
1994, 27 de marzo (d)	02:00	03:00	2	"
1994, 25 de sept. (d)	03:00	02:00	1	"
1995, 26 de marzo (d)	02:00	03:00	2	O. 16 dic 1994
1995, 24 de sept. (d)	03:00	02:00	1	"
1996, 31 de marzo (d)	02:00	03:00	2	"
1996, 27 de octubre (d)	03:00	02:00	1	"
1997, 30 de marzo (d)	02:00	03:00	2	"
1997, 26 de octubre (d)	03:00	02:00	1	"
1998, 29 de marzo (d)	02:00	03:00	2	O. 27 nov. 1997
1998, 25 de octubre (d)	03:00	02:00	1	"
1999, 28 de marzo (d)	02:00	03:00	2	"
1999, 31 de octubre (d)	03:00	02:00	1	"
2000, 26 de marzo (d)	02:00	03:00	2	"
2000, 29 de octubre (d)	03:00	02:00	1	"
2001, 25 de marzo (d)	02:00	03:00	2	"
2001, 28 de octubre (d)	03:00	02:00	1	"
2002, 31 de marzo (d)	02:00	03:00	2	RD 1 de marzo
2002, 27 de octubre (d)	03:00	02:00	1	"
2003, 30 de marzo (d)	02:00	03:00	2	"
2003, 26 de octubre (d)	03:00	02:00	1	"
2004, 28 de marzo (d)	02:00	03:00	2	"
2004, 31 de octubre (d)	03:00	02:00	1	"
2005, 27 de marzo (d)	02:00	03:00	2	"
2005, 30 de octubre (d)	03:00	02:00	1	"
2006, 26 de marzo (d)	02:00	03:00	2	"
2006, 29 de octubre (d)	03:00	02:00	1	"
2007, 25 de marzo (d)	02:00	03:00	2	O. 4 jul. 2006
2007, 28 de octubre (d)	03:00	02:00	1	"
2008, 30 de marzo (d)	02:00	03:00	2	"
2008, 26 de octubre (d)	03:00	02:00	1	"
2009, 29 de marzo (d)	02:00	03:00	2	"
2009, 25 de octubre (d)	03:00	02:00	1	"

(*) Los cambios indicados corresponden a la hora oficial en España exceptuando las islas Canarias, cuya hora oficial difiere de la anterior en una hora desde el 1 de marzo de 1922.

Cambios de hora oficial en España, 2010-2016 (*)

Fecha y hora oficial del cambio (día de la semana: lmxjvsd)	Nueva hora of.	TO-TU (h)	Decreto u Orden	Fecha de publicación
2010, 28 de marzo (d)	02:00	03:00	2	O. 4 julio 2006
2010, 31 de octubre (d)	03:00	02:00	1	"
2011, 27 de marzo (d)	02:00	03:00	2	"
2011, 30 de octubre (d)	03:00	02:00	1	"
2012, 25 de marzo (d)	02:00	03:00	2	O. 6 oct. 2011
2012, 28 de octubre (d)	03:00	02:00	1	"
2013, 31 de marzo (d)	02:00	03:00	2	"
2013, 27 de octubre (d)	03:00	02:00	1	"
2014, 30 de marzo (d)	02:00	03:00	2	"
2014, 26 de octubre (d)	03:00	02:00	1	"
2015, 29 de marzo (d)	02:00	03:00	2	"
2015, 25 de octubre (d)	03:00	02:00	1	"
2016, 27 de marzo (d)	02:00	03:00	2	"
2016, 30 de octubre (d)	03:00	02:00	1	"

(*) Los cambios indicados corresponden a la hora oficial en España exceptuando las islas Canarias, cuya hora oficial difiere de la anterior en una hora desde el 1 de marzo de 1922.

El futuro de la hora oficial

Consideraciones astronómicas

A lo largo del año la declinación del Sol va variando, entre $-23^{\circ}44'$ y $+23^{\circ}44'$, dando lugar a las estaciones climáticas y a una variación de la duración del día y la noche, que depende fuertemente de la latitud geográfica (ϕ) del lugar. En el ecuador ($\phi = 0^{\circ}$) no hay variación mientras que entre los círculos polares y los polos ($66^{\circ}56' < \phi < 90^{\circ}$) la diferencia entre la noche más corta y la más larga es de 24 horas. En lo que sigue vamos a considerar que nos hallamos en el hemisferio norte de la Tierra.

En la figura 6 se muestran las horas de salida y puesta del Sol para lugares hipotéticos situados a distintas latitudes en el meridiano central de un huso horario, como por ejemplo en el meridiano de Greenwich. En cada panel se comparan los tiempos suponiendo que se aplica (línea continua) o no (línea discontinua) la hora de verano tal como se hace actualmente en la Unión Europea. Como puede observarse, a latitudes bajas no tiene sentido establecer una hora de verano, pues no se ve suficientemente compensada por la variación de la duración del día, que es escasa. Podemos considerar que la menor latitud para la que tiene sentido establecer una hora de verano sería aquella en que la hora media de salida del Sol en el periodo de hora de verano es tan temprana como durante el periodo de otoño e invierno en que rige el horario normal. Tal latitud es de 26° , inferior a cualquier punto de las islas Canarias (la latitud mínima corresponde a la Punta de la Restinga, en la isla de El Hierro, donde $\phi = 27^{\circ}64'$) por lo que la aplicación de la hora de verano en dicho archipiélago tiene sentido.

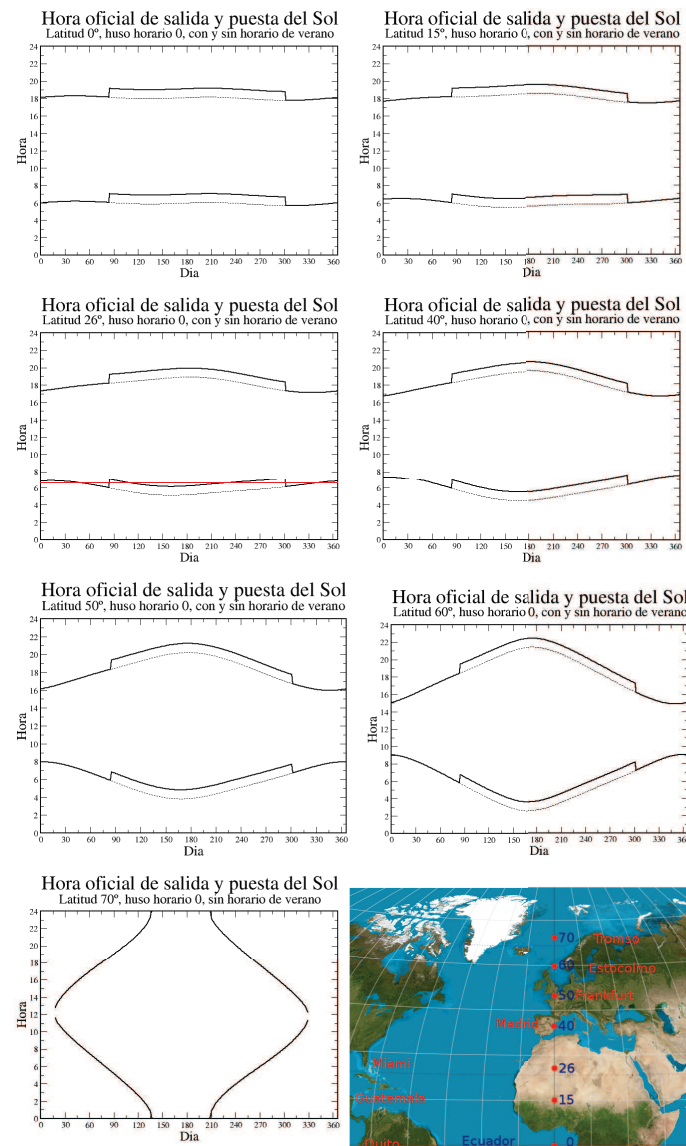


Figura 6: Horas de salida y puesta del Sol a diversas latitudes (0°, 15°, 26°, 40°, 50°, 60° y 70°) en el meridiano de Greenwich considerando o no aplicada la hora de verano vigente en la UE. En el panel correspondiente a 70° se ha representado solo la hora normal.

A latitudes muy elevadas carece de sentido aplicar una hora de verano, pues en la parte central del año hay muchas horas de luz, por lo que no hay ganancia alguna en añadir una más. Los efectos negativos que pueda tener por la mañana (oscuridad, necesidad de calefacción) se compensan en poco tiempo por el alargamiento del día, tanto más rápido cuanto mayor es la latitud.

Resulta más difícil encontrar un criterio para determinar una latitud máxima para la que pueda tener sentido aplicar la hora de verano. Podemos considerar que sólo es efectivo en el periodo del año en que la duración de la noche es superior a las 8 horas recomendables de sueño. Si imponemos (arbitrariamente) que el número de tales noches debe exceder el 50% durante el periodo de verano, obtenemos como límite superior una latitud de 62°. Parece un límite razonable pues si, además, consideramos que durante los crepúsculos también hay algo de luz, resulta que a esta latitud solo un tercio de las noches en hora de verano tienen un periodo de oscuridad superior a 8 horas (entendido como el intervalo de tiempo que transcurre entre el crepúsculo civil vespertino y el matutino que le sigue). Como referencia, las capitales europeas Oslo y Estocolmo se encuentran justo por debajo de este límite.

En resumen, el adelanto de una hora en el periodo de primavera y verano adoptado en la Unión Europea puede tener sentido en el rango de latitudes comprendidas entre 26° y 62°.

Las discutidas justificaciones de la hora de verano

El número de países que ha adoptado la hora de verano ha ido aumentando en las sucesivas épocas de crisis que ha habido a lo largo del siglo XX. Al final de la primera guerra mundial eran unos 30 países; al final de la segunda guerra mundial unos 50 y desde 1980 hasta la actualidad más de 60. Hay que tener en cuenta también que, debido a la descolonización, ahora hay en el mundo muchos más países que hace un siglo. En general, los países que han instaurado la hora de verano suelen ser países desarrollados o recién industrializados situados a latitudes medias (ver Fig. 7) pues, como hemos visto, a latitudes cercanas al ecuador no es necesario, dado que no hay grandes variaciones horarias a lo largo del año que permitan un mayor aprovechamiento de la luz solar, y a altas latitudes las noches son muy cortas en gran parte del periodo de primavera-verano (ver Fig. 6), por lo que también resulta innecesario introducir el cambio de una hora. La razón de que algunos de esos países lo hayan instaurado (tal es el caso de los países escandinavos) es la necesidad de sincronizar los horarios con países vecinos. Hay que recordar que los efectos de la hora de verano son una salida del Sol más tardía, o sea, mañanas más oscuras al inicio y al final del periodo, y una puesta más tardía, que se traduce en unas tardes más largas durante todo el periodo.

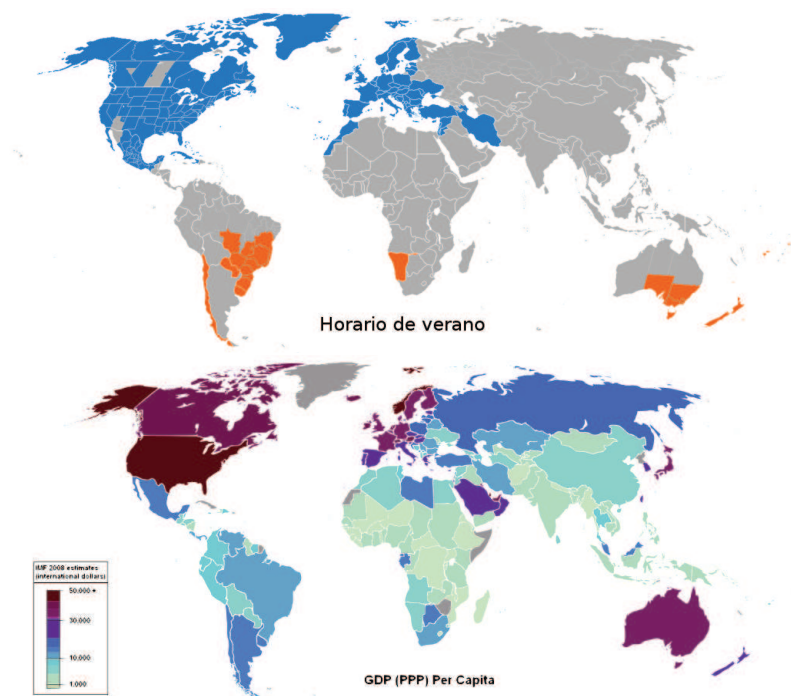


Figura 7: (a) Mapa indicando, en color, los países y territorios en los que se aplicó la hora de verano en 2011 (cortesía de TimeZones-Boy/Wikimedia Commons). (b) Mapa indicando, en gradiente de colores, la renta per cápita en 2008 según el FMI, expresada en términos de poder adquisitivo paritario, PPP (cortesía de Powerkeys-/Wikimedia Commons).

Como se ha comentado en apartados anteriores, la justificación para establecer y mantener la hora de verano es el supuesto ahorro energético. Sin embargo, cuando el constructor inglés William Willett lo propuso por primera vez, en 1907, argumentó que un horario de verano permitiría disponer de más tiempo para el ocio o practicar el deporte al aire libre tras finalizar la jornada laboral. Este argumento se ha repetido numerosas veces. En la década de 1930, en los EEUU se habló del placer de disponer de una hora más de sol antes del anochecer y del aumento en las ventas de equipos de deporte.

Hacia 1975 un estudio del Departamento de Transportes de los EEUU concluyó que había un ligero ahorro de energía, que se veía compensado parcialmente por el incremento del gasto de carburante por los desplaza-

mientos en coche para realizar compras o actividades de ocio. También se observó un aumento en la seguridad en el tráfico. En la propuesta realizada en 2005 de alargar el periodo de hora de verano en EEUU, debido al nuevo aumento de los precios del petróleo, se argumentaron como beneficios: el ahorro energético, la disminución de accidentes de tráfico y de la criminalidad, el mayor tiempo de ocio para dedicar al deporte, a la jardinería, a salir o a cualquier forma de relax, y el aumento de la actividad económica, al facilitar la realización de compras a la salida del trabajo. Desde 2007 el periodo de hora de verano en los EEUU dura casi ocho meses, desde el segundo domingo de mayo al primero de noviembre, pues así incluye la fiesta infantil de la noche de brujas (*Halloween*), que se celebra el 31 de octubre. En la Unión Europea dura siete meses.

La Unión Europea, tras adoptar la octava directiva, decidió realizar un examen profundo de las implicaciones de la hora de verano en los Estados miembros por sectores de actividad (agricultura, medio ambiente, industria, comercio, banca, salud, ocio y turismo, seguridad en el tráfico, transportes y comunicaciones). Instituciones, autoridades y expertos fueron consultados por parte de una consultora independiente que debía evaluar los efectos e implicaciones de la hora de verano tal como estaba establecida. Las conclusiones fueron presentadas en Bruselas en 1999. La primera se refiere al ocio y el turismo de corta distancia, que se ven facilitados con la hora de verano y con la progresiva reducción de horas de trabajo, pudiéndose generar más puestos de trabajo para dar tales servicios. Otras se refieren a su aceptación tácita por parte de la mayoría y a la necesidad de implantar un sistema perenne, lo que se lograría con la última directiva. En las conclusiones no hay ninguna mención referente al ahorro de energía.

En el cuerpo del documento se menciona que la mayoría de estudios efectuados concluyen que si hay ahorro de energía, este es muy modesto, siendo muy difícil distinguir el ahorro expresado en relación con el consumo total de energía, no solamente la energía eléctrica ahorrada en los hogares. Se estimó (con gran incertidumbre) que el ahorro total era de unos 200 millones de euros, o sea, de medio euro por persona y año. Además, hay que tener en cuenta que los modos de vida están cambiando mucho en las últimas décadas, como también está incrementándose mucho el precio de los combustibles para generación de energía y calefacción y el de los carburantes, de modo que se requieren actualizaciones reiteradas de tales estudios.

En 2007 la Comisión de las Comunidades Europeas publicó un compendio de comentarios recibidos de los Estados miembros. Ninguno solicitó modificación o anulación del régimen actual de hora de verano. La conclusión general fue que la hora de verano favorece la práctica de actividades de ocio aprovechando el alargamiento de las tardes, actividades que no se limitan al fin de semana, y permite un pequeño ahorro de energía. A este respecto, Francia y Bulgaria evaluaron el ahorro en el consumo de electri-

ciudad del orden de 0,01 %. Eslovenia, Estonia y Letonia comprobaron que el ahorro en electricidad era insignificante. Los demás países no presentaron resultados de nuevos estudios. La supresión de la comercialización de bombillas de incandescencia en Europa y su progresiva sustitución por las de bajo consumo tendrá como efecto una reducción del gasto más importante que la debida a la hora de verano. Por otra parte, al reducir el consumo de electricidad también se reducirá el ahorro. No es posible aún evaluar el impacto que tiene el uso del aire acondicionado en oficinas y hogares, especialmente en los países del sur de Europa, donde el alargamiento de las tardes se traduce en una mayor necesidad de aire acondicionado en las viviendas, así como en los lugares de ocio y restauración que prolongan sus servicios.

España dispone de un instituto destinado específicamente a la mejora del ahorro y la eficiencia energética. Se trata del Instituto para la Diversificación y Ahorro de la Energía (IDAE), creado en 1974 a raíz de la primera crisis energética. Desde hace al menos 10 años el IDAE cifra el ahorro potencial (en el mejor de los casos) en 90 millones de euros en los hogares españoles (2 euros por persona) y en 210 millones de euros en los edificios del sector terciario y la industria, o sea, un 5 % de la factura de iluminación en las viviendas y cerca de un 3 % del recibo correspondiente al aire acondicionado en oficinas y comercios, sin considerar aquellos que alargan su horario para dar servicio a los clientes. (El ahorro en las viviendas corresponde al consumo de cuatro bombillas de 60 vatios una hora al día durante el periodo de hora de verano.) Curiosamente la cifra total de 300 millones es mayor que la obtenida, en 1999, para toda la Unión Europea y difiere mucho de los valores estimados recientemente por otros países (por ejemplo, Francia), lo cual seguramente indica que se trata de una fuerte sobreestimación. La conclusión repetida por los medios de comunicación durante más de una década, citando el viejo informe del IDAE, de que el ahorro total es de 6 euros por persona y año es, cuando menos, equívoca.

En la práctica el ahorro es menor por la baja utilización de tecnologías de ahorro en iluminación por aprovechamiento de la luz natural en edificios del sector terciario y en industrias, por el aumento de consumo energético propiciado por el mayor tiempo de utilización del aire acondicionado en las viviendas y otros edificios, en especial los lugares de ocio y restauración, por el mayor consumo de carburantes por un mayor número de desplazamientos por las tardes, un mayor consumo de electricidad y calefacción por las mañanas al principio y el final del periodo de hora de verano (especialmente en el norte y el oeste de la península), etc. Estimaciones realizadas en la década de 1990 mostraron que el único periodo favorable en cuanto a reducción del consumo eléctrico total es mayo-septiembre, mientras que si la hora de verano se extiende fuera de este periodo se produce un incremento considerable del consumo durante los meses adicionales. Además, últimamente los picos de consumo de electricidad en España se están pro-

duciendo en verano y se achacan al uso masivo del aire acondicionado. En definitiva, no parece que se haya realizado ninguna medida o estimación del ahorro energético real total (electricidad, gas y carburantes) en España debida a la hora de verano, algo sin duda muy difícil de cuantificar por la variedad de fuentes de energía y las variaciones climáticas anuales, y que sólo parece posible realizar con una suspensión temporal de la hora de verano, como se hizo en Estonia hace una década no obteniendo resultados concluyentes.

En resumen, no queda demostrado que haya ningún ahorro significativo en el consumo total de energía. Los efectos de la hora de verano en cuanto al gasto energético son complejos y difíciles de evaluar, además de evolucionar con el tiempo a medida que cambian las costumbres de los ciudadanos, los precios de la energía y la economía de los países. La conclusión es que el balance energético es incierto y aún abierto a discusión, la cual está ampliando su alcance al considerar también los efectos medioambientales.

El futuro de la hora oficial en España

Básicamente, en otoño e invierno España tiene un horario adelantado en una hora con respecto al huso horario y de dos horas en primavera y verano, lo cual le coloca en una situación particular con respecto a otros países europeos y dificulta la comparación con ellos. La implantación de una doble hora de verano se ha llevado a cabo en algunos países temporalmente y se ha discutido repetidamente en el Reino Unido en relación a adoptar la hora centroeuropea, concluyendo recientemente con la estimación de que supondría un aumento del gasto energético (y de la consiguiente emisión de CO₂), en contradicción con informes elaborados a principios de la década de 1990 en que se concluía en que habría un ahorro significativo en electricidad. En la actualidad en Europa se encuentran en esta situación Bélgica, Francia, Holanda, Luxemburgo y España, que es el caso más extremo al hallarse más hacia el oeste. La Asamblea Parlamentaria del Consejo de Europa aprobó en 1999 la Recomendación 1432, por la que en 2001 el Comité de Ministros invitó a los gobiernos de los países citados a considerar el restablecimiento de una hora oficial que corresponda a su huso horario, lo cual permitiría su adopción por los procedimientos democráticos que la legitimarían, pues la adopción del huso +1 en estos países tuvo lugar durante la segunda guerra mundial.

Como ya hemos explicado, en el caso de España ya se aplica una hora de adelanto durante todo el año, por lo que los beneficios de la hora de verano que tienen otros países, cuyo horario normal corresponde a su huso horario, ya han sido tenidos en cuenta. No es claro que el adelanto de una segunda hora, tal como ahora se realiza, tenga tantos beneficios de ahorro energético como se le adjudican. En cambio, propicia que los horarios laborales y

escolares en España sean más tardíos que en los demás países europeos. Se viene discutiendo desde hace tiempo la armonización de nuestros horarios laborales/escolares, familiares y personales con los mayoritarios en Europa, así como la necesidad de lograr una mayor conciliación familiar y laboral, que debería redundar en una mejora en la calidad de vida y, cabe esperar, en una mayor productividad en el trabajo.

Hora oficial media de ortos y ocasos del Sol

Ciudad	Huso horario	Salida del Sol		Puesta del Sol	
		oto+inv	pri+ver	oto+inv	pri+ver
Berlín	1	7 ^h 33 ^m	5 ^h 53 ^m	16 ^h 43 ^m	20 ^h 17 ^m
Londres	0	7 ^h 24 ^m	5 ^h 50 ^m	16 ^h 41 ^m	20 ^h 09 ^m
Madrid	1	8 ^h 10 ^m	7 ^h 27 ^m	18 ^h 23 ^m	21 ^h 00 ^m
Madrid	0	7 ^h 10 ^m	6 ^h 27 ^m	17 ^h 23 ^m	20 ^h 00 ^m
Las Palmas GC	0	7 ^h 36 ^m	7 ^h 30 ^m	18 ^h 31 ^m	20 ^h 30 ^m
Las Palmas GC	-1	6 ^h 36 ^m	6 ^h 30 ^m	17 ^h 31 ^m	19 ^h 30 ^m

Hora oficial media de salida y puesta del Sol en Londres, Berlín, Madrid y Las Palmas de Gran Canaria, con aplicación de la hora de verano según la normativa actual. En el caso de Madrid y Las Palmas se consideran dos posibilidades, según el huso horario adoptado como referencia. Véanse las figuras 8a y 8b.

Las actividades de índole laboral, escolar y administrativo suelen realizarse en Europa por la mañana, mientras que las actividades de ocio se realizan predominantemente por la tarde. Veamos qué consecuencias tendría para España la armonización de sus horarios con Europa. En el huso horario 1 el país de mayor peso, por economía y población, es Alemania, país que se encuentra aproximadamente centrado en el área ocupada por la Unión Europea (véase la Fig. 3). Además, su capital, Berlín, se encuentra en el centro del huso horario correspondiente. En el huso número 0 el país de mayor peso es el Reino Unido cuya capital, Londres, se encuentra en el centro de dicho huso. En la tabla anterior podemos ver que las horas oficiales medias de salida y puesta del Sol en ambas ciudades son muy parecidas, en ambas épocas del año. Podemos considerar las 7^h30^m, la hora media de la salida del Sol durante el periodo de hora normal (que corresponde a parte de otoño e invierno), como referencia para el inicio de las actividades matutinas. Análogamente, podemos considerar las 20^h15^m, la hora media de la puesta del Sol durante la hora de verano (primavera y verano), como referencia para la finalización de las actividades de ocio. Para que las horas oficiales medias correspondientes en Madrid, situada cerca del centro de la península, se parecieran a las de Berlín y Londres sería necesario que la hora normal de la península fuera la correspondiente

al huso horario 0, lo mismo que en Las Palmas de Gran Canaria, efecto conseguido debido a la diferencia de latitud (véanse las figuras 8b y 8c). En tal caso, toda España podría tener un horario común.

Si se eliminara la aplicación de la hora de verano o bien esta se aplicara durante todo el año, la conclusión sería otra, según se puede apreciar en la tabla siguiente. La aplicación de la hora de verano durante todo el año se ha propuesto en alguna ocasión con el fin de mantener las ventajas de dicho horario que alarga las tardes, opción favorecida por muchos ciudadanos, y evita los inconvenientes de la realización de los cambios de hora. Hay que señalar que, en el caso de no realizar cambios de hora a lo largo del año, los países situados en las latitudes más elevadas seguirían disfrutando de unas largas tardes en verano (véase Fig. 6), mientras que los países meridionales, como España, tendrían la ventaja de poder adaptar mejor los horarios laboral y personal a las horas de Sol a lo largo de todo el año.

