

Grupo D. Conocimientos avanzados en Ciencias de la Tierra

- Tema 1. El campo de la gravedad terrestre. Superficies de nivel. Definición y propiedades. Representación analítica. Curvatura de las superficies de nivel. Curvatura de la línea de la plomada. Gradiente de la gravedad: Ecuación de Bruns generalizada.
- Tema 2. Desarrollo del potencial gravitatorio mediante armónicos esféricos. Variaciones en el tiempo del campo gravítico. Cambios temporales de la gravedad con origen en las mareas y con origen geodinámico.
- Tema 3. Correcciones y reducciones de la gravedad: Reducción de aire libre o de Faye. Reducción de Bouguer. Corrección topográfica. Cartas gravimétricas. La isostasia. Correcciones isostásicas. El cogeóide.
- Tema 4. Teorema de Meusnier. Radios principales de curvatura. Fórmula de Euler. Longitudes de arcos de meridiano y paralelo. Métodos numéricos de integración de la longitud de arco del meridiano.
- Tema 5. Redes de nivelación de precisión. Señalización. Instrumentos, métodos de observación, comprobaciones, correcciones. Errores sistemáticos. Cálculo y compensación de redes de nivelación. La red de nivelación de alta precisión española: REDNAP.
- Tema 6. Sistema GPS. El Segmento espacial. La señal GPS, sus características y procesamiento. El Segmento de control del Sistema GPS. Estaciones maestras y monitoras. Órbitas de los satélites. Cálculo de efemérides.
- Tema 7. Observables GPS. Captura de datos. Combinación de datos. Fuentes de error en GPS. Modelos matemáticos para el posicionamiento. Métodos de posicionamiento. Procesamiento de datos. Cálculo y compensación de una campaña GPS.
- Tema 8. Problemas directo e inverso de la Geodesia. Fundamentos de los tres tipos de soluciones: Integración del sistema de ecuaciones diferenciales de la geodésica, transferencia del triángulo polar elipsoidal a una esfera concéntrica y empleo de la esfera osculatriz de Gauss. Convergencia de meridianos.
- Tema 9. Compensación de una red geodésica por variación de coordenadas: Descripción general del método. Fórmulas diferenciales sobre el plano y sobre el elipsoide. Precisión requerida en las coordenadas aproximadas. Relaciones de observación: Direcciones, distancias y acimutes Laplace. Pesos de las observaciones. Elipse de error.
- Tema 10. Aplicación del método de mínimos cuadrados a la compensación de redes geodésicas. Análisis estadístico de los resultados. Concepto de precisión y fiabilidad. Incertidumbre de medida.
- Tema 11. Elasticidad. Ley de Hooke. Ecuaciones del movimiento en un medio elástico, homogéneo e isótropo. Funciones potencial de los desplazamientos y fuerzas. Función de Green.
- Tema 12. Ondas: La ecuación de ondas. Ondas sísmicas internas. Propagación en un medio elástico, homogéneo e isótropo. Reflexión y refracción en medios estratificados, planos y esféricos.
- Tema 13. Ondas sísmicas superficiales. Ondas superficiales en un medio semiinfinito. Ondas superficiales en una capa. Ondas Love. Dispersión de ondas. Velocidad de fase y de grupo.
- Tema 14. Resolución del problema inverso en Sismología. Inversión de dromocrónicas: Fórmula de Herglotz-Wiechert. Inversión de ondas superficiales. Inversa generalizada.
- Tema 15. Estructura de la corteza terrestre. Perfiles sísmicos. Interpretación. Sismogramas sintéticos. Método gravimétrico y magnético. Anomalías locales y regionales. Estructuras tipo de corteza.
- Tema 16. Representación de la fuente sísmica: modelos cinemáticos y dinámicos. Fuerzas equivalentes. Fuente puntual. Fracturas y dislocaciones.
- Tema 17. Parámetros focales de un terremoto. Magnitud y Energía. Tipos de fallas. Tensor momento sísmico y parámetros de fractura.
- Tema 18. Volcanología: Aspectos geodésicos y geofísicos de la actividad volcánica. Terremotos volcano-tectónicos. Tremor volcánico. Precusores sísmicos de la actividad volcánica. Otros precusores.
- Tema 19. Análisis armónico del campo geomagnético. El campo geomagnético internacional de referencia. Cartografía magnética. Anomalías magnéticas. Levantamientos magnéticos: terrestres, aéreos y marítimos.
- Tema 20. Paleomagnetismo. Propiedades magnéticas de la materia. Minerales magnéticos en rocas. Mecanismos de magnetismo remanente. Polos virtuales geomagnéticos. Migración de los polos y de los continentes. Inversiones del campo magnético.

Tema 1. Superficies de nivel. Definición y propiedades. Representación analítica. Curvatura de las superficies de nivel. Curvatura de la línea de la plomada. Gradiente de la gravedad: Ecuación de Bruns generalizada.

1.1. Superficies de nivel. Definición y propiedades

Sobre un punto de la superficie de la Tierra en reposo P, de coordenadas φ, λ , a una distancia R del centro de masas de la Tierra, el vector gravedad es la fuerza por unidad de masa, resultante de la fuerza de atracción gravitatoria f_N y de la fuerza centrífuga f_c debida a la rotación de la Tierra ω .

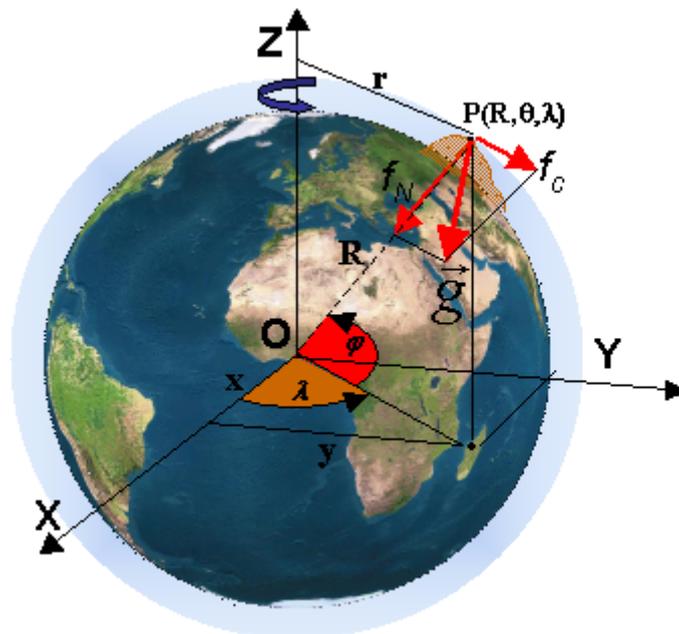


Figura 1. Vector gravedad y fuerzas que lo componen.

Tomando un sistema coordenado con origen O en el centro de masas de la Tierra (figura 1), con el eje Z en el eje de rotación medio de la Tierra, cada una de estas fuerzas se puede expresar en sus componentes, y por la teoría del potencial, dichas fuerzas pueden obtenerse a partir de un potencial gravitatorio V y un potencial centrífugo Φ , como:

$$\vec{f}_c = \omega^2 \vec{r} = (\omega^2 x, \omega^2 y, 0) = \overrightarrow{\text{grad} \Phi} = \nabla \Phi = \left(\frac{\partial \Phi}{\partial x}, \frac{\partial \Phi}{\partial y}, \frac{\partial \Phi}{\partial z} \right) = (\Phi_x, \Phi_y, \Phi_z)$$

$$\vec{f}_N = \overrightarrow{\text{grad}V} = \nabla V = \left(\frac{\partial V}{\partial x}, \frac{\partial V}{\partial y}, \frac{\partial V}{\partial z} \right) = (V_x, V_y, V_z)$$

$$\vec{g} = \vec{f}_c + \vec{f}_N = \overrightarrow{\text{grad}W} = \nabla W = \left(\frac{\partial W}{\partial x}, \frac{\partial W}{\partial y}, \frac{\partial W}{\partial z} \right) = (W_x, W_y, W_z)$$

Dado que el potencial total de la gravedad es la suma del potencial gravitatorio (integral extendida a todo el volumen de la Tierra) y el potencial centrífugo:

$$W(x, y, z) = V + \Phi = G \iiint_V \frac{\rho}{\ell} dv + \frac{1}{2} \omega^2 (x^2 + y^2)$$

Las laplacianas de estas funciones son:

$$\Delta \Phi = 2\omega^2$$

$$\Delta V(x, y, z) = -4\pi G\rho$$

$$\Delta W(x, y, z) = -4\pi G\rho + 2\omega^2 = (W_{xx}, W_{yy}, W_{zz})$$

que es la ecuación diferencial de Poisson generalizada. En el espacio exterior la densidad es nula (despreciando la densidad del aire) y se convierte en la ecuación de Laplace generalizada:

$$\Delta W(x, y, z) = 2\omega^2$$

El lugar geométrico de los puntos que cumplen la expresión:

$$W(x, y, z) = W_o = \text{cte.}$$

se denominan superficies equipotenciales, superficies de nivel o superficies geopotenciales. Al diferenciar el potencial total de la Tierra o potencial gravífico W y expresarlo en notación vectorial en coordenadas cartesianas como un producto escalar de dos vectores obtenemos:

$$dW = \frac{\partial W}{\partial x} dx + \frac{\partial W}{\partial y} dy + \frac{\partial W}{\partial z} dz = \overrightarrow{\text{grad} W} \cdot \vec{ds}$$

donde $ds=(dx,dy,dz)$ es un elemento diferencial entre superficies equipotenciales. Tomando el vector ds a lo largo de la superficie equipotencial, el potencial es constante y, por tanto, $dW=\mathbf{g}\cdot ds=0$. Esto quiere decir que el vector gravedad es normal a la superficie equipotencial que pasa por el mismo punto. Así pues, las denominaciones vertical, línea de la plomada y dirección del vector gravedad son sinónimas.

Las superficies de nivel son superficies de equilibrio, pues no se realiza trabajo para un desplazamiento a lo largo de la superficie de nivel. Las superficies equipotenciales tienen una expresión matemática muy compleja. Las superficies equipotenciales que se encuentran totalmente fuera de la Tierra son analíticas, y por tanto, no tienen discontinuidades en sus primeras ni segundas derivadas. Sin embargo, aunque el geoide como superficie equipotencial W_o es continua y sin aristas, no es analítica, pues presenta discontinuidades de densidad en su derivada segunda.

Las superficies de nivel no son paralelas, se juntan a medida que nos acercamos al polo. El decremento relativo entre dos superficies de nivel desde el ecuador a los polos es de 0,005.

Las líneas de la plomada son curvas espaciales que intersectan perpendicularmente a las superficies de nivel.

Si ahora tomamos el vector normal a la superficie, $d\vec{n}$, que es tangente a la línea de la plomada, y hallamos el producto escalar de ambos:

$$dW = \vec{g} \cdot d\vec{n} = g \, dn \cos(\vec{g}, d\vec{n}) = g \, dn \cos 180 = -g \, dH$$

se deduce la importante relación que existe entre la geometría y la física

$$g = -\frac{\partial W}{\partial H}$$

el vector aceleración de la gravedad (o simplemente gravedad) es el gradiente del potencial total de la gravedad. En su componente vertical, es el gradiente vertical negativo del gradiente del potencial W . En este caso la H es la altitud ortométrica, que se mide a lo largo de la línea de la plomada. Es la relación entre una cantidad física (la diferencia de potencial) y una cantidad geométrica (la diferencia de altura) entre superficies de nivel cercanas (figura 2).

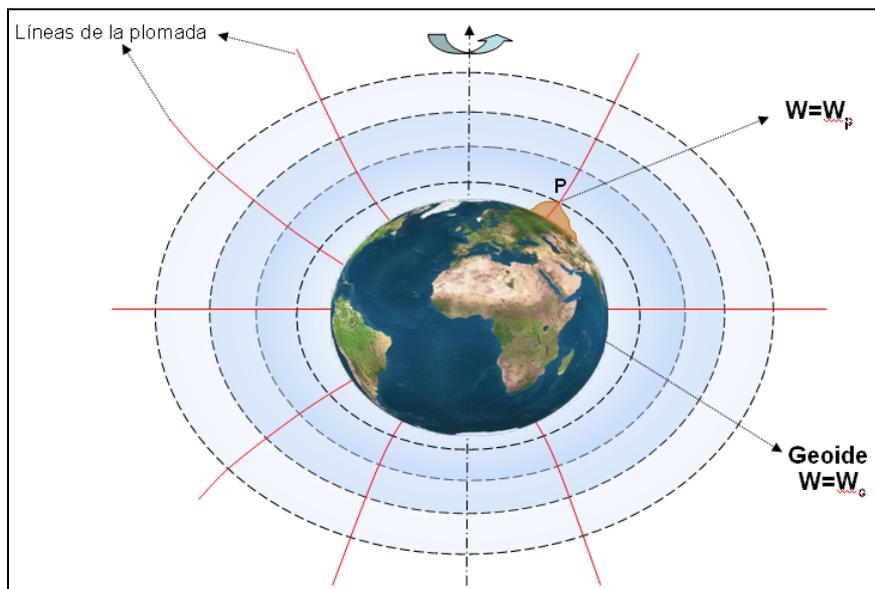


Figura 2. Superficies equipotenciales y líneas de la plomada.

1.2. Representación analítica

Fuera de la tierra las $W(x,y,z)=cte.$ son analíticas y continuas todas sus derivadas. Dentro de la Tierra no son continuas las derivadas segundas cuando hay cambios de densidad.

Situando un sistema cartesiano local en el que el eje Z coincide con la línea de la plomada y los ejes X e Y determinan el plano horizontal del entorno del punto P, tangente a la superficie equipotencial, podemos desarrollar en serie de Taylor en el entorno del punto P la ecuación diferencial de la superficie equipotencial. Si desarrollamos por serie de Taylor, podemos expresar la ecuación diferencial de la superficie de nivel en función del potencial total como:

$$W = W_p + W_x x + W_y y + W_z z + \frac{1}{2}(W_{xx} x^2 + W_{yy} y^2) + W_{xy} xy + \dots = 0$$

Como podemos sustituir $W_x=W_y=0$, $W_z=-g$ y $W=W_p$, entonces la ecuación analítica será:

$$-gz + \frac{1}{2}(W_{xx} x^2 + W_{yy} y^2) + W_{xy} xy + \dots = 0$$

1.3. Curvatura de las superficies de nivel

El radio de curvatura R_A de la sección normal para pequeñas distancias es:

$$\frac{1}{R_A} = -\frac{2z}{s^2}$$

Utilizando coordenadas polares planas:

$$x = s \cdot \cos A$$

$$y = s \cdot \sen A$$

Y sustituyendo en la ecuación analítica obtenemos:

$$\frac{1}{R_A} = -\frac{1}{g}(W_{xx} \cos^2 A + 2W_{xy} \sen A \cos A + W_{yy} \sen^2 A)$$

La curvatura normal adopta sus valores extremos en las dos direcciones de la curvatura principal, que son perpendiculares. Finalmente obtenemos las curvaturas en las direcciones $-x$ ($A=0^\circ$) y $-y$ ($A=-90^\circ$) donde R_x y R_y son los radios de curvatura correspondientes:

$$\kappa_1 = \frac{1}{R_x} = -\frac{W_{xx}}{g}$$

$$\kappa_2 = \frac{1}{R_y} = -\frac{W_{yy}}{g}$$

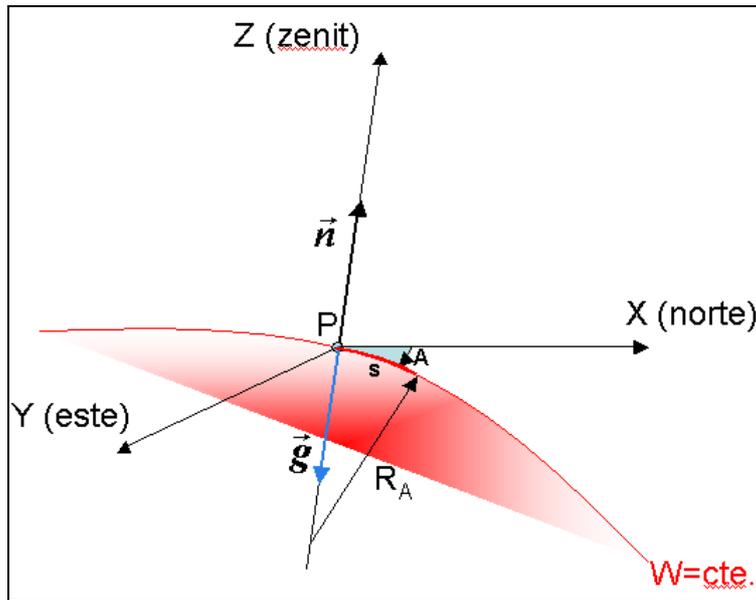


Figura 3. Superficies equipotenciales y líneas de la plomada.

1.4. Curvatura de la línea de la plomada

Una línea de la plomada puede definirse como una curva cuyo elemento de arco tiene la dirección del vector gravedad. La ecuación que define el factor de proporcionalidad entre el arco de curva de la plomada $ds = (dx, dy, dz)$ y las componentes del vector gravedad $\mathbf{g} = (W_x, W_y, W_z)$ que se deriva del potencial total de la gravedad es:

$$\frac{dx}{W_x} = \frac{dy}{W_y} = \frac{dz}{W_z}$$

y por tanto $\frac{dx}{dz} = \frac{W_x}{W_z}$

Si derivamos la expresión anterior obtendremos

$$\frac{d^2x}{dz^2} = -\frac{W_{xz}}{W_z}$$

Análogamente, para las otras dos componentes $\frac{dy}{W_y} = \frac{dz}{W_z}$ y por tanto $\frac{dy}{dz} = \frac{W_y}{W_z}$

obteniendo también:

$$\frac{d^2y}{dz^2} = -\frac{W_{yz}}{W_z}$$

Si consideramos el eje Z como la vertical del lugar, (teniendo en cuenta que $y=0$, $dx/dz=0$; $W_x=W_y=0$) la curvatura de las líneas proyectadas sobre los planos correspondientes XZ e YZ tienen las expresiones:

$$\kappa_1 = \kappa_x = \frac{d^2 x}{dz^2} = -\frac{1}{g} \frac{\partial g}{\partial x}$$

$$\kappa_2 = \kappa_y = \frac{d^2 y}{dz^2} = -\frac{1}{g} \frac{\partial g}{\partial y}$$

de acuerdo con la geometría diferencial, la componente cuadrática de ambas es la curvatura total de la línea de la plomada:

$$\kappa = \sqrt{\kappa_1^2 + \kappa_2^2} = \frac{1}{g} \sqrt{g_x^2 + g_y^2}$$

La curvatura media de la superficie de nivel J será:

$$J = \frac{1}{2}(\kappa_1 + \kappa_2) = -\frac{W_{xx} + W_{yy}}{2g}$$

Como

$$\Delta W(x, y, z) = W_{xx} + W_{yy} + W_{zz} = -4\pi G\rho + 2\omega^2$$

entonces podemos obtener la fórmula de Bruns,

$$\frac{\partial g}{\partial z} = \frac{\partial g}{\partial H} = -2gJ + 4\pi G\rho - 2\omega^2$$

que expresa la variación de la gravedad con la altitud, es decir, el gradiente de la gravedad en función de la curvatura de la superficie de nivel, la densidad de las masas que provocan esa superficie de nivel y la velocidad de rotación de esas masas sobre su eje.

1.5. Gradiente de la gravedad: Ecuación de Bruns generalizada

El gradiente de la gravedad en función del potencial gravífico es:

$$\overrightarrow{\text{grad } g} = \frac{\partial g}{\partial x} \overrightarrow{dx} + \frac{\partial g}{\partial y} \overrightarrow{dy} + \frac{\partial g}{\partial z} \overrightarrow{dz} = -(W_{xz} \overrightarrow{dx} + W_{yz} \overrightarrow{dy} + W_{zz} \overrightarrow{dz})$$

siendo dx , dy , dz los vectores unitarios en la dirección de los tres ejes cartesianos.

La ecuación de Bruns generalizada es la expresión del vector gradiente de la gravedad en sus tres componentes con relación a la curvatura de la superficie. Las componentes horizontales del gradiente de la gravedad constituyen el denominado gradiente horizontal y determinan la curvatura de la línea de la plomada. La componente vertical expresa la variación de la gravedad con respecto a la altitud. Se obtiene al combinar la ecuación diferencial de Poisson generalizada con la expresión de la curvatura media:

$$\overrightarrow{\text{grad } g} = \frac{\partial g}{\partial x} \overrightarrow{dx} + \frac{\partial g}{\partial y} \overrightarrow{dy} + \frac{\partial g}{\partial z} \overrightarrow{dz} = g \cdot \kappa_x \overrightarrow{dx} + g \cdot \kappa_y \overrightarrow{dy} + (-2gJ + 4\pi G\rho - 2\omega^2) \overrightarrow{dz}$$

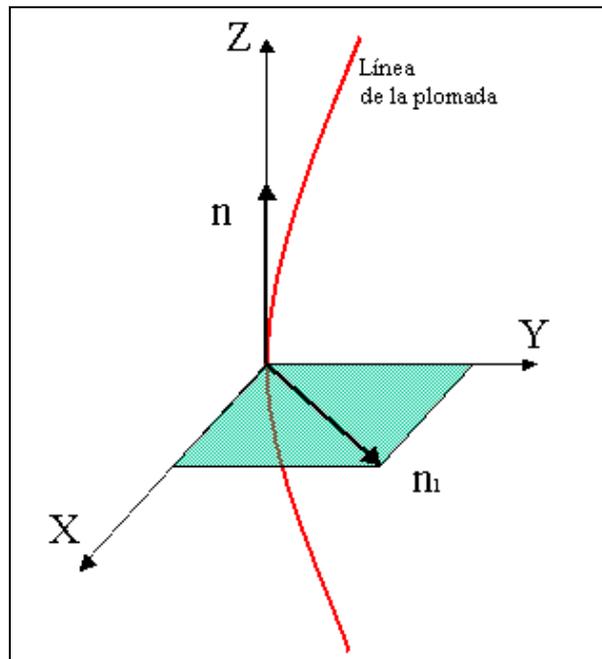


Figura 4. Línea de la plomada.

La ecuación que H. Bruns dedujo en 1878 denominada ecuación de Bruns generalizada es una ecuación vectorial:

$$\overrightarrow{\text{grad } g} = (-2gJ + 4\pi G\rho - 2\omega^2) \vec{n} + g\kappa \vec{n}_1$$

En donde las componentes \mathbf{n} y \mathbf{n}_1 se expresan como $\vec{n} = (0,0,1)$ y $\vec{n}_1 = (\cos A, \sin A, 0)$ donde A es el ángulo entre la normal principal y el eje x o acimut astronómico.

Bibliografía

- [1] Heiskanen, W, Moritz, H. Geodesia Física. IGN. 1985.
- [2] Torge, W. Gravimetry. Walter de Gruyter. Berlín, 1989.
- [3] Udías, A. y Mézcua, J. Fundamentos de Geofísica. Editorial Alianza, 1996.
- [4] Torge, W. Geodesy. Walter de Gruyter. Berlin, 2001.

Tema 2. Desarrollo del potencial gravitatorio mediante armónicos esféricos. Variaciones en el tiempo del campo gravífico. Cambios de gravedad con origen en las mareas y con origen geodinámico.

2.1. Desarrollo del potencial gravitatorio mediante armónicos esféricos

El potencial total de la gravedad se expresa como la suma de un potencial gravitatorio y un potencial centrífugo mediante:

$$W = V + \Phi = \iiint_{\text{tierra}} G \frac{\rho dv}{\ell} + \frac{1}{2} \omega^2 (x^2 + y^2)$$

Tanto uno como otro se pueden expresar por medio de funciones armónicas esféricas. La ecuación de Laplace $\Delta V=0$, que caracteriza el campo gravitatorio en el exterior en ausencia de masa externas, tiene una solución especial que es el desarrollo o expansión del potencial gravitatorio terrestre en armónicos esféricos y que se muestra útil para la resolución de problemas relacionados con el campo de la gravedad global. Debido a que la función densidad es desconocida no se puede calcular directamente el potencial gravitatorio. Por el método de separación de variables sabemos que las funciones solución a la ecuación de Laplace son de la forma

$$V(r, \theta, \lambda) = f(r) \cdot Y(\theta, \lambda) = r^n Y(\theta, \lambda)$$

$$V(r, \theta, \lambda) = f(r) \cdot Y(\theta, \lambda) = \frac{Y(\theta, \lambda)}{r^{n+1}}$$

Las primeras funciones se denominan armónicos esféricos sólidos y las $Y(\theta, \lambda)$ armónicos esféricos de superficie. Las primeras corresponden al potencial gravitatorio interno y las segundas corresponden al potencial gravitatorio externo. Como la ecuación de Laplace es lineal y conocemos varias soluciones, entonces la suma es también solución, por lo que:

$$V(r, \theta, \lambda) = f(r) \cdot Y(\theta, \lambda) = \sum_{n=0}^{\infty} r^n Y(\theta, \lambda)$$

$$V(r, \theta, \lambda) = f(r) \cdot Y(\theta, \lambda) = \sum_{n=0}^{\infty} \frac{Y(\theta, \lambda)}{r^{n+1}}$$

Sin embargo, es mejor desarrollar $1/l$ (l es la distancia entre el punto generador del campo gravitatorio y el punto afectado por dicho campo) por lo que desarrollado mediante una serie convergente, resulta ser en función de los polinomios de Legendre:

$$\frac{1}{\ell} = (r^2 + r'^2 - 2rr' \cos \psi)^{-1/2} = \frac{1}{r} \left(1 + \frac{r'}{r} - 2 \frac{r'}{r} \cos \psi\right)^{-1/2} = \frac{1}{r} \sum_{n=2}^{\infty} \left(\frac{r'}{r}\right)^n P_n(\cos \psi) = \sum_{n=0}^{\infty} \left(\frac{r'^n}{r^{n+1}}\right) P_n(\cos \psi)$$

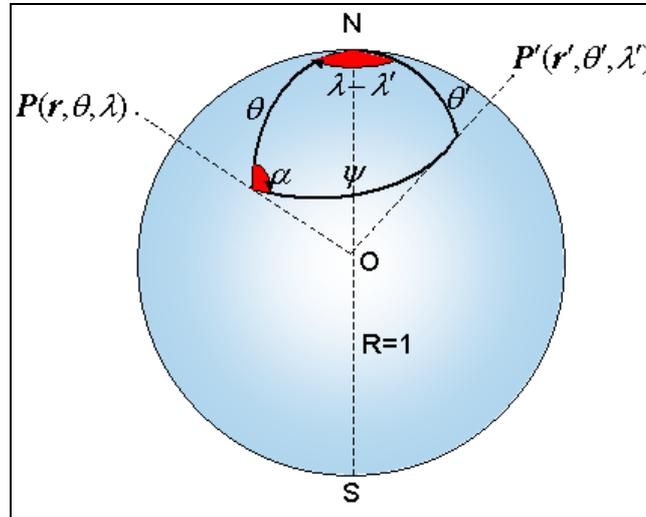


Figura 1. Punto fuente P' y punto campo P: triángulo esférico.

Del triángulo de la figura 1 se deduce:

$$\cos \psi = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos(\lambda' - \lambda)$$

y la fórmula de descomposición:

$$\begin{aligned} P_n(\cos \psi) &= P_n(\cos \theta)P_n(\cos \theta') + \\ &+ 2 \sum_{m=0}^n \frac{(n-m)!}{(n+m)!} [P_{n,m}(\cos \theta) \cos m\lambda P_{n,m}(\cos \theta') \cos m\lambda' + P_{n,m}(\cos \theta) \sin m\lambda P_{n,m}(\cos \theta') \cos m\lambda'] = \\ &= P_n(\cos \theta)P_n(\cos \theta') + 2 \sum_{m=0}^n \frac{(n-m)!}{(n+m)!} [R_{n,m}(\theta, \lambda) R_{n,m}(\theta', \lambda') + S_{n,m}(\theta, \lambda) S_{n,m}(\theta', \lambda')] \end{aligned}$$

en donde se han introducido las abreviaturas:

$$R_{n,m}(\theta, \lambda) = P_{n,m}(\cos \theta) \cos m\lambda$$

$$S_{n,m}(\theta, \lambda) = P_{n,m}(\cos \theta) \sin m\lambda$$

aunque otros autores utilizan la notación:

$$C_{n,m}(\theta, \lambda) = P_{n,m}(\cos \theta) \cos m\lambda$$

$$S_{n,m}(\theta, \lambda) = P_{n,m}(\cos \theta) \sin m\lambda$$

Las funciones asociadas de Legendre (funciones armónicas esféricas) para el argumento $t = \cos \psi$ vienen dadas por la expresión:

$$P_{l,m}(t) = (1-t^2)^{m/2} \frac{d^m}{dt^m} P_{l,0}(t)$$

en donde los polinomios de Legendre vienen dados por la expresión:

$$P_{l,0}(t) = P_l(t) = \frac{1}{2^\ell \ell!} \frac{d^\ell}{dt^\ell} (t^2 - 1)^\ell$$

Si ahora tenemos en cuenta la definición los armónicos esféricos, la inversa de la distancia ℓ será en función de los armónicos esféricos convencionales:

$$\frac{1}{\ell} = \sum \frac{P_n(\cos\theta)}{r^{n+1}} r'^n P_n(\cos\theta') + 2 \sum_{m=0}^n \frac{(n-m)!}{(n+m)!} \left[\frac{R_{n,m}(\theta, \lambda)}{r^{n+1}} r'^n R_{n,m}(\theta', \lambda') + \frac{S_{n,m}(\theta, \lambda)}{r^{n+1}} r'^n S_{n,m}(\theta', \lambda') \right]$$

Si en vez de armónicos convencionales utilizamos armónicos esféricos normalizados, que denotaremos con una raya superior, obtendremos una expresión más simple para el inverso de la distancia:

$$\frac{1}{\ell} = \sum_{n=0}^{\infty} \sum_{m=0}^n \left[\frac{\bar{R}_{n,m}(\theta, \lambda)}{r^{n+1}} r'^n \bar{R}_{n,m}(\theta', \lambda') + \frac{\bar{S}_{n,m}(\theta, \lambda)}{r^{n+1}} r'^n \bar{S}_{n,m}(\theta', \lambda') \right]$$

Introduciendo las fórmulas anteriores del inverso de la distancia entre los puntos fuente y los puntos campo, el potencial gravitatorio V :

$$V = G \iiint_{tierra} \frac{dm}{\ell} = \sum_{n=0}^{\infty} \left(\frac{1}{r^{n+1}} \right) G \iiint_{tierra} r'^n P_n(\cos\psi) dm$$

Entonces, sustituyendo la fórmula de descomposición en el potencial gravitatorio V , su desarrollo se puede expresar de dos maneras:

I. Con armónicos esféricos normalizados:

$$V = \sum_{n=0}^{\infty} \sum_{m=0}^n \left[A_{nm} \frac{\bar{R}_{nm}(\theta, \lambda)}{r^{n+1}} + B_{nm} \frac{\bar{S}_{nm}(\theta, \lambda)}{r^{n+1}} \right]$$

$$(2n+1) \bar{A}_{n,m} = G \iiint_{tierra} r'^n P_n(\cos\theta') \cos m\lambda' dm = G \iiint_{tierra} r'^n \bar{R}_{nm}(\theta', \lambda') dm$$

$$(2n+1) \bar{B}_{n,m} = G \iiint_{tierra} r'^n P_n(\cos\theta') \sen m\lambda' dm = G \iiint_{tierra} r'^n \bar{S}_{nm}(\theta', \lambda') dm$$

II. Con armónicos esféricos convencionales:

$$V = \sum_{n=0}^{\infty} \sum_{m=0}^n \left[A_{nm} \frac{R_{nm}(\theta, \lambda)}{r^{n+1}} + B_{nm} \frac{S_{nm}(\theta, \lambda)}{r^{n+1}} \right]$$

donde los coeficientes de los armónicos esféricos $A_{n,m}$ (ó $C_{n,m}$) y $B_{n,m}$ (ó $S_{n,m}$) son integrales de masa del campo gravitatorio que tienen la forma:

$$A_{n,0} \equiv C_{n,0} = G \iiint_{tierra} r'^n P_n(\cos\theta') dm$$

$$\left. \begin{aligned} A_{n,m} &\equiv C_{n,m} = \frac{2}{M} \frac{(n-m)!}{(n+m)!} G \iiint_{\text{tierra}} r'^n R_{nm}(\theta', \lambda') dm \\ B_{n,m} &\equiv S_{n,m} = \frac{2}{M} \frac{(n-m)!}{(n+m)!} G \iiint_{\text{tierra}} r'^n S_{nm}(\theta', \lambda') dm \end{aligned} \right\} m \neq 0$$

El campo gravitatorio total se expresa frecuentemente, en conexión con la observaciones realizadas por los satélites artificiales, utilizando M como la suma de la masa total de la Tierra incluida la atmósfera, y a el radio ecuatorial de la Tierra:

$$V = \frac{GM}{r} \left\{ 1 - \sum_{n=2}^{\infty} \left(\frac{a}{r}\right)^n J_n P_n(\cos \theta) + \sum_{n=2}^{\infty} \sum_{m=1}^n \left(\frac{a}{r}\right)^n P_n^m(\cos \theta) (A_{nm}(\cos m \lambda) + B_{nm}(\sin m \lambda)) \right\}$$

y también en la forma:

$$V = \frac{GM}{r} \left\{ 1 - \sum_{n=1}^{\infty} \sum_{m=0}^n \left(\frac{a}{r}\right)^n [J_{nm} R_{nm}(\theta, \lambda) + K_{nm} S_{nm}(\theta, \lambda)] \right\}$$

$$\text{con} \quad \begin{aligned} A_{n,m} &= -GMa^n J_{n,m} \\ B_{n,m} &= -GMa^n K_{n,m} \end{aligned}$$

Generalmente, en geodesia mediante satélites, se suelen utilizar las denominaciones de los coeficientes armónicos con la notación: $J_n = -C_n$; $J_{n,m} = -C_{n,m}$; $K_{n,m} = -S_{n,m}$. Entonces la fórmula del potencial V es:

$$V = \frac{GM}{r} \left\{ 1 + \sum_{n=1}^{\infty} \sum_{m=0}^n \left(\frac{a}{r}\right)^n P_{nm}(\cos \theta) (C_{nm}(\cos m \lambda) + S_{nm}(\sin m \lambda)) \right\}$$

En particular, los coeficientes armónicos de segundo grado son función de los momentos de inercia de la Tierra:

$$\begin{aligned} A &= J_{xx} = \iiint_{\text{tierra}} (Y'^2 + Z'^2) dm & B &= J_{yy} = \iiint_{\text{tierra}} (X'^2 + Z'^2) dm \\ C &= J_{zz} = \iiint_{\text{tierra}} (X'^2 + Y'^2) dm \end{aligned}$$

Y de los productos de inercia

$$D = J_{yz} = \iiint_{\text{tierra}} Y'Z' dm \quad E = J_{xz} = \iiint_{\text{tierra}} X'Z' dm \quad F = J_{xy} = \iiint_{\text{tierra}} X'Y' dm$$

En particular los coeficientes:

$$\begin{aligned} C_{2,0} &= \frac{1}{a^2 M} \left(\frac{J_{xx} + J_{yy}}{2} - J_{zz} \right) \\ C_{2,1} &= \frac{1}{a^2 M} J_{xz} & S_{2,1} &= \frac{1}{a^2 M} J_{yz} \end{aligned}$$

$$C_{2,2} = \frac{J_{yy} - J_{xx}}{4a^2 M} J_{xz} \quad S_{2,2} = \frac{1}{2a^2 M} J_{xy}$$

El coeficiente $J_2 = -C_{2,0}$ es el llamado aplanamiento geométrico polar de la Tierra y se suele denominar coeficiente dinámico, que tiene un valor aproximado de $1.08 \cdot 10^{-3}$. Los coeficientes $J_{10} = J_{11} = K_{11} = 0$, expresan que en realidad el desarrollo del potencial comienza en $n=2$. También $J_{21} = K_{21} = 0$.

La expresión del campo gravitatorio en función de los momentos de inercia principales de la Tierra y de los productos de inercia se escriben según la fórmula de MacCullagh:

$$V = \frac{GM}{r} + \frac{G}{r^3} \left[\frac{1}{2} \left(C - \frac{A+B}{2} \right) (1 - 3 \cos^2 \theta) + \frac{3}{4} (B - A) \sin^2 \theta \cos 2\lambda \right] + O \left[\frac{1}{r^4} \right]$$

donde los términos de orden $O(r^{-4})$ y superiores se desprecian por ser infinitésimos. Además de representar cantidades físicas los coeficientes de los términos armónicos esféricos, el significado geométrico de los armónicos esféricos en el desarrollo del potencial distingue entre:

- Armónicos zonales: $m=0$, son los que no dependen de la longitud geográfica. Por tanto la simetría en longitud, o simetría en rotación de la Tierra anularía todos los términos del desarrollo del potencial que dependieran de m . En la realidad estos coeficientes m distintos de cero son muy pequeños.
- Armónicos teserales: m distinto de cero.
- Armónicos sectoriales: $m=n$, desaparece la dependencia de la latitud.

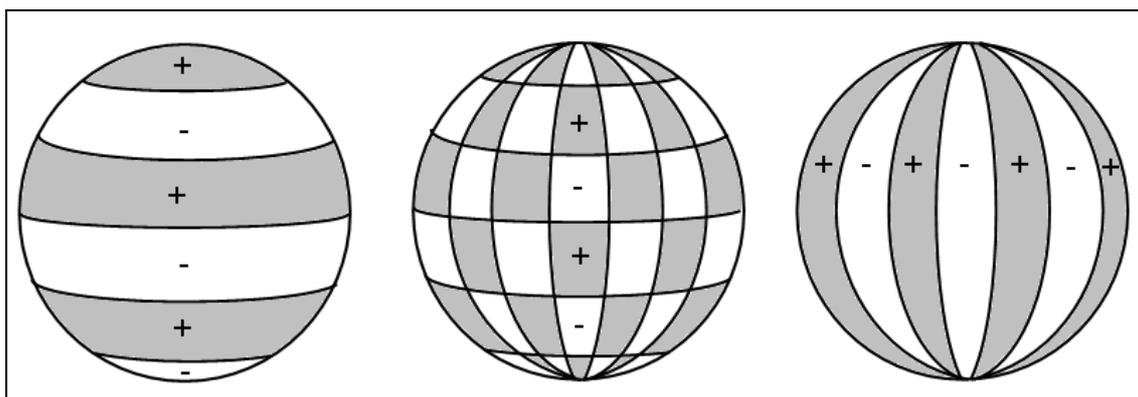


Figura 2. Signo de los armónicos esféricos zonales, teserales y sectoriales.

La expansión armónica converge fuera de una esfera de radio $r=a$ y no puede ser válida para el interior de las masas atractivas, puesto que entonces ya no sería válida la ecuación de Laplace, sino la ecuación de Poisson.

2.2. Variaciones en el tiempo del campo gravífico

2.2.1. Cambios de gravedad con origen en las mareas

Las observaciones directas de la gravedad están afectadas de las mareas terrestres, del movimiento del polo, de la atracción gravitatoria y carga procedente de la carga oceánica de las mareas oceánicas (efecto indirecto oceánico), y de los cambios de gravedad debidos a las variaciones de presión atmosférica.

Las aceleraciones de las mareas dependientes del tiempo son ocasionadas por las fuerzas de atracción gravitatoria ejercidas por la Luna y el Sol y que actúan en las diferentes zonas de la tierra que se traslada y gira sobre sí misma. También se producen aceleraciones por la rotación de la Luna alrededor de la Tierra y de la Tierra alrededor del Sol.

Las descomposiciones del potencial de la marea de la Luna y el Sol fueron realizadas por Laplace, y posteriormente por Doodson en una suma de oscilaciones armónicas de frecuencias y amplitudes de marea. Una de las expansiones del potencial de marea o modelo de marea más utilizados en las últimas décadas ha sido el de Cartwright, Tayler y Edden (1973), en donde se desarrollaron 505 ondas de marea. Las variaciones máximas debidas a la suma de todas las contribuciones de las ondas de marea terrestre son de $300 \mu\text{Gal}$ (3000 nms^{-2}), aunque en la figura se observan menos de $250 \mu\text{Gal}$ de máximo a mínimo. En el día se producen dos máximos y dos mínimos.

Además de las mareas y sus efectos indirectos, el campo gravitatorio terrestre está afectado por los desplazamientos de masas en el interior de la Tierra o en el exterior (en la atmósfera). Aunque los cambios son muy pequeños, los instrumentos actuales comienzan a ser capaces de detectar y medir variaciones temporales que no son dependientes de las mareas. Las variaciones en la aceleración centrífuga y las variaciones en la velocidad de rotación de la tierra manifestadas en la variación de la longitud del día producen cambios relativamente periódicos y seculares en el valor de la gravedad para un lugar determinado.

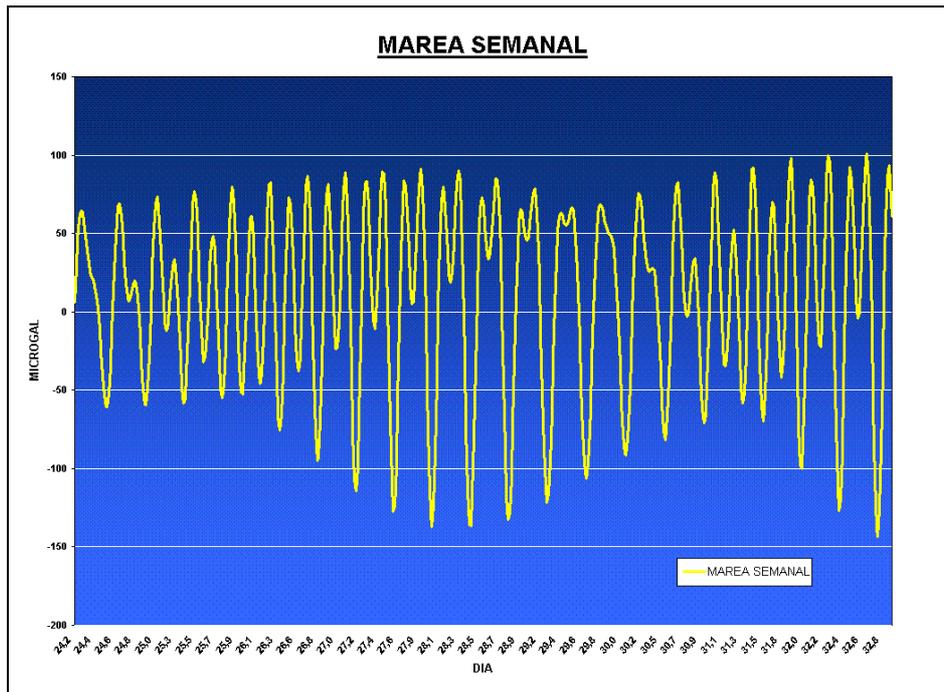


Figura 3. Marea terrestre en una semana experimentada en un punto de la superficie de la Tierra.

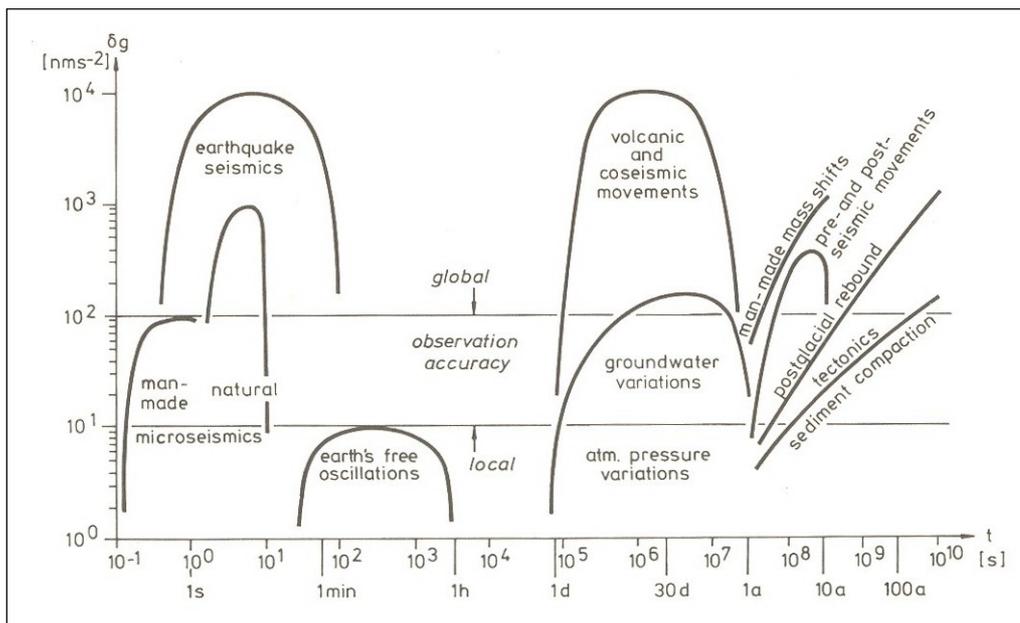


Figura 4. Frecuencias, períodos y amplitudes en las variaciones de gravedad con el tiempo (Torge, 1989).

2.2.2. Cambios de gravedad con origen geodinámico

En general, las variaciones temporales de la gravedad tienen una amplitud y frecuencia que depende del fenómeno considerado. Dependiendo del área de conocimiento objeto de estudio, unos fenómenos serán considerados como ruido, o bien como señal. En el caso de ruido deberán ser corregidos, acotados o despreciados según afecten o no a la señal objeto de estudio. Algunos de ellos se superponen en periodo, por lo que no es sencillo el poder diferenciarlos (i.e. carga oceánica y cambios del nivel freático). En la figura 4 se muestran las variaciones de gravedad con el tiempo en el eje de ordenadas, mientras que en el eje de abscisas se muestran los periodos.