Hora oficial media hipotética de ortos y ocasos del Sol

Ciudad	Huso	Salida del Sol		Puesta del Sol	
Berlín	1	5 ^h 58 ^m		18 ^h 15 ^m	
Berlín (*)	2	6 ^h 58 ^m		19 ^h 15 ^m	
Londres	0	5 ^h 53 ^m		18 ^h 08 ^m	
Londres (*)	1	6 ^h 53 ^m		19 ^h 08 ^m	
Madrid	0	6 ^h 09 ^m		18 ^h 21 ^m	
Madrid (*)	1	7 ^h 09 ^m		19 ^h 21 ^m	
Las Palmas GC	-1	5 ^h 57 ^m		18 ^h 06 ^m	
Las Palmas GC (*)	0	6 ^h 57 ^m		19 ^h 06 ^m	

Hora oficial media de salida y puesta del Sol en Londres, Berlín, Madrid y Las Palmas de GC, en el caso de que no se aplicara la hora de verano o de que se aplicara durante todo el año (*). En el primer caso el huso horario es el correspondiente geográficamente.

En definitiva, en aras de armonizar los horarios con Europa, se podrían considerar las siguientes opciones:

1. Si se mantiene la hora de verano en Europa, toda España podría referir su horario normal al huso 0.
2. Si se decidiera eliminar los cambios, España podría favorecer el permanecer en los husos actuales a fin de aprovechar los posibles beneficios energéticos y modificar solo moderadamente los horarios y hábitos actuales.

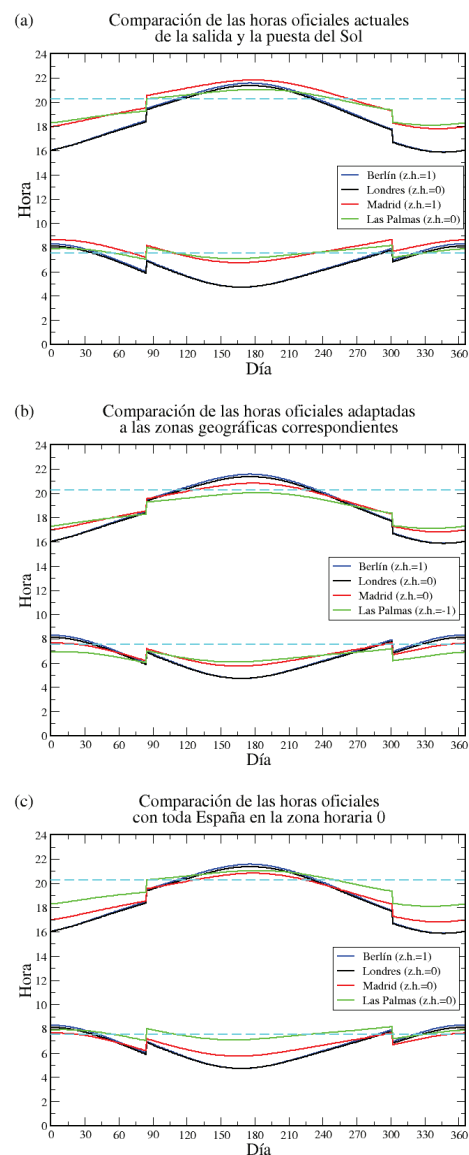


Figura 8: Comparación de las horas oficiales de salida y puesta del Sol en varios supuestos discutidos en el texto. Las líneas azules discontinuas corresponden a las 7^h30^m y 20^h15^m, horas citadas como referencia para las actividades de la mañana y la tarde.

Conclusiones

La implantación de un sistema de horas oficiales nacionales en todo el mundo, ligadas a una única referencia (el tiempo universal coordinado, UTC) mediante un sistema de husos horarios ha sido altamente satisfactoria al facilitar las comunicaciones de todo tipo, los transportes y la sincronización de actividades, que cada vez más se realizan a nivel internacional, cuando no mundial. España fue uno de los últimos países europeos independientes en adoptar ese sistema, en 1901, estableciendo como tiempo oficial el correspondiente al meridiano de Greenwich. No se ha logrado aún restringir las horas oficiales de los países del mundo a los 24 husos horarios; en la actualidad hay 40 zonas horarias.

Las ventajas de la implantación de la hora de verano no son tan claras, pues no está demostrado que en la actualidad haya un ahorro energético significativo. La mayor justificación para mantenerlo parece ser el incremento de tiempo para el ocio y el turismo, por la mayor actividad económica que ello conlleva al incrementar el consumo de bienes y servicios. Por otra parte, genera algunos inconvenientes a tener en cuenta. La implantación de una hora de verano por parte de menos de un tercio de los países del mundo y con periodos de aplicación distintos, incluso entre los países de un mismo hemisferio, tiene efectos negativos en las comunicaciones a nivel internacional, provocando desajustes en las conexiones aéreas y en la coordinación de los trabajos de empresas o proyectos multinacionales, cada vez más comunes debido a la mundialización de la actividad productiva y del comercio. Dentro de un mismo país, se generan problemas con los horarios de los trenes nocturnos y de otras actividades nocturnas en las dos fechas de cambio. Los principales inconvenientes a nivel individual quizás sean los desórdenes del sueño sufridos durante unos días y la necesidad de cambiar la hora en la multitud de relojes y aparatos de los que disponemos. A pesar de ello, encuestas realizadas muestran que el nivel de aceptación ciudadana de la hora de verano en la Unión Europea supera el 50%.

En Europa, con la adopción de los husos horarios referidos a un meridiano de referencia y la implantación de una hora de verano común ha habido dos armonizaciones de los horarios, pero aún está pendiente una tercera, la de los horarios laborales, escolares y administrativos y, consiguientemente, los familiares y sociales. En mi opinión, el problema del mantenimiento de la doble hora de verano actualmente en vigor en España debe ser abordado en términos de esta tercera armonización. Incluso podría llegarse a unificar la hora oficial en España, adoptando la de las islas Canarias.

* * *

Las listas de fechas de los cambios de hora legal habidos en España desde el inicio del siglo XX suelen ser incompletas o contener algún error. Para confeccionar las tablas de este artículo hemos tratado de encontrar todas las órdenes, decretos y notas publicados en el BOE y en las gacetas que le precedieron que hicieran referencia al establecimiento o cambios de la hora oficial o legal en España. Agradeceríamos que se nos comunicara cualquier error u omisión con el fin de poner en el futuro a disposición de todos unas tablas completas y correctas en las páginas web del Observatorio Astronómico Nacional (<http://www.oan.es>).

Referencias

Para ampliar conocimientos sobre la hora oficial en España:

M. M. Gámez Mejías, 2008, “Origen, evolución y futuro de la determinación normativa de la hora legal en España y la Unión Europea”, *Revista de Administración Pública*, núm. 177, págs. 377-417. [Visión legalista muy completa e interesante.]

M. Aranda Mendíaz y E. Galván Rodríguez, 1997, “Una hora menos en Canarias: apunte histórico-jurídico”, *Tebeto. Anuario del Archivo Histórico Insular de Fuerteventura*, núm. 10, págs. 207-216. [Detallada descripción de la situación anterior a 1922 y de la solución adoptada.]

Para ampliar conocimientos sobre la implantación de la hora estandarizada en el mundo:

I. Bartky, 2007, “One time fits all: the campaigns for global uniformity”, Stanford University Press, Stanford. [Muy bien documentado, posiblemente la mejor referencia sobre el tema, aunque escasa en referencia a Europa.]

D. S. Prerau, 2005, “Saving the daylight: why we put the clocks forward”, Granta Books, Londres. [Repaso histórico, muy entretenido por sus numerosas anécdotas, pero que está muy sesgado al mundo anglosajón y contiene inexactitudes.]

Información adicional sobre los husos horarios, ortos, ocasos y crepúsculos puede encontrarse en los capítulos dedicados al tiempo y al Sol en este mismo ANUARIO.

PUESTA AL DÍA DEL TEMA

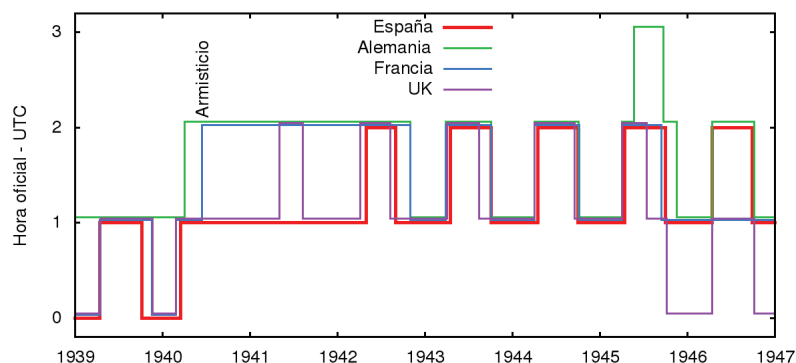
¿Hora franquista, republicana, alemana, británica, francesa...?

En los cuatro años transcurridos desde la publicación del artículo **La hora oficial en España y sus cambios** la posible modificación del horario oficial español se ha convertido en un asunto de actualidad política. En 2013 el Congreso de los Diputados publicó un texto acordado por la Subcomisión para el estudio de la Racionalización de Horarios, la Conciliación de la Vida Personal, Familiar y Laboral y la Corresponsabilidad. De todo lo tratado en los 90 folios que tiene el documento, llamó la atención la recomendación de que se valorara el coste y consecuencias que podría suponer la vuelta al huso horario de Greenwich. En los años sucesivos, tres de los principales partidos políticos se manifestaron a favor de tal medida. A finales de 2016 el Gobierno anunció la promoción de un pacto nacional por la conciliación y la racionalización de horarios y el estudio de las consecuencias del cambio de hora oficial.

Un argumento implícito en la propuesta del cambio de la hora oficial es que, por razones históricas, actualmente ésta no se ajusta al huso horario geográfico. Además, uno de los mensajes mediáticos que se ha esgrimido en favor del cambio de hora oficial ha consistido en calificar la hora actual de “hora franquista”, afirmando que la adopción del horario actual (UTC+1) tuvo lugar en 1940 o en 1942, al inicio de la dictadura. Veamos lo que dice la documentación oficial disponible. En 1942 se publicaron dos órdenes, una en mayo por la que se adelantaba la hora oficial y otra en septiembre, por la que se retrasaba. En ninguna de ellas, como tampoco en la de 1940, se habla de cambio de huso horario de referencia. Es más, en las de 1942 ni siquiera se hace referencia a los países vecinos, sino a que la jornada laboral se adapte lo mejor posible a la solar. En la Orden de 1940 sí se hace referencia a los países vecinos, no en vano el Reino Unido, Portugal y Francia habían cambiado la hora el 25 de febrero. España lo hizo el 16 de marzo y Alemania no lo haría hasta el 1 de abril. O sea que al referirse la Orden de 1940 al “adelanto temporal” para “que el horario nacional marche con los de otros países europeos”, se está refiriendo a Francia, Portugal y Reino Unido, no a Alemania. Durante el periodo que va de principios de 1940 hasta mayo de 1942 el horario de España (UTC+1) difirió del de Alemania (UTC+2) en una hora, por lo que no se puede defender que la Orden de 1940 supusiera la adopción de la hora alemana. Eso sí se puede afirmar en el caso de Francia, cuyos territorios ocupados en 1940 adoptaron

el horario alemán (UTC+2). Curiosamente, la hora vigente en España a partir de 1940 se corresponde con la que el gobierno de la República adoptó en abril de 1938, en plena guerra civil. Y es la misma que adoptó el Reino Unido en 1940, un cambio de carácter premeditadamente permanente hasta el fin de la guerra. Tal era también la idea en Francia antes de la invasión. La prensa española se hizo eco de ambas noticias, pero no de que España pudiera haber tomado una decisión análoga.

Evolución de las horas oficiales

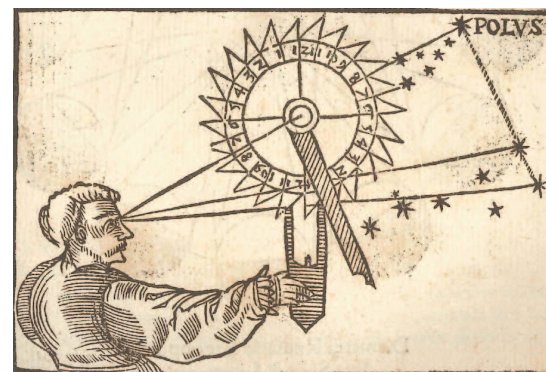


Otra cuestión a considerar es qué ocurrió con las horas oficiales una vez acabada la guerra mundial. A finales de 1945 el Reino Unido recuperó la hora correspondiente a su huso horario, mientras que España y Francia conservaron el horario adquirido durante la guerra, con la diferencia de que Francia dejó inmediatamente de aplicar el adelanto veraniego, mientras que España aún lo retuvo en 1946. En definitiva, España siguió con el mismo horario oficial de los últimos años que, posiblemente, había adquirido carácter de normalidad. De hecho, en las órdenes de 1945 y 1946 no hay ninguna mención a los países vecinos, sino que su redacción sugiere continuidad.

En conclusión, no parece que pueda afirmarse que durante la guerra España adoptara el horario alemán (por influencia o simpatía, pues en el período 1940 a 1942 los horarios español y alemán difirieron, y posteriormente todos, incluido el del Reino Unido, se parecían), ni el francés tras 1945. No he encontrado documentación pública que avale lo uno ni lo otro. Los países eran soberanos en cuanto a qué hora oficial adoptar y cada uno hacía lo que creía conveniente en cada momento, especialmente durante la guerra. Y después de ella, adaptada ya la ciudadanía a determinada correspondencia entre el horario personal/laboral y el solar, no ha habido presión pública suficiente en España para plantear un cambio, hasta recientemente.

Pere Planesas Bigas, enero 2017

RETOS TECNOLOGICOS DE LA ASTRONOMIA ACTUAL



NAVEGACIÓN INTERPLANETARIA CON VLBI, USANDO LAS ESTRELLAS COMO REFERENCIA

María José Rioja Capellán

*Observatorio Astronómico Nacional
Instituto Geográfico Nacional*

Abstract

Exploration by spacecraft provides us with fascinating insights into, and increases of our knowledge of, the solar system. Since the first launch, in 1957, the history of astronautics and planetary exploration is one of space missions with ever increasingly ambitious goals. The development of technology and of navigation techniques are key factors in supporting this advance. In particular the application of ultra-precise space navigation techniques related to Very Long Baseline Interferometry (VLBI), such as Δ DOR and the “astronomical” configurations, provides the most accurate, inertial reference frame, orbit determination estimates. VLBI is a multi-disciplinary technique which historically stands out for providing the most accurate measurements of positions. This article is concerned with the usage of this technique in space navigation protocols, in particular the achievements in a number of recent space missions.

Introducción

La exploración espacial nos acerca al misterio encerrado en los mundos desconocidos, y a menudo las noticias despiertan gran expectacion más allá del círculo de profesionales trabajando en este campo. Este interés generalizado por el “más allá” representa la curiosidad innata al ser humano y su empeño por expandir los confines del conocimiento, en todos los campos de la ciencia.

Las misiones espaciales resultan un tema fascinante tanto por su vertiente tecnológica y la pericia requerida en la navegación espacial de una nave a millones de kilómetros de distancia de la Tierra, como por los nuevos conocimientos adquiridos sobre la Tierra, los planetas y el sistema solar. Estos últimos han proporcionado respuesta a una inmensidad de preguntas, y han generado no menos incógnitas científicas, contribuyendo así al avance del conocimiento.

La historia de la exploración espacial es muy reciente. La era espacial y la astronáutica arrancan con el lanzamiento de los primeros satélites artificiales, el *Sputnik 1* por la Unión Soviética el 4 de octubre de 1957, y el *Explorer 1* por Estados Unidos el 31 de enero de 1958. Estos acontecimientos supusieron el pistoletazo de salida de la carrera espacial entre las potencias de la época. En las dos décadas siguientes se llegaron a lanzar más de 1.600 naves espaciales de todo tipo, la mayoría destinadas a orbitar nuestro planeta, con diversos fines, entre ellos: investigación científica, comunicaciones, militares, meteorológicos, de reconocimiento fotográfico, sondas lunares y planetarias; se han enviado astronaves tripuladas en viajes de ida y vuelta a la Luna; los *Voyager 1* y *2* y por los *Pioneer 10* y *11* han enviado a la tierra fotografías apasionantes que incrementan inmensamente lo que se conocía sobre los planetas y sus satélites; hay dos robots desplazándose por la superficie de Marte; estamos esperando el retorno de una misión con muestras de material de un asteroide ... Mirando hacia atrás uno tiene la sensación de que los logros en la conquista del espacio siguen un ritmo vertiginoso y se suceden sin solución de continuidad.

Pero la historia de la exploración planetaria no está libre de “víctimas”, o misiones que no han llegado a cumplir su objetivo. En particular, Marte cuenta con una amplia casuística de misiones fracasadas, bien porque la nave no llegó a abandonar la Tierra, o porque sufrió una avería en la fase interplanetaria de aproximación a su objetivo, o pasó de largo por un error fatal en la navegación, o dejó de funcionar inmediatamente después de llegar a su destino. En concreto, el año 1999 fué especialmente malo. El *Mars Climate Orbiter*, supuestamente se abrasó al adentrarse unos 20 kilómetros más allá de la distancia de seguridad en la atmósfera de Marte, debido a un error en la conversión de unidades de longitud inglesas a metros. Tres meses más tarde, el *Mars Polar Lander*, su satélite gemelo, se estrelló contra la superficie del planeta porque los motores de frenado durante el descenso se apagaron antes de tiempo. Aunque, probablemente, una de las razones por las que ha habido tantos fallos es porque también ha habido muchos intentos. Marte, por su proximidad a la Tierra, es un destino favorito en las misiones espaciales.

Aún con todo, el número de misiones que han satisfecho, y a menudo excedido, sus objetivos científicos y tecnológicos superan con creces al de los fracasos.

Sin duda, el desarrollo de nuevas técnicas que garantizan una navegación más precisa de las naves espaciales ha jugado un papel fundamental en el avance en la conquista del espacio profundo. Su implementación, además de ayudar a evitar algunos de estos accidentes, ha permitido aumentar el grado de complejidad de las misiones, al permitir realizar maniobras cada vez más complicadas.

Navegación espacial

El objetivo de la navegación espacial es conducir la nave hacia su destino ajustándose a una trayectoria predefinida. Esta labor es responsabilidad de los miembros del equipo de vuelo de la misión, quienes diseñan las maniobras pertinentes para que la trayectoria real se ajuste a la planeada. Así pues, la determinación precisa de la órbita o trayectoria real seguida por la nave es de vital importancia para la navegación.

La navegación espacial se lleva a cabo desde las antenas parabólicas en los centros de control, o estaciones de seguimiento de satélites, de las agencias espaciales en Tierra, ya que la nave no es visible para el ojo humano, ni siquiera utilizando telescopios. Estos centros constituyen el segmento terrestre de la misión espacial. Por su parte las naves están equipadas con múltiples antenas para utilizar en diferentes fases de la misión y que permiten la comunicación en todas direcciones, y con otros satélites. Dichas antenas abren un canal de intercambio de información entre los sistemas a bordo de la nave y los miembros del equipo de vuelo y permiten el control remoto desde Tierra. El análisis de las señales emitidas por la nave y recibidas en Tierra, usando técnicas tradicionales de telemetría (medidas “Doppler” y medidas de distancia o “ranging”) y otras incorporadas en los últimos años, permite conocer la posición y velocidad de la nave en todo momento, junto con otros parámetros dinámicos tales como la presión de radiación solar. Esta información es esencial para conducirla a su destino.

El servicio de navegación obtiene estimaciones de la velocidad a partir de medidas de la frecuencia de las radio señales – o más apropiadamente, del cambio de frecuencia entre la emitida desde la nave y la recibida en Tierra. Este fenómeno, conocido como “efecto Doppler”, es similar al cambio escuchado en el tono de la bocina de un tren mientras se acerca o se aleja de nosotros. Las medidas Doppler en telemetría reflejan la velocidad relativa entre el transmisor a bordo de la nave y el receptor en la Tierra y permiten determinar la dirección con respecto a la antena en Tierra. Las medidas de la distancia se realizan mediante algo así como el envío de una señal desde la Tierra hasta la nave, y la recepción de su “eco”. El tiempo empleado en el viaje de ida y vuelta determina la distancia a la nave con una precisión de unos pocos metros, ¡aunque la nave pueda estar a varios cientos de millones de kilómetros de distancia!

La incorporación de la técnica Δ DOR (Delta Differential One-way Ranging) a los protocolos de navegación espacial es más reciente. Esta técnica se viene utilizando en las dos últimas décadas y usa observaciones de cuásares – objetos celestes muy lejanos – para la determinación de la posición de la nave (Thornton y Border, 2003). Para ello se realizan observaciones desde dos antenas simultáneamente, alternando entre la nave y el cuásar. El observable fundamental en la aplicación de la técnica Δ DOR es la (“doble”) diferencia entre los tiempos de llegada de las señales de la nave y de la emisión natural del cuásar a dos antenas separadas por una gran distancia. Las diferencias “simples” en los tiempos de llegada se calculan mediante el procesamiento conjunto de las señales recibidas en ambas antenas, utilizando el algoritmo de correlación para el caso de la emisión del cuásar; y mediante la comparación de los tiempos de emisión, desde el satélite, y recepción, en Tierra, para la señal de la nave – similar al “ranging”, pero con un sólo viaje de ida. Cada una de ellas permite estimar la dirección de procedencia de la señal correspondiente con respecto a las antenas. La “doble diferencia” entre las medidas del satélite y el cuásar produce la cancelación de muchos errores comunes a ambas medidas, inducidos por la atmósfera y la instrumentación, y proporciona un observable mucho más preciso que los individuales. El observable “doble diferencia” permite estimar la separación angular entre la nave y el cuásar con alta precisión. La técnica Δ DOR podría verse como una configuración particular de la técnica radioastronómica de VLBI, que se considera en cierto detalle más adelante, y que destaca por proporcionar medidas ultra precisas de posiciones.

Mucho más reciente es la aplicación de la configuración “astronómica” de VLBI en la navegación espacial; quizás por ello su uso no es tan frecuente, comparado con la configuración Δ DOR, en los protocolos de navegación. Ambas configuraciones representan el “estado del arte” de las técnicas de navegación espacial. Más adelante se incluyen las conclusiones de un estudio comparativo entre ellas.

Para permitir el seguimiento continuado de los satélites con antenas que están fijas sobre la Tierra en rotación, la Agencia Espacial Norteamericana (NASA) dispone de tres estaciones de seguimiento estratégicamente distribuidas por el globo terrestre, en Europa (Robledo de Chavela, Madrid, España), Estados Unidos (Goldstone, California) y Australia (Canberra). Estas estaciones constituyen la denominada *red de espacio profundo* de NASA (Deep Space Network, DSN) y garantizan la visibilidad permanente de cada uno de los satélites desde al menos una de ellas (*ver fig. 1*). Las antenas se van turnando en la tarea de seguimiento de los satélites, de manera similar a los corredores de un equipo que se van pasando el relevo mientras dura la carrera, y proporcionan la información esencial para conducir la nave a su destino.



Figura 1: *Izquierda*: Distribución geográfica de las estaciones de seguimiento de satélites de la NASA (Agencia Espacial Norteamericana). *Derecha*: Antena de seguimiento de la ESA (Agencia Espacial Europea).

Trayectorias de las naves, o las carreteras en el espacio

Durante los viajes espaciales casi nunca se avanza en línea recta. Las trayectorias de vuelo se planean para minimizar la demanda de consumo energético, con el fin de prolongar el tiempo de vida y el retorno científico de la misión.

Las órbitas de transferencia de Hohmann, así denominadas en honor al ingeniero alemán Wolfgang Hohmann que las presentó por primera vez en 1925, son la opción energéticamente más favorable para llevar a una nave desde la Tierra a otro planeta cercano, como Venus o Marte. Son órbitas elípticas cuyo perihelio (punto más cercano al Sol) está en la órbita de la Tierra y su afelio (punto más distante del Sol) en la de Marte; o viceversa en el caso de Venus. Para insertar una nave en una de estas órbitas se necesita “ajustar” la velocidad orbital de la nave aumentando (o disminuyendo) sólo su componente tangencial. Esto se realiza con la asistencia de cohetes en el momento del lanzamiento.

Si el objetivo de la misión es llegar al planeta, y no sólo a su órbita, deberá ocurrir que la llegada de la nave a la órbita del planeta coincida con la llegada del planeta a la misma ubicación. Para calcular la posición propicia del planeta relativa a la Tierra en el momento del lanzamiento se utilizan las leyes de Kepler, propuestas por el astrónomo alemán Johannes Kepler en el siglo XVII para describir el movimiento de los planetas en el sistema solar. Además, para que la nave se coloque en órbita alrededor de Marte (es decir, sea capturada por Marte) necesitará disminuir su velocidad con respecto al planeta, bien encendiendo los motores de frenado o por otro medio; el frenado deberá ser aún mayor si lo que se pretende es descender hasta su superficie, para conseguir que el punto más bajo de la nueva órbita coincida con la superficie del planeta.

Los cohetes proporcionan los cambios en la velocidad necesarios para

inyectar una nave espacial dentro de las órbitas de transferencia, o para saltar entre órbitas. Por ejemplo, para ir a Marte, se necesita aumentar en algo menos de 3 kilómetros por segundo la velocidad de la nave, con respecto a la velocidad orbital de la Tierra. Sin embargo, para llegar a planetas más lejanos se necesita una velocidad adicional para vencer la atracción gravitatoria del Sol. La gran potencia del cohete necesario para suministrarla supone un serio inconveniente en el uso de estas órbitas para alcanzar los planetas distantes. Afortunadamente, existen otras maniobras alternativas que proporcionan el impulso de velocidad necesario sin necesidad de consumir combustible.