Cambios globales de gravedad

Podemos citar como cambios globales en el valor de la gravedad en todas las partes del globo aquellos que tienen su origen en circunstancias de una extensión suficientemente grande como para cambiar el campo gravitatorio total o crear algún tipo de variación global. Afectan por tanto a los coeficientes armónicos del desarrollo en serie del campo gravitatorio. Si la masa de la Tierra es M , G la constante de gravitación universal, a el radio ecuatorial de la Tierra, ϑ la colatitud y λ la longitud, el potencial gravitatorio terrestre U en coordenadas esféricas (r, ϑ, λ) con origen en el centro de masas de la Tierra se puede expresar como:

$$U = \frac{GM}{r} \left\{ 1 - \sum_{n=2}^{\infty} \left(\frac{a}{r} \right)^n J_n P_n(\cos \theta) + \sum_{n=2}^{\infty} \sum_{m=1}^n \left(\frac{a}{r} \right)^n P_n^m(\cos \theta) (A_{nm}(\cos m \lambda) + B_{nm}(\sin m \lambda)) \right\}$$

siendo las funciones $P_n(\cos \vartheta)$ los polinomios de Legendre de orden n y $P_n^m(\cos \vartheta)$ los polinomios asociados de Legendre (soluciones particulares de la ecuación de Laplace), respectivamente. Los momentos de gravitación, números adimensionales J_n , A_{nm} y B_{nm} , dan peso a los términos de armónicos esféricos (longitud de onda).

La distinta orientación del eje de rotación del manto con respecto al núcleo terrestre el NDFW (*Nearly Diurnal Free Wobble*), así como del núcleo externo respecto del núcleo interno (FCN, *Free Core Nutation*), pueden dar lugar a variaciones globales que se encuentran por debajo del μGal , detectables con los gravímetros superconductores. Estos son los modos normales del núcleo, cuyo primer modo se denomina triplete de Schlichter.

Por ejemplo, los desplazamientos de masas convectivas (agua, atmósfera, manto) de muy largo período en el interior de la Tierra pueden afectar a los primeros armónicos del potencial gravitatorio (tabla 1), especialmente a J_2 , o producir variaciones del vector de rotación de la

Tierra. Las anomalías de la gravedad de muy larga longitud de onda pueden ser causadas por las plumas ascendentes o descendentes de material del manto convectivo.

\dot{J}_2	\dot{J}_4	\dot{J}_6
-3		
$-2,6 \pm 0,6$		
$-2,5 \pm 0,3$	$0,3 \pm 0,6$	
$-2,8 \pm 0,4$		
$-2,4 \pm 0,2$		
$-2,8 \pm 0,3$	$0,2 \pm 1,5$	
$-3,0 \pm 0,5$	$-0,8 \pm 1,5$	
$-2,7 \pm 0,4$	$-1,4 \pm 1,0$	$0,3 \pm 0,7$
$-2,9 \pm 0,2$	$0,6 \pm 0,5$	$0,3 \pm 0,3$

\dot{J}_3	\dot{J}_5	\dot{J}_{impar} $(\dot{J}_3 + 0,837 \dot{J}_5)$	\dot{J}_{impar} $(\dot{J}_3 + 0,9 \dot{J}_5)$
		$1,6 \pm 0,4$	
$-1,3 \pm 0,5$	$2,1 \pm 0,6$	0,5	0,6
			$0,5 \pm 0,2$

Tabla 1. Variaciones seculares (10^{-11} /año) en los coeficientes zonales pares e impares bajos del campo gravitatorio terrestre a partir de medidas de SLR.

Otro ejemplo interesante lo encontramos en las variaciones inducidas por terremotos muy grandes en la variación del coeficiente $J_2 = -2 \cdot 10^{-12}$ entre los años 1977-1985. Las variaciones detectadas mediante gravímetros criogénicos han permitido evaluar de nuevo los modos normales de oscilación de la Tierra. Incluso se ha escrito sobre su posible influencia sobre el eje instantáneo de rotación de la Tierra, especialmente con motivo del terremoto de magnitud 9 de Sumatra de 2004, que produjo el tsunami más devastador de la historia. Las observaciones absolutas y relativas de la gravedad también detectan estas perturbaciones.

Las últimas observaciones por satélite del campo gravitatorio terrestre han sido posibles gracias a los proyectos GRACE (“Gravity Recovery and Climate Experiment”), CHAMP (“Challenging Mini-satellite Payload”). Las anomalías de gran longitud de onda y su variación con el tiempo son detectables, aunque con menor exactitud que las mediciones realizadas sobre la superficie con los gravímetros absolutos y superconductores. A partir de 14 meses de observaciones de satélite, la Universidad de Texas ha realizado un modelo de campo gravitatorio global con 200 términos denominado “GRACE Gravity Model 02” (GGM02C), combinando datos de satélites gemelos (GRACE) y terrestres. En Europa, el GFZ (“GeoForschungZentrum”) de Potsdam ha desarrollado modelos de campos gravitatorios desde el EIGEN-GL01C hasta el nuevo modelo de campo gravitatorio terrestre EIGEN-GL04C (2006), que resuelve anomalías en el geoide y gravedad de hasta 110 km de longitud de onda y que contiene términos de desarrollo del potencial hasta el orden y grado 360, uniendo datos del satélite GRACE, Lageos y datos de

medida de gravedad en superficie. El modelo previo, EIGEN-GL03C, que unía datos de GRACE y CHAMP, tiene exactitudes de 30 cm y 8 mGal en resolución de longitudes de onda de 100 km.

El proyecto GOCE (Gravity Field and steady-state Ocean Circulation Explorer), previsto su comienzo para 2007, pretende obtener anomalías de la gravedad con exactitudes de 1 mGal (10^{-5} ms⁻²) y un geoide con exactitud de 1-2 cm, ambos en resoluciones espaciales mejores de 100 km.

Cambios regionales de gravedad

Los cambios regionales en el tiempo de la gravedad son los que afectan a regiones entre cien y varios miles de kilómetros. Tienen también un carácter de muy largo período o tendencia clara, aunque en los procesos de las dorsales oceánicas sean más rápidos. Se producen en los límites de placas, con la acumulación de esfuerzos en mucho tiempo y su liberación. También se producen en las zonas intraplaca debidos al rebote post-glacial, compactación sedimentaria y la neotectónica regional. En zonas volcánicas o de terremotos cambios locales pueden superponerse con los cambios regionales.

En lugares tan activos como las dorsales oceánicas es natural observar grandes variaciones de la gravedad en amplitud y en relativamente corto período de tiempo, asociadas a fenómenos de creación de corteza litosférica (“rifting”). Levantamientos regionales y disminución de gravedad en los flancos del rift, subsidencia local e incremento de gravedad en la zona central de la fisura fueron caracterizados para Islandia con medidas entre 1965 y 1980. Variaciones en esta zona de hasta 80 μ Gal en la fisura de Krafla a lo largo de 15 km.

La tectonofísica de un área activa muestra que cambios de gravedad en el tiempo entre 50 y más de 200 μ Gal son posibles. Los movimientos verticales en la corteza terrestre (levantamiento o subsidencia) son generalmente un orden de magnitud menor que los horizontales. Ocurren en los límites de placas y en el interior de las mismas, siendo las mayores subidas asociadas a tectónica de colisión entre continente-océano o continente-continente.

Geológicamente, ninguna parte de la corteza terrestre parece haber estado libre de deformación; en general, la deformación horizontal es del orden de 10 cm al año y la vertical del orden de algunos mm al año, proceso que ocurre entre 10^6 y 10^8 años. Las estimaciones del rebote postglacial en Canadá y Escandinavia muestran variaciones verticales de alrededor de 1 cm al año, proceso que ha ocurrido en los últimos 10^4 años. Sin embargo, se puede afirmar que existen regiones tectónicamente más activas que otras. En la Caldera de Long Valley se observaron una disminución de la gravedad debida a un levantamiento entre 1980 y 1983 después de la serie de terremotos de enero de 1983, y que probablemente fue seguida de adición de masa en profundidad. Los modelos propuestos incluyen inyección de magma a profundidades entre 3 y 5

km, e implican volúmenes inyectados de hasta $0,2 \text{ km}^3$. La gravedad y sus anomalías pueden cartografiar las distribución de densidad en profundidad e inferir la geología del subsuelo, y los cambios temporales de la gravedad pueden permitir evaluar la tectónica y la actividad volcánica. Se correlacionaron máximos de anomalías residuales isostáticas con la región de sismicidad de los enjambres de terremotos en 1983.

Sin embargo, los dos ejemplos más claros de áreas de levantamiento de la corteza terrestre son la región alrededor del mar Báltico y de la bahía de Hudson. Con datos provenientes de las investigaciones paleoclimatológicas conocemos que estas zonas estuvieron cubiertas de hielo durante la última glaciación. Ahora se elevan progresivamente y este efecto es interpretado como un rebote de la corteza después de que el hielo se ha derretido hace 10.000 años. Estudios desde 1939 han confirmado los desplazamientos posteriormente con técnicas actuales en ambos lugares. El rebote post-glacial ha sido estudiado también utilizando datos GPS, modelos de Tierra viscoelásticos obtenidos por inversión de observaciones del nivel del mar, o por observaciones de gravedad. Variaciones de dos μGal para levantamientos de 1 cm al año son los obtenidos en este último estudio utilizando observaciones relativas entre 1966 y 1993. A partir de la variación anual de $1,0 \pm 0,14 \mu\text{Gal/año}$ (junto con medidas de nivelación o GPS) se obtuvo la relación empírica $\dot{g}/\dot{H} = -2,0 \pm 0,6 \mu\text{Gal/cm}$. Utilizando diversos datos de gravímetros absolutos en Finlandia, se han obtenido en diversos lugares series temporales con tasas de variación de $-2,2$, $-3,0$ y $-1,1 \mu\text{Gal}$ al año, respectivamente.

En Canadá se han detectado variaciones totales de unos unos $22 \mu\text{Gal}$, y 14 cm en 12 años, que suponen tasa de variación del orden de los $2 \mu\text{Gal/año}$. Los gravímetros absolutos son especialmente adecuados para este tipo de estudios de largo período, pues no tienen deriva como el resto de instrumentos utilizados para la medida de la gravedad.

Las observaciones de gravedad y GPS se pueden comparar con la respuesta elástica existente en la región antártica. Se contemplan dos modelos que puedan causar movimientos verticales: un fenómeno de rebote elástico post-glacial de mayor largo plazo y otro de rebote elástico producido por el balance más actual de hielo. A partir de estas medidas se han obtenido observaciones repetidas con tasas de variación de gravedad en el continente antártico, ya que en el continente las condiciones de observación son difíciles con la instrumentación actual. Las tasas de variación obtenidas oscilan entre $-0,4 \pm 1,5 \mu\text{Gal/año}$ y $0,5 \pm 0,5 \mu\text{Gal/año}$.

En otros lugares como EEUU se han observado variaciones pico de $10 \mu\text{Gal}$ en varios meses, y en Alemania de $2 \mu\text{Gal}$ medidas en unos dos años. En Europa, en Italia se han obtenido variaciones en gravedad de $+4 \text{ nms}^{-2}/\text{año}$. Las variaciones medidas en Alemania son de $+4 \text{ nms}^{-2}/\text{año}$, una vez que han sido eliminadas las señales procedentes del nivel freático y otras de corto periodo. Las variaciones pico a pico que se muestran son en Italia de 70 nms^{-2} y en Alemania de casi 200 nms^{-2} . Estas variaciones están probablemente causadas por fenómenos no

modelados de período semianual y anual como las variaciones en el nivel freático u otras. En Bélgica, observaciones de gravedad admiten variaciones del orden de $-0,4 \mu\text{Gal/año}$. En la región de Gran Bretaña y norte de Francia las variaciones de gravedad deducidas han sido de $0,2 \pm 0,2$ y $0,1 \pm 0,2 \mu\text{Gal/año}$, respectivamente.

Las tasas anuales de variación de gravedad son mínimas y menores en comparación a las obtenidas en Europa continental (entre 0 y $0,4 \mu\text{Gal/año}$). El modelo a utilizar para los cambios temporales de gravedad de origen geodinámico por movimientos verticales de placas tectónicas es un modelo lineal de la forma:

$$\Delta g = \Delta g(r, t) + \Delta g_e(r, t) + \varepsilon$$

donde Δg_e representa las variaciones por nivel freático (periódicas), otras variaciones de corto y medio período y ε , todas consideradas como ruido aleatorio en el muy largo plazo.

Cambios locales de gravedad

Los cambios locales de gravedad pueden estar relacionados con actividad sismotectónica y volcánica en los límites de las placas continentales y también en zonas interiores de la placas. Su extensión se puede cifrar entre 10 y 100 km.

Barnes en 1966 atribuye las variaciones en gravedad observadas antes y después del terremoto de Alaska de 1964 a un incremento neto de la masa total en las regiones epicentrales. Kisslinger (1975) observa levantamientos de 22 cm entre 1965 y 1966 atribuidos a disminución de densidad en la corteza que relaciona con fenómenos de dilatancia. Efectos de levantamientos de 4 cm e incrementos de gravedad de $14 \mu\text{Gal}$ entre 1978 y 1981 fueron observados en Canadá. Una vez descartados los efectos del rebote postglacial, se atribuyeron a expansión elástica de fracturas los primeros, y a un transporte lateral de agua en el subsuelo el segundo.

Hagiwara informó de incrementos seculares de $0,02 \text{ mGal/año}$ superpuestos con ciclos de tres años con variaciones de $0,04 \text{ mGal}$. También detectó cambios de $-31 \mu\text{Gal}$ en 15 meses, y 15 cm de movimiento vertical en la Península de Izu (Japón). Cambios de gravedad en diferentes estaciones de California muy diversos se han analizado. Variaciones entre $+21$ y $-18 \mu\text{Gal}$ se han deducido a partir de varios tipos de instrumentos.

Son muchos los estudios de variaciones de gravedad relacionados con fenómenos volcánicos. El fenómeno a analizar es altamente dependiente del tipo de volcán y las anomalías de gravedad (junto con otros estudios tomográficos, sísmicos, etc) pueden ayudar a describir la estructura a la hora de definir un nivel de referencia del volcán a caracterizar. Estructuralmente, anomalías positivas de longitud de onda menores de 20 km están asociadas, con amplitudes del orden de 30 mGal , a material más máfico y denso que la media del material que lo circunda. Anomalías negativas mayores que 60 mGal se producen en áreas volcánicas más grandes, con materiales piroclásticos altamente silíceos de cenizas y pumitas; también pueden ser debidos a relleno de

una caldera no compactada silícea y cuerpos de magma de baja densidad. Cambios pequeños se asocian a respiraderos basálticos de material volátil pobre y a zonas de “rift” extensional, mientras que grandes cambios se refieren a explosivos, estratovolcanes de zonas de subducción y ricos en material volcánico andesítico.

En ciertos casos, en los pasos previos a la inminente erupción volcánica, se producen deformaciones superficiales considerables por la obstrucción de fisuras cercanas a la chimenea volcánica, así como variaciones en el campo gravitatorio próximo al volcán, como ocurrió previamente a la erupción explosiva del volcán del Monte Santa Elena (EEUU) en 1981. Sin embargo, el proceso de alimentación y realimentación con material del manto de la cámara magmática, en los volcanes que la poseen, puede generar variaciones importantes si se realiza la medición en el lugar adecuado.

En los volcanes activos sin cámara magmática o que se encuentran en zonas de emisión de lava en dorsales oceánicas (por ejemplo Krafla, Islandia) los cambios son grandes y más evidentes. Se distinguen entre variaciones pre-eruptivas y post-eruptivas. Para el volcán de Krafla, en las primeras, variaciones de 40 μGal y de 5-10 $\mu\text{Gal/año}$ son factibles. Las segundas detectaron 13 μGal y $-4,5 \mu\text{Gal/año}$.

Se puede incluso predecir la magnitud de la erupción a partir de los cambios de gravedad precursores. Es de destacar la importancia de las calderas intranquilas y de la medida del gradiente vertical de la gravedad en ellas. Asumiendo el modelo de Mogi, se pueden asociar incrementos de gravedad a lugares lejanos al centro de deformación y también por migración de agua en el sistema hidrotermal del volcán. También se asocian decrementos de gravedad al proceso de drenaje de la cámara de magma poco profunda o a procesos de deflación del volcán.

Bibliografía

- [1] Udías, A. y Mézcua, J. Fundamentos de Geofísica. Editorial Alianza, 1996.
- [2] Torge, W. Gravimetry. Walter de Gruyter. Berlín, 1989.
- [3] Torge, W. Geodesy. Walter de Gruyter. Berlin, 2001.
- [4] Heiskanen, W, Moritz, H. Geodesia Física. IGN. 1985.

Tema 3. Correcciones y reducciones de la gravedad: Reducción de aire libre o de Faye. Reducción de Bouguer. Corrección topográfica. Cartas gravimétricas. La isostasia. Correcciones isostáticas. El cogeode.

3.1. Correcciones y reducciones de la gravedad

Desde el punto de vista geodésico, el objetivo de las reducciones de la gravedad es obtener el valor de la componente vertical de la gravedad en el geoide, como si la medida se hubiese realizado allí. Eliminando las masas externas al geoide nos encontramos en las condiciones adecuadas para, con una gravedad normal, determinar unas anomalías suficientemente pequeñas que permitan mediante la fórmula de Stokes la determinación del geoide, y en definitiva, la determinación de la forma de la Tierra a través de su más representativa superficie equipotencial. En geodesia, las reducciones tienen por objeto cumplir las condiciones de contorno en el geoide, minimizando el efecto indirecto, para poder determinar la forma de la Tierra y sus superficies equipotenciales. Desde el punto de vista geofísico, estas reducciones se realizan con el fin de determinar la estructura de la corteza terrestre, en el sentido de defectos y excesos de masa por la composición de la corteza en relación a una tierra ideal. En geofísica y sus aplicaciones, la finalidad de las correcciones y reducciones de la gravedad es eliminar el efecto de atracción gravitatoria que ejerce la forma del terreno (topografía) y las masas por encima del nivel de referencia para investigar la estructura y las masas anómalas de la corteza terrestre.

Tanto desde un punto de vista geofísico como geodésico, las anomalías de la gravedad se realizan y se refieren respecto un cierto nivel de referencia o de un campo gravitatorio de referencia, como es el campo normal de la gravedad. Para la determinación del geoide por el método gravimétrico, así como para la representación cartográfica, se requieren anomalías de la gravedad que deban ser utilizadas como valores frontera en el geoide. Por tanto, podemos resumir que la reducción de la gravedad es necesaria para realizar:

- La determinación del geoide.
- Interpolación y extrapolación de la gravedad.
- Investigación de la estructura de la corteza terrestre.

El uso de la fórmula integral de Stokes requiere que las anomalías de la gravedad representen valores de contorno en el geoide:

- Debe referirse al geoide.
- No debe haber masas fuera del geoide.

Si $S(\psi)$ es la función de Stokes, Δg las anomalías de la gravedad y N la ondulación del geoide, podemos determinar N conociendo las anomalías de la gravedad mediante la fórmula integral de Stokes (integral sobre la superficie de la Tierra) se escribe como:

$$N = \frac{R}{4\pi G} \iint_{\sigma} \Delta g S(\psi) d\sigma$$

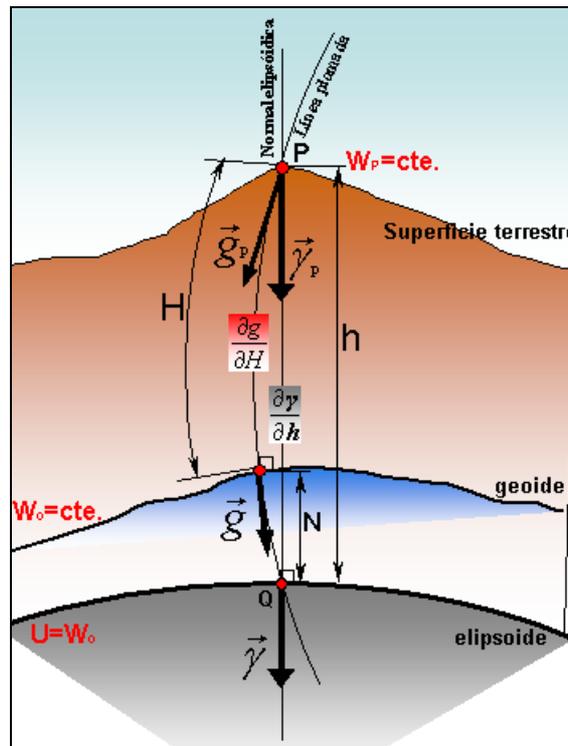


Figura 1. Esquema de superficies.

Si tenemos en cuenta la fórmula de Bruns $T = \gamma \cdot N$, la fórmula de Stokes es la solución de la ecuación diferencial que relaciona el potencial anómalo T (diferencia entre el potencial del geoide W y del elipsoide en un punto U) y las anomalías de la gravedad:

$$\frac{\partial T}{\partial h} - \frac{1}{\gamma} \frac{\partial \gamma}{\partial h} T + \Delta g = 0$$

La gravedad observada en un punto P es reducida al geoide o punto Q mediante un gradiente vertical de la gravedad normal teórico $\partial\gamma/\partial h$ sobre la línea normal elipsóidica, aunque debería realizarse a partir del gradiente vertical de la gravedad $\partial g/\partial H$ existente a lo largo de la línea de la plomada. Como este gradiente es difícil de obtener en el interior de la tierra, realizamos la aproximación $\partial\gamma/\partial h \approx \partial g/\partial H$, y los efectos de utilizar las altitudes geoidales H (ortométricas) o elipsoidales (h) son a efectos del gradiente vertical de la gravedad despreciables.

Por tanto es necesario reducir la gravedad observada en la superficie terrestre al geode. Las tres reducciones para llevar a cabo la traslación del valor de la gravedad en la superficie al geode son:

3.2. Reducción aire-libre o Faye

Se utiliza para corregir el valor de la gravedad observado en la superficie terrestre al que observaríamos de estar en el geode, por el simple hecho de trasladarnos en el espacio. Es siempre positiva, y su denominación es en memoria de H. Faye (1814-1902).

La fórmula lineal, teniendo en cuenta la aproximación del gradiente vertical de la gravedad, es:

$$C_{AL} = -\left(\frac{\partial g}{\partial h}\right) H_N \approx -\left(\frac{\partial \gamma}{\partial h}\right)_o H_N$$

La fórmula de segundo orden es:

$$C_{AL} = \left(\frac{\partial \gamma}{\partial h}\right)_o H_N + \frac{1}{2} \left(\frac{\partial^2 \gamma}{\partial h^2}\right)_o H_N^2$$

La fórmula de tercer orden es:

$$C_{AL} = \left(\frac{\partial \gamma}{\partial h}\right)_o H_N + \frac{1}{2} \left(\frac{\partial^2 \gamma}{\partial h^2}\right)_o H_N^2 + \frac{1}{6} \left(\frac{\partial^3 \gamma}{\partial h^3}\right)_o H_N^3$$

A efectos prácticos y poca precisión se utiliza la fórmula $C_{AL}=+0.3086 \cdot H$, que corresponde al gradiente normal de la gravedad, y nos da la corrección en mGal cuando la altitud H viene expresada en metros. La incertidumbre en esta fórmula es de $2.7 \cdot 10^{-5} \text{ ms}^{-2}$ para $H=10\text{km}$. Como $g_{AL}=g_{ob}+C_{AL}$, se obtiene la anomalía de aire libre escalar en la forma:

$$\Delta g_{AL} = g_{AL} - \gamma$$

Si tenemos en cuenta los efectos de elipsoidales, la corrección aire-libre en la teoría moderna de Molodensky se puede expresar como (Heiskanen y Moritz, 1967):

$$C_{AL} = \left(\frac{\partial \gamma}{\partial h}\right) = -\frac{2\gamma_e}{a} (1 + f + m - 2f \text{sen}^2 \varphi) h - \frac{3\gamma_e}{a^2} h^2$$

donde γ_e es la gravedad normal en el Ecuador, φ es la latitud geodésica del lugar, $m = \omega^2 a / \gamma_e$ es la relación entre la fuerza centrífuga en el Ecuador y la gravedad en el Ecuador, y f es el aplanamiento geométrico.

3.3. Reducción por lámina de Bouguer

Esta reducción, que se denomina así en memoria de P. Bouguer (1698-1758), se puede realizar mediante la asimilación de la masa entre el punto de observación y la superficie del geode como una lámina semi-infinita de Bouguer o una sección esférica. El efecto es de restar al valor

observado y se calcula aplicando la fórmula de atracción gravitatoria de un cilindro de radio infinito y altura H, según la expresión:

$$C_B = 2\pi\rho GH$$

A efectos prácticos y poca precisión se utiliza la fórmula $C_B = -0.1119 \cdot H$, que nos da la corrección en mGal cuando la altitud H viene expresada en metros, al tomar como densidad media para la corteza terrestre $\rho = 2670 \text{ kg}\cdot\text{m}^{-3}$. Si suponemos una lámina esférica (Sevilla, 1994), entonces la corrección de Bouguer sería $C_B = 2\pi\rho Gh + c_e$, en donde:

$$c_e = 0.0001464 \cdot H + 0.0000003182 H^2$$

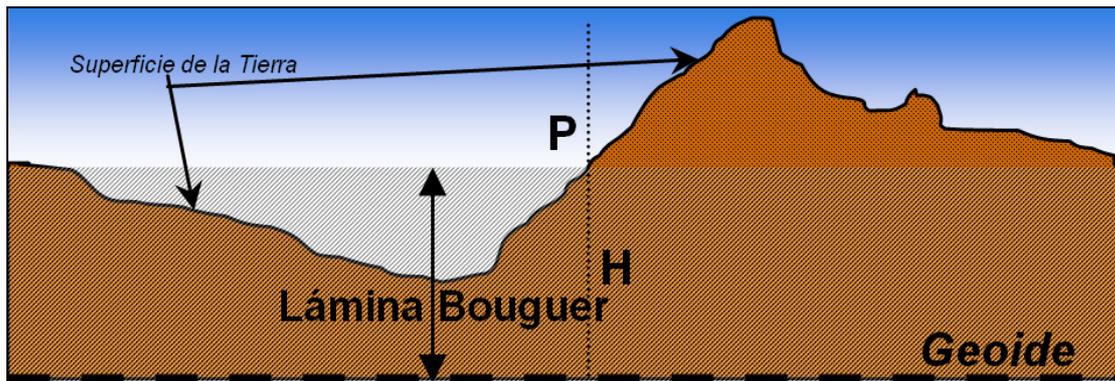


Figura 2. Esquema reducción de lámina de Bouguer.

3.4. Corrección topográfica (o de terreno)

En la reducción por una placa de Bouguer se ha considerado un plano semi-infinito en el entorno de la estación. En la realidad no ocurre de esta manera, sino que existen masas fuera de la lámina de Bouguer y huecos o defectos de masa dentro de la placa de Bouguer (ver figura 1). Esta corrección es siempre positiva, ya que el efecto de las masas externas a la lámina de Bouguer es disminuir la gravedad, y el efecto de los huecos es también su disminución. La incertidumbre en la determinación de la reducción topográfica está determinada por los errores de la hipótesis de densidad, al considerar una densidad promedio en la corteza. Los valores entre los que oscila esta corrección en áreas planas son $0.1-1.0 \cdot 10^{-5} \text{ ms}^{-2}$, y entre $10-100 \cdot 10^{-5} \text{ ms}^{-2}$, si consideramos una densidad media para la corteza terrestre de $2670 \text{ kg}\cdot\text{m}^{-3}$.

Antiguamente se corregía este efecto considerando zonas o compartimentos circulares alrededor del punto P donde se ha observado la gravedad. Los modelos digitales de elevación (MDE o DEM, acrónimo inglés) permiten en la actualidad aliviar el cálculo. Antiguamente la corrección de terreno se materializaba mediante las plantillas o compartimentos esféricos de Hayford y Bowie (1912) que nos proporcionaban utilizando las tablas de Lejay y de Cassinis, para ciertos compartimentos circulares concéntricos a la estación, las correcciones de terreno hasta distancias de 166.7 km, teniendo en cuenta que estos compartimentos tenían una altura media.

También las tablas de Hammer (1939,1982) han sido ampliamente utilizadas. Métodos espectrales se han utilizado también (Sideris, 1985; Fosberg, 1985).

Actualmente la corrección de terreno se materializa mediante los modelos digitales de elevaciones y las correspondientes fórmulas de atracción directa gravitatoria de los prismas que rodean la estación, en una malla con una rejilla determinada que corresponde al modelo digital del terreno en altitud.

La dificultad principal de esta corrección estriba en la formulación no sencilla en la determinación de la atracción gravitatoria de un prisma sobre un punto situado fuera del prisma. El algoritmo de Nagy (1966) ha sido ampliamente utilizado y programado para el cálculo automático mediante ordenador. Para un prisma rectangular de dimensiones límite $x_1, x_2, y_1, y_2, z_1, z_2$, la atracción newtoniana se calcula desarrollando la fórmula ($r=(x^2+y^2+z^2)^{1/2}$):

$$(\delta g_i)_T = G \Delta \rho \left\| \left\| -x \ln(y+r) - y \ln(x+r) + z \arctan \frac{xy}{zr} \right\|_{x_1, y_1, z_1}^{x_2, y_2, z_2} \right.$$

La suma de todas las contribuciones de los elementos en forma de prisma alrededor de la estación cuyo valor observado es objeto de esta corrección será:

$$C_T = \sum_i \delta g_i$$

Otras correcciones son:

- Corrección esférica y elipsoidal: para más precisión a zonas lejanas de la estación en la que se mide la gravedad de más de 400 km se ha de tener en cuenta los efectos esféricos (ya considerados en la lámina de Bouguer) o elipsoidales.
- Corrección atmosférica: La gravedad normal tiene en cuenta la atracción de la atmósfera, ya que se ha derivado de observaciones de satélite, mientras que la observada en la superficie de la Tierra no ha tenido en cuenta este efecto.

En resumen, la denominada gravedad reducida (para obtener la anomalía de Bouguer completa) se puede expresar mediante la fórmula:

$$g_B = g_{ob} + C_{AL} - C_B + C_T$$

De esta manera se obtiene la anomalía de Bouguer escalar completa en la forma:

$$\Delta g_B = g_B - \gamma$$

Existen tres maneras de definir la desviación de la vertical definidas por Helmert, Pizetti, y Molodensky. También, dependiendo del punto donde estén calculadas la gravedad normal o reducida la gravedad observada, existen varias maneras de calcular la anomalía gravimétrica. En el sentido clásico, la anomalía gravimétrica es el resultado de la diferencia entre la gravedad observada g_p en el geoide y una gravedad normal γ_Q calculada en el elipsoide (figura 1), ambos puntos con la misma latitud geodésica geocéntrica:

$$\Delta g = g_P - \gamma_Q$$

En la teoría moderna, según Molodensky, la anomalía gravimétrica es el resultado de la diferencia entre la gravedad observada g_P en la superficie de la Tierra y una gravedad normal γ_Q calculada en el teluroide (superficie no equipotencial):

$$\Delta g = g_P - \gamma_Q$$

por lo que debemos trasladar la gravedad normal en el elipsoide al teluroide (donde $W_P = U_Q$, y Q' está en la normal al elipsoide que pasa por P):

$$\gamma_{Q'} = \gamma_o + \left(\frac{\partial \gamma}{\partial h}\right)_o H_N + \frac{1}{2} \left(\frac{\partial^2 \gamma}{\partial h^2}\right)_o H_N^2 + \frac{1}{6} \left(\frac{\partial^3 \gamma}{\partial h^3}\right)_o H_N^3$$

La forma habitual en que se expresa la gravedad normal (en la superficie del elipsoide) es mediante el desarrollo en serie:

$$\gamma_o = \gamma_e (1 + \beta \text{sen}^2 \varphi - \beta_1 \text{sen}^2 2\varphi + \dots)$$

siendo γ_e = la gravedad normal en el Ecuador.

β = aplanamiento gravimétrico.

β_1 = factor que depende del aplanamiento geométrico = $-1/8 f^2 + 5/8 f m$

φ = Latitud geodésica del lugar.

$m = \omega^2 a / \gamma_e$ = fuerza centrífuga en el Ecuador/gravedad en el Ecuador = $\omega^2 a^2 b / GM$

f = aplanamiento geométrico.

FÓRMULA DE...	Referencias	γ_e (ms ⁻²)	β	β_1	f
HELMERT (1901)	Primera fórmula de gravedad normal	9,780300	0,005302	0,000007	1/298,3
BOWIE (1917)	US COAST AND GEODETIC SURVEY	9,780390	0,005294	0,000007	1/297,4
CASSINIS (1930)	Fórmula internacional de la gravedad 1930	9,780490	0,005288	0,0000059	1/297,0
GRS67	IAG1971	9,780318	0,005302	0,0000059	1/298,247
GRS80	Moritz 1984	9,780327	0,005302	0,0000058	1/298,257

Tabla 1. Coeficientes de las fórmulas de gravedad normal.

También se puede encontrar en la literatura gravimétrica la fórmula cerrada de Somigliana (1929), que es exacta:

$$\text{Originalmente: } \gamma_o = (a \gamma_e \cos^2 \varphi + a \gamma_p \text{sen}^2 \varphi) / (a^2 \cos^2 \varphi + b^2 \text{sen}^2 \varphi)^{1/2}$$

$$\text{Actualmente } \gamma_o = \gamma_e (1 + k \text{sen}^2 \varphi) / (1 - e \text{sen}^2 \varphi)^{1/2}$$

$$\text{donde } k = (b \gamma_p - a \gamma_e) / a \gamma_e$$

En el año 1980, se aprueba el Geodetic Reference System 1980 (GRS80), y con él la fórmula de la gravedad normal GRS80:

$$(\varphi = 0^\circ) \gamma_e = 978032.67715 \text{ mGal}$$

Con precisión de $0.1 \text{ mgal} = 1 \mu\text{ms}^{-2}$:

$\gamma_{o 1980} = \gamma_e (1 + \beta \text{sen}^2 \varphi - \beta_l \text{sen}^2 2\varphi) = 9.780 327 (1 + 0.005 3024 \text{sen}^2 \varphi - 0.000 0058 \text{sen}^2 2\varphi)$
en ms^{-2} .

Con precisión de $0.1 \mu\text{gal} = 1 \text{ nms}^{-2}$:

$$\gamma_{1980} = 9.7803267715 \cdot (1 + 0.0052790414 \cdot \text{sen}^2 \varphi + 0.0000232718 \cdot (\text{sen}^4 \varphi) + 0.0000001262 \cdot (\text{sen}^6 \varphi) + 0.0000000007 \cdot (\text{sen}^8 \varphi)) \quad \text{en } \text{ms}^{-2}.$$

Para el sistema WGS84, la gravedad normal viene dada por la fórmula cerrada de Somigliana:

$$\gamma_o = \gamma_e (1 + k \text{sen}^2 \varphi) / (1 - e^2 \text{sen}^2 \varphi)^{1/2}$$

$$\text{donde } k = (b \gamma_p - a \gamma_e) / a \gamma_e = b \gamma_p / a \gamma_e - 1$$

y sus valores son:

$$a = 6378137,0 \text{ m}$$

$$b = 6356752,3142 \text{ m}$$

$$e^2 = 6,69437999014 \cdot 10^{-3}$$

$$\gamma_p = 9,8321849378 \text{ m s}^{-2} = \text{gravedad en el polo.}$$

$$\gamma_e = 9,7803253359 \text{ m s}^{-2} = \text{gravedad en el ecuador.}$$

$$\varphi = \text{latitud geodésica.}$$

3.5. Cartas gravimétricas

Las cartas gravimétricas son la representación cartográfica de las anomalías del campo gravitatorio sobre la superficie terrestre. Cuando nos referimos a mapas o cartas gravimétricas queremos decir mapas de líneas isoanómalas gravimétricas. Las medidas sobre la superficie de la Tierra son reducidas a una superficie llamada geoide, materializada por la superficie de los océanos en calma y en ausencia de mareas, y se comparan con los valores calculados de gravedad normal en un elipsoide de referencia determinado. Las diferencias entre el campo normal y el campo real observado y reducido son pequeñas y se pueden considerar lineales (Heiskanen, 1985). Son las denominadas anomalías de la gravedad. A lo largo de la historia son muchos los elipsoides y fórmulas de gravedad normal determinados. En un punto de la superficie de la Tierra, la anomalía de la gravedad será la diferencia entre la gravedad teórica en el elipsoide para su latitud y la gravedad observada reducida al geoide. Las cartas gravimétricas son la representación cartográfica de las anomalías del campo gravitatorio sobre la superficie terrestre. Cuando nos referimos a mapas o cartas gravimétricas queremos decir mapas de líneas isoanómalas gravimétricas.

Los mapas gravimétricos se construyen a partir de una red de orden cero o fundamental convenientemente densificada en una densidad suficiente para la escala o hecho a representar.

Por lo tanto requieren conocer el valor de las anomalías de:

- Una malla de puntos obtenidos por medida directa.
- Una interpolación entre los puntos.
- Dibujo de curvas de nivel o isoanómalas.

El intervalo entre curvas de nivel ha de ser equidistante, permitiendo la lectura clara, al igual que ocurre en la cartografía convencional. La equidistancia es por tanto marcada por la escala del mapa, la precisión y exactitud con que se han calculado las anomalías de la gravedad y la estructura gravimétrica de la zona a representar.

Un primer criterio puede clasificar los mapas de anomalías gravimétricas en marinos, terrestres, aéreos y por satélite. Una segunda clasificación puede atender a la escala de representación, por lo que se distinguen mapas a escala regional (nacional) y mapas locales de estudios más específicos (por ejemplo para prospección geofísica).

Aparte de las universidades (Universidad Complutense, entre otras), las instituciones que más cartografía gravimétrica han editado han sido el Instituto Geológico y Minero de España (IGME), Instituto Geográfico Nacional (IGN), el Real Observatorio de la Armada (ROA), Instituto Hidrográfico de la Marina (IHM). El IGN ha sido pionero en estos estudios gravimétricos en tierra (figura 3), incluso para prospección geofísica a finales del siglo XIX y principios del XX con los métodos pendulares y la balanza de torsión de Eötvös-Schweydar.



Figura 3. Mapas de Anomalías de Aire Libre realizado por D. Guillermo Sans Huelín y presentados en la Asamblea de Geodesia y Geofísica Internacional celebrada en Octubre de 1924 en Madrid.

Numerosos estudios gravimétricos, esencialmente de anomalías de aire libre, Bouguer e isostáticas, existen en la región de la Península Ibérica, Baleares, Canarias y sus alrededores. Cabe citar en orden cronológico los de Sans Huelín (1924), Inglada (1927), Vening-Meinesz

(1932), Sans Huelín (1944), San Huelín y Lozano Calvo (1948) realizan el mapa de anomalías de aire libre y Bouguer con el conjunto de estaciones pendulares completado. Lozano Calvo (1948) realiza bajo la hipótesis de Airy el primer mapa de anomalías isostáticas de España, continuando su labor hasta los años 60, y elabora en 1961 el mapa para la corrección topoisostática según la teoría de Airy-Heiskanen y profundidades de 20 y 30 km. En 1964 realiza cartografía de aire libre, Bouguer-isostática, a escala 1:2000000.

A partir de las estaciones de la red fundamental, las observaciones para elaborar los mapas provinciales de anomalías de Bouguer y aire libre se extienden entre finales de los años cincuenta hasta finales de los años sesenta por Lozano Calvo, editando el IGN mapas provinciales de cartografía gravimétrica, a escala 1:200000. Las provincias observadas y publicadas con memoria son Ávila (107), Segovia (141), Santander (506), Gerona (266); Vicaya, Guipúzcoa, Álava, (total 303); Huelva (139), Navarra (1190), Burgos (1067), Palencia (877), Ciudad Real (845), Toledo (578), y Madrid, que se publicó en el año 1980. El BGI edita también junto con el norte de África mapas de anomalías, como Corón (1957, 1973), y Bureau Gravimétrico Internacional (1964). Otros mapas son los de Woollard (1965), Gaibar-Puertas (1970, 1972, 1973), Allan y Morelli (1971).

El Instituto Geográfico Nacional publicó el Avance del Mapa Gravimétrico de la Península Ibérica (Alonso San Millán, 1972), y el Mapa gravimétrico de la Península Ibérica e Islas Baleares, en el sistema geodésico de 1967 a escala 1:1000000 (Alonso San Millán, 1974). Otras publicaciones de cartografía gravimétrica son también Demnati (1972), Bonini et al. (1973), Van den Bosch (1974, 1981), Finetti y Morelli (1973), Morelli et al. (1975), Suriñach (1976), Hatzfeld (1978), Hatzfeld y Frogneux (1980), Liu (1993), Casas y Carbó (1990).

El “Estudio gravimétrico de la Península Ibérica y Baleares” en el sistema geodésico GRS80 por Mezcua et al. (1996) es uno de los más completos editados hasta ahora ya que incluye la corrección topográfica en una cartografía a escala 1:1000000 ; cabe citar los más actuales de Seber et al. (1996), Makris (IBCN,1998), Torné et al. (2000), Álvarez (2002).

El IGME ha planificado y observado tres categorías en la cartografía gravimétrica:

- Gravimetría estructural o Cobertura gravimétrica, a escala 1:50000 (298 hojas) (105598 puntos) Conjunto de medidas de gravedad realizadas con un objetivo estructural, generalmente realizada por itinerarios, con una densidad de estaciones 1 a 5 puntos por km².
- Gravimetría regional 1:200000 (46 hojas, 13600 puntos). Conjunto de medidas de gravedad realizadas con un objetivo regional (grandes anomalías), generalmente realizada por itinerarios, con una densidad de estaciones de 0.5 a 1 punto/10 km². En algunos casos la planimetría se realiza por posicionamiento "de visu" en plano y la altimetría con altímetro.

La corrección topográfica puede extenderse hasta la corona M de Hammer o hasta un radio de 167 Km.

- Gravimetría de detalle, minera 1:50000 (29 hojas, unos 115865 puntos). Conjunto de medidas de gravedad realizadas con un objetivo minero (áreas locales), generalmente realizada por perfiles o mallas, con una separación de estaciones del orden de 250 metros. En muchos casos la gravedad es relativa y la corrección topográfica se extiende a coronas inferiores a la M de Hammer.

El Ministerio de Defensa (IHM, IO, ROA, en colaboración con la Universidad Complutense) ha editado mapas marinos de Anomalías de Bouguer y de Aire Libre de Canarias y Baleares a escala 1:200000. También la universidad Complutense ha colaborado en la producción del mapa gravimétrico de España en el Marco del “West-East Europe Gravity Project (WEEGP).

Las anomalías de Bouguer son negativas en la mayor parte de la Península Ibérica. La región costera cantábrica y atlántica refleja anomalías positivas, incluso adentrándose en el interior del sur de Portugal, sierra de Huelva y Sevilla. La franja de costa mediterránea desde Cataluña hasta más al sur de Estepona y las islas Baleares se caracteriza por anomalías gravimétricas de Bouguer positivas.

3.6. La Isostasia

Se esperaba que las anomalías de Bouguer serían pequeñas y fluctuarían alrededor de 0. Pero las anomalías de Bouguer son en general sistemáticamente negativas en las zonas continentales montañosas y sistemáticamente positivas en la corteza oceánica. El defecto de masas en las zonas montañosas y el exceso de masa en las costeras mostrado por las anomalías de Bouguer condujeron a suponer que todos los excesos o defectos de masas por encima o por debajo del nivel del geode están compensados. Las masas topográficas son compensadas de alguna manera. Esta afirmación tiene en cuenta que los materiales se encuentran en equilibrio hidrostático, considerados solamente los esfuerzos verticales. Los modelos isostáticos son idealizaciones de la realidad, ya que no tienen en cuenta la existencia de esfuerzos tangenciales en el interior de la corteza. También parten de la condición de que cada columna vertical es independiente de las que le rodean. Existen dos hipótesis de compensación: Hipótesis de Pratt-Hayford, e hipótesis de Airy-Heiskanen. Estos dos modelos isostáticos son muy locales, por lo que también existe un tercer modelo regional de Vening-Meinesz, híbrida entre las dos primeras.

3.6.1. Hipótesis de Pratt-Hayford

La hipótesis de J. H. Pratt de 1854, puesta en forma matemática por J. F. Hayford en 1859, utiliza una capa de densidad constante y un nivel de compensación (figura 4), que suele estar situado a unos 100 km de profundidad. Todos los bloques isostáticos tienen el mismo peso y lo único que varía es su densidad y altura.

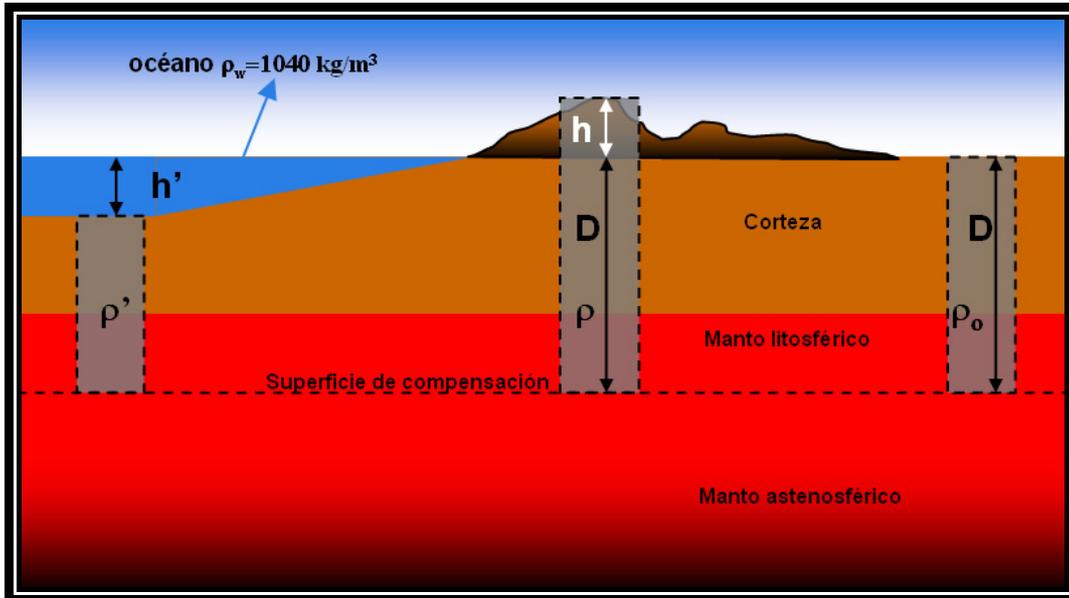


Figura 4. Esquema de modelo de Pratt-Hayford de compensación isostática.