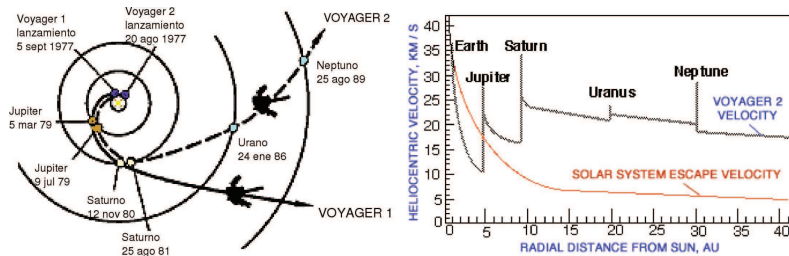


Figura 2: Izquierda: Vuelo interplanetario de las misiones *Voyager 1* y *2*, donde aparecen indicadas las maniobras de asistencia gravitatoria realizadas para llegar a sus destinos. Derecha: Gráfico con las variaciones de la velocidad de la nave durante su vuelo interplanetario. Los picos indican los impulsos de velocidad resultado de los encuentros con cada uno de los planetas.

Una maniobra muy útil en vuelos interplanetarios es la denominada maniobra de impulso gravitatorio, o de asistencia gravitatoria, o con el término inglés “fly-by”. En ella se aprovecha el paso de la nave por las inmediaciones de un planeta en movimiento para transferir parte de la energía cinética del planeta a la nave, aumentando así su velocidad a la vez que se cambia su rumbo sin necesidad de consumir combustible. La primera vez que se puso en práctica fue durante el vuelo del *Mariner 10*, lanzado en 1973 hacia Venus, y que usó la gravedad del planeta para prolongar su órbita hacia Mercurio. Desde entonces supone una parte esencial del protocolo de navegación. También, las aproximaciones de los *Voyager* a cada uno de los planetas sirvieron, además de para su exploración, para hacer posible el “gran tour” hasta Neptuno, en el borde exterior del sistema solar. El encuentro con Júpiter abasteció al *Voyager 2* con la velocidad necesaria para alcanzar Saturno; su encuentro con Saturno hizo posible su siguiente trecho de viaje hacia Urano, y así sucesivamente. Los planetas en movimiento suministraron la energía (ver fig. 2). Finalmente, también se han usado encuentros con la Tierra con

el mismo fin. Por ejemplo la sonda *Galileo*, en su camino hacia Júpiter, fue primero a Venus y luego retornó a la Tierra para ganar impulso y tomar rumbo definitivo a Júpiter. Todo esto recuerda a un juego de billar cósmico, donde además el satélite transporta un pequeño cohete que le permite corregir su curso cuando lo necesita y así hace posible jugadas complejas.

Una de las fases más críticas de las misiones espaciales es la de aproximación planetaria. El nivel de precisión requerido para realizar maniobras de asistencia gravitatoria, o para la inserción en órbita alrededor del planeta, o en una órbita de impacto para lanzar una sonda hasta un determinado punto de la superficie, aumenta con la complejidad de los objetivos de las modernas misiones espaciales, que en algunos casos parecen más cercanas ¡al guión de una película de ciencia ficción que a la realidad! La incorporación de la técnica interferométrica de muy larga base (VLBI) a las técnicas tradicionales de telemetría ha supuesto una revolución en la navegación espacial: proporciona alta resolución en un sistema de referencia cuasi inercial.

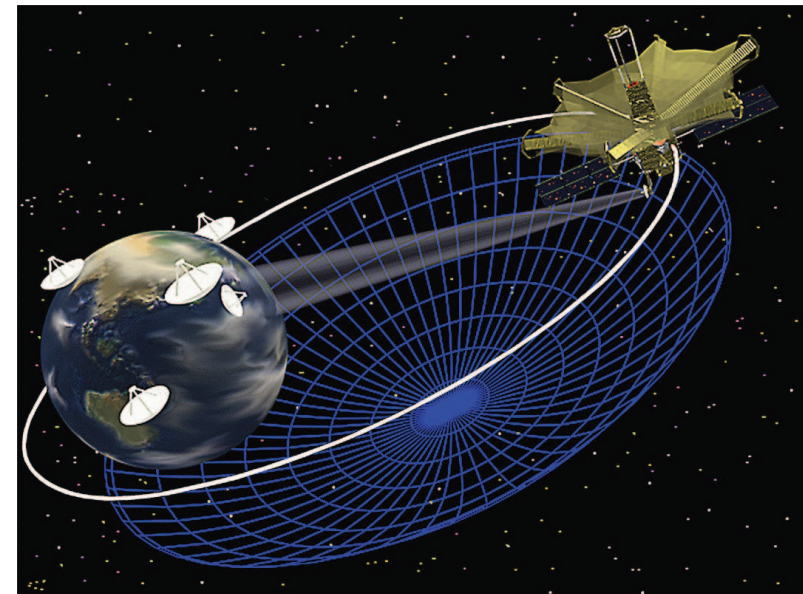


Figura 3: Impresión artística del radiotelescopio gigante con resolución equivalente a la que se obtiene con observaciones conjuntas entre antenas en la Tierra y la antena en órbita, HALCA, con la técnica de VLBI. (Cortesía de ISAS, Japón.)

VLBI: una técnica multidisciplinaria y ultra precisa

VLBI (Very Long Baseline Interferometry) es una técnica interferométrica que conecta múltiples antenas, separadas por grandes distancias entre sí, para actuar como un único gran radiotelescopio, y proporciona muchas de las ventajas de tener una antena parabólica gigante tan grande como la distancia que separa los telescopios. Entre ellas, un elevadísimo poder de resolución (Rogers et al., 1983). Este término hace referencia a la separación mínima angular entre dos objetos para apreciarse como separados, la cual es inversamente proporcional a la apertura del instrumento (y a la frecuencia a la que se realizan las observaciones). La resolución es algo así como la “agudeza visual”, o la precisión angular del instrumento. Las antenas que participan en VLBI están separadas por grandes distancias que pueden llegar a superar el tamaño de la Tierra (*ver fig. 3*); en 1997, el grupo liderado por el Prof. Hirabayashi del Instituto Astronáutico y de Ciencias del Espacio (ISAS) de Japón puso en órbita una antena de 8 metros de diámetro (VSOP/HALCA, VLBI Space Observatory Program), para participar en las observaciones de VLBI. La órbita seguida por HALCA se determinó con una precisión entre 2 y 5 metros (Porcas y Rioja, 2000).

La técnica de VLBI supone la recepción simultánea de radioseñales o frentes de onda procedentes de un objeto en el cielo en cada una de las antenas participantes, donde cada una de ellas funciona de forma independiente usando un reloj atómico de alta precisión para controlar el tiempo y la frecuencia. Posteriormente, las señales recibidas en cada antena se envían a un centro de correlación de datos de VLBI donde se procesan de forma conjunta: la diferencia en el tiempo de llegada de los frentes de ondas a cada par de antenas se determina con altísima precisión, de unos pocos picosegundos (1 picosegundo = 1 billonésima de segundo = 10^{-12} segundos). El análisis de los datos correlacionados arroja precisiones sin precedente en la medida de posiciones, del orden del milímetro en la Tierra, y de decenas de microsegundos de arco (1 microsegundo de arco = 10^{-6} segundos de arco) en el cielo. Esto es equivalente a la resolución de un telescopio que desde la Tierra nos permitiera ver nítidamente todos los rasgos de la cara de un astronauta en la Luna. En cambio, los mejores radiotelescopios individualmente consiguen imágenes con una resolución de unos pocos segundos de arco.

Los comienzos del VLBI están ligados a la radioastronomía, donde destaca por proporcionar imágenes de la más alta resolución, y las medidas más precisas de distancias, posiciones y velocidades, en objetos situados a distancias cósmicas. Sin embargo, su aplicación en otros campos ha proporcionado resultados extremadamente importantes que la han convertido en una técnica multidisciplinaria. Destaca en su contribución en geodesia, geofísica, al establecimiento de sistemas de referencia, creación de redes dedicadas a monitorizar las deformaciones en zonas geológicamente ac-

tivas, terremotos, fallas, ..., (*ver fig. 4*) y más recientemente estudios de climatología, planetología y navegación interplanetaria.

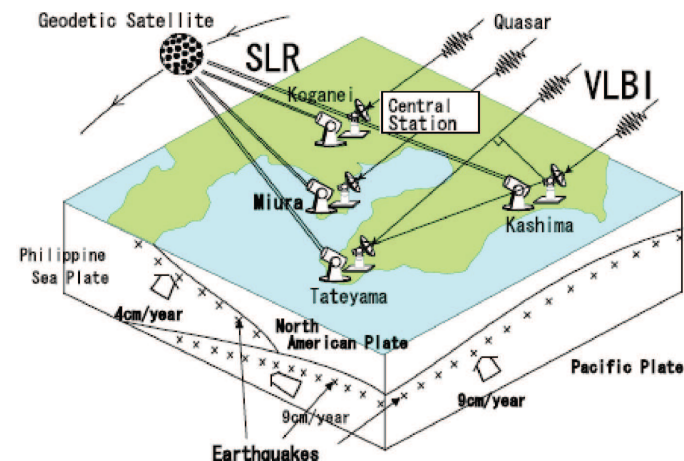


Figura 4: Proyecto *Keystone* (1996-2002) para monitorizar movimientos de tierra en el área metropolitana de Tokio (Japón) usando tres técnicas, VLBI, SLR y GPS, para la prevención de catástrofes naturales en esta zona de elevado riesgo, por la confluencia de tres placas tectónicas (Norteamericana, Pacífico y Filipina) y la alta densidad de población. La componente de VLBI la constituyen 4 antenas conectadas con el centro de procesamiento por fibra óptica, lo que permite la correlación en tiempo real (*e-VLBI*). Todo el proceso está automatizado y las estimaciones de los parámetros geodésicos están disponibles en tan sólo ¡30 minutos después de las observaciones! Desarrollado por “Communications Research Laboratory” (CRL, Kashima, Japón).

La aplicación de VLBI al campo de la geodesia, desde mediados de los años 70, ha producido resultados científicos muy interesantes relacionados con la estructura, actividad y orientación de la Tierra, utilizando observaciones astronómicas. Ha proporcionado la primera medida directa del movimiento relativo de las placas tectónicas, con el programa *Crustal Dynamics Project* (CDP), liderado por el grupo del Centro de Vuelos Espaciales de Goddard (GSFC) de NASA, y validado los modelos teóricos basados en datos geofísicos (Ryan y Ma, 1998, y otras referencias en su artículo). Se han medido movimientos seculares de hasta $63,5 \pm 0,8$ milímetros por año, entre antenas separadas 5700 kilómetros, situadas en Hawaii y Japón, que se asocia al movimiento relativo de la placa del Pacífico con respecto a Japón. También se han medido movimientos discontinuos (*ver fig. 5*) asociados con actividad volcánica y terremotos.

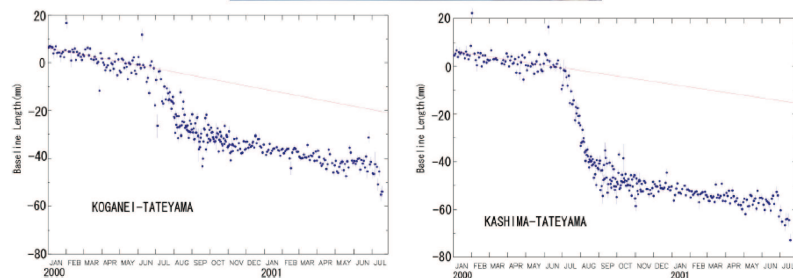


Figura 5: *Arriba*: Imagen de la erupción del volcán en la isla Miyake, situada en el archipiélago Izu (Japón) tomada el 10 de agosto de 2000. (Cortesía de Asia Air Survey Co., Ltd.) *Abajo*: Evolución temporal de la distancia entre dos pares de antenas que participan en el proyecto *Keystone*, medida con VLBI. La tendencia secular se ve claramente alterada por la erupción volcánica en la isla Miyake, del archipiélago Izu. *Créditos*: Gráfico obtenido de una publicación de Taizoh Yoshino y colaboradores, del Communications Research Laboratory, Japón.

En la actualidad, VLBI es una técnica fundamental en geodesia que colabora en la definición y mantenimiento de un sistema de referencia terrestre (ITRF, International Terrestrial Reference Frame) materializado por un conjunto de más de 200 antenas en la Tierra, cuyas posiciones y velocidades se conocen con altísima precisión, del orden del milímetro y del milímetro por año, respectivamente.

Es además la única técnica geodésica espacial utilizada para la definición y mantenimiento del sistema de referencia inercial celeste (ICRF, International Celestial Reference Frame), materializado por las posiciones de unos 300 objetos en el espacio, los cuásares, cuyas coordenadas se conocen con precisiones de unas pocas centenas de micro segundos de arco.

También es única en su capacidad de medir simultáneamente todos los parámetros que caracterizan la orientación de la Tierra en un marco de referencia inercial, una información indispensable para la navegación

espacial. A modo de ejemplo, VLBI puede detectar cambios en la longitud del día, debidos a irregularidades en la rotación diurna, de hasta unos pocos microsegundos.

Los cambios en la orientación de la Tierra responden al efecto integrado del intercambio de momento angular entre la tierra sólida, la atmósfera, los océanos y el núcleo fluido, así como a cambios en la forma de la Tierra debido a cargas glaciares, oceánicas y atmosféricas. Las observaciones ultraprecisas con VLBI proporcionan una importante base de datos para estudiar los modelos de los distintos componentes de la Tierra propuestos por los geofísicos. Quizás una de las aplicaciones más sorprendentes sea al estudio del interior de la Tierra; en este campo las observaciones con VLBI han revelado la existencia de un núcleo con una forma más elíptica de lo que se pensaba.

Por supuesto, la extremada precisión alcanzable con las técnicas de VLBI no ha pasado inadvertida entre la comunidad astronáutica. Los objetivos de las nuevas misiones espaciales requieren una gran precisión en los sistemas de control para la navegación interplanetaria. Por ello VLBI, en su configuración Δ DOR, se ha incorporado al protocolo para la navegación espacial interplanetaria.

Misiones

En esta sección se hace referencia explícitamente a 5 misiones espaciales que están íntimamente relacionadas con la aplicación de técnicas ultraprecisas (VLBI) para la navegación espacial; podemos decir que representan el “estado del arte” de las técnicas de navegación espacial.

Mars Express (MEX)

La Agencia Espacial Europea (ESA) lanzó su primera misión a Marte, el *Mars Express* (MEX), el 2 junio de 2003, desde el cosmódromo de Kazajastán. La nave llegó a su destino el 25 de diciembre de 2003 y desde entonces permanece en órbita exploratoria alrededor del planeta. Durante este tiempo los instrumentos a bordo de la nave han realizando experimentos que han permitido ampliar nuestro conocimiento sobre la atmósfera de Marte, proporcionando mapas globales de alta resolución de la composición atmosférica y de su circulación global. Merecen especial mención las imágenes en color de la superficie de Marte tomadas con su cámara de alta resolución (HRSC, High Resolution Stereo Camera). Dicha cámara, además, permite captar imágenes en varias direcciones a la vez que dan lugar a espectaculares perspectivas en 3 dimensiones (*ver fig. 6*). Otro de los instrumentos a bordo de la MEX es un radar para el sondeo del subsuelo y la ionosfera (MARSIS, Mars Advanced Radar for Subsurface

and Ionosphere Sounding). Uno de los principales objetivos de este radar es buscar agua, en forma líquida o helada, en el subsuelo marciano. La vida de la misión ha sido recientemente prolongada por dos años más, hasta finales de 2007, para dar tiempo a completar los proyectos de exploración.

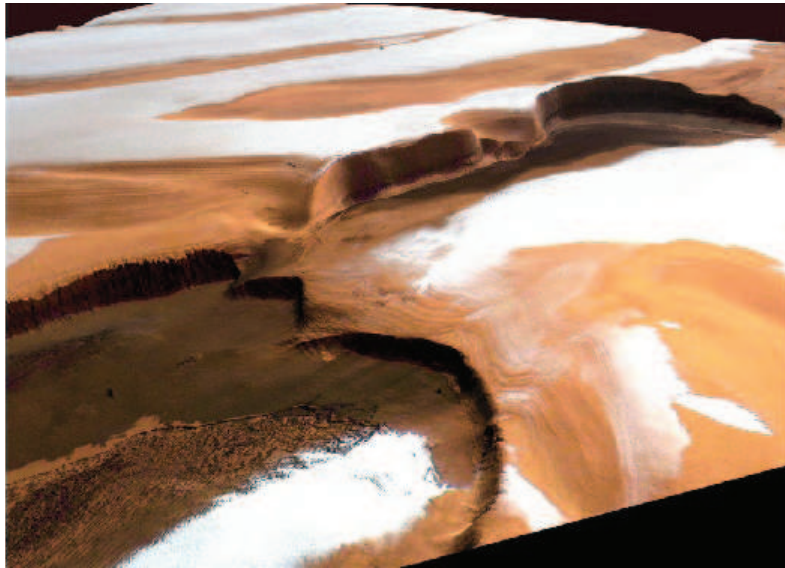


Figura 6: Perspectiva del casquete de hielo en el polo Norte de Marte captada por la cámara de alta resolución a bordo de la nave *Mars Express* (MEX). Se aprecian acantilados de unos 2 kilómetros de altura. El material oscuro en la parte más baja podrían ser cenizas volcánicas (Cortesía de ESA/DLR/FU Berlín (G.Neukum).

La misión transportaba una sonda con un robot, el *Beagle 2*, para la exploración “in-situ” de la superficie marciana, al cual se le perdió la pista antes de llegar a su destino. Existen especulaciones sobre el destino final del *Beagle 2*, entre otras la posibilidad de que quedara atrapado dentro de un cráter, pero lo cierto es que no se tiene ninguna evidencia observacional más allá del momento en que la sonda se separó de la nave principal. En la Tierra, la esperada señal que emitiría el robot al llegar a la superficie, planeada para el 24 de diciembre de 2003, no se llegó a detectar.

La pérdida de la sonda no ha empañado el éxito de la misión, que también desde un punto de vista técnico ha supuesto un gran logro por superar los desafíos de navegación necesarios para realizar las complicadas maniobras.

En total, durante la fase de navegación interplanetaria, se llevaron a cabo 55 observaciones de 90 minutos de duración con la técnica Δ DOR, cada 3 o 4 días, utilizando las antenas del DSN (con observaciones simultáneas entre las antenas en Goldstone y Canberra, y entre Goldstone y Madrid).

La fase de aproximación a Marte fué sin duda una de las mas críticas de la misión. Durante las semanas anteriores al encuentro se intensificó la campaña de observaciones, hasta dos observaciones por día, y se redujo el tiempo empleado en el procesamiento de las observaciones Δ DOR hasta 12 horas. En esta fase se efectuaron complicadas maniobras para dirigir la nave al punto idóneo desde donde soltar la sonda *Beagle 2*, y acto seguido redirigirse desde la trayectoria de impacto hasta una que permitiría a la nave principal entrar en órbita alrededor de Marte. La precisión en la determinación de la órbita fué crucial para predecir la trayectoria de la nave, y el diseño de las maniobras. La precisión en las medidas de las posiciones de la nave superó las expectativas, resultando ser mejores que 0,1 nanosegundos, que corresponde a un error de algo más de medio kilómetro a la distancia a la que se encuentra Marte (Dongsuk Han et al., 2004).

MER

En las mismas fechas que la ESA enviaba su misión MEX a Marte, la NASA preparaba el lanzamiento de su misión *Mars Exploration Rover* (MER) para la exploración del planeta rojo mediante 2 robots móviles, el *Spirit* y el *Opportunity*, depositados en su superficie (ver fig. 7). El objetivo de la MER era la exploración geológica de la superficie marciana en busca de evidencias que aporten respuestas a la cuestión abierta sobre la existencia de agua en Marte, y para determinar si se dan las condiciones favorables para preservar pruebas de cualquier tipo de posible vida en el pasado. Los robots se controlan desde Tierra directamente, mediante el envío y recepción de comandos a través de las antenas del DSN, o utilizando como intermediarios otras naves de NASA, el *2001 Mars Odyssey* y el *Mars Global Surveyor*, que se encuentran orbitando alrededor de Marte. A día de hoy, ambos robots han superado las expectativas de supervivencia y descubrimientos en Marte, y continúan en buen estado (página web de la misión MER).

La misión MER tiene también gran interés desde el punto de vista de la comparación de técnicas de navegación espacial de alta precisión. Durante el vuelo interplanetario se han utilizado las dos técnicas de navegación relacionadas con VLBI: en su configuración Δ DOR, tradicionalmente usada por NASA; y usando las técnicas de referencia de fase tradicionalmente usadas en radioastronomía, realizado por el grupo compuesto por el Dr. Lanyi y sus colaboradores. Las observaciones con Δ DOR se llevaron a cabo con la red de antenas de espacio profundo (DSN) de NASA; las obser-



Figura 7: Impresión artística de uno de los dos robots transportados por la misión *Mars Exploration Rover* (MER) para la exploración “in-situ” de la superficie de Marte (cortesía de NASA).

vaciones de referencia de fase con la red de antenas de VLBI del NRAO (VLBA, Very Long Baseline Array). Ambas herramientas de navegación espacial están íntimamente relacionadas con la técnica de VLBI, aunque difieren en la estrategia observacional, y en los observables utilizados en el análisis. Ambas miden la posición relativa de la nave con respecto a un sistema inercial predeterminado en el espacio, o dicho de otro modo, proporcionan una navegación utilizando como referencia las posiciones de las “estrellas” fijas (o mejor, de los “cuásares”).

El estudio comparativo presentado por el Dr. Lanyi y colaboradores (Lanyi et al., 2005) concluye que las precisiones proporcionadas utilizando la configuración “astronómica” de referencia de fase superan a las de Δ DOR, en las observaciones a las frecuencias utilizadas hoy en día para la comunicación con satélites (banda X, a 8,4 GHz). En este caso, el error formal en las posiciones estimadas a partir del análisis de las observaciones con las antenas del VLBA es la mitad del correspondiente a las observaciones con la DSN. También concluyen que se espera que ambas técnicas proporcionen precisiones equivalentes usando observaciones a frecuencias más altas, por ejemplo a 32 GHz (banda Ka), a las que migrarán las comunicaciones con satélites en un futuro próximo.

Nozomi

Nozomi (=“esperanza”, en japonés) fué la primera sonda desarrollada y enviada por la Agencia Espacial Japonesa (ISAS) a Marte, el 4 de julio de 1998. Su objetivo científico era estudiar el efecto del viento solar en la

atmósfera de Marte y también desarrollar nuevas tecnologías para futuras misiones al planeta.

Su llegada al planeta rojo sufrió un importante retraso debido a un problema en la propulsión. A raíz de un excesivo consumo de combustible durante una maniobra en su primer giro alrededor de la Tierra, los expertos en navegación decidieron realizar una laboriosa reestructuración del plan de vuelo. Se optó por colocar la nave en una órbita heliocéntrica a la espera de poder realizar un par de maniobras de asistencia gravitatoria alrededor de nuestro planeta, que le transfirieran el suficiente impulso de velocidad para tomar rumbo definitivo hacia Marte. Finalmente llegó a su destino en enero de 2004.

La relevancia de esta misión, relativa al contenido de este artículo, reside en su contribución al desarrollo de la aplicación de la técnica de VLBI en su configuración “astronómica” para la navegación espacial, en tiempo real. Durante su trayectoria se llevaron a cabo más de 30 observaciones con VLBI de las misiones *Nozomi* y *Hayabusa* (ver más adelante), entre septiembre de 2002 y noviembre de 2003, usando mayormente una red de 9 antenas distribuidas por territorio japonés (Ichikawa et al. 2003, Sekido et al. 2003). Los resultados fueron validados por comparación con medidas telemétricas.

Japón cuenta con una larga historia en las aplicaciones de VLBI a la navegación espacial. El Dr. Asaki y colaboradores (Asaki et al., 1998) realizaron el primer intento satisfactorio de determinación ultraprecisa de la posición de un satélite utilizando VLBI en su configuración “astronómica”, es decir, con observaciones alternadas del satélite y de cuásares cercanos, y explotando al máximo el potencial de VLBI. Aunque hubo grupos anteriores que utilizaron esta configuración con éxito, en las dos misiones soviéticas *Vega* (McElrath et al., 1988) y *Phobos* (Hildebrand et al. 1994), el grupo del Dr. Asaki es el primero que utilizó el observable fase interferométrica, el más preciso de los observables de VLBI, en el análisis.

Japón lidera asimismo el desarrollo tecnológico para la realización de observaciones de VLBI en “tiempo real”, utilizando fibra óptica que permite la transmisión de datos a alta velocidad desde cada antena hasta el centro de procesamiento de datos de VLBI en Kashima, cerca de Tokio (Japón). Esto permite obtener estimaciones de las posiciones en un tiempo récord, de vital importancia para la aplicación de esta técnica a la navegación espacial.

Hayabusa

... o la misión “halcón” japonesa, que mantiene en vilo a los miembros del equipo de navegación mientras se confecciona este artículo, ocupados en la ejecución de complicadas maniobras de aproximación de la nave a un

asteroide para su exploración, incluídos el descenso de robots y toma de muestras de material ¡para su posterior retorno a la Tierra!

Los asteroides están compuestos por material sin procesar, que ha dado lugar a los planetas. Por eso su estudio puede ayudar a resolver muchos misterios sobre la génesis del sistema solar. Son, por decirlo de alguna manera, cápsulas que han escapado al paso del tiempo y a partir de su estudio se puede recrear el pasado.



Figura 8: Imagen del asteroide *Itokawa* tomada desde la cámara a bordo de la nave *Hayabusa*, en órbita alrededor del asteroide (cortesía de ISAS/JAXA).