Para una zona montañosa de altitud h la ecuación de los pesos es:

$$D\rho_o = (D+h)\rho$$

donde ρ es la densidad del bloque de altitud h y ρ_o es la densidad del bloque de altitud $h=0$ (nivel del mar). Se obtiene la densidad de la columna al despejar en la anterior ecuación:

$$\rho = \frac{D}{(D+h)}\rho_o$$

Para una zona oceánica de altitud h la ecuación de los pesos es:

$$D\rho_o = h'\rho_w + (D-h')\rho'$$

donde ρ' es la densidad del bloque de profundidad h' y ρ_w la densidad de la columna de agua.

De la misma manera se obtiene la densidad de la columna al despejar en la anterior ecuación:

$$\rho' = \frac{D\rho_o - h'\rho_w}{(D-h')}$$

3.6.2. Modelo de Airy-Heiskanen

La hipótesis de G. B. Airy en 1855, supone que existen dos densidades constantes, ρ_c para la corteza y ρ_M para el manto. Los espesores de la corteza son los que varían, existiendo debajo de los continentes raíces y debajo de los mares antiraíces (figura 5). Existe una superficie de separación de ambas regiones de distinta densidad. Por lo tanto, el mecanismo de compensación a igualdad de densidad consiste en un diferente volumen de las columnas, lo que se traduce en distinta altitud h y profundidad de la corteza t . La profundidad t es lo que se denomina raíz.

Para una zona continental:

$$(H + h + t) \rho_c = H \rho_c + t \rho_M$$

Y despejando la profundidad de la raíz t se obtiene:

$$t = \frac{\rho_c}{\rho_M - \rho_c} h$$

Para una zona oceánica se debe tener en cuenta la densidad ρ_w de la columna de agua:

$$h' \rho_w + (H - h' - t') \rho_c + t' \rho_M = H \rho_c$$

Y despejando t' se obtiene:

$$t' = \frac{\rho_c - \rho_w}{\rho_M - \rho_c} h'$$

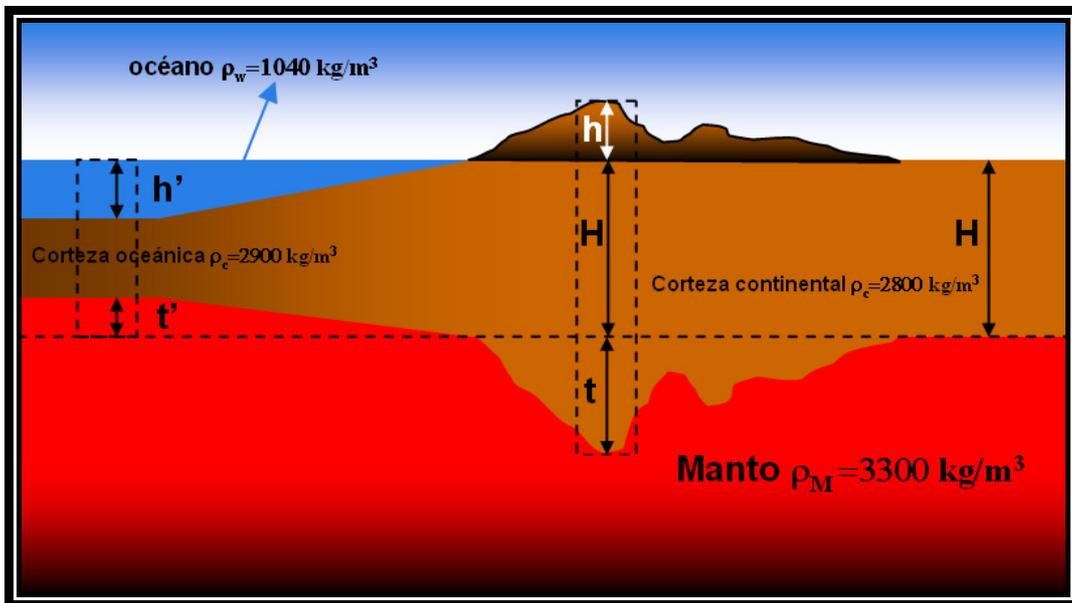


Figura 5. Esquema del modelo de Airy-Heiskanen de compensación isostática.

3.7. Correcciones isostáticas

En las reducciones isostáticas la topografía es eliminada junto con una compensación de acuerdo al modelo isostático elegido. Los pasos para obtener la denominada gravedad reducida isostáticamente son:

- Eliminación de la topografía (corrección topográfica).
- Corrección isostática.
- Reducción aire-libre al geoide.

La reducción isostática se calcula mediante la atracción de compartimientos alrededor de la estación mediante tablas calculadas a tal efecto (Cassinis, 1939; Heiskanen, 1938; en España, Lozano, 1965), o mediante la utilización de los modelos digitales de elevación (como en la corrección de terreno) con atracción de prismas (Nagy, 1966), en cuyo caso todas las contribuciones de los distintos cilindros o prismas se puede expresar mediante la fórmula:

$$C_I = \sum_i \delta g_i$$

donde δg_i se obtiene haciendo uso de las fórmulas auxiliares de atracción de un cilindro vertical de radio a , altura b , y la distancia c del punto donde queremos calcular el efecto a la base del cilindro:

$$\delta g_i = 2\pi G \Delta\rho \left[b + \sqrt{a^2 + (c-b)^2} - \sqrt{a^2 + c^2} \right]$$

que para ambas hipótesis, en zona continental, son:

- En la Pratt-Hayford $b=D$; $c=D+h$; densidad $\Delta\rho = \rho_o \cdot h/D$
- En la Airy-Heiskanen $b=t$; $c=t+H+h$; densidad $\rho-\rho_c$

Y en zona oceánica son:

- En la Pratt-Hayford $b=D-h'$; $c=D$; densidad $\Delta\rho = \rho-\rho_o$
- En la Airy-Heiskanen $b=t'$; $c=T$; densidad $\rho_c-\rho_w$

En resumen, la denominada gravedad reducida isostáticamente (para obtener la anomalía isostática mediante algún modelo de compensación de la isostasia) se puede expresar mediante la fórmula:

$$g_I = g_{ob} + C_{AL} - C_B + C_T + C_I$$

De esta manera se obtiene la anomalía isostática en la forma:

$$\Delta g_I = g_I - \gamma$$

Si alguna de las hipótesis isostáticas fueran rigurosamente ciertas, las anomalías isostáticas serían nulas. En la realidad, las anomalías isostáticas son pequeñas y suaves, aunque no son nulas, y aleatoriamente son ligeramente positivas o negativas. Si son positivas en una zona montañosa, la corteza está infracompensada, si son negativas la corteza está sobrecompensada.

3.8. Cogeoide

El llamado efecto indirecto de la gravedad es el cambio del potencial del geoide al aplicar las reducciones citadas de la gravedad. Al eliminar, cambiar o transportar las masas externas al geoide al interior del geoide estamos variando:

- Las superficies de nivel y las curvas equipotenciales.
- El campo gravitatorio y su potencial: la gravedad.

Por tanto, lo que estamos haciendo es cambiar el geoide, que ahora se denominará geoide regularizado, compensado o cogeoide (figura 6). Además estamos considerando una determinada distribución media de masas, que evidentemente, está sujeta a errores. Por lo tanto, la superficie calculada con la fórmula de Stokes no es el geoide sino el cogeoide. En realidad la fórmula de Stokes para determinación de la forma de la Tierra se aplica a las anomalías sobre el cogeoide, y la superficie determinada es el cogeoide. De acuerdo con el teorema de Bruns, la diferencia entre el geoide y el cogeoide δN ($N=N^C + \delta N$) se puede determinar a partir del cambio de potencial δW y la gravedad normal:

$$\delta N = \frac{\delta W}{\gamma}$$

El efecto indirecto es grande con anomalías de Bouguer simples, por lo que no son útiles para el cálculo del geoide, ya que aproximadamente es de 10 N (unos 440 m). Esto es debido a que la Tierra está isostáticamente compensada. Las anomalías isostáticas si sirven para realizar un buen cálculo del geoide, pues tienen una variación de unos 10 m, aproximadamente. Antes de aplicar la fórmula de Stokes las anomalías isostáticas deben ser reducidas del geoide al cogeoide mediante una corrección del tipo aire-libre. Para calcular un geoide con una precisión de 0.1 m y mejor se necesita tener en cuenta la masa atmosférica, la reducción de terreno y la aproximación esférica en las reducciones de la gravedad sustituida por una aproximación elipsoidal.

Superficie	Definición	Tipo de superficie	Potencial gravitatorio	Altitud
Terrestre	Relieve o topografía de la Tierra	Real	W_p	Geométrica
Teluroide	Lugar geométrico de los puntos que cumplen que el potencial normal es igual al potencial del punto $U=W_p$ de la superficie terrestre, en la normal al elipsoide	Esferopotencial	$U=W_p$	Normal, anomalía de altura
Geoide	Superficie materializada por los mares en calma sin los efectos de marea	Real, Equipotencial	W_o	Ortométrica (y geopotencial)
Cogeoide	Resultado del efecto indirecto de las reducciones de la gravedad	Equipotencial (o de nivel)	$W^C=W-\delta W$	Ortométrica (y geopotencial)
Cuasigeoide	Resultado de situar la anomalía de altura sobre el elipsoide.	Esferopotencial	-	Normal, anomalía de altura
Elipsoide	Superficie geométrica de revolución	Esferopotencial	$U=W_o$	Elipsoidal

Tabla 2. Potenciales gravitatorios de las superficies que definen la Tierra (campo de la gravedad real W y campo de la gravedad normal U).

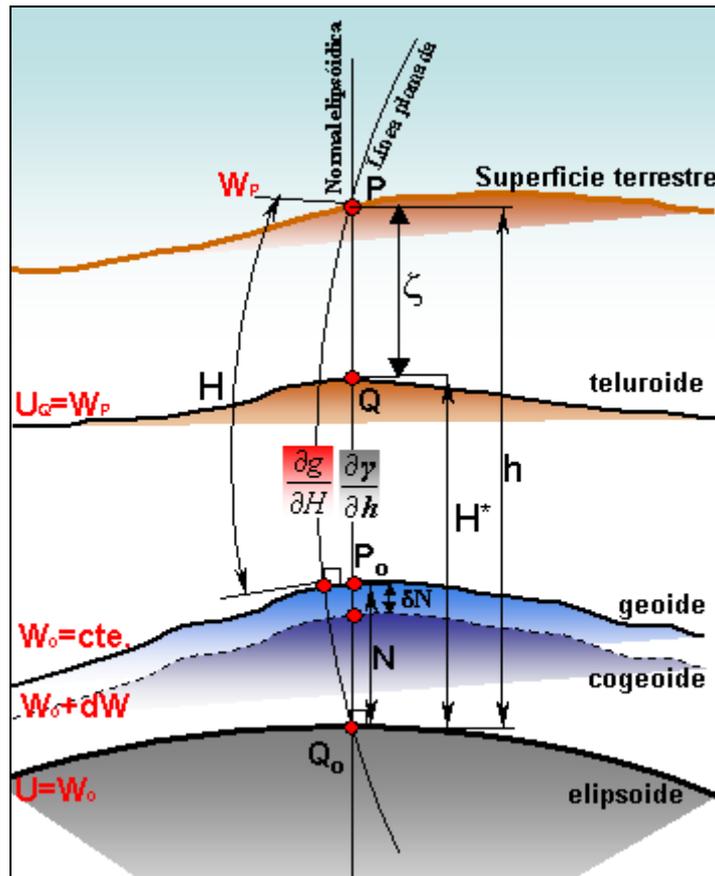


Figura 6. Esquema de superficies consideradas al efectuar las reducciones de la gravedad.

Mientras que el geoide, el cogeide son superficies equipotenciales o de nivel, el cuasigeoide, teluroide y la superficie terrestre no lo son. El cuasigeoide, que no tiene sentido físico, y al igual que el teluroide refleja la topografía de la superficie terrestre. La anomalía de altitud, altitud normal, altitud elipsoidal, gravedad normal media, gravedad observada reducida media están relacionadas mediante las expresiones:

$$h - H - \zeta = N - \zeta = \frac{\bar{g} - \bar{\gamma}}{\bar{\gamma}} H = H^* - H$$

Bibliografía

- [1] Udías, A. y Mézcua, J. Fundamentos de Geofísica. Editorial Alianza, 1996.
- [2] Torge, W. Gravimetry. Walter de Gruyter. Berlín, 1989.
- [3] Torge, W. Geodesy. Walter de Gruyter. Berlin, 2001.
- [4] Heiskanen, W, Moritz, H. Geodesia Física. IGN. 1985.
- [5] Sevilla, M. Análisis de Observaciones Gravimétricas y Cálculo de Anomalías. 1994. IAG.

**Tema 4. Teorema de Meusnier. Radios principales de curvatura.
Fórmula de Euler. Longitudes de arcos de meridiano y paralelo.**

4.1. Teorema de Meusnier

El Teorema de Meusnier es necesario abordarlo, para su comprensión, desde la teoría de la representación analítica diferencial de superficies.

Dada una superficie S que contiene una curva C en la que se encuentran dos puntos P y Q , tendremos:

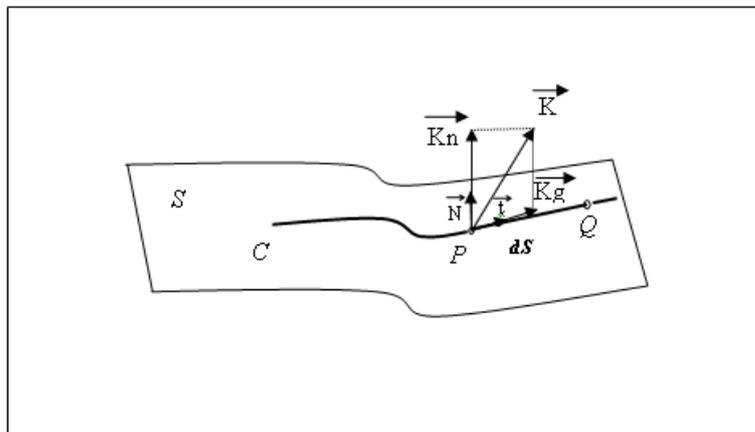


Figura 1.

Toda curva queda determinada sobre una superficie de la forma;

$$\begin{aligned} u &= u(t) \\ v &= v(t) \end{aligned} \quad \text{Respecto a un parámetro } t.$$

Siendo el vector \vec{t} ;

$$X_t = \frac{\partial \vec{x}}{\partial S} = \vec{X}_u \cdot \frac{\partial u}{\partial S} + \vec{X}_v \cdot \frac{\partial v}{\partial S}$$

Tangente a la curva C , y por lo tanto a la superficie.

Por otro lado, tendremos que la distancia entre los dos puntos P y Q , queda determinada por;

$$dS^2 = (X_u \cdot \partial u + X_v \cdot \partial v) \cdot (X_u \cdot \partial u + X_v \cdot \partial v)$$

$$dS^2 = (X_u \cdot X_u) \partial u^2 + 2 \cdot (X_u \cdot X_v) \partial u \partial v + (X_v \cdot X_v) \partial v^2$$

$$dS^2 = E \partial u^2 + 2 \cdot F \partial u \partial v + G \partial v^2$$

E , F y G son funciones de u y v . La distancia entre P y Q será;

$$S = \int_t^{t_0} \sqrt{E \partial u^2 + 2 \cdot F \partial u \partial v + G \partial v^2}$$

En donde $E \partial u^2 + 2 \cdot F \partial u \partial v + G \partial v^2$ es la expresión de la *primera forma fundamental* de la teoría de superficies, y por ser dS una longitud, será siempre positiva.

Si ahora particularizamos para el caso del elipsoide, tendremos:

$$\bar{X}_u = \left(\frac{\partial x}{\partial \varphi}, \frac{\partial y}{\partial \varphi} \right)$$

$$\bar{X}_v = \left(\frac{\partial x}{\partial \lambda}, \frac{\partial y}{\partial \lambda} \right)$$

Entonces;

$$E = \left(\frac{\partial x}{\partial \varphi} \right)^2 + \left(\frac{\partial y}{\partial \varphi} \right)^2$$

$$F = \left(\frac{\partial x}{\partial \varphi} \cdot \frac{\partial x}{\partial \lambda} + \frac{\partial y}{\partial \varphi} \cdot \frac{\partial y}{\partial \lambda} \right)$$

$$G = \left(\frac{\partial x}{\partial \lambda} \right)^2 + \left(\frac{\partial y}{\partial \lambda} \right)^2$$

$$S = \sqrt{E \partial \varphi^2 + 2 \cdot F \partial \varphi \partial \lambda + G \partial \lambda^2}$$

Si los meridianos y paralelos son perpendiculares entre sí, entonces $F = 0$, y la expresión quedará de la forma;

$$S = \sqrt{E \partial \varphi^2 + G \partial \lambda^2}$$

Retomando el vector \vec{t} , tangente unitario a la curva C en un punto P , tendremos que;

$$\vec{K} = \frac{d\vec{t}}{dS}$$

Donde \vec{K} es el vector curvatura de C en P .

Si \vec{N} es el vector unitario de la normal a la superficie en P , el vector \vec{K} se descompondrá en:

$$\vec{K} = \vec{K}_n + \vec{K}_g \quad \text{Donde;}$$

\vec{K}_n ; Componente sobre la normal. Vector curvatura normal. $\vec{K}_n = K_n \cdot \vec{N}$

\vec{K}_g ; Componente sobre la tangente. Vector curvatura geodésica o tangencial.

Se verifica que $\vec{N} \cdot \vec{t} = 0$

Derivando;

$$\frac{d\vec{t}}{dS} \cdot \vec{N} + \vec{t} \cdot \frac{d\vec{N}}{dS} = 0$$

$$\frac{d\vec{t}}{dS} \cdot \vec{N} = -\vec{t} \cdot \frac{d\vec{N}}{dS} = -\frac{d\vec{x}}{dS} \cdot \frac{d\vec{N}}{dS} = -\frac{d\vec{x} \cdot d\vec{N}}{dS^2}$$

$$K_n = -\frac{d\vec{x} \cdot d\vec{N}}{dS^2}$$

Siendo;

$$d\vec{N} = \vec{N}_u \cdot \partial u + \vec{N}_v \cdot \partial v$$

$$d\vec{x} = \vec{x}_u \cdot \partial u + \vec{x}_v \cdot \partial v$$

$$K_n = -\frac{(\vec{X}_u \cdot \vec{N}_u)\partial u^2 + (\vec{X}_u \cdot \vec{N}_v + \vec{X}_v \cdot \vec{N}_u)\partial u\partial v + (\vec{X}_v \cdot \vec{N}_v)\partial v^2}{E\partial u^2 + 2 \cdot F\partial u\partial v + G\partial v^2}$$

La expresión $-(\vec{X}_u \cdot \vec{N}_u)\partial u^2 + (\vec{X}_u \cdot \vec{N}_v + \vec{X}_v \cdot \vec{N}_u)\partial u\partial v + (\vec{X}_v \cdot \vec{N}_v)\partial v^2$ se denomina *segunda forma fundamental*, y se puede expresar de la forma:

$$e\partial u^2 + 2f \cdot \partial u\partial v + g\partial v^2$$

Donde;

$$e = -(\vec{X}_u \cdot \vec{N}_u)$$

$$2f = -(\vec{X}_u \cdot \vec{N}_v + \vec{X}_v \cdot \vec{N}_u)$$

$$g = -(\vec{X}_v \cdot \vec{N}_v)$$

Por lo tanto;

$$K_n = -\frac{e\partial u^2 + 2f \cdot \partial u\partial v + g\partial v^2}{E\partial u^2 + 2 \cdot F\partial u\partial v + G\partial v^2} = \frac{\text{II f f}}{\text{I f f}}$$

De esta última expresión, podemos deducir que:

- e, f, g, E, F, G en un punto son constantes
- K_n sólo depende de u , v y de du , dv .
- “Todas las curvas que pasan por un punto P de la superficie y son tangentes a la misma dirección tienen el mismo vector curvatura normal.”

Y de esta forma queda enunciado el *Teorema de Meusnier*.

Otros autores, enuncian el Teorema de Meusnier de la siguiente forma;

“Si se considera el haz de planos que pasa por la tangente a la superficie en un dirección no asintótica, los círculos osciladores de sus intersecciones con la superficie, están sobre una esfera”

En donde definiremos;

- *Plano oscilador*: Plano determinado por la recta tangente en un punto de la curva de una superficie y de un punto próximo.

- *Dirección no asintótica*: Aquellas direcciones en las que $K_n \neq 0$

Siendo K la curvatura, y $R = 1/K$ el radio de curvatura.

Si $K_n \neq 0 \Rightarrow K \neq 0$

Para que esto ocurra; $0 \leq \varphi \leq \Pi/2$ siendo φ el ángulo que forman \vec{N} (normal a la superficie) y \vec{n} (normal principal de la curva)

$$\boxed{\vec{K} \cdot \cos \varphi = \vec{K}_n} \quad \text{Teorema de Meusnier}$$

Este teorema no se cumple en las direcciones asintóticas, puesto que la segunda forma fundamental K_n se anula; $K_n = e\partial u^2 + 2f \cdot \partial u \partial v + g\partial v^2 = 0$

4.2. Radios principales de curvatura

Como hemos visto, si tomamos la inversa de la curvatura K en un punto P , obtenemos el “radio de curvatura” de la curva en ese punto, o bien “radio del círculo oscilador” correspondiente a ese punto P .

En general, se llama *radio principal de curvatura* en un punto A de una superficie, *al correspondiente a la sección producida por un plano normal a la misma, tal que el radio de curvatura correspondiente sea el máximo o el mínimo entre todos los posibles.*

Normalmente, en una superficie habría 2 secciones principales. Todas las demás producidas por planos que pasen por la normal en el punto A , tendrán radios de curvatura comprendidos entre ambos.

Concretándonos al elipsoide, se llaman secciones principales a:

- Una *elipse meridiana*; cuya curvatura es máxima, y por tanto, su radio de curvatura ρ es mínimo.
- Otra producida por un plano *que contuviera a la normal en el punto A, y que fuese perpendicular al plano meridiano*, cuya curvatura es mínima y con radio de curvatura máximo N .

Por tanto, y según las dos secciones principales tendremos:

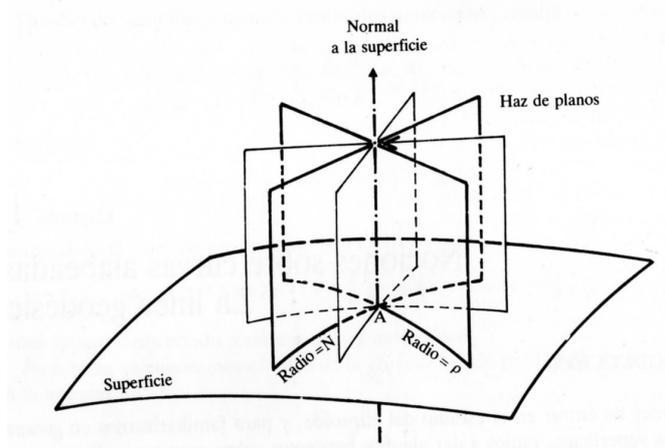


Figura 2. Fuente “Geodesia y Cartografía Matemática”

4.2.1. Normal principal, o gran normal

Observando la figura 3 vemos que, la gran normal, o normal principal es el segmento comprendido entre el punto M considerado y la intersección de la normal en él con el eje menor de la elipse meridiana. Este segmento se representa por la letra N .