El 9 de mayo de 2003 la Agencia Espacial Japonesa (JAXA, que engloba a la antigua ISAS) lanzó al espacio una nave para explorar asteroides, y desarrollar la tecnología que permita recoger muestras y traerlas de vuelta a la Tierra, por primera vez en la historia. Cuenta con una serie de instrumentos a bordo, cámaras para obtener imágenes del asteroide, altímetros láser para medir el relieve y espectrómetros para analizar la composición de sus minerales. Esta misión tiene muchos aspectos innovadores, entre ellos el retorno de muestras del suelo y rocas del asteroide para estudiar en laboratorios ubicados en la Tierra; también, la utilización de un novedoso sistema de propulsión a base de un motor de iones de gas xenón.

El 12 de septiembre de 2005 *Hayabusa* llegó a su destino, el asteroide 25143 *Itokawa* (que toma su nombre del padre del programa espacial japonés, el Dr. Hideo Itokawa), que se encuentra a 320 millones de kilómetros de la Tierra. El plan de la misión se está desarrollando según lo previsto, con alguna que otra variación. La nave, tras llegar a su destino, ha permanecido en órbita alrededor del asteroide para obtener fotos de su superficie, desde unos 7 kilómetros de distancia. Sus primeras imágenes

muestran que el asteroide mide unos 600 metros de longitud y tiene forma muy elongada (ver fig. 8). La misión transporta dos robots equipados con cámaras que, una vez soltados, se desplazarán por el asteroide durante varios días para enviar imágenes del paisaje.

Finalmente, la nave se acercará y se posará en la superficie para recoger muestras. En diciembre de 2005 la nave abandonará el asteroide *Itokawa* y emprenderá su regreso a la Tierra, donde se espera que llegue en junio de 2007 (ver la página web de la misión *Hayabusa*).

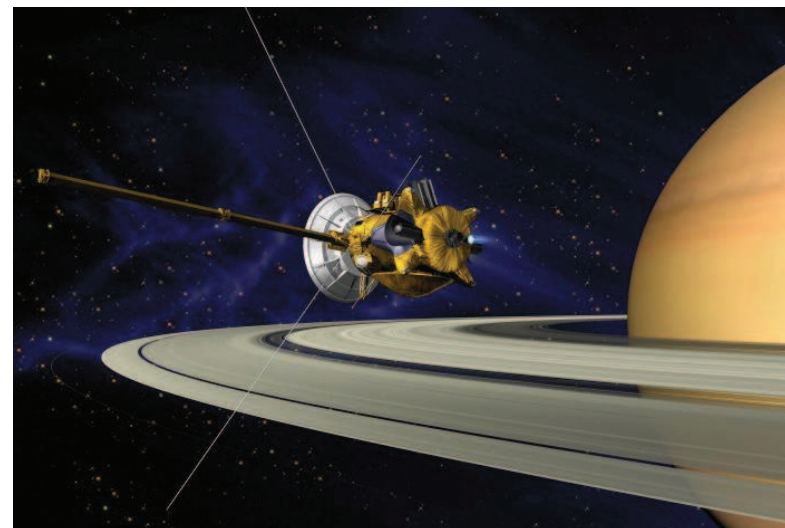


Figura 9: Impresión artística de la misión *Cassini-Huygens* en el momento de su llegada al sistema de Saturno (cortesía de NASA-ESA). La sonda *Huygens* viajó a lomos de la *Cassini* durante los casi 7 años de travesía interplanetaria. Es el instrumento con forma de cono que se aprecia en la imagen. Finalmente se separó de la nave principal para iniciar un descenso atravesando la densa atmósfera de Titán hasta su superficie.

Cassini-Huygens

Sin duda, uno de los logros del año en el campo de VLBI ha sido la contribución de esta técnica a la misión planetaria *Cassini-Huygens*. Las precisiones alcanzadas en la determinación de posiciones y de trayectorias han sido realmente espectaculares. Se puede decir que ha entrado en el libro de records de VLBI por ser el aterrizaje efectuado a mayor distancia, por control remoto, desde la Tierra; y por la extensa presencia institucional, con

la participación de 12 organizaciones pertenecientes a 7 países diferentes. Este puesto seguirá vigente durante varios años más ya que no hay planeada ninguna misión espacial para el futuro que pueda equiparar estos logros.

En julio del año pasado la nave *Cassini* entró en órbita alrededor de Saturno después de una travesía interplanetaria de casi 7 años y 3.500 millones de kilómetros recorridos (ver fig. 9). La nave empezó a enviar imágenes de alta calidad del planeta y sus anillos y de sus más de 30 lunas; éste era el principio de una misión científica de reconocimiento del planeta Saturno y sus lunas que durará cuatro años. Esta misión está capacitada para llevar a cabo una exploración mucho más completa que sus predecesores, puesto que está equipada con un sofisticado laboratorio de instrumentos científicos. Entre ellos, la sonda *Huygens*, una especie de cono de tres metros de diámetro, cuya misión es sumergirse en la densa y desconocida atmósfera de Titán, la luna más grande de Saturno. El plan de la misión era que *Cassini* la colocara en una trayectoria de impacto, a partir de donde comenzaría un descenso de dos horas y media hacia la superficie, con la ayuda de un paracaídas; su objetivo: llevar a cabo experimentos durante su descenso, además de realizar medidas de la temperatura, presión, densidad y balance de energía en la atmósfera, y de tomar fotografías de las nubes y de la superficie. El retorno de datos de *Huygens* a la Tierra se organizó, durante la planificación de la misión, en dos fases: en primer lugar toda la información se transmitiría al cercano orbitador *Cassini*, y en una segunda fase desde *Cassini* a la Tierra, usando su antena principal de 4 metros de diámetro, al final de la misión.

Entre los experimentos a realizar por *Huygens* figura uno relacionado con el contenido de este artículo, el *Doppler Wind Experiment* (DWE), cuyo investigador principal es el Dr. Michael Bird (Universidad de Bonn, Alemania). El objetivo científico del DWE es obtener medidas precisas del perfil de los vientos a diferentes alturas en la atmósfera de Titán, mediante medidas Doppler de la señal de *Huygens* a lo largo de su trayectoria de descenso – estabilizada con la ayuda de alas aerodinámicas que generan un giro controlado a baja velocidad. Los vientos en la atmósfera afectarán la velocidad horizontal durante el descenso dando lugar a cambios en la frecuencia de la señal recibida por *Cassini*, procedente del oscilador ultraestable a bordo de *Huygens*. A partir de estas medidas Doppler se pueden estimar las variaciones de la velocidad del viento con la altura en Titán; más concretamente, de la componente de la velocidad en la dirección *Huygens-Cassini*.

La nave *Cassini* emprendió el largo viaje hacia Saturno el 15 de octubre de 1997 desde la estación de lanzamiento de cohetes en Cabo Cañaveral (Florida, Estados Unidos) y es el resultado de una colaboración entre tres agencias espaciales, americana, europea e italiana, NASA, ESA y ASI, respectivamente. Durante la fase de crucero se ha acercado a Venus y a la Tierra, en 1998, y posteriormente también a Júpiter, con el fin de realizar

maniobras de asistencia gravitatoria y adquirir la velocidad suficiente para poner rumbo a Saturno. Durante los 6 meses que *Cassini* estuvo en las cercanías de Júpiter coincidió con otro satélite de la NASA, el *Galileo*, en órbita desde 1995 hasta 2003, lo que proporcionó la oportunidad poco frecuente de estudiar este planeta desde 2 perspectivas diferentes al mismo tiempo. Este encuentro enriqueció los estudios de la influencia del viento solar en la magnetosfera, sus lunas, anillos y nubes.

El 25 de diciembre de 2004, la sonda europea *Huygens* se separó de la nave principal y emprendió su camino en solitario. Tres semanas más tarde, el 14 de enero de 2005, comenzaría su descenso hacia la superficie de Titán, convirtiéndose en el primer caso de “alunizaje” en la luna de otro planeta. Este apasionante acontecimiento, sin precedentes en la historia de la astronáutica, fué seguido por numerosos miembros de la comunidad científica y de los medios de comunicación, que se desplazaron hasta alguno de los centros de seguimiento de naves espaciales para seguir en directo el desarrollo de los acontecimientos. Durante las dos horas y media que duró el descenso hubo momentos de euforia, alternados con momentos de gran consternación y de esperanza.

La monitorización de la fase de descenso de *Huygens* (ver la figura que encabeza el artículo) contó con el apoyo de las observaciones, desde Tierra, con un arsenal de antenas situadas en Estados Unidos, Australia, Japón, China y Europa. Un total de 20 antenas para complementar la estrategia de retorno de datos de *Huygens*, inicialmente planeada únicamente a través de la antena principal de *Cassini*. Es digno de recalcar que estas antenas observaron la débil señal enviada por *Huygens* a *Cassini* (¡desde una distancia de 1.200 millones de kilómetros!).

Las observaciones terrestres tenían dos objetivos: por una parte conocer en “casi” tiempo real si la misión estaba operacional y transmitiendo (con apenas 67 minutos de retraso introducidos por la propagación, frente a las 4 o 5 horas de retraso contempladas inicialmente); por otra, aportar medidas de la componente de velocidad en la dirección *Huygens-Tierra* para el experimento DWE, para complementar las medidas de la única componente, *Huygens-Cassini*, contempladas en la misión original.

Encabezando la lista de potentes antenas participantes en el programa estaba el Telescopio Robert C. Byrd en Green Bank (GBT) del NRAO, en Virginia del Oeste, Estados Unidos, y el Radiotelescopio Parkes del CSIRO, en Nueva Gales del Sur, Australia, con 110 y 64 metros de diámetro, respectivamente. Para las medidas Doppler durante esta oportunidad única se utilizó instrumentación especialmente diseñada para la detección de señales débiles. La detección inicial, llevada a cabo con los “Receptores de RadioCiencia” ultrasensibles del Laboratorio de Propulsión a Chorro (JPL, California) proporcionó la primera prueba inequívoca de que *Huygens* había sobrevivido a la fase de entrada y había comenzado su transmisión de radio señales hacia *Cassini*. Esta sofisticada instrumentación permitió

monitorizar en “casi” tiempo real los momentos principales de la misión, cuándo se abrió el primer paracaídas, el cambio de paracaídas tras unos 15 minutos de vuelo atmosférico, y determinar el momento del impacto en Titán, después de 2 horas y media de descenso a una velocidad de unos 5 metros por segundo. Estos eventos dejaron su impronta característica, claramente identificable, en las medidas Doppler de la débil señal. También permitieron seguir en contacto con la sonda horas después del aterrizaje, proporcionando datos para medidas precisas del movimiento de Titán, y para la calibración de las medidas tomadas durante el descenso.

Otro equipo liderado por el Dr. Leonid Gurvits, científico del Instituto Conjunto Europeo para VLBI (JIVE), en Holanda, coordinaron las observaciones de VLBI con una red de 16 radiotelescopios, que grabaron la señal de *Huygens* y la de un cuásar cercano en el cielo, en configuración “astronómica”; el objetivo siendo la determinación precisa de su trayectoria de descenso, usando técnicas de referencia de fase durante el análisis. Las antenas participantes, distribuidas por toda la superficie terrestre, en Norteamérica (red VLBA), Australia (antenas en Mopra, Ceduna, Hobart), China, Japón y Europa, se fueron incorporando a las observaciones a medida que el planeta se alzaba en sus horizontes. Para aumentar la sensibilidad se utilizaron las más altas velocidades de grabación utilizadas en la historia de esta técnica, de 1 Gigabit por segundo, en discos magnéticos de alta capacidad.

La gran expectación reinante durante las dos horas y media que pasaron entre la primera toma de contacto con *Huygens* y su llegada a la superficie de Titán se vio incrementada por la inesperada larga supervivencia de la sonda tras su alunizaje, que en ningún caso se esperaba que superase los 30 minutos. Sin embargo, aunque la señal enviada por la sonda seguía siendo recibida alta y clara en Tierra, Saturno se iba acercando al horizonte de la antena australiana de Parkes, lo que marcaba el fin de su visibilidad. Cuando parecía claro que *Huygens* seguiría transmitiendo durante más tiempo del que podría verse desde Hawaii, Australia y el este de Asia, se llevó a cabo una labor encomiable que culminó con la organización urgente de observaciones “ad hoc” con antenas en Europa: Westerbork (Holanda), Onsala (Suecia), Medicina (Italia) y Wettzell (Alemania), abandonaron sus ocupaciones, se dirigieron hacia Titán y sintonizaron sus receptores para recibir la señal de *Huygens*.

Para contribuir a la excitación del momento, poco después se tuvo conocimiento de un fallo en la configuración de uno de los receptores en *Cassini* que afectaba de forma fatal a la recogida de la señal ultraestable de *Huygens*. Esto impediría obtener las medidas Doppler entre la sonda y *Cassini*, lo que significaba el fracaso total del experimento DWE, y causó gran desánimo entre el grupo de científicos ansiosos de tener acceso a estos datos, después de una espera que se ha prolongado ¡durante décadas!

En este ambiente de desolación, los científicos del DWE sabían que el

éxito de la componente VLBI del seguimiento sería lo único que podría dar una segunda oportunidad a la misión. No se ahorraron esfuerzos para agilizar el envío de los discos magnéticos grabados en cada una de las antenas al centro de procesamiento de datos de VLBI en Dwingeloo (Holanda), donde llegaron en cuestión de horas. Las buenas noticias no se hicieron esperar y llegaron poco después, en tiempo record, con la confirmación de la detección de franjas de interferencia, desde Dwingeloo (Holanda).

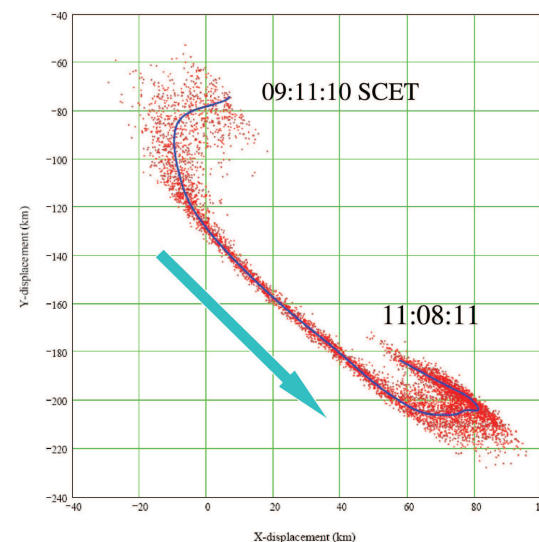


Figura 10: Impresión artística de la alta precisión obtenida con VLBI en la determinación de la trayectoria de la sonda *Huygens*, comparada con otras técnicas de navegación espacial. La alta resolución de VLBI permite determinar la trayectoria de impacto con una precisión mejor que 1 kilómetro, a una distancia de 1200 millones de kilómetros (figura cedida por el Dr. Leonid Gurvits).

El análisis de los datos de VLBI se está llevando a cabo utilizando un nuevo paquete de programas diseñado especialmente para este proyecto, cuyo objetivo es determinar la posición de la sonda con una precisión de 1 kilómetro en Titán, ¡a una distancia de 1.200 millones de kilómetros de la Tierra! (ver fig. 10). Esta resolución angular permitiría distinguir, desde Tierra, la pelota en un partido de ping-pong jugado en la Luna (cita literal del Dr. Leonid Gurvits).

Ya existen resultados preliminares (Bird et al., 2005) de las variaciones del viento con la altura en Titán a partir del análisis de los datos de VLBI

y de las medidas Doppler (*ver fig. 11*). Estos han permitido determinar que los vientos en Titán soplan en la dirección de la rotación de Titán (de Oeste a Este) a casi todas las alturas. La máxima velocidad encontrada, de unos 120 metros por segundo, fue medida minutos después del comienzo del descenso, a una altura de unos 120 kilómetros. Cerca de la superficie los vientos son flojos y se incrementan poco a poco con la altura hasta llegar a los 60 kilómetros, donde se observa un cambio del comportamiento y grandes variaciones en las medidas Doppler, un resultado inesperado y todavía sin explicación. También existen estimaciones de VLBI preliminares de la trayectoria bidimensional (proyección de la trayectoria en el plano del cielo), con una precisión de 1 kilómetro (*ver fig. 12*). El estudio combinado de las medidas Doppler y las medidas de VLBI permitirá a los científicos reproducir la posición y trayectoria de la sonda, en 3 dimensiones, en su descenso hacia la superficie con altísima precisión.

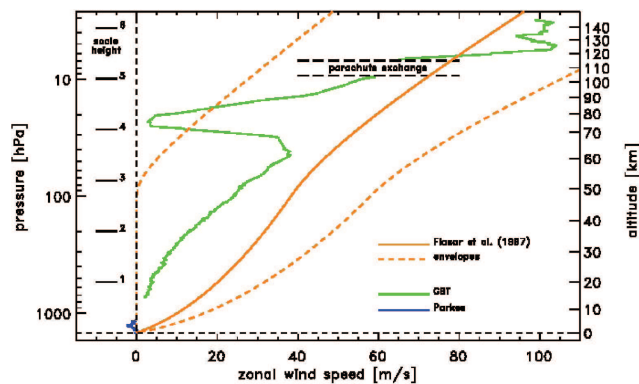


Figura 11: Resultados preliminares del perfil de los vientos a diferentes alturas en la atmósfera de Titán, obtenidas a partir de las observaciones con VLBI y las medidas “Doppler” (*verde*). También se muestran las predicciones de los modelos (*rojo*). La máxima velocidad se midió momentos después del comienzo del descenso, y se observa una disminución progresiva con la disminución de la altitud, a excepción de un comportamiento anómalo y todavía sin explicación a una altura de unos 60 kilómetros. (Figura cedida por el Dr. Leonid Gurvits. Material elaborado por el Dr. Bird y colaboradores para su publicación en Nature, 2005.)

Lo aprendido en Titán durante su descenso servirá a los planetólogos, además de para conocer este lejano y misterioso cuerpo, para ampliar los conocimientos sobre la atmósfera primitiva de nuestro propio planeta, ya que existen argumentos para pensar que Titán es una copia “helada” de la Tierra en sus comienzos.

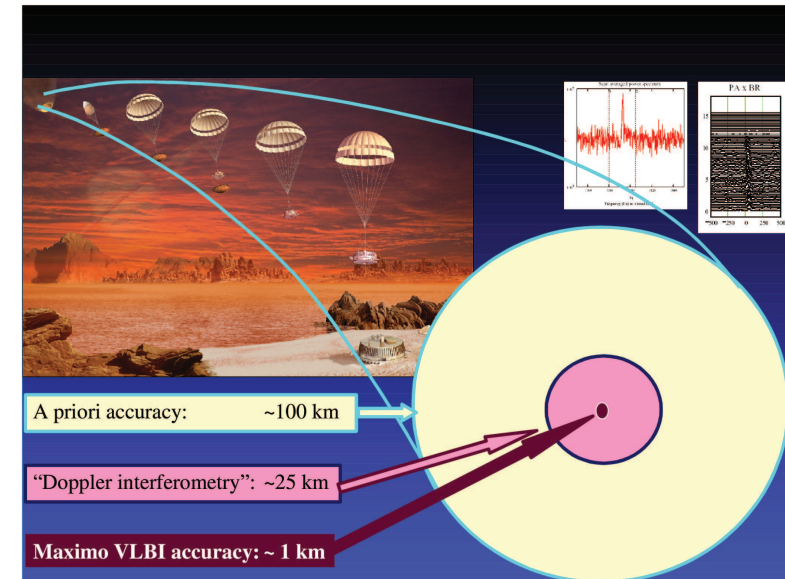


Figura 12: Reconstrucción de la trayectoria de descenso a través de la atmósfera de Titán seguida por la sonda *Huygens* a partir de medidas ultra precisas con la técnica VLBI. Los puntos representan las posiciones medidas y la línea su trayectoria. Son posiciones relativas a un cuásar, resultado del análisis con técnicas de referencia de fase que proporcionan las mayores precisiones alcanzables. *Nota:* Conversión entre unidades angulares y lineales a la distancia de Titán: 1 milisegundo de arco equivale a 6 kilómetros. (Figura cedida por el Dr. Leonid Gurvits. Material elaborado por el Dr. Bird y colaboradores para su publicación en Nature, 2005.)

Futuro ...

Sin lugar a duda, el desarrollo de nuevas tecnologías permitirán la exploración del espacio con más detalle todavía, mediante la obtención de imágenes de mayor resolución, descensos precisos a la superficie, movilidad en la superficie con mayor alcance, retorno de las misiones con información para su estudio en la Tierra, nuevos métodos más efectivos para ganar impulso, ejecución de maniobras más complejas, o sistemas de navegación autónomos, entre otros. Todo ello orientado a poder navegar por el espacio tan cómodamente como nos desplazamos hoy en día por la Tierra con ayuda de un receptor GPS!

En particular, en el campo de VLBI, se está dedicando un gran esfuerzo al desarrollo de nuevas tecnologías para permitir observaciones en tiempo real (*e*-VLBI), utilizando conexiones con fibras ópticas, y para la adquisición de datos a mayor velocidad. Una vez en funcionamiento la dotarán de un carácter más dinámico, al reducir el tiempo de espera entre las observaciones y los resultados, y crecerá su campo de “acción” para incluir objetivos más débiles. También, el futuro de VLBI está ligado a la integración con otras técnicas espaciales (SLR, GPS y Galileo, entre otras).

La misión *Cassini-Huygens* ha proporcionado un ejemplo de la sinergia entre técnicas experimentales desarrolladas para su uso en astronomía y basadas en radiofísica, electrónica y procesamiento de señales digitales, con la ciencia planetaria y la exploración del sistema solar con sondas de espacio profundo. Sin duda, es un primer paso en la aplicación futura de VLBI a las misiones planetarias en las próximas décadas. Asimismo representa un magnífico ejemplo de la efectividad de la auténtica cooperación científica a escala global.

Referencias

- Asaki, Y., et al., *RadioScience*, 33, 1297, 1998.
- Bird, M.K., et al., The vertical profile of winds on Titan, *Nature* 2005 (en prensa).
- Bird, M.K., et al., American Astronomical Society, DPS meeting #37, 2005.
- Bird, M.K., et al., The Cassini/Huygens Doppler Wind Experiment: Results from the Titan Descent, 36th Annual Lunar and Planetary Science Conference, in League City, Texas, abstract no.1620, 2005.
- Dongsuk Han et al., “18th International Symposium on Space Flight Dynamics”, celebrado en Munich, Alemania, 2004.
- Hildebrand, C.E. et al., “Radio-Planetary Frame Tie From Phobos-2 VLBI Data”, TDA Progress Report 42-119, 1994.
- Ichikawa et al., An evaluation of VLBI observations for the positioning of NOZOMI Spacecraft and the future direction in research and development of the deep space tracking using VLBI, *IVS CRL/TDC News*, No. 23, pp 31-33, 2003.
- Lanyi et al., Determination of Angular separation between spacecraft and quasars with the Very Long Baseline Array, *IPN Progress Report* 42-162, 2005.
- McElrath, T.P. and Bhat, R.S., “Determination of the Inner Planet Frame Tie Using VLBI Data”, paper 88-4234, AIAA/AAS Astrodynamics Conference, Minneapolis, Minnesota, 1988.
- Porcas, R.W., and Rioja, M.J., “Space VLBI Astrometry and Phase-

Reference Mapping with VSOP”, *Proc. of the 14th Working Meeting of European VLBI for Geodesy and Astrometry*, p.139, 2000.

Rogers, A.E.E, et al., VLBI: The MarkIII system for geodesy, astrometry and aperture synthesis, *Science*, 219, 51-54, 1983.

Ryan, J.W., and Ma, C., NASA-GSFC’s Geodetic VLBI Program: a twenty year retrospective, *Phys. Chem. Earth*, Vol. 23, No. 9-10, pp. 1041-1052, 1998.

Sekido et al., VLBI Application for Spacecraft Navigation (NOZOMI) - follow-up on Model and Analysis, *IVS CRL/TDC News*, No. 23, pp 34-35, 2003.

Thornton, C.L., Border, J.S., “Radiometric Tracking Techniques for Deep-Space Navigation”, Wiley-Interscience, 2003.

Páginas web:

Misión MER, <http://marsrovers.jpl.nasa.gov/mission/>

JAXA, <http://www.isas.jaxa.jp/en/>

Misión Hayabusa, http://global.jaxa.jp/projects/sat/muses_c/

EL OBSERVATORIO ESPACIAL HERSCHEL

Pere Planesas

*Observatorio Astronómico Nacional
y
Centro Nacional de Información Geográfica
Instituto Geográfico Nacional - Ministerio de Fomento*

Javier Graciá Carpio

*Fractal SLNE
y
Observatorio Astronómico Nacional*

Abstract

The Herschel Space Observatory will be the first astronomical mission to study the cold Universe at far-infrared and submillimeter wavelengths. Its main goal is to look at the formation of stars and galaxies. It will carry the largest telescope ever launched and three complementary scientific instruments. In this paper we describe the telescope, instruments, operation and main scientific objectives, as well as the OAN/CNIG participation in the design, construction and scientific core program of the heterodyne instrument HIFI.

Introducción

El Observatorio Espacial Herschel, conocido también por sus siglas en inglés HSO y coloquialmente como “Herschel”, es la siguiente gran misión científica en astronomía que la Agencia Europea del Espacio (ESA) va a

lanzar, acompañada de Planck, una misión para el estudio de la radiación cósmica de fondo. Herschel es un vehículo espacial de gran tamaño equipado con un telescopio de 3,5 m de diámetro destinado a observar el universo durante más de 3 años en un rango espectral prácticamente inexplorado hasta la fecha, al no ser observable desde la superficie terrestre y al requerir sistemas de medida que no se han desarrollado hasta muy recientemente. Sus tres instrumentos científicos cubren las regiones del espectro electromagnético conocidas como infrarrojo lejano (FIR, en sus siglas en inglés) y submilimétrico o, más precisamente, detectará radiación emitida o absorbida por los astros entre las longitudes de onda de 55 y 672 μm ¹. Tras su lanzamiento, actualmente previsto en 2008, funcionará como un observatorio astronómico multiusuario disponible para la comunidad científica mundial.