Para su cálculo se considera la elipse meridiana. Partiendo de la ecuación de la elipse:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

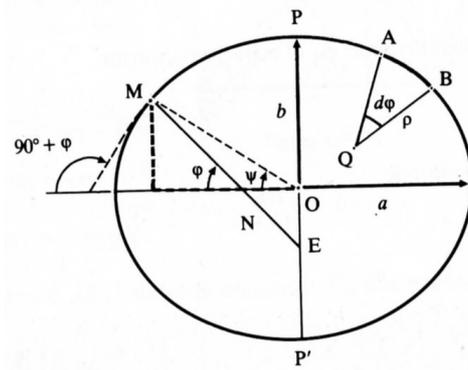


Figura 3. Fuente "Geodesia y Cartografía Matemática"

El elipsoide se genera al girar la elipse alrededor del "eje del mundo". Diferenciando tendremos;

$$2x/a^2 + 2y \cdot y'/b^2 = 0$$

$$y' = -b^2/a^2 \cdot x/y = \text{tg}(\varphi + 90^\circ) = -\text{ctg}\varphi$$

$$\text{tg}\varphi = -a^2/b^2 \cdot y/x$$

Considerando la excentricidad;

$$e^2 = \frac{a^2 - b^2}{a^2} = 1 - b^2/a^2 = 1 - \alpha$$

$$1 - e^2 = b^2/a^2 \Rightarrow 1 - e^2 = b^2/a^2 \Rightarrow b = a \cdot \sqrt{1 - e^2}$$

Sustituyendo en $\text{tg}\varphi$;

$$\text{tg}\varphi = \frac{a^2}{a^2 \cdot (1 - e^2)} \cdot y/x \Rightarrow \boxed{\text{tg}\varphi = \frac{1}{1 - e^2} \cdot y/x} \quad (1) \Rightarrow \boxed{y = x \cdot (1 - e^2) \cdot \text{tg}\varphi} \quad (2)$$

Sustituyendo b e y en la ecuación de la elipse;

$$\begin{aligned} \frac{x^2}{a^2} + \frac{x^2 \cdot (1-e^2)^2 \cdot \operatorname{tg}^2 \varphi}{a^2 \cdot (1-e^2)} &= 1 \Rightarrow x^2 + x^2 \cdot (1-e^2)^2 \cdot \operatorname{tg}^2 \varphi = a^2 \Rightarrow x^2 [1 + (1-e^2) \cdot \operatorname{tg}^2 \varphi] = a^2 \\ x^2 &= \frac{a^2}{1 + (1-e^2) \cdot \operatorname{tg}^2 \varphi} = \frac{a^2}{1 + (1-e^2) \cdot \frac{\operatorname{sen}^2 \varphi}{\operatorname{cos}^2 \varphi}} = \frac{a^2 \cdot \operatorname{cos}^2 \varphi}{\operatorname{cos}^2 \varphi + (1-e^2) \operatorname{sen}^2 \varphi} \\ x^2 &= \frac{a^2 \cdot \operatorname{cos}^2 \varphi}{\operatorname{cos}^2 \varphi + \operatorname{sen}^2 \varphi - e^2 \operatorname{sen}^2 \varphi} = \frac{a^2 \cdot \operatorname{cos}^2 \varphi}{1 - e^2 \operatorname{sen}^2 \varphi} \Rightarrow \boxed{x = \frac{a \cdot \operatorname{cos} \varphi}{\sqrt{1 - e^2 \operatorname{sen}^2 \varphi}}} \quad (3) \\ \Rightarrow \frac{x}{\operatorname{cos} \varphi} &= N \end{aligned}$$

Luego;

$$\boxed{N = \frac{a}{\sqrt{1 - e^2 \operatorname{sen}^2 \varphi}}}$$

4.2.2. Radio de curvatura de la elipse meridiana

Observando de nuevo, la fig. 3, sea una curva plana situada sobre el elipsoide. Se define la curvatura a partir de las normales a dos puntos próximos, A y B . Ambas normales se cortarán en un punto Q . Cuando B , tiende a A , es decir cuando $\widehat{AB} = dS$ tiende a cero, el punto hacia el cual tiende Q , es el llamado centro de curvatura en el punto A .

Por tanto, el radio de curvatura de la curva ρ quedará definido por el cociente $dS/d\varphi = \rho$, siendo $d\varphi$ el ángulo que forman las normales. Al círculo, cuyo radio es ρ , se le llama círculo osculador. Y como ya hemos visto, a la inversa del radio de curvatura, $d\varphi/dS$ se le llama curvatura de la curva en el punto. Este cociente nos demuestra que entre dos curvas con el mismo arco dS , tendrá mas curvatura (estará mas curvada) la que tenga mayor $d\varphi$

Ciñéndonos de nuevo al elipsoide, tendremos que si la curva considerada es la elipse meridiana, o intersección del elipsoide por un plano que pasa por los polos, el límite del cociente

$\lim_{\Delta\varphi \rightarrow 0} \frac{\Delta S}{\Delta\varphi}$ es a lo que se llama radio de curvatura de la elipse meridiana, en el punto considerado, y se representa con la letra ρ .

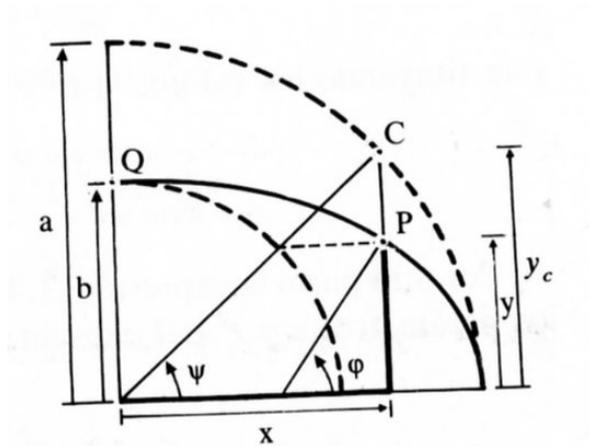


Figura 4. Fuente "Geodesia y Cartografía Matemática"

Para su cálculo, consideremos el círculo principal de radio a (fig 1.) de donde deducimos que:

$$\frac{y_c}{y} = \frac{a}{b}$$

$$\frac{a \cdot \text{sen } \psi}{y} = \frac{a}{b}$$

$$x = a \cdot \text{cos } \psi \quad (4)$$

$$y = b \cdot \text{sen } \psi \quad (5)$$

Diferenciando las dos últimas expresiones tendremos;

$$dx = -a \cdot \text{sen } \psi \cdot d\psi$$

$$dy = b \cdot \text{cos } \psi \cdot d\psi$$

Como hemos visto, por definición;

$$\rho = dS/d\varphi$$

Siendo;

$$ds = \sqrt{dx^2 + dy^2} = \sqrt{a^2 \cdot \text{sen}^2 \psi + b^2 \cdot \text{cos}^2 \psi} d\psi \quad (6)$$

Por otra parte la expresión (3) proporciona un valor de x en función de φ y de los parámetros del elipsoide a y e^2 . De esta expresión y de la establecida en (2) tendremos que:

$$y = \frac{a \cdot (1 - e^2) \text{sen } \varphi}{\sqrt{1 - e^2 \text{sen}^2 \varphi}}$$

Si empleamos la notación;

$$W^2 = 1 - e^2 \operatorname{sen}^2 \varphi$$

Entonces; $x = \frac{a \cdot \cos \varphi}{W}$ (7)

$$y = \frac{a \cdot (1 - e^2) \operatorname{sen} \varphi}{W} = \frac{b^2}{a} \cdot \frac{\operatorname{sen} \varphi}{W} \quad (8)$$

Por otra parte, sustituyendo (4) y (5) en (1) tendremos que;

$$\operatorname{tg} \varphi = \frac{1}{1 - e^2} \cdot \frac{y}{x} = \frac{1}{1 - e^2} \cdot \frac{b}{a} \cdot \operatorname{tg} \psi$$

Como; $\frac{1}{1 - e^2} = \frac{a^2}{b^2} \Rightarrow \operatorname{tg} \varphi = \frac{a}{b} \cdot \operatorname{tg} \psi$

Diferenciando;

$$\frac{d\varphi}{\cos^2 \varphi} = \frac{a}{b} \cdot \frac{d\psi}{\cos^2 \psi} \Rightarrow d\psi = \frac{b}{a} \cdot \frac{\cos^2 \psi}{\cos^2 \varphi} d\varphi \quad (9)$$

Por otro lado, igualando convenientemente (4)=(7) y (5)=(8) tendremos:

$$x = a \cdot \cos \psi = \frac{a \cdot \cos \varphi}{W} \Rightarrow \cos \psi = \frac{\cos \varphi}{W} \quad (10)$$

$$y = b \cdot \operatorname{sen} \psi = \frac{b^2}{a} \cdot \frac{\operatorname{sen} \varphi}{W} \Rightarrow \operatorname{sen} \psi = \frac{b}{a} \cdot \frac{\operatorname{sen} \varphi}{W}$$

Y sustituyendo estos valores en (6) resulta:

$$ds = \sqrt{a^2 \cdot \frac{b^2}{a^2} \cdot \frac{\operatorname{sen}^2 \varphi}{W^2} + b^2 \cdot \frac{\cos^2 \varphi}{W^2}} d\psi$$

Ayudándonos de (9) obtendremos:

$$ds = \sqrt{a^2 \cdot \frac{b^2}{a^2} \cdot \frac{\operatorname{sen}^2 \varphi}{W^2} + b^2 \cdot \frac{\cos^2 \varphi}{W^2}} \cdot \frac{b}{a} \cdot \frac{\cos^2 \psi}{\cos^2 \varphi} d\varphi$$

Y de (10) tendremos que:

$$ds = \sqrt{\frac{b^2}{W^2} \cdot \frac{b}{a} \cdot \frac{1}{W^2}} d\varphi$$

Y llevando este valor a la expresión de:

$$\rho = \frac{ds}{d\varphi} = \frac{b^2}{a} \cdot \frac{1}{W^3} = \frac{a^2(1-e^2)}{a} \cdot \frac{1}{W^3}$$

Se obtiene finalmente:

$$\rho = \frac{a(1-e^2)}{(1-e^2 \operatorname{sen}^2 \varphi)^{3/2}}$$

Por tanto, tanto N como ρ dependen de φ .

Estas dos curvaturas en el elipsoide, corresponden a direcciones ortogonales en el elipsoide.

Al tratar los radios principales de curvatura, conviene realizar una serie de consideraciones:

- Las líneas de curvatura de la superficie forman dos sistemas de curvas ortogonales que cubren la superficie una solo vez, salvo en aquellos puntos en los que $N = \rho$, que reciben el nombre de *puntos umbilicales*. Así, en una esfera todos los puntos son umbilicales, puesto que el radio es único. En un elipsoide, los puntos umbilicales serían los puntos de tangencia de la circunferencia inscrita en él.
- Por otro lado, conocidos los radios de curvatura principales en un punto, se define como *curvatura media* a la expresión;

$$K_m = 1/R_m = 1/\sqrt{N\rho}$$

- El radio de curvatura media será:

$$R_m = \sqrt{N\rho}$$

- Por su parte, la *curvatura total*, será;

$$K_T = 1/R_T = 1/N\rho$$

- Si en un punto A del elipsoide conocemos el acimut de una sección normal z , el radio de curvatura correspondiente a esa sección, queda determinado por;

$$1/R_z = \frac{\cos^2 z}{\rho} + \frac{\operatorname{sen}^2 z}{N}$$

Este sería el denominado *Teorema de Euler*. Así;

- Para $z = 0^\circ$, resultará que $R_z = \rho$ radio de curvatura del meridiano.
- Para $z = 90^\circ$, resultará que $R_z = N$ radio de curvatura del primer vertical.

4.3. Longitudes de arco de meridiano y de paralelo

Si tenemos en cuenta que el radio del paralelo que pasa por un punto;

- En la esfera sería: $r_p = R \cdot \cos \varphi$
- En el elipsoide sería: $r_p = N \cdot \cos \varphi$

Por otro lado, el arco de paralelo dS_p queda determinado;

- En la esfera como: $dS_p = R \cdot \cos \varphi \cdot d\lambda$
- En el elipsoide como: $dS_p = N \cdot \cos \varphi \cdot d\lambda$

Análogamente, el arco de meridiano dS_m queda determinado;

- En la esfera como: $dS_m = R \cdot d\varphi$
- En el elipsoide como: $dS_m = \rho \cdot d\varphi$

Si ahora, lo que pretendemos es obtener las longitudes de los arcos de meridiano y de paralelo sobre el elipsoide, tendremos;

4.3.1. Longitud de arco de paralelo

Partiendo de la expresión del arco de paralelo expresada anteriormente;

$$dS_p = N \cdot \cos \varphi \cdot d\lambda = r_p \cdot d\lambda$$

$$S_p = \int_{\lambda_0}^{\lambda_1} r_p \cdot d\lambda = \int_{\lambda_0}^{\lambda_1} r_p \cdot (\lambda_1 - \lambda_0) \quad \text{Expresando } (\lambda_1 - \lambda_0) \text{ en radianes.}$$

$$\boxed{S_p = r_p \cdot [\lambda_1 - \lambda_0]}$$

4.3.2. Longitud de arco de meridiano

Partiendo ahora de la expresión del arco de meridiano, tendremos:

$$dS_m = \rho \cdot d\varphi$$

$$S_m = \int_{\varphi_0}^{\varphi_1} \rho \cdot d\varphi = \frac{a \cdot (1 - e^2)}{\sqrt{(1 - e^2 \sin^2 \varphi)^3}} d\varphi = a \cdot (1 - e^2) \int_{\varphi_0}^{\varphi_1} \frac{d\varphi}{W^3}$$

La integral $\int_{\varphi_0}^{\varphi_1} \frac{d\varphi}{W^3}$ es una integral elíptica, que no puede calcularse por medio de funciones elementales. Sólo puede obtenerse con una aproximación dada, integrando una serie:

$\frac{1}{W^3} = (1 - e^2 \operatorname{sen}^2 \varphi)^{-3/2}$ Y desarrollando en una serie binómica:

$$\begin{aligned} (1 - e^2 \operatorname{sen}^2 \varphi)^{-3/2} &= 1 - \binom{-3/2}{1} \cdot e^2 \operatorname{sen}^2 \varphi + \binom{-3/2}{2} \cdot e^4 \operatorname{sen}^4 \varphi - \binom{-3/2}{3} \cdot e^6 \operatorname{sen}^6 \varphi + \\ &+ \binom{-3/2}{4} \cdot e^8 \operatorname{sen}^8 \varphi - \binom{-3/2}{5} \cdot e^{10} \operatorname{sen}^{10} \varphi + \dots \\ &= 1 + \frac{3}{2} \cdot e^2 \operatorname{sen}^2 \varphi + \frac{15}{8} e^4 \operatorname{sen}^4 \varphi + \frac{35}{16} \cdot e^6 \operatorname{sen}^6 \varphi + \frac{315}{128} e^8 \operatorname{sen}^8 \varphi + \frac{699}{256} \cdot e^{10} \operatorname{sen}^{10} \varphi + \dots \end{aligned}$$

Teniendo en cuenta:

$$\begin{aligned} \operatorname{sen}^2 \varphi &= \frac{1}{2} - \frac{1}{2} \cos 2\varphi \\ \operatorname{sen}^4 \varphi &= \frac{(3 - 4 \cos 2\varphi + \cos 4\varphi)}{8} \\ \operatorname{sen}^6 \varphi &= \frac{(10 - 15 \cos 2\varphi + 6 \cos 4\varphi - \cos 6\varphi)}{32} \end{aligned}$$

Sustituyendo y reordenando:

$$\begin{aligned} \frac{1}{W^3} &= (1 + \frac{3}{4} e^2 + \frac{45}{64} e^4 + \frac{350}{512} e^6 + \dots) - (\frac{3}{4} e^2 + \frac{15}{16} e^4 + \frac{525}{512} e^6 + \dots) \cos 2\varphi + \\ &+ (\frac{15}{64} e^4 + \frac{210}{512} e^6 + \dots) \cos 4\varphi - (\frac{35}{512} e^6) \cos 6\varphi + \dots \end{aligned}$$

$$\frac{1}{W^3} = A - B \cos 2\varphi + C \cos 4\varphi - D \cos 6\varphi \quad \text{Siendo:}$$

$$A = 1 + \frac{3}{4} e^2 + \frac{45}{64} e^4 + \frac{350}{512} e^6 + \dots$$

$$B = \frac{3}{4} e^2 + \frac{15}{16} e^4 + \frac{525}{512} e^6 + \dots$$

$$C = \frac{15}{64} e^4 + \frac{210}{512} e^6 + \dots$$

$$D = \frac{35}{512} e^6$$

Integrando término a término, tendremos que:

$$S_m = a(1 - e^2) \cdot \left\{ A(\varphi - \varphi_0) - \frac{B}{2} (\operatorname{sen} 2\varphi - \operatorname{sen} 2\varphi_0) + \frac{C}{4} (\operatorname{sen} 4\varphi - \operatorname{sen} 4\varphi_0) - \frac{D}{6} (\operatorname{sen} 6\varphi - \operatorname{sen} 6\varphi_0) + \dots \right\}$$

Cuando $\varphi_0 = 0$, se obtiene la distancia desde el punto de latitud φ , a lo largo de la elipse meridiana.

Por otro lado, la expresión del arco de meridiano, pone de manifiesto, que la longitud de este arco es función de los parámetros del elipsoide de referencia considerado, así como de las latitudes geodésicas de los extremos.

Además, supuestos calculadas tanto φ_0 , como φ , como S_m podría llegar a determinarse mediante *mm.cc.* los parámetros del elipsoide a y e , siempre y cuando se tuvieran un conjunto superabundante de medidas de latitudes y arcos. Este ha sido el proceso seguido para determinar los diversos elipsoides de referencia.

Bibliografía

[1] Martín Asín, F., “Geodesia y Cartografía Matemática”, Editorial Paraninfo, pág. 15 a 23

[2] Calero, E. “Apuntes de Geodesia de la E.P.E.S.”

Tema 5. Redes de nivelación de precisión. Señalización. Instrumentos, métodos de observación, comprobaciones, correcciones. Errores sistemáticos. Cálculo y compensación de redes de nivelación. La red de nivelación de alta precisión española: REDNAP.

5.1. Redes de nivelación de precisión

En la XVII Conferencia de la Asociación Geodésica Internacional (Hamburgo 1912) se acordó clasificar como “Nivelación de Alta Precisión” a *toda línea o red de líneas, niveladas dos veces en sentidos opuestos, con un operador en cada sentido y en fechas distintas, y cuyo error accidental sea $e \leq 1,5mm\sqrt{K}$, siendo K el número de Km. de la línea.*

Las aplicaciones de la NAP pueden dividirse en dos grandes familias: redes grandes, normalmente oficiales (redes nacionales), y pequeñas redes para aplicaciones concretas, generalmente de control de deformaciones (presas, viaductos, túneles...)

En todos los países desarrollados hay una red oficial de NAP que es la base altimétrica de la cartografía nacional. Se comienza estableciendo una superficie fundamental como plano de comparación con altitud cero. En España está dado por el nivel medio del mar en Alicante. A partir de esta superficie fundamental, midiendo desniveles se establece en todo el territorio la altitud de una serie de señales fijas y numeradas que constituyen la Red.

Las pequeñas redes son las que se emplean en trabajos de control o industriales, como en presas, en estudios geodinámicos, montajes de estructuras o maquinaria, etc.

Obtener toda la precisión que pueden alcanzar los equipos utilizados en estas observaciones exige el cumplimiento estricto de una serie de normas y el fiel seguimiento de una desarrollada metodología.

El método utilizado para la nivelación de alta precisión (N.A.P) tiene su fundamento en el empleo de la nivelación geométrica o por alturas.

5.2. Señalización

5.2.1. Trabajos previos

Previamente a la señalización de una red NAP, es necesario llevar a cabo una serie de labores tales como;

Recopilación de la documentación necesaria y preparación del material

Hojas de mapas (MTN25, MTN50, etc.), anteproyecto de itinerario, reseñas, credenciales, pliegos de prescripciones técnicas, martillo perforador, señales de nivelación, chalecos reflectantes, cámaras fotográficas, etc.

Reconocimiento de posibles itinerarios

Sobre la cartografía, se realiza en gabinete un anteproyecto del trazado de la línea de nivelación, que siempre será susceptible de mejora en campo. Por ello será necesario recorrer aquellas vías de comunicación próximas al trazado planteado y que puedan servir de alternativas.

Además, deben tenerse en cuenta los siguientes condicionantes:

- Debe de ser un trazado llano.
- Evitar la circulación abundante de vehículos.
- También debe evitarse el paso por los núcleos de población (que quedarán enlazados mediante ramales secundarios)
- No son aconsejables las vías de comunicación que presenten grandes variaciones climatológicas (frecuentes pasos de solanas a umbrías, de arbolado a raso).
- Investigar si están previstos cambios en los sitios en los que se proyecte colocar las señales, como nuevas edificaciones importantes, modificaciones o ampliaciones de trazados de viales, grandes obras, etc.

5.2.2. Determinación del trazado definitivo

Tipos de señales

Las señales usadas en NAP se fijan de forma permanente en perforaciones realizadas en sitios consolidados, como en suelo firme, rocas nativas o partes estables de estructuras. Con ello las señales quedan situadas en los lugares que presenten mayores garantías de inamovilidad y permanencia.

Las señales, o clavos de nivelación, son metálicas y siempre presentan una superficie semiesférica hacia arriba, donde se puede apoyar el talón de la mira. El punto sobre el que se mide el desnivel es el de tangencia superior.

En el suelo, o en una superficie horizontal, el clavo se coloca vertical, por lo que se llama “señal vertical”. En una superficie vertical, como la pared de una edificación, el clavo se coloca horizontal, asomando su cabeza de forma que sobre ella se pueda colocar la mira, por lo que se llama “señal horizontal”.

Aunque pueden existir señales específicas para determinados trabajos especiales, generalmente (y normativamente en redes oficiales) las señales o clavos de nivelación son de dos clases:

- Señal principal (SP): Situadas en grupos de dos o tres. Tienen forma de casquete esférico.
- Señal secundaria (CS): Son más pequeñas y de distinto metal.

Colocación de la señalización y distribución de las señales

El proyecto puede constar de una o varias líneas, que pueden ser abiertas o cerradas, formando anillo, o ser una red o malla, con líneas uniendo los nudos. La línea está compuesta de un cierto número de tramos. Se llama tramo al espacio existente entre dos señales, que pueden nivelarse sin interrupción. En líneas oficiales los tramos se separan con señales secundarias. Entre estaciones secundarias hay una distancia aproximada de un kilómetro. En las ciudades y pueblos se colocan en los edificios principales.

Para colocar las señales, se hace un orificio, con ayuda del martillo perforador, quedando al descubierto solamente la parte esférica. La señal se fija en el orificio por medio de mortero de cemento de arena fina. Entre la señalización y la observación es necesario esperar más de tres días para que puedan consolidarse debidamente las señales.

La distribución teórica sería:+

- Un clavo secundario -CS- cada 800 m (unos 1000 m en zonas llanas y 600 m en alta montaña)
- Una señal principal -SP- (compuesta por la agrupación de, al menos, 2 clavos primarios -CP-) cada 4 o 5 km
- Como máximo a 100m del eje de la vía de comunicación, sobre rocas nativas o edificios de construcción sólida
- Cuando se señalice un grupo, ha de tenerse en cuenta que es deseable poder hacer la nivelación entre dos señales cualquiera de las que lo compongan con una sola estación de nivel, lo que limita tanto el desnivel como la distancia existente entre señales.
- Queda terminantemente prohibido poner señales principales en alcantarillas y puentes de construcción moderna.

Confección de las reseñas

Una vez colocada la señal, se hace la reseña correspondiente, tanto literal como gráfica, acompañada de fotografías de su emplazamiento.

5.3. Instrumentos, métodos de observación, comprobaciones, correcciones

5.3.1. Instrumentos

En general, el equipamiento que cada operador (con un mínimo de cuatro peones) debe llevar consigo será:

- 1 nivel automático de alta precisión con su micrómetro, o un nivel digital.
- Trípode.
- 2 miras invar de 3 metros con sus planchas y puntales.
- 1 cinta métrica.
- 1 nivel esférico para mira, de repuesto.
- 1 anilla de mira (o pieza especial de giro), de repuesto.
- 1 esfera.
- Señales de tráfico, chalecos reflectantes y cascos (1 por persona).
- Cuadernillos de observación y fichas de cálculo.
- Vehículo adecuado para transporte de personal y equipamiento a la zona de trabajo.

Las características que debe tener un nivel para alta precisión:

- Estabilización de la línea de colimación del orden de $\pm 0,1''$.
- Error medio por kilómetro en doble nivelación con placa plano-paralela: de $\pm 0,4$ a $\pm 0,2$ mm.
- Aumento del antejo: de 30A a 50A. y distancia mínima de enfoque: 1,5m.
- Campo visual del antejo: de 12,8 a 1,00
- Exactitud de estabilidad del compensador, para los automáticos: 0,5''
- Estos niveles llevan además un micrómetro de placa plano-paralelas con lectura óptica de la desviación, permitiendo leer directamente la décima de milímetro y a estima la centésima de milímetro. El micrómetro es, normalmente, de 10mm de barrido.
- El retículo está formado por un doble hilo en forma de cuña

Las características de las miras serán:

- Miras de 3 metros de longitud y 5,5 Kg. de peso.
- Las miras llevan dos escalas centimétricas de invar, desplazadas una respecto de la otra y de numeración distintas. De este modo se obtienen dos lecturas independientes entre sí, cuya diferencia debe ser constante. Esta constante sirve de control para eliminar errores de lectura groseros.

5.3.2. Métodos de observación

Señalización del itinerario

Previamente a la nivelación, debe marcarse el itinerario a nivelar, para lo cual, con ayuda de una cinta métrica o cuerda de nylon, se señalan, con una cruz, los puntos de estación del aparato y, con un trazo, las estaciones de mira, ajustándose a las siguientes condiciones:

- Las distancias aparato-mira no sobrepasarán nunca los 25 metros, admitiéndose, en cada estación de *nivel automático (lectura por el operador)*, una diferencia de ± 0.2 m entre las distancias a la mira de *frente* y a la de *espalda*. Esta tolerancia puede ampliarse hasta 0,50 m. siempre que, en el total del tramo, la diferencia entre las sumas de las distancias de espalda y frente sea inferior a $\pm 0,50$ m. Para los *niveles digitales*, la tolerancia en la diferencia de distancias entre las sumas de las niveladas de frente y espalda será de 0.50 m. en el total del tramo, siendo este mismo error el admitido para cada estación.
- Queda prohibido leer en la mira por debajo de 0,50 m. y por encima de 2,50 m., salvo que la distancia nivel-mira sea inferior a 10 m, en cuyo caso los límites serán 0,20 m. y 2,80 m., respectivamente.
- La señalización debe efectuarse por el lado de la vía de comunicación que presente mayor uniformidad, tanto climática como topográfica, y más adecuado firme. Cuando sea necesario atravesar la calzada, se hará por una zona uniforme y bajo las mismas condiciones para las visuales a las dos miras.
- El número de estaciones de aparato entre señales fijas debe ser par, a fin de empezar y terminar el tramo sobre la misma mira, eliminando así el posible error de talón de mira.

Observación

Una vez marcada la línea, se procede a la observación de la misma, para lo cual se divide en *tramos* de longitud igual al trabajo estimado como de posible realización diaria (entre 1 y 2 km según el terreno y el operador), delimitados por señales fijas, haciéndose la nivelación del tramo en sentido contrario por cada operador (ida y vuelta) con el fin de eliminar los errores sistemáticos ligados al sentido de la marcha.

Salvo circunstancias excepcionales como pueden ser fuertes vientos y lluvias imprevistas, queda prohibido el empleo de *esferas* para marcar la finalización del trabajo diario.

Para los *niveles automáticos*, las lecturas se hacen en el siguiente orden, siendo *A* la mira de espalda y *B* la de frente:

1. - Oscilación del compensador
2. - Mira *A*, hilos estadimétricos sobre la escala I (escala I de 3 a 6 m.)

3. - Mira *A*, hilo nivelador, escala I
4. - Mira *B*, hilo nivelador, escala I
5. - Mira *B*, hilos estadimétricos sobre la escala I
6. - Mira *B*, hilo nivelador, escala II (escala II de 0 a 3 m.)
7. - Mira *A*, hilo nivelador, escala II
8. - Oscilación del compensador

Anotándose todas ellas en el cuaderno o en el colector de datos, apreciándose hasta 0.01 mm, si bien la última cifra deberá ser par. Las lecturas deberán repetirse siempre que la diferencia entre las escalas I y II sea superior en 0.1 mm a la constante de la mira; que el promedio de los estadimétricos discrepe en más de 1.0 mm de la lectura de la escala I; y que los desniveles entre las miras *A* y *B*, obtenidos sobre la escala I y la escala II, discrepen en más de 0.4 mm.

Para los *niveles digitales*, el programa operativo será:

1. - lectura mira *A*
2. - lectura mira *B*
3. - lectura mira *B*
4. - lectura mira *A*

En cada lectura, el número de medidas realizadas por el aparato no será inferior a tres. La tolerancia por estación (diferencia entre las dos medidas efectuadas), a introducir en el programa de lectura, será de 0.4 mm.

Además, deberán tenerse presentes las siguientes consideraciones generales:

- La observación de una estación debe durar el menor tiempo posible. Por tanto, si por cualquier motivo justificado fuere necesario efectuar una interrupción, resultaría imprescindible comenzar de nuevo la observación completa de la estación, pese a que esta precaución no evitaría el hundimiento de la mira que permanece atrás.
- Las observaciones deben comenzar con una altura del Sol de unos 20° y ser interrumpidas cuando comiencen las oscilaciones en el campo del anteojo que puedan impedir la obtención de las tolerancias fijadas.
- Si en el instante de la observación se aprecian pequeñas vibraciones en el aparato motivadas por agentes externos, podría ser posible eliminar su efecto tocando con la mano de forma delicada una de las patas del trípode, con lo que el cuerpo del operador absorbe las citadas vibraciones.
- Los portamiras no retirarán las miras de su emplazamiento hasta que el observador haya comprobado la aceptable calidad de las lecturas efectuadas.

- Como los hilos del compensador pueden no ser antimagnéticos, es necesario tener una gran precaución al atravesar campos magnéticos, ya que pueden generarse fuertes errores. Es conveniente no estacionar a menos de 10 metros de un posible campo magnético.
- El giro de 180° que debe sufrir la mira de frente al cambiar de estación el nivel debe ser realizado por el propio jefe de equipo, sujetando la mira por la parte central con gran precaución y suavidad para evitar cualquier desplazamiento en su base.
- Si en el transcurso de las observaciones en una estación de nivel, se produce alguna anomalía (ligeros golpes a miras o aparatos, asentamiento imprevisto de las planchas, etc.), es necesario comenzar de nuevo el trabajo a partir de la última señal permanente NAP.
- Es muy importante el anotar en el cuaderno de campo todas aquellas incidencias que se aprecien al comenzar, en el transcurso del trabajo y al finalizar, tanto meteorológicas como accidentales, tales como día, hora, tiempo soleado o nublado, viento en calma o suave, etc. Estas anotaciones se efectuarán con bolígrafo, al igual que el croquis detallado que ha de figurar en la contraportada.

5.3.3. Comprobaciones y correcciones

Antes de iniciar los trabajos de la campaña, así como durante el transcurso de los mismos, es necesario llevar a cabo las siguientes comprobaciones y correcciones.

Nivel esférico

Normalmente, debe comprobarse una vez por semana y, desde luego, siempre que se aprecien anomalías en el desarrollo de las operaciones, procediendo, si es necesario, a su inmediata corrección. Para ello, se estaciona el nivel en un terreno bien firme y al abrigo de perturbaciones externas (vientos, tráfico, etc.); se cala la burbuja con lentitud actuando sobre los tornillos nivelantes; se gira el antejo 180°, lo que, si el nivel está descorregido, dará lugar a una desviación de la burbuja; se corrige la mitad de esta desviación con los tornillos nivelantes y la otra mitad con los de reglaje del nivel esférico; se torna a girar el antejo, y la burbuja debe quedar en posición correcta. Caso contrario, deberá repetirse la operación completa.

Es preceptivo realizar esta corrección siempre que se aprecie una desviación superior a 0.5 mm en la burbuja entre las visuales de frente y espalda.

Horizontalidad de la visual

Esta comprobación debe verificarse una vez por semana y siempre que se observen irregularidades en la nivelación, una vez finalizado el tramo que se está nivelando.

Para efectuar esta comprobación, es conveniente disponer en la zona de trabajo de una base estaquillada en un terreno llano.

Para los *niveles automáticos*, se marcan 3 puntos, A, B, e I_2 , aproximadamente alineados y dispuestos de forma que $AB = BI_2 \cong 30$ m; se estaciona el aparato en un punto I_1 , aproximadamente alineado con los anteriores y tal que $20 < I_1A < 30$ m.; las dos miras se colocan en A y B, como se indica en la figura; se realizan las lecturas a_1 y b_1 en las dos escalas de las miras; a continuación, se estaciona el aparato en I_2 y se hacen las lecturas a_2 y b_2 , anotándolas en el estadillo siguiente; se calcula la lectura $a_2 = a_1 - b_1 + b_2$, que correspondería si el aparato no estuviese descorregido; si este a_2 calculado difiere del valor a_2 leído en más de 1 mm, es necesario efectuar una corrección, para lo cual el aparato dispone de un dispositivo que permite llevar el hilo del retículo a coincidir con el trazo correspondiente de la mira, poniendo previamente la lectura a_2 en el micrómetro. Una vez efectuada esta corrección, se estaciona en el punto medio de AB y se observa el desnivel entre ambos puntos; a continuación, se estaciona a unos 2 m de A observando de nuevo; si la corrección ha sido la adecuada, este nuevo desnivel debe coincidir con el anterior.

Para los *niveles digitales* se aplicará al instrumento el test de comprobación al completo, debiendo quedar registrado en el correspondiente fichero.

Horizontalidad del retículo

Será necesario comprobar que el retículo no se encuentra torcido, es decir, que el hilo horizontal realmente lo sea, lo que es fácil de confirmar apuntando a un sitio cualquiera y girando lentamente el instrumento alrededor de su hilo horizontal.

Barrido del micrométrico

Visando una mira se pone en el micrómetro el valor 00. Subiendo y bajando el instrumento con los tornillos nivelantes se hace la coincidencia de la cuña con un trazo, asegurándose de que la lectura del micrómetro es 00. Dándole la vuelta al micrómetro se hace coincidencia en el trazo interior de la mira, en cuyo momento el micrómetro deberá marcar 100. La diferencia debe de ser menor que la tolerancia establecida para la constante.

Nivel esférico de la mira

Se corrige como cualquier nivel esférico, teniendo cuidado de centrar la anilla del talón de la mira en el pivote de la plancha, a fin de mantener el mismo eje de giro de la mira. Aun estando

corregido, puede el nivel no estar calado como consecuencia del calentamiento solar. Esta comprobación debe efectuarse semanalmente, al mismo tiempo que las otras antes expuestas.

Comparación de miras

Al ser varias las miras que intervienen en los trabajos, es necesario conocer la longitud de cada una de ellas con exactitud, así como las variaciones de longitud. Estas comprobaciones se deben realizar tanto al salir a campo, como al volver. Si bien las variaciones (de temperatura, humedad....) de las miras invar, son pequeñas. Estas comparaciones, que antiguamente se realizaban con patrones de iridio, hoy en día se llevan a cabo con interferómetros láser.

Ortogonalidad del talón de la mira

Si el rectángulo de apoyo que presenta la mira no es ortogonal a la cinta, o lo que es equivalente, si no es horizontal cuando la mira es vertical (podría hasta no ser plano) las lecturas dependerán del punto en el que se apoye. El uso de la anilla minimiza la repercusión accidental de este error, pero cuando se estacota sobre señales permanentes no puede usarse normalmente la anilla, estacionando a veces en el borde del talón, como, por ejemplo, cuando se estaciona en una señal horizontal.

Trípode

El trípode debe mantenerse en perfecto estado de conservación. Se considera que sus articulaciones están correctas cuando, suspendido en el aire cogido por su plataforma superior y con las patas abiertas, éstas no se cierran, pero pueden moverse con facilidad. Es fundamental vigilar con frecuencia las bridas de ajuste, especialmente ante fuertes cambios meteorológicos.

5.4. Errores sistemáticos

Para la nivelación de alta precisión se utiliza el método del punto medio, de esta manera se eliminan ciertos errores y se puede alcanzar la precisión exigida. Los errores sistemáticos que podrán ser:

5.4.1. Proporcional a la longitud

Estos errores se acumulan en función de l (longitud de nivelada), n (nº de tramos), y t (tiempo). Todos estos errores se manifiestan en los errores de cierre (e_c) de las nivelaciones, y para reducirlos o eliminarlos, se emplea el método del punto medio y la nivelación doble.

Es el caso de los errores de:

- *Esfericidad o curvatura:*

$$\varepsilon = \frac{D^2}{2R} = \frac{100^2 * 10^3}{2 \cdot 6370 * 10^6} \approx 1mm$$

Esto supone que para un tramo de 100m de longitud, el error cometido es de 1mm. Por ello en la nivelación de precisión no se sobrepasan los 50m en la longitud de los tramos.

- *Enfoque del objetivo:* Cuando no se efectúa un enfoque correcto de este.

5.4.2. Proporcional a la altitud

Si son constantes, no se manifiestan en nivelación doble, ni en itinerarios cerrados. Son los errores debidos a:

- *Longitud de la mira.*
- *Falta de verticalidad de la mira.*
- *Refracción desigual en las pendientes:* En cuyo caso, los puntos mas cercanos al suelo, tienen menor refracción.

5.4.3. Proporcional a la latitud

Si son constantes, no se manifiestan en nivelación doble, ni en itinerarios cerrados. Son los errores debidos a:

- *Variación de la iluminación solar:* Que provoca errores al realizar la bisección con el hilo del retículo
- *Itinerarios NS:* Es necesario evitarlos.

5.5. Cálculo y compensación de redes de nivelación

La altitud final dada para un punto que pertenece a una red de nivelación de precisión, se trata de una altitud ortométrica. Para obtenerla es necesario seguir una serie de operaciones a partir de las lecturas tomadas en campo.

5.5.1. Cálculo de desniveles

El desnivel entre los dos extremos de un tramo, puntos extremos vendrá dado por:

$$\Delta h = \sum h_i = \sum a_i - \sum b_i$$

Siendo; a_i el sumatorio de las lecturas de frente.

b_i el sumatorio de las lecturas de espalda.

Error de cierre y tolerancia

En cada lectura se comete un cierto error, que depende de:

- La precisión en la división de la mira.
- Precisión del aparato.
- Error de verticalidad.
- Error de lectura.

Si se considera que en cada nivelada se comete un error μ y que es el mismo para todas las lecturas, se deduce que:

$$m_{\Delta H} = \mu\sqrt{2n}$$

Donde $m_{\Delta H}$, es el error medio cuadrático y n es el número de tramos de la línea de nivel.

Si llamamos L a la longitud total del itinerario entre dos puntos A y B, y l a la distancia constante entre el nivel y la mira, en cada nivelada, la longitud total será igual a:

$$L = 2l \cdot n, \text{ luego: } 2l = L/n$$

Y sustituyendo en la expresión anterior, tenemos:

$$m_{\Delta H} = \mu\sqrt{\frac{L}{n}}$$

Las tolerancias para alta precisión son: $T = 1,5mm \cdot \sqrt{K}$, siendo K la distancia en Kilómetros.

Compensación de una red

En las nivelaciones de precisión no suelen seguirse itinerarios simples, sino formando polígonos cerrados, de forma que se obtendrían cierres iguales a cero, pero debido a los errores, estos cierres no son ceros.

De forma que en cada polígono se cumplirá:

$$h_1 \pm h_2 \pm h_3 \pm h_4 \pm h_4 \dots = \sigma_i$$

Donde las h_i son los desniveles de cada lado del polígono y las σ_i , los errores de cierre para cada polígono.

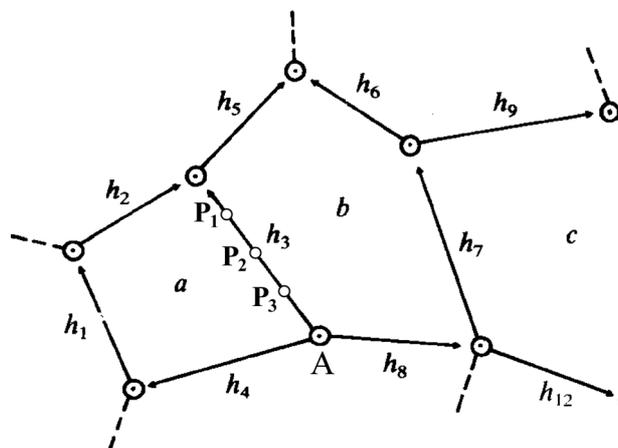


Figura 1.

Para compensar los resultados obtenidos, a cada h_i , habrá que aplicarle una corrección desconocida v_i de manera que para cada polígono se tendrá:

$$(h_1 + v_1) \pm (h_2 + v_2) \pm (h_3 + v_3) \pm (h_4 + v_4) \pm \dots = 0,$$

Restando esta expresión de la anterior queda:

$$v_1 + v_2 + v_3 + v_4 + \dots + v_i = 0 \text{ (una para cada polígono)}$$

Por lo tanto se tendrá un sistema con tantas ecuaciones como polígonos. Esto se resuelve por mínimos cuadrados, obteniendo los valores más probables de las correcciones v_i .

5.5.2. Cálculo de geopotenciales

Para la obtención de las cotas geopotenciales, será necesario realizar una serie de medidas de la gravedad.

Se obtienen medidas de gravedad en todos los nodos (cruces de líneas) y en una serie de señales principales intermedias.

Observando la Fig. 1, siendo A y B, dos nodos consecutivos de una línea de nivelación, tendríamos medidas de gravedad tanto en A, como en B, como en los puntos P_1 , P_2 y P_3 que pertenecen a clavos principales de la red.

El incremento de cota geopotencial entre A y B, vendrá dado por la expresión:

$$C_A^B = W_B - W_A = \sum_B^A g_i \cdot \Delta h_i$$

De esta forma obtendríamos las cotas geopotenciales C_A , C_{P1} , C_{P2} , C_{P3} , C_B .

5.5.3. Compensación de geopotenciales

Obtenidas las cotas geopotenciales de todos los nodos, en primer lugar, se procede a la compensación de geopotenciales de la red de nodos. Los pesos se establecen en función de la distancia, de tal forma que:

$$P_i = \frac{1}{L_i(Km.)}$$

Con ello se obtienen los nuevos valores C_A' , C_B' para los nodos, que nos servirán para recalcular los geopotenciales en los puntos intermedios C_{P1}' , C_{P2}' , C_{P3}' .

Los valores de cotas geopotenciales para los puntos en las señales de nivelación en los que no se ha efectuado medidas de gravedad reales, se obtienen, en función de los desniveles, por promedio respecto a las señales más próximas en las que sí que se han llevado a cabo las medidas.

5.5.4. Cálculo de altitud ortométrica

Una vez se tienen cotas geopotenciales compensadas de todos los puntos de la red, se procede al cálculo de la altitud ortométrica de todos ellos (altitud Helmert H) con ayuda de la fórmula de la reducción de la gravedad de Pointcaré-Prey:

$$H = \frac{C}{\bar{g}}$$

$\bar{g} = g + 0.0424 \cdot H$ Siendo; \bar{g} el valor medio de la gravedad sobre la línea de la plomada.

g la gravedad medida en el terreno.

$$0.0424 \cdot H^2 + g \cdot H - C = 0 \quad C \text{ en u.g.p.}$$

g en gales

H en Km.

Resolviendo esta ecuación de segundo grado, en la que sólo hay un valor válido, obtenemos la altitud ortométrica.

Para aquellos puntos en los que no se han realizado medidas de la gravedad real, hay que estimar su valor utilizando la Anomalía de Bouguer.

$$\Delta g^B = g - \gamma + (3,086 - 4,191 \cdot \rho) \cdot h$$

Que para un valor de $\rho = 2,67 \text{ gr/cm}^3$, nos daría:

$$\Delta g^B = g - \gamma + 1,967 \cdot h$$

5.6. La red de nivelación de alta precisión española: REDNAP

Desde mediados del siglo XIX han tenido lugar en el territorio español sucesivas nivelaciones utilizando en cada una de ellas el más reciente instrumental existente en cada momento. Los datos de altitudes así obtenidos han resultado de gran utilidad científica y técnica, pese a que su precisión diste de la alcanzable en la actualidad.

La REDNAP es competencia del Instituto Geográfico Nacional, al igual que el proyecto, monumentación, observación, cálculo, compensación y mantenimiento de las otras Redes Permanentes: Geodésica, Gravimétrica, Mareográfica y de Estaciones GPS. La REDNAP ha sido la más dañada de todas ellas, como consecuencia de la desaparición de un gran número de señales por sucesivas reformas viarias que, en gran parte, han tenido lugar en las últimas décadas. Por ello, el IGN consideró de alto interés el urgente establecimiento de una nueva REDNAP que recubra todo el territorio español, con superior densidad y precisión que la anterior.

La nueva Red de Nivelación de Alta Precisión (REDNAP) es un proyecto del I.G.N. de carácter plurianual que, comenzado en 2001, y completado en 2007 en la Península y en 2008 va a ser completada en todo el territorio nacional. Su financiación es totalmente estatal, con cargo a los Presupuestos Generales del Estado.

Ya desde 1997, mediante adecuados convenios de colaboración, el Instituto Geográfico Nacional llevó a cabo la implantación de Nivelación de Precisión (NP), con características muy próximas a Alta Precisión, en las islas Canarias (1100 Km.), trabajo que finalizó en febrero del año 2001. Igualmente al amparo del proyecto INTERREG II de la Unión Europea, durante el año 2000 se finalizó la Red NAP de los Pirineos (1500 Km.).

El resto del país se dividió en 7 zonas de trabajo. En el año 2001, se inició la señalización de la Zona 1, que enlazando con la zona pirenaica, cubre la cuenca del Ebro con un desarrollo de unos 2600 Km. Cada una de las zonas corresponde a un año de trabajo.