Este vehículo fue bautizado como Observatorio Espacial Herschel en la reunión *The promise of FIRST* que tuvo lugar en Toledo (España) en diciembre de 2000. Inicialmente denominado *Far-InfraRed and Submillimeter Space Telescope* (Telescopio Espacial para el Infrarrojo Lejano y Submilimétrico), se decidió entonces nombrarlo con el apellido del descubridor de la radiación infrarroja, justo dos siglos antes, el músico y astrónomo Friedrich Wilhelm (William) Herschel (1783-1822).

El Observatorio Espacial Herschel, con su gran telescopio de 3,5 m, sus sensibles detectores y su elevada resolución espectroscópica, superará en mucho la capacidad de los telescopios espaciales en infrarrojo lanzados hasta la fecha, alcanzándose así la mayoría de edad de la astronomía del infrarrojo lejano, pues se dispondrá de espectrometría en todo este rango y una elevada resolución espacial. En 1983 se lanzó el satélite IRAS (*InfraRed Astronomical Satellite*), equipado con un telescopio refrigerado de 57 cm de diámetro y que observó el universo durante diez meses a las longitudes de onda de 12, 25, 60 y 100 μm . En 1995 la ESA lanzó el satélite ISO (*Infrared Space Observatory*), equipado con un telescopio de 60 cm, que fué operativo hasta 1998 observando el universo a longitudes de onda entre 2,5 y 240 μm con detectores entre 4 y 1000 veces más sensibles que IRAS y con una cierta capacidad espectroscópica de baja resolución. El telescopio espacial Spitzer, lanzado por la NASA en 2003 y operativo al menos hasta 2008, está equipado con un telescopio de 85 cm de diámetro y es capaz de obtener imágenes a varias longitudes de onda entre 3,6 y 160 μm y espectroscopía de baja resolución entre 5 y 38 μm . El satélite japonés Akari lanzado en 2006 ha dejado de funcionar en frío el pasado 26 de agosto de 2007, después de casi quince meses dedicado a cartografiar casi todo el cielo en seis bandas a longitudes de onda entre 9 y 180 μm con un telescopio de 68 cm refrigerado a -267°C . Estas observaciones son de gran interés para las misiones Herschel y Planck, por lo que la ESA y otros institutos europeos participan en su análisis.

¹Una micra es una milésima de milímetro: $1 \mu\text{m} = 10^{-6} \text{ m}$

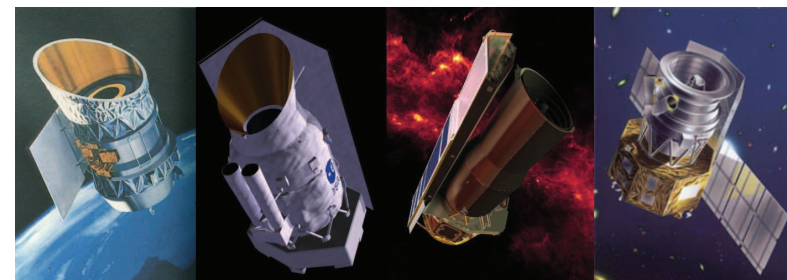


Figura 1: Imágenes artísticas de los observatorios IRAS, ISO, Spitzer y Akari (cortesía de ESA, NASA y JAXA).

Herschel está diseñado para observar el universo “frío”, es decir aquellos astros que hallándose a temperaturas menores de unos cientos de grados Kelvin² tienen su máxima emisión de ondas electromagnéticas a las longitudes de onda en que operan sus instrumentos. El énfasis de las investigaciones se pondrá en asuntos pendientes del mayor interés científico, como son, entre otros: (1) la búsqueda de las protogalaxias más antiguas, su evolución, en particular en la asociación de los intensos brotes de formación estelar y de la actividad en los núcleos de las galaxias del universo primitivo; (2) el nacimiento y evolución de las estrellas y de sus sistemas planetarios; (3) la física y química del medio interestelar, en particular mediante el estudio del agua, y (4) el estudio de los astros más fríos del sistema solar.

El observatorio espacial

La misión Herschel es liderada por Göran Pilbratt, de la ESA. Esta institución construye todo el vehículo excepto los tres instrumentos científicos, cuya integración en el observatorio y su verificación ha llevado también a cabo. Herschel mide aproximadamente 4 m de diámetro, lo máximo que permite el lanzador Ariane 5, y 7.5 m de longitud. En el momento del lanzamiento tendrá una masa de unos 3 170 kg, incluyendo los 2 160 litros (346 kg) de helio líquido superfluido destinado a la refrigeración de los instrumentos. Su progresiva pérdida será la que determine su tiempo de operación, que inicialmente se estima en 3,5 años. Herschel consta de tres partes, además de una gran pantalla, parcialmente recubierta de paneles solares, que protegerá del calentamiento solar y terrestre los elementos que deben funcionar a la menor temperatura posible, como son el telescopio y el criostato.

²La escala de grados Kelvin toma como origen la temperatura teórica más baja posible, conocida como “cero absoluto”, que corresponde a $-273,15^\circ\text{C}$, y conserva la misma dimensión para los grados que la escala habitual de grados centígrados o Celsius.

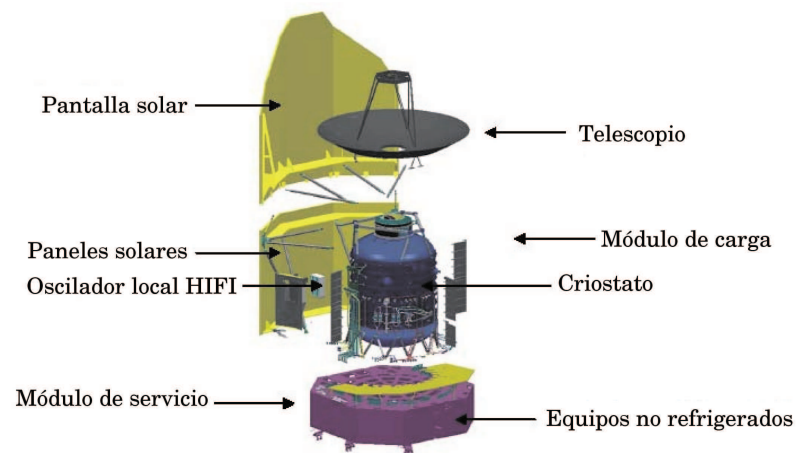


Figura 2: Elementos principales del Observatorio Espacial Herschel.

El telescopio

En un extremo se encuentra el telescopio, de tipo Cassegrain. Es decir, consta de un reflector principal de perfil parabólico que refleja la radiación incidente según su eje en un reflector secundario de perfil hiperbólico que, a su vez, refleja dicha radiación hacia el centro del reflector principal donde una abertura permite que la radiación se encamine hacia los tres instrumentos de medida. El reflector principal, o antena, tiene un diámetro de 3,5 m, el máximo que permite el lanzador para un telescopio hecho de una sola pieza, es decir que no sea desplegable. Es mayor que cualquiera de los telescopios ópticos o infrarrojos que han sido puestos en órbita hasta la fecha, incluido el Telescopio Espacial Hubble. El reflector secundario, o subreflector, es mucho menor, de 31 cm de diámetro, y está sujeto por tres pares de soportes de titanio. Dado que éstos y el subreflector se hallan enfrente del reflector principal, constituyen un obstáculo a la radiación que se intenta detectar, por lo que se ha tratado de reducir lo máximo posible su tamaño y, por lo tanto, sus efectos de bloqueo y de dispersión sobre tal radiación.

La calidad de los elementos ópticos (reflector y subreflector) debe ser tal que se comporten como un espejo para la radiación incidente. Ello significa que su rugosidad y su diferencia con respecto de la figura geométrica teórica debe ser mucho menor que la longitud de onda de trabajo. Las medidas del telescopio que se han realizado en laboratorio a la temperatura de funcionamiento han revelado que las desviaciones (en el sentido cuadrático medio, o *rms*) son de $5,5 \mu\text{m}$, lo que asegura un

funcionamiento óptimo hasta longitudes de onda de $82 \mu\text{m}$ y bueno hasta las $55 \mu\text{m}$.

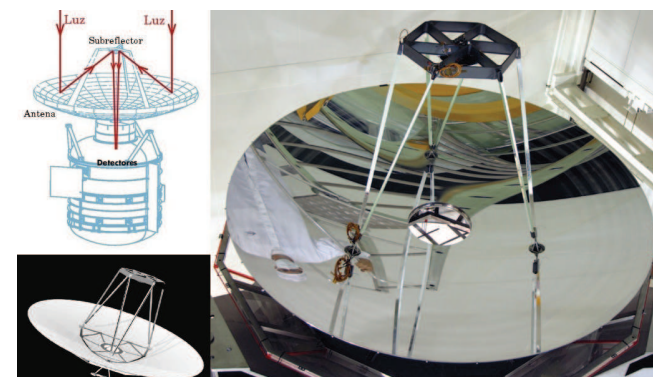


Figura 3: Antena de 3,5 m de diámetro del HSO: (1) Esquema de un sistema Cassegrain. (2) Esquema de la estructura de la antena. (3) Fotografía de la antena montada.

Otro de los fuertes requisitos impuestos al telescopio es que su temperatura debe ser lo más baja posible a fin de que su propia radiación contribuya poco a lo detectado por los instrumentos y apenas perturbe las medidas a realizar. Con tal propósito se ha construido con carburo de silicio (SiC), que además asegura su ligereza (315 kg), recubierto de aluminio reflectante y una fina capa protectora de óxido de silicio. El conjunto se ha protegido con la pantalla solar antes citada. Todo ello asegura una baja emisividad, muy bajas variaciones y bajos gradientes en su temperatura que reducirían su rendimiento, y favorece que su enfriamiento pasivo en el espacio permita alcanzar una temperatura de 80 K, o sea de -193°C . La necesidad de una temperatura lo más baja posible eliminó la posibilidad de que Herschel orbitara la Tierra como hace, por ejemplo, el Telescopio Espacial Hubble, pues en este caso la emisión térmica de la propia Tierra no permitiría enfriar la antena por debajo de 100 K ni aseguraría la estabilidad térmica necesaria.

Los dos módulos

Tras el telescopio se encuentra el Módulo de Carga cuyo elemento principal es el criostato de 2,5 m de altura y 2 m de diámetro que, enfriado hasta 1,7 K mediante el helio superfluido, contiene las correspondientes unidades del plano focal (FPU) de los tres instrumentos. Los elementos de estas FPUs, en especial los detectores, se enfrían a muy bajas temperaturas,

alcanzando en algún caso temperaturas inferiores de 0,3 K, o sea de unos -273°C . En la parte exterior del criostato se encuentran algunos equipos complementarios de los instrumentos y las sujeciones de la pantalla solar y del propio telescopio.



Figura 4: Criostato instalado en un banco de pruebas.

El resto del equipo, que corresponde a la alimentación eléctrica, la electrónica de control, los motores, la parte no refrigerada de los instrumentos científicos, el almacenamiento y procesamiento de datos, y las comunicaciones, se encuentra en el otro extremo del vehículo, constituyendo el Módulo de Servicio.

Los tres instrumentos científicos

La construcción de cada uno de los tres instrumentos científicos de los que consta Herschel ha sido llevada a cabo por un consorcio de institutos de investigación, observatorios y departamentos de universidad, cada uno de los cuales ha aportado aquella técnica en que es experto o incluso líder a nivel europeo. Para participar, cada institución ha tenido que demostrar frente al Investigador Principal del instrumento su capacidad

para diseñar el equipo correspondiente de manera que dé las mejores prestaciones, su capacidad para producirlo con características espaciales y la disponibilidad de los fondos necesarios para todo ello. Las agencias nacionales de subvención de proyectos espaciales han debido dar el apoyo necesario durante el largo periodo de tiempo que exige un desarrollo de este tipo, típicamente superior a una década. En el caso de España, han intervenido el Ministerio de Educación y Ciencia, a través del Plan Nacional del Espacio, y el CDTI (Centro para el Desarrollo Tecnológico Industrial).

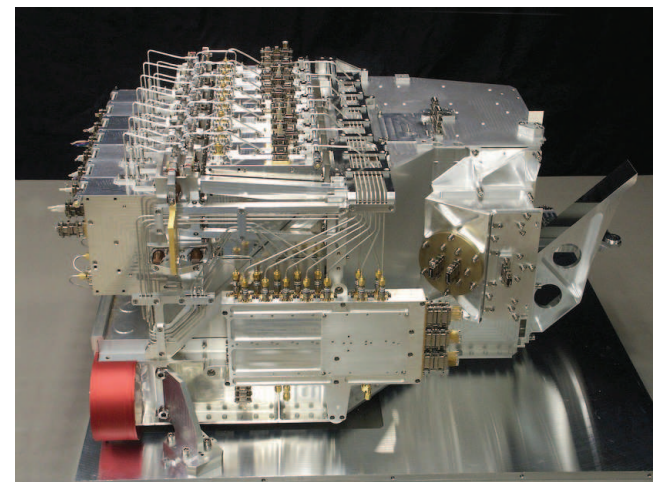


Figura 5: Unidad del plano focal (FPU) del instrumento heterodino HIFI. Las FPU son los elementos que necesitan ser refrigerados a temperaturas cercanas al cero absoluto ($-273,15^{\circ}\text{C}$) y, por ello, se instalan en un banco óptico en la parte superior del criostato.

El instrumento HIFI

Es el instrumento heterodino de Herschel (Heterodyne Instrument for the Far-Infrared), un sistema que permite analizar el espectro electromagnético a unas frecuencias mucho menores que las de observación y, gracias a ello, alcanzar una resolución espectral muy elevada, hasta $R = \lambda/\Delta\lambda \sim 10^7$, sólo alcanzable actualmente en Radioastronomía. En el caso extremo, permitirá alcanzar una resolución en velocidad tan baja como 0,02 km/s. Este fue uno de los requisitos del diseño, junto con la capacidad de realizar exploraciones espectrales de manera eficiente y con alta resolución en toda su banda, que debía alcanzar desde las transiciones del O_2 y del carbono a $\sim 610 \mu\text{m}$ hasta la transición a $158 \mu\text{m}$ del ión C^+ .

HIFI permitirá medir miles de líneas espectrales de átomos, moléculas e iones del medio interestelar y de envolturas de estrellas, en el rango de 157 a 625 μm , o sea de 480 a 1910 GHz³, un rango prácticamente inaccesible desde radiotelescopios situados en la superficie terrestre. Se realizarán así inventarios de moléculas en muy diversos entornos (regiones de formación estelar, nebulosas difusas, discos protoplanetarios, galaxias cercanas, etc), y permitirá analizar muchas transiciones de una de las moléculas más interesantes del universo, la del agua. Con ello se pretende avanzar en el conocimiento de la interacción cíclica entre las estrellas y la materia interestelar en las galaxias.

Para ello dispone de 14 receptores heterodinos, que cubren siete bandas en dos polarizaciones independientes (5 de ellos cubren de manera continua el rango de 480 a 1250 GHz y los otros dos de 1410 a 1910 GHz), lo cual redundará en una mejor sensibilidad y en redundancia ante fallos. Estos receptores se basan en tecnologías punteras, empleando mezcladores de señal tipo superconductor-aislador-superconductor (SIS) y tipo bolómetro de electrones calientes (HEB), seguidos de una etapa de amplificación (IF1) de muy bajo ruido y banda ancha, diseñada y construida en los laboratorios del Centro Astronómico de Yebes, del Observatorio Astronómico Nacional (OAN). En la construcción de HIFI han participado 20 institutos de 12 países liderados por Thijs de Graauw, del Instituto Holandés para la Investigación Espacial (SRON).



Figura 6: Instituciones participantes en el Consorcio de HIFI, típica colaboración para construir un instrumento científico espacial.

³ 1 GHz = 10⁹ Hz es una unidad de medida de la frecuencia apropiada para estas longitudes de onda. Frecuencia (ν) y longitud de onda (λ) se relacionan entre sí mediante la expresión: $\nu\lambda = 299\,792\text{ GHz } \mu\text{m}$.

El instrumento PACS

Se trata de una cámara fotométrica a la vez que espectrométrica que cubre el rango de menor longitud de onda de Herschel. Cubre con gran sensibilidad un rango esencial para el estudio de la historia de la formación estelar y la actividad nuclear en las galaxias, el estudio de envolturas circumestelares y su influencia en la química interestelar, y el nacimiento de las estrellas.

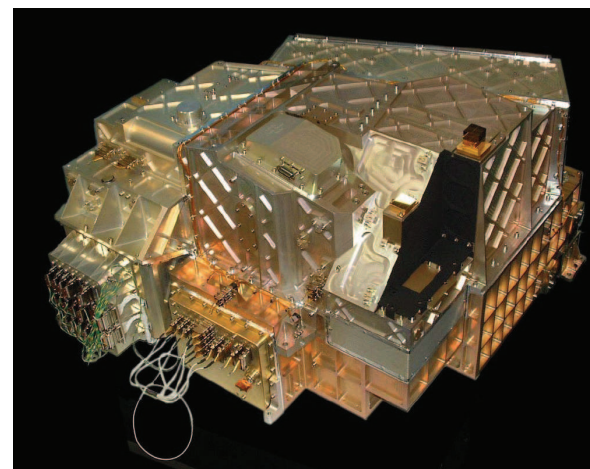


Figura 7: Unidad del plano focal (FPU) del instrumento PACS (cortesía de MPE).

El instrumento PACS (*Photodetector Array Camera and Spectrometer*) ha sido concebido, diseñado y construido por un consorcio de 14 institutos de 6 países liderados por Albrecht Poglitsch, del Instituto Max-Planck para Física Extraterrestre (Alemania). Es el instrumento que cubre las longitudes de onda más cortas, entre 55 y 210 μm . Consta de dos cámaras bolométricas de 32×16 y 64×32 píxeles que operan en tres bandas anchas (llamadas también “colores”), de 65 – 85, 85 – 130 y 130 – 210 μm , pudiendo esta última observarse simultáneamente con una de las otras dos. La cámara cubre en el cielo un campo de $1,75' \times 3,5'$.

PACS también puede usarse como cámara espectrométrica de resolución media, midiéndose 16 valores espectrales contiguos en cada una de las 5×5 posiciones del cielo que observa simultáneamente. En este modo se cubre un campo de $47'' \times 47''$ en el cielo con una resolución espectral que depende de la longitud de onda de observación. En la más baja se obtiene una resolución de $R = 5000$, mientras que a la más elevada se obtiene una resolución de 1000.

El instrumento SPIRE

Este instrumento cubre un rango del espectro electromagnético parecido a HIFI, pero hace incapié en la sensibilidad fotométrica, mientras que la resolución espectral que alcanza es baja o moderada. Es un elemento complementario tanto de HIFI como de PACS y que puede ser usado en paralelo con éste. Los objetivos principales de este instrumento son la investigación de la formación de galaxias en el universo remoto y el estudio de las primeras etapas del nacimiento de las estrellas, cuando la protoestrella está aún sumergida en el medio interestelar donde se formó.

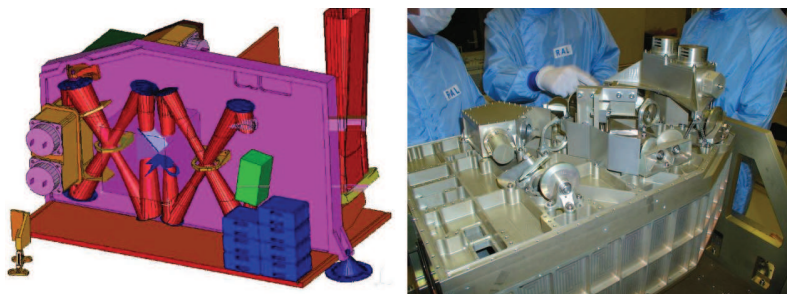


Figura 8: Esquema y montaje parcial de la unidad del plano focal (FPU) del instrumento SPIRE.

El instrumento SPIRE (*Spectral and Photometric Imaging Receiver*) ha sido concebido, diseñado y construido por un consorcio de 16 institutos en 8 países, liderados por Matt Griffin, de la Universidad de Cardiff (Reino Unido). Complementa a PACS a las longitudes de onda más largas, de 194 a 672 μm , disponiendo de tres cámaras fotométricas y de una cámara espectrométrica de baja resolución. Las tres cámaras fotométricas operan simultáneamente en las bandas de 250, 360 y 500 μm . Las tres cubren un mismo campo de $4' \times 8'$ en el cielo, aunque con píxeles de distinto tamaño. El espectrómetro en transformada de Fourier permite la observación simultánea de toda una banda desde 194 a 672 μm con baja resolución espectral (R entre 40 y 1000) cubriendo un campo en el cielo de $2' \times 2'$.

Participación del OAN/CNIG

Los astrónomos e ingenieros del OAN han estado involucrados en el proyecto Herschel desde su concepción inicial hace unos veinte años. Especialistas en las técnicas de Radioastronomía (tanto en el uso de radiotelescopios como en el desarrollo de equipos para ellos), la participación se

decantó de manera natural hacia el instrumento HIFI, que emplea técnicas y modos de observación típicos en Radioastronomía de alta frecuencia. Las investigaciones realizadas con HIFI supondrán una extensión a más altas frecuencias (submilimétricas e infrarrojo lejano) de las que actualmente se realizan en el OAN a frecuencias milimétricas utilizando radiotelescopios situados en alta montaña.

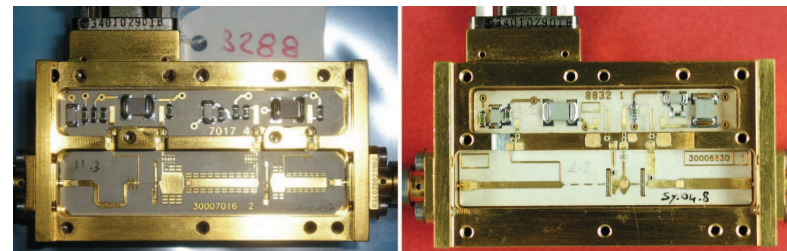


Figura 9: Amplificadores criogénicos de bajo ruido y banda ancha construidos por el OAN/CNIG para el instrumento HIFI. El de la izquierda corresponde al diseño realizado para los receptores 1 a 5 y el de la derecha para los receptores 6 y 7.

Desde el punto de vista técnico, esta colaboración suponía extender la experiencia del equipo de ingenieros en amplificadores criogénicos de bajo ruido al sector espacial. Su amplia experiencia lograda en la década de 1990 en el diseño y la construcción de amplificadores criogénicos de arseniuro de galio (GaAs) para la frecuencia intermedia de receptores que operan en ondas milimétricas y submilimétricas hizo que los amplificadores del OAN fueran considerados los mejores de Europa por sus prestaciones de ruido, ganancia, planitud de la banda, etc. Los retos de participar en HIFI consistían, primero, en demostrar su capacidad para diseñarlos y realizarlos con los novedosos dispositivos de fosforo de indio (InP) jamás usados en el espacio y raramente en tierra, superando las exigentes prestaciones requeridas por el Consorcio de HIFI, como son un bajo consumo de energía, menor a 5 mW, una banda ancha de 4 GHz y muy bajo ruido, menor a 5 K. Tras ello se formalizó el encargo, y después se transfirió este conocimiento de I+D a la industria espacial española. Ambos retos fueron superados satisfactoriamente. En total se han construido más de 50 amplificadores (prototipos, modelos de demostración, de cualificación, de vuelo y repuestos), la mayoría de ellos (40) cualificados espacialmente.

Desde el punto de vista astrofísico, la participación en HIFI supone para los astrónomos del OAN el acceso a una parte del espectro electromagnético prácticamente inexplorada, que debe permitir avanzar en el conocimiento de los diversos campos de investigación en que dichos astrónomos están involucrados desde hace años. Tales son el estudio de las regiones y

procesos de formación estelar, el de las nebulosas y envolturas alrededor de estrellas viejas, y el de las condiciones físicas y químicas del medio interestelar en otras galaxias, claves para comprender el nacimiento de estrellas y la actividad del núcleo en ellas.

Por razones operativas y de gestión, la contribución española a HIFI se realiza desde el Centro Nacional de Información Geográfica (CNIG), con fondos provistos inicialmente por el MCyT y actualmente por el MEC, dentro del Plan Nacional del Espacio, y con el soporte técnico y científico del OAN, ambas instituciones, OAN y CNIG, dependientes de la Dirección General del Instituto Geográfico Nacional. Pere Planesas es el investigador principal de esta contribución.

Otras participaciones españolas

España participa en la construcción de los tres instrumentos científicos de Herschel. La contribución en PACS y SPIRE es realizada por el Instituto de Astrofísica de Canarias bajo el liderazgo, respectivamente, de Jordi Cepa e Ismael Pérez Fournón. La mayor contribución se realiza en PACS, para el que se ha construido la unidad de procesamiento de la señal y compresión de datos, además de contribuciones a cada Centro de Control del Instrumento (ICC).

La contribución al ICC de HIFI, que se inició en el CNIG/OAN, actualmente se lleva a cabo en el Departamento de Astronomía Molecular e Infrarroja, del CSIC, liderada por J. Martín-Pintado. Por otra parte, en el CNIG se están desarrollando aplicaciones web para la difusión rápida de los resultados del programa clave liderado por el OAN, una de las exigencias de los consorcios de los instrumentos de Herschel. En las páginas web <http://www.oan.es/hifi/> se encontrará además información complementaria recopilada o redactada por Javier Graciá.

En la construcción de Herschel y sus instrumentos han participado más de una docena de empresas españolas del sector espacial o relacionadas con él, además de empresas de desarrollo de software científico, como son, entre otras: Alcatel Espacio, CRISA, EADS CASA, Fractal, GMV, INTA, MASA, RYMSA, Sener, Tecnológica y TTI Norte.

Órbita y operaciones

Herschel y Planck serán enviados al espacio simultáneamente con un lanzador Ariane 5 ECA (que permite cargas dobles) desde el Centro Espacial Guyanés situado a una latitud de 5°, cerca de Kourou, en el litoral de la Guayana Francesa, en Sudamérica. Ambos vehículos espaciales seguirán sendas trayectorias independientes que los situarán en órbita alrededor del segundo punto lagrangiano (L2) del sistema Tierra-Sol, un

punto de equilibrio situado en la prolongación de la línea recta que une el Sol y la Tierra. Este punto se encuentra a 1,5 millones de km de la Tierra, unas cuatro veces más lejos que la Luna. Herschel se encontrará, pues, en órbita alrededor del Sol pero, a pesar de encontrarse más lejos de él que la Tierra, su periodo será igual al de ésta pues, por su cercanía, ésta lo “arrastrará”. Visto desde la Tierra, Herschel parecerá describir una amplia trayectoria de Lissajous alrededor del punto L2, llegando a alejarse de él hasta 800.000 km y con un periodo de medio año. La distancia a la Tierra variará entre 1,2 y 1,8 millones de km. Al tratarse de una órbita inestable, se corregirá una vez al mes.

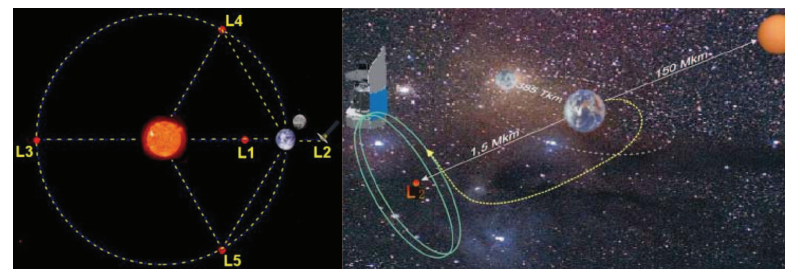


Figura 10: Puntos de Lagrange del sistema Sol-Tierra. Herschel orbitará alrededor del punto L2, tal como se indica en el esquema de la derecha.

Durante el recorrido hasta su órbita en L2, una vez estabilizado en sus tres ejes y tras la última corrección de trayectoria prevista para el día decimosegundo de viaje, se iniciarán el enfriamiento del telescopio y la verificación del funcionamiento de todos los instrumentos y equipos de Herschel, fase de *commissioning* que se prevé dure a lo sumo tres meses. La apertura de la cubierta del criostato, a un mes o dos tras el lanzamiento, supondrá la primera luz para el observatorio. Se realizarán medidas de algunas fuentes de radiación, predominantemente del sistema solar, para evaluar su rendimiento, eficiencia, apuntado, estabilidad y demás características esenciales para su buen funcionamiento, incluido el procesado y envío de datos. Al final de esta fase Herschel se encontrará en su órbita definitiva.

A continuación se dedicarán al menos dos meses a la fase de demostración científica, en que se optimizará su funcionamiento y se determinará la sensibilidad real de los instrumentos, conociéndose así las capacidades del nuevo observatorio espacial. Tras esta etapa se harán públicas las primeras imágenes y los primeros espectros, tomados hacia astros de especial interés, tanto para los científicos como para el público en general. Se realizarán reuniones de trabajo para informar a los científicos de las consecuencias de estos resultados y de la posible necesidad de variar sus estrategias

de observación para optimizar el uso del observatorio. Se planea dedicar un máximo de 6 meses tras el lanzamiento al conjunto de las tareas descritas, tras las cuales se iniciarán las operaciones científicas rutinarias que deberán durar al menos 3 años, hasta que se agote el helio necesario para su buen funcionamiento.

Durante las operaciones científicas, se dedicarán 21 horas diarias a las observaciones astronómicas, maniobras y diversas calibraciones, que se realizarán de modo autónomo. Tres horas diarias se dedicarán al contacto con el Centro de Control de la Misión (situado en el European Space Operations Control Centre en Darmstadt, Alemania) para la descarga de los datos tomados durante el día, vía la estación de Perth (Australia) o, en caso de emergencia, de la de Cebreros (España). Previamente, los datos (unos 8 Gbits por día) habrán sido comprimidos digitalmente para aprovechar al máximo el ritmo de transmisión de 130 kbps. Para la verificación de su calidad, los datos se remitirán al Centro Científico de Herschel (HSC) que se encuentra en el Centro Astronómico Espacial Europeo (ESAC), con sede en Villafranca del Castillo, cerca de Madrid (España). Su contrapartida norteamericana se encontrará en el centro de datos y científico para astronomía infrarroja IPAC en Pasadena (California).

Los procedimientos de calibración y de procesamiento estándar de los datos de cada instrumento se llevarán a cabo en los sendos Centro de Control del Instrumento (ICC) que se han organizado para tal propósito. Cada uno de ellos tiene su sede en el mismo instituto que ha liderado la construcción del instrumento en cuestión y cuenta con la colaboración de las demás instituciones de los respectivos consorcios. La reducción de datos final será llevada a cabo por los astrónomos que participan en cada una de las propuestas de observación, contando con el apoyo del ICC correspondiente al instrumento usado.

Objetivos científicos

Se considera que habrá unas 20.000 horas (mil días) de observación disponible para realizar las observaciones científicas. Una fracción de este tiempo será asignada al Centro Científico de Herschel, como compensación a los astrónomos que dedican su tiempo a las calibraciones, apuntado y planificación de observaciones, y hasta un 4 % adicional podrá ser usado de forma discrecional por el Investigador Principal (IP), asignándose a observaciones de gran interés o que no pueden esperar (un astro variable, un cometa especialmente brillante, una nueva medida sugerida por un interesante resultado inesperado, etc.).



Figura 11: Imagen en el infrarrojo lejano (a 90 y 140 μm) de la región de Cygnus X tomada por Akari en julio de 2007 (cortesía de JAXA).

Los programas clave

El tiempo de observación con Herschel se repartirá en varios bloques. Uno de ellos, equivalente al 32 % del total, corresponde a los Programas Clave de Tiempo Garantizado. Los programas clave son grandes proyectos de investigación, de más de 200 horas de observación, que hacen uso de capacidades únicas de Herschel para abordar temas científicos importantes de manera completa, en base a realizar observaciones de manera uniforme y coherente a fin de producir bases de datos de alto valor para la comunidad científica, tanto en la actualidad como en el futuro, cuando nuevos instrumentos puedan complementarlos. Aquellos programas clave que forman parte del Tiempo Garantizado corresponden a los que se han planteado en el seno de los consorcios que han construido cada uno de los instrumentos y que son el retorno al que tienen derecho por la inversión realizada (tanto en fondos como en horas de trabajo) por su participación. El día 4 de abril de 2007 se presentaron 21 Programas Clave de Tiempo Garantizado para ser evaluados por un comité de expertos (entre los que se encuentra R. Bachiller, director del OAN) establecido por el Centro Científico de Herschel (HSC), que los aprobaron total o parcialmente.

Otra fracción importante del tiempo de observación, que puede llegar a ser de otro tercio del total, corresponderá a los Programas Clave de Tiempo Abierto, es decir aquellos que, satisfaciendo los mismos requisitos que

los ya aprobados, son propuestos por equipos de científicos de todo el mundo, hayan o no participado en la construcción de los instrumentos. Se espera que se presenten decenas de propuestas que serán evaluadas siguiendo los procedimientos habituales de evaluación por un comité, asignándose tiempo de observación a las mejores. En la reunión científica sobre programas clave de tiempo abierto que tuvo lugar en la sede de la ESA en Noordwijk (Holanda) en febrero de 2007 se defendieron al menos 40 ideas para posibles propuestas. El día 28 de febrero de 2008 se hará pública la lista de propuestas de este tipo aprobadas.

El resto del tiempo de observación se dedicará principalmente a propuestas normales de tiempo abierto a toda la comunidad científica. Es decir, se tratará de propuestas de observaciones de duración corta o media que pretenden abordar estudios muy específicos. Habrá al menos dos convocatorias para presentar tales propuestas, la primera de las cuales tendrá lugar en 2008.

La participación del OAN

Varios científicos de esta institución han estado involucrados durante años en la definición de los programas clave más relacionados con su área de mayor experiencia. Asimismo, participan en los trabajos preparatorios necesarios, tales como la realización de observaciones complementarias con radiotelescopios terrestres o el modelado de la transferencia de la radiación que se va a observar en ciertos tipos de astros. Los programas clave de tiempo garantizado aprobados en los que astrónomos del OAN participan a nivel de co-Investigadores son:

- Observaciones de agua y monóxido de carbono en envolturas de estrellas AGB (rama gigante asintótica), prenebulosas planetarias y nebulosas planetarias (V. Bujarrabal (IP), J. Alcolea, P. Planesas).
- Agua en regiones de formación estelar (R. Bachiller, M. Tafalla, A. Fuente).
- El medio interestelar denso y caliente (A. Fuente, P. de Vicente).
- Condiciones físicas y químicas del medio interestelar en núcleos galácticos (S. García Burillo, P. Planesas).
- Exploraciones espectrales en regiones de formación estelar (A. Fuente, M. Tafalla).

Es decir, se participa en la mayoría de programas clave garantizados del consorcio de HIFI. Es de destacar que uno de los programas clave aprobados está liderado por un astrónomo del OAN, tanto más cuanto

la contribución española en HIFI en términos económicos no es grande. Posiblemente han influido tanto la probada experiencia de V. Bujarrabal y colaboradores en el OAN en este área de investigación, como el esfuerzo realizado por el equipo de ingenieros del Centro Astronómico de Yebes (J.D. Gallego, I. López Fernández, M. C. Díez, A. Barcia) en su continua disponibilidad y asistencia desinteresada y entusiasta a otros grupos del consorcio en el diseño y la integración de sus propios equipos en HIFI, como también la participación activa del CNIG/OAN en la resolución de crisis que han supuesto desembolsos de fondos y esfuerzos adicionales.

Está prevista la participación de otros astrónomos de plantilla y becarios pre o postdoctorales del OAN en estos proyectos científicos, como investigadores asociados, así como su participación en las propuestas de tiempo abierto que actualmente (agosto de 2007) se están preparando.



Figura 12: Despegue de un Ariane 5 (cortesía de ESA).

El lanzamiento

A fecha de hoy la ESA mantiene como fecha más tardía para el lanzamiento de Herschel el 31 de julio de 2008. Sin embargo, los retrasos habidos en la entrega y verificación de los instrumentos de Herschel, así como los habidos con Planck, pueden obligar a retrasar este lanzamiento previsiblemente en medio año, lo que obligaría a gastos adicionales que la ESA quiere evitar. Por ello, se están realizando grandes esfuerzos para mantener como fecha de lanzamiento la prevista. El plan científico, los anuncios de oportunidades para el uso de Herschel y el desarrollo de

software para el análisis de los datos de cada instrumento, se mantienen usando como referencia la fecha oficial de lanzamiento.

Se prevé que las actividades asociadas a Herschel terminen en cuanto las últimas observaciones calibradas y depuradas hayan sido almacenadas en el archivo final, lo que previsiblemente ocurrirá al menos seis años tras su lanzamiento. Dicho archivo será accesible, por ejemplo vía Observatorio Astronómico Virtual, y se espera que sea una referencia útil para los astrónomos durante décadas, como lo ha sido el archivo de su predecesor IRAS.

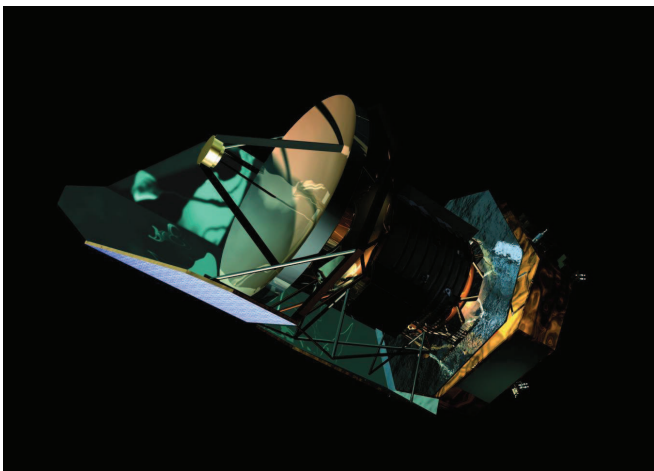


Figura 13: Imagen artística del Observatorio Espacial Herschel, HSO (cortesía de ESA).

Referencias

<http://sci.esa.int/herschel>
<http://herchel.jpl.nasa.gov>
<http://herchel.esac.esa.int>
<http://www.ipac.caltech.edu/project/herchel/>
<http://smc.cnes.fr/HERSCHEL/>
<http://www.oan.es/hifi/>

PUESTA AL DÍA DEL TEMA

Desarrollo de la misión y resultados de Herschel

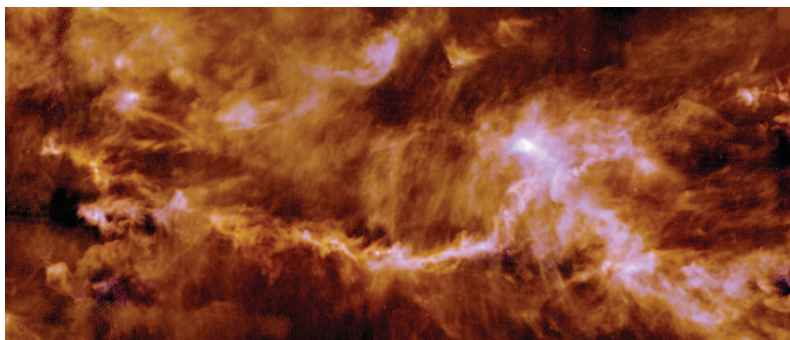
El Observatorio Espacial Herschel fue lanzado el 14 de mayo de 2009, en una órbita altamente elíptica ($e = 0,999549$) con un apogeo de casi 1.200.000 km destinado a llevarle, tras varias maniobras, hasta una amplia órbita de Lissajous alrededor del segundo punto lagrangiano (L2) del sistema Tierra-Sol, que alcanzó dos meses después. Durante las cinco semanas siguientes se realizaron la puesta en marcha y las pruebas de funcionamiento de los tres instrumentos, completadas el 21 de julio. La fase de demostración científica fue más lenta de lo previsto por la desconexión automática de HIFI en agosto, instrumento que no pudo ser reactivado con éxito hasta enero del año siguiente. A pesar de ello, las observaciones científicas se iniciaron en septiembre, con los instrumentos SPIRE y PACS funcionando independientemente o en paralelo.

La misión científica terminó tres años y medio más tarde, en abril de 2013, al agotarse el helio líquido usado para refrigerar los instrumentos. Durante varias semanas se realizaron pruebas técnicas y de software que pudieran ser útiles en misiones futuras. A mediados de mayo se transfirió a una órbita heliocéntrica segura, más alejada que la de la Tierra ($a = 1,0422$ ua), consumiendo casi todo el combustible restante, y un mes más tarde, por seguridad, se consumió el resto de combustible y se desconectaron todos sus sistemas.

Durante los 43 meses de actividades científicas, se realizaron 23.400 horas de observaciones astronómicas y 2.600 horas de observaciones para calibración de las medidas. Todos los datos son accesibles por medio del Archivo Científico de Herschel (HSA), al que se puede acceder a través de los enlaces: <http://archives.esac.esa.int/hsa/whsa/> y <http://irsa.ipac.caltech.edu/applications/Herschel/>, además de resultados altamente elaborados de los Proyectos Clave en: <https://www.cosmos.esa.int/web/herchel/user-provided-data-products>. Entre ellos se encuentran los del Proyecto HIFISTARS, que fue liderado por astrónomos del OAN. Hasta finales de 2017 se siguió trabajando en mejorar la calidad de la calibración y de los datos obtenidos con los tres instrumentos, en la fase de posoperaciones.

En los meses de julio y octubre de 2010 se publicaron dos números monográficos de la revista *Astronomy and Astrophysics*, con más de 200 artículos con los resultados de las observaciones iniciales. En 2017 se han superado los 2000 artículos científicos basados en datos de Herschel, que

abarcan casi todos los campos de la astrofísica. Las capacidades de PACS y SPIRE para obtener imágenes en colores infrarrojos se pueden apreciar en: http://www.esa.int/Our_Activities/Space_Science/Herschel/Highlights/Herschel_images y en: http://www.esa.int/Our_Activities/Space_Science/Highlights/Herschel, donde hay una selección de las imágenes que se han tomado de extensas regiones de formación estelar, de restos de supernova y de galaxias, además de estrellas y objetos del sistema solar. Por ejemplo, Herschel ha permitido identificar en la Vía Láctea numerosos filamentos de gas y polvo de gran extensión, como el filamento B211/B213 en la Nube Molecular de Tauro que se muestra en la figura (cortesía de ESA/Herschel/PACS y SPIRE), llegando alguno a exceder los 100 años luz de longitud.



Las capacidades espectroscópicas de Herschel han permitido, por ejemplo, confirmar la detección de la molécula de oxígeno en el espacio, detectar agua en muchos entornos (discos protoplanetarios, cometas, etc.) y por primera vez en el cinturón de asteroides (Ceres) y en una condensación preestelar (L1544). Se han estudiado grandes flujos de gas molecular expulsado del centro de galaxias por la acción de agujeros negros supermasivos y la compleja química que hay en el centro de la Vía Láctea, en las envolturas de estrellas evolucionadas y en otros objetos galácticos.

Los datos obtenidos por Herschel serán usados durante muchos años, pues se tardará más de una década en disponer de un observatorio espacial de características similares (cubriendo el infrarrojo lejano, FIR) que lo supere en sensibilidad y/o en resolución espacial. Varios proyectos han sido propuestos para superar las limitaciones de Herschel: SPICA (de JAXA y ESA) debería aportar más sensibilidad en el rango de longitudes de onda de 3,5 a 210 μm , mientras que las misiones interferométricas SPIRIT (de NASA, 25-400 μm) y FIRI (de ESA, 25-300 μm) estarían encaminadas a conseguir una resolución angular menor que 1".

Pere Planesas Bigas, enero 2017

La última y única vez que he colaborado en este Anuario escribiendo un artículo de divulgación fue hace catorce años, en el volumen correspondiente a 1994, el primero de esta nueva época. En aquel artículo titulado *Radiotelescopios para el fin del decenio*, repasaba los nuevos instrumentos radioastronómicos que se esperaba entrasen en funcionamiento por estas fechas. El futuro que predije hace más de diez años no dista mucho de la realidad. Algunos instrumentos no han prosperado, como el MMA que ha sido reconvertido en el proyecto ALMA, y otros que no estaban previstos han entrado en funcionamiento, como la antena de 40 m de diámetro del Centro Astronómico de Yebes. Una de las secciones estaba dedicada al Sub-Milimeter Array (SMA). La titulé *El SMA: captando pequeñas ondas en la gran isla*, tomando prestado el lema del CSO, otro telescopio submilimétrico situado en el cono del volcán Mauna Kea (isla de Hawaii, EE.UU.), *Catching small waves from the big island*. Hoy retomo el hilo de la narración donde lo dejé, intentando plasmar en las siguientes páginas qué nos deparará el próximo futuro en lo que a radiotelescopios se refiere. Esta vez no voy a abarcar instrumentos de todo tipo, sino que me ceñiré a aquellos que se dedicarán a trabajar a gran longitud de onda, por lo que haciendo un guiño a los amantes del surf, titularé este artículo

A LA CAZA DE LAS GRANDES ONDAS

Javier Alcolea

*Observatorio Astronómico Nacional
Instituto Geográfico Nacional - Ministerio de Fomento*

Abstract

In this paper, I will briefly review the main characteristics and research areas of the two major new initiatives in the field of low frequency radio-interferometry: the projects LOFAR and SKA. I will also comment on the possibility of using the Moon as a platform for the installation of an ultra-low frequency array.

Introducción

La Astronomía es una ciencia bastante peculiar. Por definición trata de todo lo que está lejos y, antes de la era espacial, inalcanzable. Para los astrónomos el laboratorio es el propio cosmos, pero éste diseña los experimentos por su cuenta y a nosotros sólo nos resta observarlos. Incluso hoy en día, el acceso que tenemos al universo es muy limitado. El hombre sólo ha llegado a poner el pie en el cuerpo astronómico más cercano, La Luna. Misiones no tripuladas han llegado a posarse en cuerpos más lejanos de nuestro sistema planetario, como Titán en el caso de la sonda Huygens. Otras han llegado a medir la gravedad y los campos electromagnéticos en los confines del Sistema Solar. Podemos medir aquí en La Tierra los efectos de la gravedad de La Luna y el Sol, y de un modo más indirecto los de los otros planetas. También llegan hasta nosotros los neutrinos provenientes del Sol y de las explosiones de supernovas cercanas, así como las partículas de alta energía (rayos cósmicos) que atraviesan La Galaxia. Y poco más, si no fuera por las ondas electromagnéticas (OEM), la luz. La luz, visible o no, es el verdadero mensajero del cosmos. Nos permite localizar objetos, estudiar sus movimientos, entender su composición y los procesos que en ellos ocurren. A veces incluso nos permite descubrir que no entendemos nada, por lo que necesitamos invocar la presencia de materia y energía oscuras. La luz nos llega desde lugares tan lejanos que nos permite ver como era el universo cuando solo tenía un 3 % de su edad actual (la era de la reionización), e inferir de lo que hoy vemos qué ocurrió en las épocas en las que todo era oscuridad (la era de la recombinación).

La naturaleza de la luz, o de las OEM (usaré los dos términos indistintamente), es muy peculiar. Se comporta como las ondas que se forman en la superficie de un estanque, interfiriendo entre sí. Al mismo tiempo, como las piedras que tiramos al estanque, la luz está cuantizada: podemos tirar una piedra o dos, pero no una y media. Las piedras electromagnéticas, los cuantos de luz, son los fotones. Como cualquier fenómeno ondulatorio, la luz nos llega con distintos colores igual que hay sonidos graves y agudos. Color y tono no son más que manifestaciones de una cualidad de las ondas, su frecuencia ν , que no es más que el número de oscilaciones que se producen en un segundo de tiempo. Por supuesto, al igual que hay sonidos tan graves (infrasonidos) y tan agudos (ultrasonidos) que no podemos oír, no debemos pensar que los colores de la luz sólo abarcan del rojo al violeta. El espectro electromagnético, que es como solemos denominar al rango completo de tonos de la luz, abarca desde las ondas radio a muy baja frecuencia, hasta los rayos γ en el extremo opuesto, pasando por las microondas, el infrarrojo lejano, medio y cercano, la luz visible, el ultravioleta y los rayos X.

Las piedras que tiramos al estanque serán una, dos o tres, pero la energía que depositen en la superficie del agua dependerá de su tamaño y la

fuerza con que las lancemos. Con la luz ocurre lo mismo, pero lo curioso es que la energía de los fotones tiene que ver con su frecuencia. La energía de un fotón no es más que $h\nu$ donde h es la constante de Planck, $6,6260755 \cdot 10^{-34} \text{ J s}^{-1}$. Como la frecuencia y la longitud de onda λ (la separación entre dos máximos consecutivos de las fuerzas electromagnéticas que se propagan) están relacionadas entre sí mediante la expresión $\nu\lambda = c$, siendo c la velocidad de la luz en el vacío, $299\,792\,458 \text{ m s}^{-1}$, la energía de un fotón también puede expresarse como hc/λ . De estas sencillas ecuaciones se desprende que los fotones de mayor energía son los de mayor frecuencia, esto es, menor longitud de onda, mientras que los menos energéticos serán los de mayor longitud de onda, y por lo tanto menor frecuencia.

Estas relaciones son muy importantes porque los fotones muy energéticos sólo pueden producirse en entornos también muy energéticos. Por ejemplo, los fotones de rayos γ se originan en reacciones nucleares a temperaturas de decenas de millones de grados o por el efecto Compton inverso¹. Por el contrario, los fotones de las ondas radio pueden llegar a ser emitidos por cuerpos con temperaturas de sólo algunas décimas de grado por encima del cero absoluto ($-273,15^\circ\text{C}$), pero también en fenómenos muy violentos. Los entornos energéticos pueden emitir fotones duros (alta frecuencia) o blandos (baja frecuencia). Por el contrario, los entornos con poca energía sólo son capaces de emitir señales de muy baja frecuencia. Además, no debemos olvidar que la luz que nos llega de objetos muy distantes sufre un corrimiento al rojo (aumento de la longitud de onda, expresado por $z = \Delta\lambda/\lambda$), debido a la expansión del universo. Así, para observar fenómenos que ocurrieron cuando el cosmos era joven, tenemos que usar ondas mayores que las que usamos para estudiar esos mismos procesos en el universo local.

En lo que sigue me ceñiré a tres instrumentos que, en un futuro más o menos lejano, van a ser un referente en Radioastronomía para ondas de frecuencia entre 10 MHz y 25 GHz (25 mil megaciclos), o longitudes de onda $30\text{ m} < \lambda < 13\text{ cm}$. Éste es el rango de ondas en el que nació la Radioastronomía allá por 1932, y aunque con el paso del tiempo ha habido siempre un ímpetu por subir en frecuencia, la cuna de esta ciencia sigue siendo un campo que depara y deparará muchas sorpresas.

LOFAR (www.lofar.org)

El nombre de este telescopio situado al norte de los Países Bajos, es un acrónimo de la expresión en inglés **LOW** Frequency **AR**ray, **I**nterferómetro

¹ Cuando un electrón relativista (que se mueve a velocidades cercanas a la de la luz) choca con un fotón y le cede energía, no puede empujarle (pues el fotón siempre se mueve a la misma velocidad) sino que le aumenta la frecuencia.

de Baja Frecuencia. Se trata de un conjunto de radiotelescopios funcionando en modo interferométrico que va a poder trabajar en las bandas HF y VHF, desde los 30 MHz ($\lambda = 10\text{m}$) hasta los 240 MHz ($\lambda = 1,25\text{m}$).

En óptica, pero no sólo en el rango visible sino desde las ondas radio a los rayos γ , el detalle de las imágenes de un telescopio tiene una limitación impuesta por la naturaleza ondulatoria de la luz. Así, la resolución espacial de un instrumento (esto es, cuán detalladas son las imágenes que proporciona) nunca puede ser mejor que $1,22\lambda/D$, donde λ es la longitud de onda de trabajo, D es el tamaño de la apertura del telescopio y la resolución espacial angular está dada en radianes. Si expresamos la resolución en segundos de arco, la ecuación que debemos usar es $250000\lambda/D$. La apertura es la zona del telescopio que capta la luz. En una cámara fotográfica el tamaño de la apertura está limitado por el diafragma, que podemos abrir más o menos. En un telescopio óptico la apertura coincide con el espejo primario, y lo mismo ocurre con los radiotelescopios con forma de paraboloide. Tomemos por ejemplo la frecuencia central de LOFAR, 120 MHz. Si queremos tener una resolución de $1''$ necesitaremos una apertura de 600 km. Desde luego es impensable construir semejante telescopio, a no ser que se trate de un interferómetro. Un interferómetro es una serie de telescopios que trabajan conjuntamente combinando las señales que reciben. En este caso, el tamaño de la apertura será la distancia entre los elementos del instrumento más apartados entre sí, pues al fin y al cabo estamos captando luz que llega a una zona con ese tamaño. Para obtener la resolución que hemos tomado por ejemplo, bastará con tener elementos separados unos 600 km. En el caso de LOFAR la separación máxima será de 100 km, con lo que el detalle alcanzable por el instrumento será $750''/\nu(\text{MHz})$, es decir, $25''$ a 30 MHz y $3''$ a 240 MHz.

Más o menos, todo el mundo se hace una idea de cómo se selecciona la región del cielo que se quiere observar con un telescopio óptico: simplemente se apunta hacia allí. Algo similar ocurre con los radiotelescopios simples, pero apuntar un radiointerferómetro requiere algo más que apuntar al mismo sitio con todos sus elementos. Volvamos a nuestro estanque de aguas tranquilas y tiremos una piedra relativamente grande. Las ondas que se generen serán de gran amplitud (la superficie subirá y bajará mucho). Tomemos ahora un puñado de piedras pequeñas que en conjunto pesen más o menos como la piedra grande de antes. Al tirarlas al estanque observaremos que no se forma una onda grande sino muchas pequeñas, que se estorban unas a otras; es el fenómeno de la interferencia. Si queremos que las piedras pequeñas produzcan el mismo efecto que la grande, tendrán que entrar en el agua *en fase*, de forma que todas las oscilaciones se sumen. Para ello no hace falta que todas las piedrecitas caigan muy juntas ni al mismo tiempo, las podríamos tirar en cualquier lugar del estanque, con tal de que las ondas de cada una de ellas se sumasen, interfiriesen constructivamente, a las de las demás. En la práctica esto es imposible, pero en teoría basta

con que la entrada en el agua de cada piedra ocurra con un retraso igual al tiempo que tardan las ondas en llegar a ese punto desde el centro; tenemos que hacer funcionar cada elemento con el desfase adecuado.



Figura 1: Elementos del interferómetro LOFAR. A la izquierda los dipolos de baja frecuencia (30–90 MHz), a la derecha los de alta frecuencia (110–240 MHz). Fotografías cortesía de www.lofar.org.

Igual ocurre en un interferómetro. Para que las señales que provienen de un punto se sumen constructivamente, hay que compensar las diferencias que existan entre los tiempos de llegada de un mismo rayo de luz a los distintos elementos del instrumento. Esto depende de la orientación del interferómetro con respecto a la zona del cielo que queremos visualizar. Haciendo uso de esta peculiaridad de los radiointerferómetros, LOFAR consigue apuntar a cualquier región del cielo sin mover nada, simplemente introduciendo los retrasos correspondientes. De hecho LOFAR no tiene partes móviles. Este tipo de telescopios se denomina *interferómetro por software*. La idea es muy simple. Como elementos individuales se utilizan dipolos o conjuntos de dipolos. Aislados, estos dipolos tienen la propiedad de detectar señales provenientes de todo el cielo. En cambio, el conjunto de dipolos sólo detecta la señal proveniente de la zona del cielo seleccionada mediante el conjunto adecuado de retrasos. En la práctica, las señales detectadas por cada uno de los elementos se envían por fibras ópticas a un centro de cálculo, donde se graban en discos duros. En este sitio, un gran computador correla (multiplica las señales de cada par de antenas) introduciendo un determinado grupo de retrasos, obteniendo la transformada de Fourier de la imagen de la zona del cielo seleccionada. Posteriormente se deshace esa transformación de Fourier y se limpia el mapa². De hecho usando los mismos datos y un grupo de retrasos diferentes, puede hacerse el mapa de otra región del cielo. En teoría se podría obtener una imagen de todo el cielo completo si se dispusiese de suficiente capacidad de cálculo.

²Para entender el proceso de generación de imágenes en un interferómetro se puede consultar el artículo publicado por F. Colomer en el Anuario de 1998.

Como vemos se trata de un proyecto en el que las tecnologías de la comunicación y de la computación son parte integrante del telescopio en sí y no meras herramientas auxiliares. Esto tiene la gran ventaja de que conforme pasa el tiempo los costes disminuyen y las capacidades aumentan; es la conocida ley de Moore, que se aplica no sólo a los procesadores, sino también a las líneas de comunicación y sistemas de almacenamiento de datos.

Antes de describir las líneas de investigación en las que se espera que LOFAR haga una importante contribución, describiré brevemente los parámetros más significativos del proyecto. El instrumento está compuesto por una estación central y 45 más repartidas en un área de 100 km de diámetro. Cada estación tiene 100 antenas de baja frecuencia (30–90 MHz) y otras tantas de alta frecuencia (110–240 MHz), excepto la estación central que está mucho más poblada (3200 elementos). Las estaciones externas se disponen a lo largo de cinco brazos espirales en nueve anillos, cuya distancia al centro aumenta en progresión geométrica de 1,2 km hasta los 50 km. Cada elemento producirá datos a razón de 10 Gbps (10 000 Mbps) por sensor, por lo que el centro de proceso de datos debe soportar un flujo de 900 Gbps y deberá tener una potencia de cálculo de 160 billones de operaciones por segundo. Además, esta infraestructura se va a aprovechar para instalar sismómetros y detectores de infrasonidos (para escuchar el subsuelo y analizar los movimientos que en él se producen), detectores útiles para llevar a cabo agricultura de precisión (control de plagas y condiciones ambientales), y sensores de otro tipo. Como vemos se va a utilizar LOFAR como una red neuronal de captación de datos astronómicos, geofísicos o de otra índole.

LOFAR avanza poco a poco. La estación central ya está lista y realiza observaciones de prueba con regularidad. El despliegue de la totalidad de las estaciones se espera que finalice en 2008. Una vez concluido, el potencial de LOFAR para hacer observaciones que cubran un gran campo visual será enorme. Por ejemplo, suponiendo que se observen cinco zonas del cielo al tiempo, a 30 MHz se espera que en el plazo de un año de observaciones se pueda cubrir un área de 3000 grados cuadrados (toda la bóveda celeste ocupa 41 250 grados cuadrados) con la mayor sensibilidad posible (hasta alcanzar el límite de confusión,— ver más adelante —), detectando en este tiempo más de 4 millones de objetos astronómicos.

De acuerdo con nuestro presente entendimiento de la evolución del universo, medio millón de años después del Big Bang la temperatura típica descendió por debajo de los 3000 grados, y el hidrógeno se recombinó con los electrones para quedar en estado neutro. Durante esta *Era Oscura* el universo siguió enfriándose hasta que las primeras estrellas, protogalaxias y cuásares volvieron a ionizar el gas neutro con su intensa radiación ultravioleta. Por las observaciones de galaxias muy remotas (en el tiempo y el espacio) y los resultados de los satélites COBE y WMAP, se cree

que la Época de la Reionización (EdR), cuando surgieron esas primeras fuentes de luz, tuvo que suceder para valores de corrimiento al rojo, z , entre 15 y 6. La mejor manera de observar el hidrógeno neutro atómico que llenaba el universo durante la *Era Oscura* y hasta el final de la EdR es mediante su transición de 21 cm de longitud de onda. Pero para z entre 6 y 15 estos 21 cm se transforman en 1,5 y 3 m respectivamente, es decir, entre 100 y 200 MHz. Mediante LOFAR se podrá observar cómo ese hidrógeno atómico neutro se fue ionizando con el paso del tiempo, y de aquí entender mejor cuáles fueron esas fuentes de ionización y el porqué de la estructura actual del cosmos.

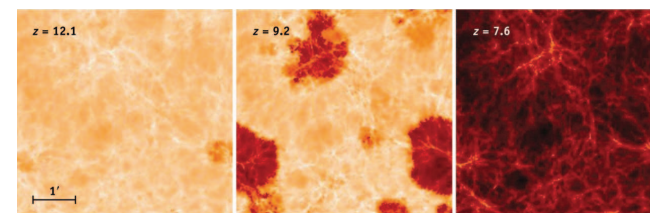


Figura 2: Simulación de la estructura del universo durante la época de la reionización. A la izquierda el universo es fundamentalmente neutro (predominan los tonos claros), mientras que a la derecha la componente más importante es el gas ionizado (tonos oscuros). Imagen cortesía de S. Furlanetto et al. y www.skatelescope.org.

Acabada la EdR, comienza la denominada Era de los Cuásares, $7 > z > 1,5$. Es en esta época en la que se establece la jerarquía de estructuras que vemos en la actualidad, estrellas, galaxias, cúmulos de galaxias y supracúmulos de galaxias. Pero, por ejemplo, se desconoce si las galaxias se formaron por la presencia de un agujero negro supermasivo o viceversa. Durante ese tiempo las galaxias tenían una tasa de formación estelar mucho más alta que en la actualidad (de 4 a 10 veces mayor), y las que presentaban brotes de intensa formación estelar (galaxias *starburst*) o núcleos activos (cuásares) eran mucho más abundantes que en la actualidad (en el universo local a $z < 1,5$). Los núcleos activos de galaxias (AGNs) se deben a la presencia en el centro de las galaxias de un agujero negro supermasivo (centenares de millones de masas solares). Este agujero negro puede ser muy voraz, alimentándose del gas y estrellas circundantes, pero es incapaz de tragar tanta energía gravitacional como se libera con la caída de material. El resultado es la eyección de chorros de partículas a velocidades próximas a las de la luz, en dos direcciones opuestas. Los electrones acaban frenándose por choques con el material galáctico o intergaláctico, lo que hace que emitan una intensa radiación a bajas frecuencias, denominada radiación sincrotrón. Cuanto más tiempo pasa más baja es la frecuencia de

emisión, por lo que las estructuras más antiguas de estos chorros (y las más grandes), sólo se detectan a muy baja frecuencia. Se piensa que LOFAR será capaz de detectar más de 1000 de estas radiogalaxias a distancias mayores que la más distante detectada hasta la fecha. Además, LOFAR será capaz de observar galaxias *starburst* a distancias sin precedentes, gracias a la intensa radiación sincrotrón emitida por estos objetos y que se debe a los numerosos restos de supernova producidos por la ingente muerte de estrellas de muy alta masa (mueren a un ritmo muy alto porque también nacen a un ritmo muy elevado). En particular LOFAR podrá detectar objetos similares a la galaxia *starburst* Arp 220 (situada a $z = 0,018$) a distancias de $z = 3,3$. LOFAR también será capaz de detectar la emisión difusa que rodea los cúmulos de galaxias, permitiendo estudiar la influencia de estos halos en la formación y evolución de dichos cúmulos.

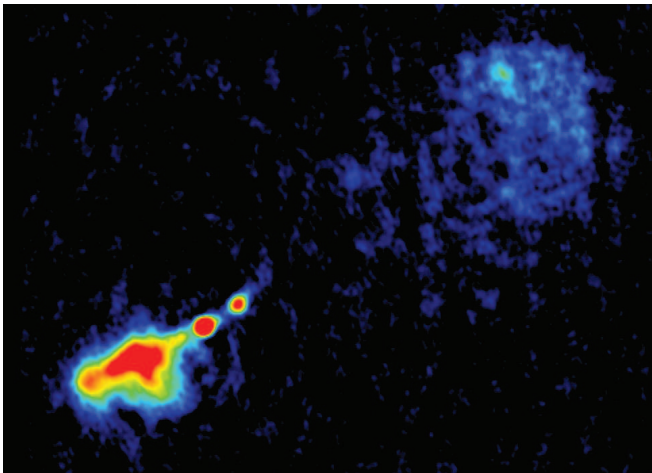


Figura 3: Imagen de la radiogalaxia 3C 236, probablemente el objeto individual más grande del universo (4 Mpc de tamaño). Se trata de una galaxia con núcleo activo, que se situaría a mitad de camino de los puntos calientes cerca del centro del cuadrante inferior izquierdo de la figura. La nebulosidad compacta e intensa situada más abajo y a la izquierda, y la más grande y difusa arriba a la derecha, trazan la presencia de electrones ultrarelativistas emitidos por el agujero negro supermasivo. La diferencia de intensidad entre ambos lóbulos se debe al carácter fuertemente directivo de la radiación sincrotrón, siendo el lóbulo inferior-izquierdo el que apunta hacia nosotros. Imagen cortesía de R. Schilizzi y www.evlbi.org.

Se espera que LOFAR también contribuya a esclarecer el problema del origen de los rayos cósmicos de muy alta energía. Los posibles candidatos

que se barajan son los choques en los chorros de partículas de radiogalaxias particularmente intensas o las fuentes de rayos γ . La observación en ondas radio de baja frecuencia de la radiación sincrotrón de estos dos tipos de objetos será sin duda crucial en este tema. Por otra parte, LOFAR también será capaz de medir la entrada de estos rayos cósmicos en la atmósfera. Trabajando a modo de radar, podrá detectar los ecos que se generarán en las zonas ionizadas de la atmósfera que surgen debido a los impactos de esas partículas tan energéticas. Usando únicamente su nodo central, LOFAR será capaz de observar el cielo visible desde el norte de Europa (2/3 del total del cielo) diariamente, lo que permitirá la detección casi instantánea de objetos que muestren una inusual actividad. Estos objetos podrán ser estudiados instantáneamente con más detalle usando el instrumento en su totalidad.

A distancias más próximas, en el universo local, La Galaxia y el propio Sistema Solar, son muchos los proyectos que se podrán abordar. Se espera que se pueda completar el censo de supernovas en nuestra Galaxia, y compararlo con el que se realizará en otros objetos con más alta tasa de formación estelar como la galaxia *starburst* M 82. Se podrán estudiar las propiedades del medio interestelar, separando la emisión térmica (gas caliente) y sincrotrón, y se descubrirán más de 1500 nuevos púlsares. Finalmente se espera poder comprender la emisión de muy baja frecuencia que se origina en la magnetosfera de Júpiter, e incluso tratar de usar este tipo de emisión como otro método para detectar la presencia de planetas gigantes alrededor de estrellas cercanas similares a nuestro Sol.

Aunque el detalle que alcanzarán las imágenes tomadas por LOFAR a baja frecuencia será mucho mejor que lo que hoy en día se puede hacer, una resolución de $25''$ a $10''$ plantea un problema fundamental; enseguida se llega al límite de confusión. Cuando se realizan mapas de alta sensibilidad pero poca resolución espacial ocurre que llega un punto en el que se observa un ruido de fondo en las imágenes, que no desaparece por mucho que se aumente el tiempo de exposición. Este ruido de fondo no es ruido sino la emisión de multitud de radiofuentes que, al estar más próximas entre sí que la resolución del telescopio, no llegan a distinguirse. Cuando se alcanza este punto, la sensibilidad del instrumento no puede mejorarse aumentando su superficie colectora o disminuyendo el ruido de sus receptores. Lo único que se puede hacer es mejorar su resolución espacial, lo que en el caso de un interferómetro quiere decir añadir más estaciones y situarlas más lejos del centro del instrumento. Es por esto que aún no habiéndose completado la instalación del LOFAR original, ya se está trabajando para aumentar un factor 10 la separación máxima entre sus estaciones. Se pasaría de 100 a 1000 km, y por lo tanto se aumentaría la resolución de las imágenes también en un factor 10, es decir, el problema de la confusión se reduciría en un factor 100 en número de fuentes.

Para conseguir esto las autoridades de los Países Bajos ya han iniciado

conversaciones, y hasta conseguido acuerdos, con los países de su entorno. Francia, Italia, el Reino Unido, Suecia y Polonia han demostrado su interés por unirse al proyecto. Este LOFAR extendido consistiría en unas 16 estaciones de observación adicionales, que sólo requerirían de un aumento de cálculo en el correlador y conexiones de internet adicionales (el flujo de datos aumentaría hasta 1 200 Gbps). Normalmente las estaciones se situarían en otros observatorios con lo que la existencia de canales de datos de alta velocidad está garantizada. El aumento de la potencia de cálculo necesario tampoco debe representar ningún problema, porque la ley de Moore garantizaría la disponibilidad de una mayor capacidad de cómputo para la fecha en que el interferómetro extendido entrase en funcionamiento, hacia 2010.



Figura 4: Posible distribución de las estaciones LOFAR en el proyecto extendido. Imagen cortesía de www.lofar.org.

Además de disminuir drásticamente el problema de la confusión de fuentes, la disponibilidad de una resolución espacial diez veces mayor también amplía el horizonte de los objetos y fenómenos que se podrían explorar con semejante instrumento. En el Sistema Solar se podría estudiar el comportamiento del propio Sol, tratando de entender cómo surgen y se desarrollan las tormentas solares. Estas eyecciones de material se propagan en el medio interplanetario afectando a los sistemas de comunicación por satélite y a la propia ionosfera de La Tierra. Se espera que mediante la observación continua del Sol durante unos segundos cada varios minutos, podamos ser capaces de predecir el tiempo solar (tormentoso, estable, variable), y adoptar medidas para mitigar sus efectos en La Tierra. También

podrán estudiarse los cinturones de radiación de Júpiter e investigar su existencia en Saturno. Estos cinturones de radiación no son más que partículas de alta energía (rayos cósmicos) que quedan atrapadas por el campo magnético de los planetas, y que emiten ondas radio por debajo de los 300 MHz. Dada la alta capacidad destructiva de los rayos cósmicos, la habilidad de un planeta para capturar esas partículas en cinturones de radiación resulta fundamental para el desarrollo de la vida. Precisamente, la entrada de los rayos cósmicos por los polos de la Tierra (la zona más desprotegida) es lo que da origen a las auroras boreales, que también podrían ser estudiadas en Júpiter. Incluso podría llegar a detectarse la emisión radio de las descargas eléctricas atmosféricas, es decir, los rayos, en las atmósferas de otros planetas del Sistema Solar.

SKA (www.skatelescope.org)

La idea que subyace detrás del proyecto SKA es muy sencilla, lo dice su propio nombre. SKA es el acrónimo del inglés Square Kilometer Array. Es decir, Interferómetro de un Kilómetro Cuadrado, y eso es de lo que se trata. Este proyecto, del que se empezó a hablar allá por los años 1990, pretende la construcción de un conjunto de radiotelescopios cuyas áreas de apertura sumen un total de un kilómetro cuadrado. Así de sencillo y así de ambicioso. Un kilómetro cuadrado es la superficie de un cuadrado que tenga un kilómetro de lado, o la de un círculo de 1128 m de diámetro, o si lo prefieren una parcela de 100 hectáreas. En la actualidad el radiotelescopio simple más grande es todavía la antena fija de Arecibo³ en Puerto Rico (1963); un círculo de 305 m de diámetro con una superficie de apertura de 73 000 m², es decir, un 7,3 % de lo que sería el SKA. Las 50 antenas del proyecto ALMA⁴ de 12 m de diámetro sólo suman 5650 m², mientras que las 27 antenas de 25 m del VLA⁵ representan el 1,3 % del SKA. Como podemos ver el salto cuantitativo no es nada despreciable. De hecho hacia 1990 semejante instrumento representaba una mejora en un factor 100 para estudios de líneas espectrales, y de un factor 1000 para observaciones del radiocontinuo. Aunque hoy en día esos factores no son tan grandes, debido a proyectos como LOFAR y la mejora de los telescopios ya existentes, la construcción del SKA seguirá permitiendo observar objetos de 10 a 100 veces más débiles que lo que puede hacerse con la instrumentación presente.

Podemos imaginar el SKA como una versión ampliada de LOFAR, o a LOFAR como un mini-SKA. Ambos comparten el modo de funcionamiento de interferómetro por software, con capacidad para observar distin-

³<http://www.naic.edu>

⁴Para más detalles sobre ALMA ver el artículo de R. Bachiller en el Anuario de 2001.

⁵<http://www.vla.nrao.edu>

tos puntos del cielo al mismo tiempo, dependiendo del conjunto de retrasos introducidos entre las distintas estaciones. En el interferómetro SKA tampoco se van a instalar todas las antenas de un modo contiguo, ocupando un área de un kilómetro cuadrado, sino que se repartirán sobre una superficie mucho mayor, para así aumentar el poder de resolución del telescopio. El 50 % de la superficie colectora se concentrará en un área de unos 5 km de diámetro, el 37,5 % en un área de hasta 150 km, mientras que el resto (12,5 %) se distribuirá sobre distancias de al menos 3000 km. En total SKA agrupará unas 2000 antenas distribuidas en unas 100 estaciones de observación.

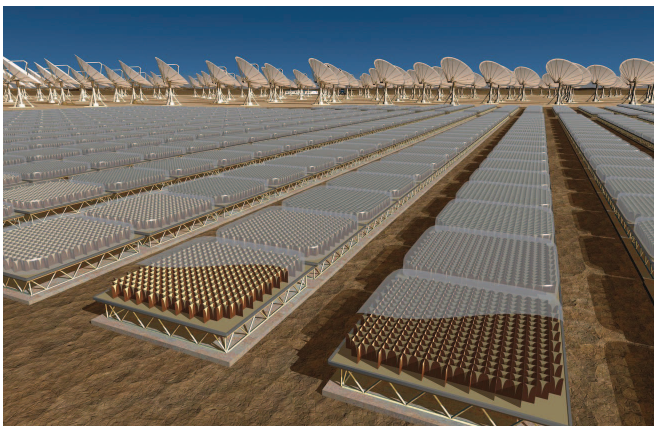


Figura 5: Recreación gráfica del aspecto del núcleo central del SKA. En primer plano las antenas de baja frecuencia. Al fondo las antenas que se podrían usar a frecuencias más altas. Imagen cortesía de www.skatelescope.org y www.xilostudios.com.

Aunque conceptualmente LOFAR y SKA son muy similares, el segundo resulta un proyecto mucho más complejo. En primer lugar se trata de un telescopio más grande y en segundo lugar se pretende que trabaje entre 100 MHz y 25 000 MHz (ó 25 GHz), es decir, $\lambda = 12 \text{ mm} - 3 \text{ m}$. La resolución espacial máxima esperada será pues de $0''.25$ a 100 MHz y $0''.001$ a 25 GHz (50 veces mejor que la del Telescopio Espacial Hubble). Además de este excelente detalle, SKA será capaz de producir imágenes gigantescas que cubran una gran región del cielo (hasta 50 grados cuadrados a 1 GHz con una resolución de $0''.25$, es decir, imágenes de 500 mil millones de píxeles). La potencia de cálculo exigida para generar estos mapas y procesarlos es impresionante, se necesitan realizar centenares de billones de operaciones por segundo. Además, se necesitan líneas de datos con capacidades de millones de Mbps para transmitir la información desde

las estaciones al centro de proceso. Nuevamente se confía en que siga cumpliéndose la ley de Moore, que duplica la velocidad de cálculo cada año y medio, la capacidad de almacenamiento de datos cada año, y la velocidad de transmisión cada 9 meses.

A $\lambda 3 \text{ m}$ la tecnología a usar será similar a la de LOFAR, pero a $\lambda 12 \text{ mm}$ tiene que ser radicalmente distinta pero, eso sí, manteniendo un coste total ajustado, entorno a los 1300 millones de euros (más 70 M€ al año de gastos de funcionamiento). El SKA representa un reto científico, tecnológico, industrial y logístico nada trivial. A fecha de hoy no está decidido qué tipo de antenas se usarán en el rango de mayor frecuencia. Existen varias opciones pero está por demostrar su viabilidad técnica y económica. De hecho está previsto que entre 2009 y 2012 se construyan distintas versiones del SKA a escala de un 1 %, a fin de comprobar fehacientemente cuál de las distintas tecnologías es la más adecuada. Tampoco se ha tomado una decisión definitiva sobre el emplazamiento del instrumento, aunque en 2006 la lista de posibles lugares se ha reducido a dos: Australia y el sur de África.

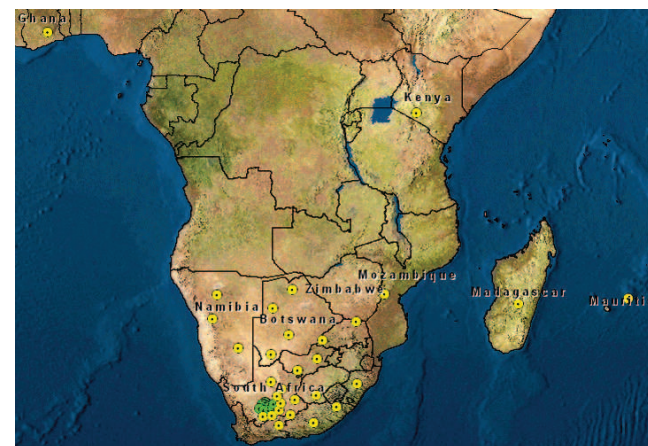


Figura 6: Localización de los nodos del interferómetro SKA propuestos para la opción sudafricana. Imagen cortesía de www.ska.ac.za.

Por las características del instrumento, la ciencia que se espera hacer con el SKA es en gran medida similar (aunque de mayor alcance) a la que he descrito para el caso de LOFAR. Aunque el SKA contribuirá de un modo muy significativo en numerosas campos de la Radioastronomía y la Astronomía en general, se han definido hasta cinco proyectos claves, suficientemente importantes como para condicionar incluso los parámetros fundamentales que caracterizan este telescopio. A estos proyectos científicos se les ha exigido que cumplan al menos tres condiciones: que se

centren en problemas fundamentales para la Física y/o la Astrofísica, que sean áreas en las que la Radioastronomía sea la mejor manera de abordar el problema, bien por sí sola o en combinación con otras técnicas de observación a otras longitudes de onda, y que sean problemas que preocupen a una gran comunidad científica, no sólo a expertos en Radioastronomía de bajas frecuencias. Estos proyectos clave son: la Cuna de la Vida, Gravitación en Condiciones Extremas, el Origen y Evolución de los Campos Magnéticos Cósmicos, el Origen y Evolución de las Galaxias, y por último la Observación de la Era Oscura. Además, no podemos olvidarnos del hecho de que con cualquier instrumento significativamente mejor que los anteriormente existentes, siempre se descubren objetos o fenómenos simplemente impensables hoy en día.

La Cuna de la Vida. Las estrellas similares a nuestro Sol, antes de empezar a quemar hidrógeno en su núcleo, pasan por una fase denominada de estrella T Tauri, en la que su luz se debe a que el gas que forma la estrella todavía se está compactando y por lo tanto liberando energía gravitacional. Las estrellas T Tauri están rodeadas de un disco de polvo y gas que gira a su alrededor: es en este disco donde se forman los planetas que orbitarán entorno a la estrella y que, dependiendo de sus condiciones físicoquímicas, serán capaces o no de albergar vida tal y como nosotros la entendemos. Se cree que los planetas se forman por agregación de esos granos de polvo. Por debajo de un milímetro es muy difícil que un choque entre dos granos los destruya, por lo que formar partículas de ese tamaño a partir de granos mucho más pequeños⁶ no plantea ningún problema. Objetos de un metro o mayores ya son capaces de aglutinar otras partículas, gracias a su atracción gravitatoria, y seguir creciendo hasta alcanzar tamaños de miles de kilómetros; es decir, hasta alcanzar el tamaño de un planeta como La Tierra. En cambio, lo que ocurre con los objetos entre un milímetro y un metro no está claro. Fundamentalmente estos granos emiten luz a longitudes de onda un poco más pequeñas que su tamaño, es decir, en el rango de alta frecuencia del SKA. Con su excelente sensibilidad y resolución angular, SKA podrá observar como crecen esos granos en objetos situados hasta distancias de 150 pc (unos 500 años luz), lo que incluye las poblaciones de estrellas T Tauri de las nubes de Tauro, Ofioco y Camaleón.

Una vez que los planetas se forman, especialmente los de tipo gigante como Júpiter, su presencia altera la estructura de esos discos circunestelares, creándose huecos, al igual que la influencia de las lunas de Saturno crea discontinuidades en el anillo de polvo de este planeta. El SKA también podrá detectar la presencia de uno de estos planetas en formación gracias a la discontinuidad generada en el disco. De hecho se podría llegar a detectar el giro del protoplaneta, puesto que para estrellas de tipo solar, el periodo orbital de los planetas cercanos será del orden de años. Repitiendo la obser-

⁶El tamaño típico de un grano de polvo interestelar es de unas milésimas de milímetro.

vación de un objeto cada pocas semanas, es posible que se distinga el giro de las irregularidades que produciría la atracción gravitatoria del planeta en el gas y el polvo del disco.

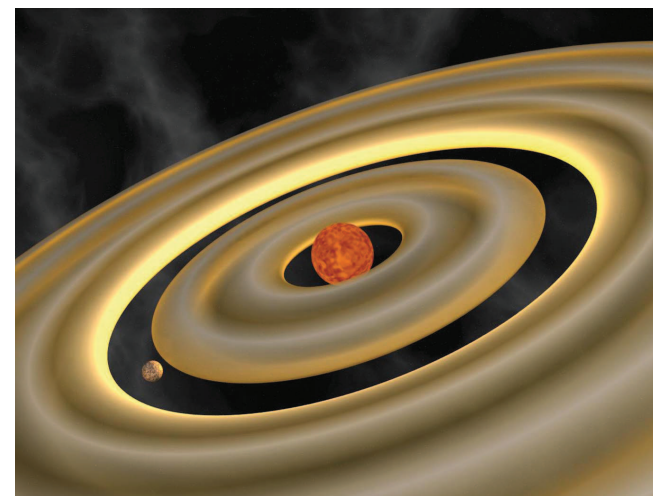


Figura 7: Recreación gráfica de la formación de un planeta gigante en el disco circunestelar de una estrella T Tauri. Imagen cortesía de Michael Kramer y www.skatelescope.org.

Claro está que detectar un planeta con condiciones habitables no es sinónimo de haber encontrado vida extraterrestre, y mucho menos vida inteligente. Aunque se supone que moléculas como el ozono y el metano son buenos indicadores de la existencia de vida (en este campo ALMA va a contribuir sin duda con notables descubrimientos), estos compuestos pueden originarse en entornos carentes de vida. En cierto sentido, es más fácil detectar vida extraterrestre inteligente. La sensibilidad del SKA será tal que si una civilización extraterrestre dispusiera de un sistema de radar similar al existente en Arecibo, podría llegar a detectarse a una distancia de entre 5 000 y 15 000 pc: es decir, a lo largo de todo el plano galáctico. Además, es precisamente entre 1 y 10 GHz donde cabría esperar este tipo de emisiones intencionadas, pues a estas frecuencias ni la emisión sincrotrón ni la de los granos de polvo las enmascararía. Como argumentaba Carl Sagan ya hace tiempo, para que dos civilizaciones se encuentren, tan importante es escuchar como transmitir. No obstante, incluso si una civilización extraterrestre no emitiera señales con el propósito de darse a conocer, sería posible detectar otro tipo de transmisiones, TV, radares de aeropuerto, etc., pero en un rango de distancias más pequeño, hasta centenares de años luz.

Gravitación Extrema. La teoría de la Relatividad General de Einstein ¿es 100 % correcta? Como es fácilmente comprensible, la respuesta a esta pregunta es de un enorme interés, no sólo para los astrofísicos sino para la Ciencia en general. De momento la respuesta es sí, incluso en los casos más extremos que se han podido medir, pero esto podría no ser así en el futuro. De hecho los físicos teóricos que trabajan en la búsqueda de un modelo unificado para todas las fuerzas, y que pasaría necesariamente por la cuantización de la gravedad, piensan que la última palabra no está dicha. Newton falla estrepitosamente en el caso de Mercurio porque este planeta está demasiado cerca del Sol como para despreciar las correcciones de la Relatividad General. De igual forma, los físicos teóricos piensan que las leyes de la Teoría de la Relatividad no se cumplen al 100 % cuando se trata de sistemas que se mueven a alta velocidad y en presencia de campos gravitatorios muy intensos; por ejemplo, un cuerpo orbitando muy cerca de un agujero negro.

Hasta la fecha no se han encontrado este tipo de sistemas binarios. Pero con la gran potencia detectora del SKA se espera que se detecten una buena parte de todos los púlsares⁷ de nuestra galaxia, unos 20 000, de los cuales más de 1 000 tendrían periodos del orden del milisegundo. En estas circunstancias no sería de extrañar que se detecten sistemas compuestos por un púlsar y un agujero negro estelar (surgido de la explosión supernova de una estrella de más de 20 masas solares), por un púlsar y un agujero negro de tipo medio (como los que parecen existir en algunos cúmulos de estrellas), e incluso por un púlsar y un agujero negro supermasivo (como el del centro de La Galaxia). La observación sistemática de estos sistemas, con un preciso control del periodo del púlsar, permitirá medir con total exactitud la órbita del púlsar alrededor del agujero negro. De esta manera se podrá verificar la validez o no de la Relatividad General en condiciones límite.

Por otra parte, observando repetidas veces un grupo numeroso de púlsares de periodo muy corto (milisegundos), se puede construir un gran detector de ondas gravitacionales. Una vez tenidos en cuenta los efectos del movimiento relativo entre un púlsar y La Tierra, y el secular pero predecible aumento del periodo del púlsar debido a su continua pérdida de energía, otras anomalías detectadas podrían deberse al paso de una onda gravitacional. Estas ondas han sido predichas pero no directamente

⁷Un púlsar es el resultado de la explosión supernova de estrellas entre 10 y 20 masas solares. En la explosión se expulsa toda la estrella excepto un pequeño núcleo, formado casi exclusivamente por neutrones. Estas estrellas de neutrones rotan hasta 1000 veces por segundo. Es tal la masa del objeto y la velocidad de giro, que esta velocidad permanece muy estable, simplemente gracias a la gran inercia del sistema. En realidad, la velocidad decrece con el tiempo pero de una forma predecible. Estos objetos emiten chorros de electrones a lo largo de las dos direcciones que definen su eje de giro. Estos electrones emiten una radiación de tipo sincrotrón que sólo es visible desde La Tierra si el haz nos enfoca, como ocurriría con un faro costero. De esta manera, midiendo la frecuencia con que nos llegan esos pulsos de energía se calcula la frecuencia de giro de la estrella de neutrones con gran precisión.

observadas hasta la fecha⁸. Si una de estas ondas atravesase el espacio entre nosotros y alguno de esos púlsares, las oscilaciones de la fuerza de la gravedad retrasarían y adelantarían la llegada de las señales emitidas por el púlsar de un modo fácilmente reconocible.

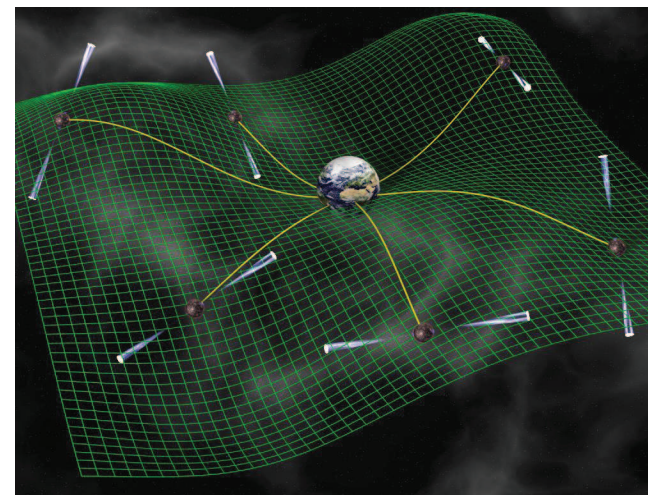


Figura 8: Mediante la observación continuada y con altísima precisión de la velocidad de giro de un conjunto de púlsares, se podría detectar el paso de una onda gravitatoria gracias a las perturbaciones inducidas en la estructura del espaciotiempo. Imagen cortesía de Michael Kramer (Sterne und Weltraum) y www.skatelescope.org.

Origen y Evolución de los Campos Magnéticos Cósmicos. No se puede tener una comprensión acertada del universo sin entender el papel que desempeñan los campos magnéticos. Si dejamos de un lado la fuerza oscura, sólo existen dos tipos de fuerza que actúan a gran distancia, la Gravedad y el Electromagnetismo. El universo a gran escala es neutro (puede estar ionizado, pero la suma de iones positivos y negativos se cancela), por lo que la fuerza eléctrica puede despreciarse. Gravedad y Magnetismo son pues las dos únicas fuerzas (junto con la fuerza oscura) que controlan la estructura del cosmos y su evolución. Todavía no se entiende bien el origen del magnetismo a gran escala. El SKA, por sus frecuencias de trabajo y su facultad para detectar la polarización de las

⁸Se sabe de su existencia por los estudios realizados por Russell Alan Hulse y Joseph Hooton Taylor en el sistema binario PSR B1913+16 (J1915+1606), que les valieron el premio Nobel de Física en 1993. En este sistema binario compuesto por dos púlsares, la evolución de las órbitas de los objetos sólo se explica si se tiene en cuenta la pérdida de energía que genera la emisión de ondas gravitacionales.

emisiones recibidas, está perfectamente capacitado para arrojar luz sobre este asunto. La intensidad de la radiación sincrotrón en las radiogalaxias mide la intensidad del campo magnético, y su ángulo de polarización su orientación. Este ángulo cambia según la frecuencia de observación debido a que el medio interestelar e intergaláctico están parcialmente ionizados. Pero precisamente la medida de este ángulo de rotación nos informa del valor medio del campo magnético a lo largo del camino recorrido por la luz. También se puede medir el campo magnético mediante el desdoblamiento de líneas espectrales de algunas moléculas como el radical hidroxilo, OH. Este tipo de observaciones serán capaces de determinar el valor del campo magnético en nubes moleculares neutras, por oposición a las medidas anteriormente citadas que trazan el gas ionizado. Resultará particularmente interesante estudiar la presencia de campos magnéticos muy extensos en galaxias de alto z . Si estos campos se originan por efecto dinamo (por el giro de un cuerpo conductor eléctrico, como en La Tierra) no deberían ser intensos en galaxias jóvenes, pues se sabe que el efecto dinamo tarda del orden de 1000 millones de años en desarrollar un campo magnético a gran escala.

Origen y Evolución de las Galaxias y Observación de la Era Oscura. No me voy a detener mucho en estos dos temas estrella para el SKA, pues en gran medida son compartidos por LOFAR y su importancia ya ha sido explicada páginas atrás. Respecto al primero de estos temas, sólo comentar que la posibilidad de observar máseres de vapor de agua con el SKA, permitirá medir la constante de Hubble con precisión del 1 % (un factor 10 mejor que las presentes medidas). Se han observado estos máseres girando alrededor de los agujeros negros supermasivos de algunas galaxias. La gran sensibilidad y resolución angular del SKA permitirá medir los movimientos orbitales de las zonas de emisión. Gracias al efecto Doppler se puede medir la velocidad de esas emisiones, mientras que la repetición de las observaciones muestra el movimiento de esas mismas zonas de emisión en unidades de ángulo subtendido por unidad de tiempo. Evidentemente unas y otras medidas están relacionadas entre sí mediante la distancia al objeto. De esta forma se pueden determinar distancias trigonométricas a galaxias lejanas, y de aquí calcular la constante de expansión del universo con mucha mejor precisión, mediante el estudio de un gran número de estos máseres.

En el segundo de los temas, añadir que la sensibilidad del SKA puede permitir la detección directa de las primeras fuentes de ionización, estrellas supermasivas (más de 200 masas solares) y los primeros agujeros negros. Nótese también que a $z=10$, la transición $J=2-1$ del monóxido de carbono (la molécula más abundante en el universo después del hidrógeno molecular), que normalmente se observa a 230 GHz, cae en la banda de 20 GHz. Esto posibilita que el SKA estudie el contenido molecular de las protogalaxias en formación antes del fin de la EdR.

Este instrumento impresionante sólo será posible aunando los esfuerzos de todos los radioastrónomos. Hoy en día, el SKA es un proyecto internacional en el que participan 17 países, representados por más de 50 instituciones, entre las que se encuentra el Observatorio Astronómico Nacional. Todavía queda mucho camino que recorrer, pero el futuro es esperanzador. Se espera que en 2010 se tome una decisión sobre el emplazamiento definitivo. En 2012 deberá estar decidida la tecnología a usar en cada longitud de onda. Para 2014–2016 se espera tener construido el 10 % de este interferómetro, iniciándose su uso científico. SKA estaría finalizado hacia 2020–2022.

FSMBULFRIA

FSMBULFRIA es un nombre de mi invención que de seguro no se corresponderá con ningún telescopio futuro, pero que describe perfectamente el último caso que quiero presentar brevemente: el Radiointerferómetro de Muy Baja Frecuencia de la Cara Oculta de La Luna (Far Side of the Moon Based Ultra-Low Frequency Radio Interferometric Array).

¿Un radiotelescopio en La Luna? En primer lugar cabe preguntarse para qué hace falta semejante instrumento, habida cuenta de lo compleja que sería su instalación y funcionamiento. La atmósfera, esa delgada capa que nos permite respirar y nos protege de los rayos X y ultravioletas, tiene su lado negativo; enturbia las observaciones astronómicas que se realizan desde tierra. Este efecto es bien conocido en luz visible, es el titileo de las estrellas, pero afecta en mayor o menor medida a todas las longitudes de onda. En las capas superiores de la atmósfera, entre 50 y 400 km de altura, se encuentra la ionosfera. Esta es una zona en la que los rayos X y ultravioletas ionizan los átomos y las moléculas, arrancándoles electrones. Estos electrones libres hacen que las ondas radio por debajo de una determinada frecuencia sean reflejadas. El valor en ciclos de esta frecuencia de corte es $9 \cdot 10^3 \sqrt{N}$, siendo N la densidad de electrones libres por cm^3 . Aunque fluctúa con la cantidad de radiación solar incidente, normalmente se alcanzan valores de N de $10^6 \text{ e}^-/\text{cm}^3$, por lo que la frecuencia de corte es de unos 10 MHz. Por esto los radioaficionados usan frecuencias más bajas, la banda HF, pues al rebotar sus emisiones en la ionosfera pueden alcanzar lugares mucho más distantes. La misma razón impide que las señales de objetos exteriores lleguen hasta nosotros, a no ser que instalemos un radiotelescopio a más de 500 km de altura. Tal radiotelescopio debería estar protegido de la emisión de la propia Tierra, por lo que un lugar ideal para su instalación sería la cara oculta de La Luna.

Establecer un radiotelescopio en La Luna plantea muchos problemas que no pasarán desapercibidos al lector. Hay que transportar los materiales hasta allí. Hay que transportar a los técnicos (humanos o robóticos) que lo instalen y lo mantengan en funcionamiento. Habrá también que enviar

los resultados hasta La Tierra, pues doy por hecho que ningún astrónomo va a subir hasta allí para realizar sus observaciones. Como no soy un experto en cosmonáutica no voy a tratar esos aspectos. Sí diré que hoy por hoy se piensa que para su instalación y funcionamiento se necesitará personal humano. Se habla de una posible base lunar, de igual forma que existe una base en el Polo Sur terrestre que da soporte allí a numerosos experimentos. Para que el telescopio no quedase cegado por el Sol, sólo se realizarían observaciones durante la noche lunar, de 14 días de duración. Pero precisamente durante este periodo no se podrían usar placas solares para la obtención de energía. Llegamos pues a la conclusión de que el instrumento debería funcionar con baterías durante dos semanas seguidas. El transporte de los datos plantea otra dificultad, pues desde la cara oculta no se pueden enviar directamente a La Tierra. Habría que disponer de un satélite de comunicación que orbitase La Luna y transmitiese los datos en diferido.

La instalación de un radiotelescopio en La Luna es una noticia recurrente, que recientemente ha cobrado más interés tras el anuncio por parte de la administración estadounidense, de su intención de volver a pisar nuestro satélite hacia 2020. De hecho la NASA ha seleccionado esta idea como una de las mejores para llevar a cabo en este hipotético regreso a La Luna. ¿Cómo sería este telescopio? Un diseño similar al de LOFAR resultaría muy ventajoso. Es simple, no tiene partes móviles y está bien adaptado para trabajar a bajas frecuencias. No parece casualidad pues que gigante aeroespacial europeo EADS y Astron, el instituto que lidera LOFAR, hayan firmado en 2006 un acuerdo para el desarrollo conjunto de un prototipo. No se trataría de un radiointerferómetro muy grande, pues el medio interplanetario emborrona mucho las imágenes a baja frecuencia, y por tanto la resolución espacial necesaria no es muy alta. A 10 MHz ($\lambda = 30\text{m}$) se pueden obtener imágenes con $1'$ de arco de resolución a pesar del centelleo interplanetario, lo que se alcanza con un interferómetro de 100 km de tamaño (a 1 MHz sólo se requieren 10 km para alcanzar la resolución máxima que permite el centelleo). Estas pobres resoluciones permiten construir un telescopio que patrullase todo el cielo, haciendo mapas cada muy poco tiempo.

¿Qué se estudiaría con semejante instrumento? En primer lugar, nos asomariamos al cosmos por una ventana jamás antes abierta. Detectaríamos la emisión ciclotrón (una versión menos energética de la emisión sincrotrón) de los cinturones de radiación de los planetas, del Sol y de radiofuentes intensas (cuásares, púlsares, etc.). Estudiaríamos las propiedades del propio medio interplanetario. Y como no, también buscaríamos vida inteligente más allá del Sistema Solar.

¿Se construirá el FSMBULFRIA? A lo mejor dentro de otros catorce años les doy la respuesta.

PUESTA AL DÍA DEL TEMA

A la caza de las grandes ondas diez más tarde

Hace casi diez años escribí un artículo dedicado a la nueva generación de radiotelescopios para emisiones de muy baja frecuencia (longitud de onda muy grande); en concreto sobre tres nuevos grandes instrumentos que estaban en proceso de construcción, diseño, o en fases aún más tempranas (definición de proyecto y búsqueda de fuentes de financiación, etc.). Aquí describiré brevemente los avances que ha habido en los últimos diez años en dos de esos desarrollos, LOFAR y SKA. Sobre el tercero de los desarrollos que entonces comentaba, la construcción de un telescopio de ondas radio de muy baja frecuencia en la cara oculta de la Luna, no ha habido ningún avance significativo; se sigue hablando del tema como una de las prioridades científicas para el establecimiento de una base lunar permanente, pero el proyecto sigue sin tener ni diseño, ni presupuesto, ni plazos de ejecución concretos.

De los tres instrumentos de los que se hablaba en el artículo publicado en 2008, solo uno de ellos se encuentra operativo, LOFAR, que ya entonces era el que estaba en una fase más avanzada de su desarrollo. El telescopio LOFAR se encuentra operativo desde 2010, pero sus capacidades siguen en evolución. A fecha de hoy, este enjambre de antenas está constituido por la estación central y 24 estaciones remotas, todas ellas localizadas en los Países Bajos, y 12 estaciones internacionales (seis en Alemania, tres en Polonia, y una en Francia, en Reino Unido y en Suecia). En la actualidad se está construyendo una estación adicional en Irlanda, y está previsto que en 2019 se construyan tres más en Letonia. La inclusión de estas nuevas estaciones, más separadas del núcleo que las existentes, aumentará la sensibilidad del instrumento, pero sobre todo aumentará su poder de resolución espacial (será capaz de obtener imágenes más nítidas). Durante los escasos años transcurridos desde su entrada en funcionamiento, LOFAR ha hecho improtantes contribuciones en distintos campos de la radioastronomía y astronomía en general. Merecen ser destacados sus observaciones del Sol (erupciones solares y viento solar), de rayos cósmicos (a través de la emisión de baja frecuencia de las partículas generadas en la atmósfera), de púlsares de muy corto periodo (observado la mayor muestra jamás estudiada), y su trabajo en el estudio de grandes muestras de galaxias, agujeros negros supermasivos, microcuásares, etc. LOFAR está realizando una cartografía completa del hemisferio norte celeste, a frecuencias de 50 y 150 MHz, que ha dado como resultado el descubrimiento de numerosas nuevas radiofuentes.

La evolución del gigantesco telescopio de 1 km^2 de superficie, el proyecto SKA, ha sufrido importantes retrasos. Hace diez años se pensaba que para estas fechas ya estaría construida la denominada fase 1, o SKA-1, que debía ser un 10% de la superficie total, pero éste no ha sido el caso. La decisión sobre el emplazamiento definitivo del instrumento se tomó en 2012; a la recta final llegaron dos propuestas, Sudáfrica y Australia, adoptándose el acuerdo salomónico de construir en cada emplazamiento uno de los dos telescopios que forman el SKA. Las antenas de media y alta frecuencia se instalarán en Sudáfrica mientras que las de baja frecuencia se emplazarán en Australia. Esta decisión está muy en la línea de las iniciativas pre-SKA que se han desarrollado en ambos países (MeerKAT en Sudáfrica y ASKAP en Australia). Un año más tarde se fundó el SKAO, la organización encargada del desarrollo del proyecto, con sede en el observatorio de Jodrell Bank, cerca de Manchester (R.U.). En la actualidad, el SKAO está participado por 10 países: Australia, Canadá, China, India, Italia, Nueva Zelanda, Países Bajos, Reino Unido, Sudáfrica y Suecia. En 2015 astrónomos de toda España redactaron un Libro Blanco para promover la adhesión de nuestro país al proyecto (2015arXiv150603474A). Se espera que este año finalicen y se aprueben los diseños para construcción de la fase 1, para la que se ha estimado un coste de 650 millones de euros. Está previsto que la construcción de SKA-1 se inicie en 2018, prolongándose hasta 2023, si bien las primeras observaciones podrían llevarse a cabo en 2020. Para la fase 2 todavía no se tienen costes precisos ni un calendario definitivo.

Para acabar, comentar brevemente que en 2016 se terminó la construcción del mayor radiotelescopio de antena única, no orientable. FAST (*Five-hundred-meter Aperture Spherical radio Telescope*) está situado en el sudeste de China y posee una superficie colectora de 500 m de diámetro, si bien no puede usar más que 300 m a un mismo tiempo. El funcionamiento de este instrumento es similar al de Arecibo (305 m de diámetro, de los que sólo se usan 225 m al tiempo). Las frecuencias de trabajo de FAST van de los 70 MHz a los 3 GHz¹.

Javier Alcolea Jiménez, marzo 2017

Este
libro ha sido
maquetado enteramente
usando el formato de L^AT_EX,
desarrollado originalmente por
Leslie Lamport, basado en el
sistema tipográfico T_EX de Donald
Knuth. El libro ha sido convertido
en PDF mediante pdfTeX (TeX
Live 2015/dev/Debian). La
versión final fue generada
el día 2 de junio de
2017.



¹ 1 GHz = 1000 MHz, 1 MHz = 10^6 Hz.