REDNAP consta de siete zonas, seis en la Península y una en Baleares. Las zonas peninsulares son similares en cuanto a su desarrollo lineal total (unos 2.500 kilómetros en promedio). La longitud total de la red balear se estima en 600 kilómetros. El total nacional de líneas, una vez finalizado el proyecto, será de aproximadamente 16.500 kilómetros en la Península, 600 en Baleares y 1.100 en Canarias. Toda la red discurre a lo largo de carreteras, autopistas y autovías. El trabajo correspondiente a cada zona se realiza en un plazo de dos años, dedicando el primero al proyecto reconocimiento y señalización de las líneas NAP y el segundo a los trabajos de observación propiamente dichos (nivelación geométrica, gravimetría y GPS) y al control y cálculo de las observaciones.

Cada año y para cada zona, previamente a los trabajos de campo, se proyecta el trazado de las nuevas líneas a lo largo de la red de carreteras, procurando informarse previamente acerca del estado de las mismas, obras futuras, densidad de tráfico, etc.; Se planificó que las actuaciones en cada zona podrían estar sometidas a futuras modificaciones en función de las obras que, en su momento, pudieran realizarse en las vías de comunicación. Así mismo, se definen los nodos (puntos de corte de tres o más líneas) y se numeran las líneas resultantes. El trabajo de campo se divide en varias fases:

- Reconocimiento y señalización
- Observación GPS y gravimétrica
- Nivelación geométrica
- Control y cálculo de observaciones.

Las principales aplicaciones de la RED-NAP son las siguientes:

- Determinación muy precisa del relieve del territorio.
- Origen indispensable para las grandes obras públicas.
- Establecimiento de la cartografía, a cualquier escala, necesaria para la ordenación del territorio.
- Conocimiento muy preciso del nivel medio del mar y de sus variaciones a causa del efecto invernadero.
- Detección de posibles variaciones verticales de la corteza en zonas con previsible fenómenos sísmicos y volcánicos.
- Determinación muy precisa de las ondulaciones del geoide con relación al elipsoide utilizado por la geodesia de satélites (GPS, EGNOS, GALILEO, etc.).

Bibliografía

- [1] Instituto Geográfico Nacional , “Red Española de Nivelación de Alta Precisión –REDNAP- “. Boletín Informativo del Instituto Geográfico Nacional (núm 6), Abril-Junio 2001.
- [2] Instituto Geográfico Nacional, Servicio de Geodesia. Instrucción Técnica nº 3.1: “Instrucciones para los trabajos de señalización en nivelaciones de precisión” Instituto Geográfico Nacional , Abril 1986
- [3] Instituto Geográfico Nacional, Servicio de Geodesia. Instrucción Técnica nº 3.2: “Instrucciones para los trabajos de observación en nivelaciones de precisión” Instituto Geográfico Nacional , Abril 1986
- [4] Instituto Geográfico Nacional, Servicio de Geodesia. Instrucción Técnica nº 3.3: “Instrucciones para las especificaciones generales para levantamientos gravimétricos terrestres” Instituto Geográfico Nacional , Abril 1986
- [5] Balbuena Durán, J. L. y Vara Jaén, M^a. D., “Instrumentación y metodología empleadas en las técnicas altimétricas”, Topografía y Cartografía. Revista del Colegio Oficial de Ingenieros Técnico en Topografía, nº 74. Mayo-Junio 1996.
- [6] Martín Asín, F., “Geodesia y Cartografía Matemática”, Editorial Paraninfo, pág. 147 a 157
- [7] Instituto Geográfico Nacional, Normas para la nivelación geométrica de precisión (Anexo al Pliego de Prescripciones Técnicas), Instituto Geográfico Nacional, Julio 2007.

Tema 6. Sistema GPS. El Segmento espacial. La señal GPS, sus características y procesamiento. El Segmento de control del Sistema GPS. Estaciones maestras y monitoras. Órbitas de los satélites. Cálculo de efemérides.

6.1. Introducción

Una nueva era en el posicionamiento comenzó cuando el 22 de Febrero de 1978, se lanzó el primer satélite del Sistema de Posicionamiento Global, *Global Position System* (GPS) ó *NAVigation System with Time and Ranging Global Positioning System* (NAVSTAR GPS) que es el nombre oficial completo del GPS. Siendo utilizado para fines geodésicos prácticamente desde 1983. La configuración final del sistema fue alcanzada en 1994. Desde entonces se ha convertido en el principal sistema (o al menos el más versátil) para posicionamiento de precisión. El primer propósito del GPS era de naturaleza militar de hay sus características previas, dar servicio las 24 horas en todo el mundo y con cualesquiera sean las condiciones meteorológicas. Pero muy pronto los beneficios civiles potenciales fueron evidentes. El predecesor inmediato del GPS fue el sistema militar TRANSIT, en servicio desde 1967. Su funcionamiento se basaba en el llamado efecto Doppler, que es la variación aparente de la frecuencia en función de la velocidad de acercamiento o alejamiento de la fuente emisora. Ya no esta operativo. En el lado soviético existía el Tsikada.

Cuando la Geodesia hace uso de las emisiones de objetos no ligados a la superficie terrestre para conseguir sus fines, entonces la llamamos Geodesia Espacial. Los enlaces geodésicos más largo que se podían realizar desde tierra sólo llegaban hasta los 200 Km. Los primeros precedentes en la Geodesia Espacial ,aparte de la observación a estrellas llamada Astronomía Geodésica, los encontramos en la observación de globos para enlaces geodésicos de masas de Tierra distantes, como el enlace de la Isla de Córcega con Francia. La historia de la moderna Geodesia espacial comienza motivada por la carrera espacial entre la URSS y los EEUU con la puesta en órbita del primer satélite artificial de la Tierra por parte de la antigua URSS el 4 de Octubre de 1957, el Sputnik I. Se observó que analizando el desplazamiento Doppler de las señales radiodifundidas desde el Sputnik I y recibidas en estaciones de posición conocida, era posible establecer la órbita del satélite. Evidentemente esto se podía realizar a la inversa y obtener la posición del receptor. Este sistema no daba la precisión idónea pero ponía de manifiesto la viabilidad de esta aplicación. Desde ese momento se han lanzado miles de satélites artificiales de los que nos

hemos aprovechado los geodestas, aunque algunos no estuviesen previstos para este uso en principio.

En la antigua URSS se desarrolló paralelamente el sistema GLONASS, similar al GPS. La Unión Europea esta a punto de desplegar el nuevo sistema GALILEO, con fundamentos muy similares a los del GPS. Además estos sistemas conviven con otros también útiles para la Geodesia Espacial pero que no tienen su origen y principal función en la navegación como son la interferometría de muy larga base (VLBI) ó la medición de distancias con láser a satélites (SLR).

Las publicaciones generales más recomendables sobre GPS son [3], [4], [5], [8] y una descripción más técnica encontramos en [2].

6.2. Sistema GPS

El GPS un sistema de radio navegación por satélite que provee a usuarios de coordenadas precisas de posicionamiento tridimensional e información sobre navegación y tiempo. Aunque diseñado para uso militar en principio, pasó a ser un sistema de uso civil con ciertas restricciones, siendo utilizado para fines geodésicos prácticamente desde 1983. La configuración final del sistema fue alcanzada en 1994, con 24 satélites utilizables. Desde entonces se ha convertido en el principal sistema (o al menos el más versátil) para posicionamiento de precisión.

Para el estudio del sistema, lo dividiremos en sus tres principales segmentos o sistemas, el espacial, el de control y el de usuario. Sobre el segmento de usuario nos referiremos a la hora de ver como es procesada la señal por el receptor.

6.3. El segmento espacial

Está constituido por la constelación de satélites NAVSTAR. La constelación está pensada para dar cobertura a cualquier hora del día y en cualquier parte del mundo. El segmento espacial proporciona cobertura con 4 a 8 satélites por encima de cualquier horizonte de cualquier lugar de la Tierra. Si la máscara de elevación en la observación se reduce a 10° , se pueden llegar a observar hasta 10 satélites. Si la máscara se reduce a 0° , se puede ver la mitad de la constelación, 12 satélites.

Las características orbitales más importantes son:

- 6 órbitas casi circulares a 20180 km de altitud, nombradas A, B, C, D, E, F, con 55° de inclinación.
- 4 satélites por órbita, 24 en total.
- Periodo de 12 horas de tiempo sidereo.
- Visibilidad de cada satélite: 5 horas.
- La configuración se repite 4 minutos antes cada día solar.
- Existen hasta 4 satélites desactivados y disponibles como reserva “spares”.

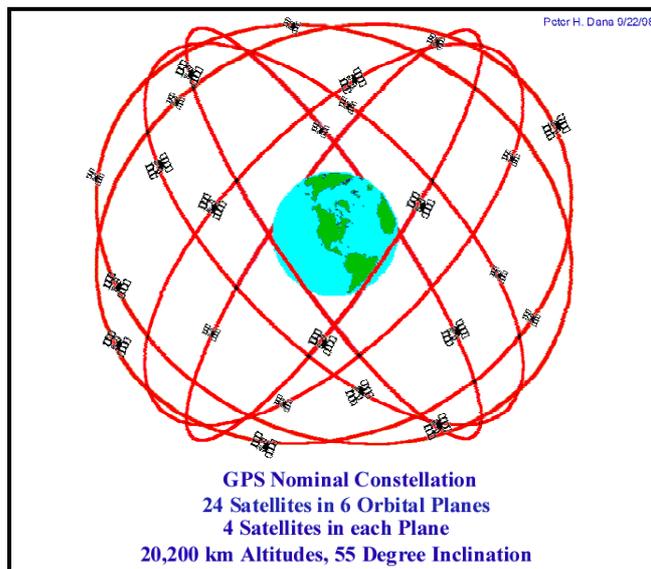


Figura 1. Segmento espacial GPS

Los satélites de diferente generación se agrupan en los llamados "bloques", existiendo por el momento los bloques I, II, IIA, IIR, IIF (y III). Como se establece en [2].

- Bloque I.- Los satélites de este bloque fueron pioneros, aportaron experiencia para el resto de los bloques. Actualmente no queda ninguno en funcionamiento. Fue experimental, estos satélites fueron lanzados desde la Base de Vandenburg en California, entre los años 1978 y 1985.
- Bloque II.- Para reducir costes se decidió poner en órbita los satélites del bloque II mediante vehículos transbordadores espaciales. Se ponían en órbita tres satélites en cada viaje. El desastre del transbordador Challenger en 1986 retrasó el programa. Actualmente se están lanzando con el cohete MLV Delta2. Los satélites de este bloque se reparten en cuatro generaciones sucesivas: Bloque II, II-A, II-R y II-F.

- Bloque III.- Está en proyecto, puede que se sustituyan sus satélites por el modelo II-R, evolución de los del Bloque II. El primer satélite está previsto para el 2009 y toda la constelación para el 2030. Se prevé la incorporación de dos nuevas portadoras para uso civil. Posiblemente irán provistos de relojes con osciladores de hidrógeno y Tendrán mejor capacidad antiinterferencia para las portadoras L1 y L2.

Los últimos satélites son mucho más fiables que los de la primera generación. Tienen sistemas alternativos computables desde tierra. Un satélite queda fuera de servicio por avería o envejecimiento de los paneles solares, falta de capacidad de los acumuladores, averías no computables en los sistemas eléctricos o agotamiento del combustible de maniobra y recuperación de órbita. Entonces decimos que ha acabado su “vida útil”.

6.3.1. Puesta en órbita

Hasta ahora la puesta en servicio de cada satélite se ha realizado con dos tipos de cohetes los Atlas F y Delta 2. El lanzamiento se realiza en distintas fases. La primera y segunda etapa sitúan al satélite en “*órbita de aparcamiento*”, una órbita elíptica con apogeo y perigeo a 870 y 180 Km. de altitud respectivamente, quedando la tercera etapa del cohete y su carga útil (satélite). La tercera etapa se sitúa en “*órbita de transferencia*” a 20.180 Km., con fuerte excentricidad. Los cohetes de maniobra del satélite retocan hasta establecer la órbita de servicio definitiva. Desde el control en tierra evidentemente se puede manipular el funcionamiento, permitiendo corregir la órbita o incluso cambiar la posición dentro de la propia órbita.

6.3.2. Relojes u osciladores

Los relojes son la fuente de la frecuencia, patrón de generación de las frecuencias emisoras L1 y L2. Los tipos de reloj pueden ser de oscilador de cristal de cuarzo, rubidio, cesio y más de hidrógeno.

Por bloques, las características de los relojes es la siguiente:

- Los satélites del Bloque I pesaban más de 400 Kg. De los 10 satélites que lo componían, 4 llevaban oscilador de cuarzo, 3 de rubidio y 3 de cesio.
- Los del Bloque II, pesan unos 800 Kg. Llevan 2 paneles solares de 7,2 m² y 700 vatios de potencia. Todos incorporan osciladores atómicos, de cesio o de cesio y rubidio. Normalmente tienen dos patrones de frecuencia cesio y dos patrones de frecuencia rubidio.

- Los del bloque III, posiblemente llevarán osciladores de máser de hidrógeno.

Tipo de reloj	Frecuencia de oscilación (GHz)	Estabilidad por día ($\Delta f/f$)	Tiempo para perder un segundo
Oscilador de Cristal de Cuarzo	0.05	10^{-9}	30 años
Rubidio	6834682613	10^{-12}	30000 años
Cesio	9192631770	10^{-13}	300000 años
Maser de Hidrógeno	1420405751	10^{-15}	30000000 años

Tabla 1. Precisión de los relojes u osciladores.

6.3.3. Identificación del satélite

La identificación de los satélites se puede hacer por varios sistemas:

- Por el número NAVSTAR (SVN). Es el número de orden de lanzamiento.
- Por el número de órbita.
- Por la posición que ocupa en la órbita
- Por el número de catálogo NASA
- Por la identificación internacional constituida por el año de lanzamiento, el número de lanzamiento en el año y una letra según el tipo.
- Por el número IRON. Número aleatorio asignado por la Junta de Defensa Aérea Norteamericana (NORAD) de Estados Unidos y Canadá.

Pero la forma generalizada de identificarlos es por su PRN o ruido pseudoaleatorio, que es característico de la señal de cada satélite NAVSTAR.

LAUNCH ORDER	PRN	SVN	LAUNCH DATE	FREQ STD	US PLANE	LAUNCH ORDER	PRN	SVN	LAUNCH DATE	FREQ STD	US PLANE		
*II-1		14	14	FEB 1989		IIA-23	4	34	26	OCT 1993	Rb	D4	
*II-2		13	10	JUN 1989		IIA-24	6	36	10	MAR 1994	Rb	C1	
*II-3		16	18	AUG 1989		IIA-25	3	33	28	MAR 1996	Rb	C2	
*II-4		19	21	OCT 1989		IIA-26	10	40	16	JUL 1996	Cs	E3	
*II-5		17	11	DEC 1989		IIA-27	30	30	12	SEP 1996	Cs	B2	
*II-6		18	24	JAN 1990		IIA-28	8	38	6	NOV 1997	Cs	A3	
*II-7		20	26	MAR 1990		**IIR-1		42	17	JAN 1997			
*II-8		21	2	AUG 1990		IIR-2	13	43	23	JUL 1997	Rb	F3	
*II-9		15	1	OCT 1990		IIR-3	11	46	7	OCT 1999	Rb	D2	
IIA-10	32	23	26	NOV 1990	Rb	E6	IIR-4	20	51	11	MAY 2000	Rb	E1
IIA-11	24	24	4	JUL 1991	Cs	D6	IIR-5	28	44	16	JUL 2000	Rb	B3
IIA-12	25	25	23	FEB 1992	Rb	A5	IIR-6	14	41	10	NOV 2000	Rb	F1
*IIA-13		28	10	APR 1992			IIR-7	18	54	30	JAN 2001	Rb	E4
IIA-14	26	26	7	JUL 1992	Rb	F2	IIR-8	16	56	29	JAN 2003	Rb	B1
IIA-15	27	27	9	SEP 1992	Cs	A4	IIR-9	21	45	31	MAR 2003	Rb	D3
IIA-16	1	32	22	NOV 1992	Cs	F6	IIR-10	22	47	21	DEC 2003	Rb	E2
*IIA-17	29	29	18	DEC 1992	Rb	F5	IIR-11	19	59	20	MAR 2004	Rb	C3
*IIA-18		22	3	FEB 1993			IIR-12	23	60	23	JUN 2004	Rb	F4
*IIA-19		31	30	MAR 1993			IIR-13	2	61	6	NOV 2004	Rb	D1
IIA-20	7	37	13	MAY 1993	Rb	C5	IIR-14M	17	53	26	SEP 2005	Rb	C4
IIA-21	9	39	26	JUN 1993	Rb	A1	IIR-15M	31	52	25	SEP 2006	Rb	A2
IIA-22	5	35	30	AUG 1993	Rb	B4	IIR-16M	12	58	17	NOV 2006	Rb	B5

Tabla 2. Satélites de los bloques II/IIA/IIR/IIR-M en Septiembre de 2007

6.3.4. Tiempo GPS

Otra característica importante del GPS es la utilización de su propio sistema de tiempo, el Tiempo GPS (TGPS). Mantiene una desviación constante de 19 segundos Sistema Internacional con el Tiempo Atómico Internacional (TAI). Coincidió con Tiempo Universal Coordinado (UTC) en la época inicial de TGPS, enero 6,0 del año 1980. Su unidad es el segundo del Sistema Internacional. Por lo que es una escala uniforme y continua. La diferencia entre el GPST y UTC no es constante. Se relacionan a través de los “segundos intercalares”.

El control de tiempos en GPS es el principal factor de imprecisión. Por otra parte, el GPS es una herramienta muy válida para la transmisión de tiempos.

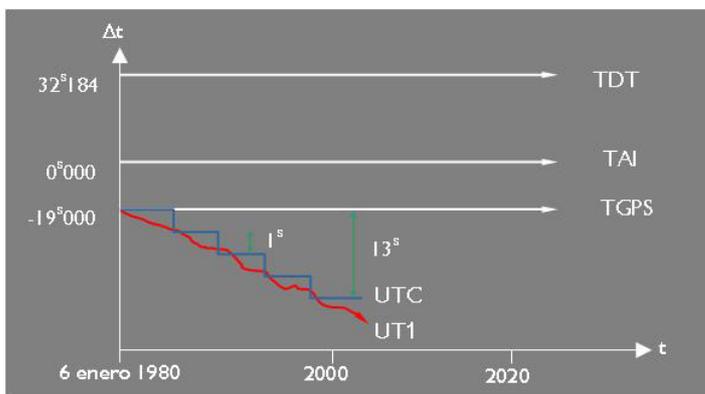


Figura 2. Relación de escalas de tiempo con tiempo GPS

A partir del Tiempo GPS se definen las llamadas unidades derivadas del tiempo GPS. Estas son, el día de Año (DOY): número del día dentro del año calculado en el intervalo 1-365, 1-366 en bisiestos, la semana GPS (GPS week): semanas transcurridas desde el 6 de enero de 1980 a 0h de UTC, el día de semana que es la numeración de días dentro de cada semana en la que corresponde el 0 para el domingo y el 6 para el sábado

Día convencional	21 de junio de 2003.
Día de Año (DOY):	172
Semana GPS (GPS week):	1223
Día de semana GPS	6 (sábado)

Tabla 3. Ejemplo de denominación de día GPS.

6.4. La señal GPS, sus características y procesamiento

El diseño de la estructura de la señal GPS, está condicionado por los objetivos que persigue el propio sistema de navegación y que distinguen al propio sistema de los anteriores sistemas de posicionamiento. Algunos de los objetivos más importantes son:

- Alta precisión en posición en tiempo real
- Navegación en tiempo real para usuarios con altas velocidades
- Cobertura mundial
- Tolerancia a las interferencias intencionadas y no intencionadas
- Conseguir una posición inicial en un tiempo razonable

A excepción de la cobertura mundial, que depende del despliegue orbital de los satélites, los otros puntos tienen relación con:

- Elección de la banda de radio frecuencias
- Utilización de dos portadoras
- Modulación de las portadoras por medio de códigos
- Señales de amplio espectro
- Diseño de las antenas

Debemos tener en cuenta que para obtener posiciones instantáneas en tiempo real, hay que realizar medidas simultáneamente de varios satélites (mínimo 4), para lo cual cada satélite va a tener asignado un código único (PRN) con el fin de identificar sus señales y además vamos a conocer la posición de cada uno de los satélites en tiempo real por medio del mensaje de navegación. Por otro lado, la obtención de velocidad (con precisión a nivel de cm/s) requiere medidas Doppler sobre portadoras de longitud de onda centimétrica.

La señal GPS va a consistir en ondas electromagnéticas pertenecientes a la banda L del espectro debido principalmente a que los efectos del retardo ionosférico son sustancialmente menores que en otras bandas y es más fácil la asignación del canal de comunicación (uso de parte del espectro electromagnético). Las señales GPS son señales de amplio espectro, es decir, el ancho de banda es superior al necesario para transmitir información, con el fin de lograr las altas precisiones en tiempo real que se requieren para la navegación, combatir el amplio número de interferencias que se producen actualmente y asegurar las comunicaciones entre el transmisor (satélite) y el receptor (observador en tierra). La definición final de las más importantes características de la señal se establece en [2].

6.4.1. Estructura de la señal

Cada satélite transmite señales en dos frecuencias, siendo estas las señales de navegación (códigos) y los datos de navegación y sistema (mensaje). Los códigos son modulados sobre la frecuencia portadora en forma de secuencias llamadas "ruido pseudoaleatorio" o "pseudo random noise" (PRN). Las dos frecuencias portadoras en la banda L (valores usados están en la banda L de radiofrecuencias que abarcan desde 1 Ghz hasta 2 Ghz) son derivadas de una frecuencia fundamental de 10,23 MHz:

- L1: $154 \times 10,23 \text{ MHz} = 1575,42 \text{ MHz}$ ($\lambda = 19,05 \text{ cm}$)
- L2: $120 \times 10,23 \text{ MHz} = 1227,60 \text{ MHz}$ ($\lambda = 24,45 \text{ cm}$)
- L5: $115 \times 10,23 \text{ MHz} = 1176,45 \text{ MHz}$ ($\lambda = 25,51 \text{ cm}$). Todavía no se usa y se empezara a utilizar en el futuro cercano.

Obviamente, las dos portadoras, L1 y L2, por si solas no aportan ningún tipo de información al receptor. La técnica que utiliza el sistema GPS es la modulación sobre estas portadoras de dos códigos y un mensaje codificado que son los que llevan la información. Estos códigos son:

- Un código C/A de fácil adquisición (*Coarse Acquisition code*) también denominado SPS (*Standard Positioning Service*), único para cada satélite de una longitud de 1023 bit, con una frecuencia $f_0/10=1,023 \text{ MHz}$ y un periodo de 1 ms. Se modula en L1 y L2.
- Un código P de precisión (*Precision code*) también denominado PPS (*Precise Positioning Service*), de una longitud de 235.469.592.765.000 bit, con una frecuencia de $f_0=10,23 \text{ MHz}$ y un periodo aproximado de 38 semanas, asignándose un segmento semanal a cada satélite. Por tanto, el código P es único y el C/A diferente para cada satélite. Sólo se modula en L2
- Un mensaje de navegación que proporciona la información necesaria para la navegación en tiempo real. Posee una velocidad de transmisión de 50 bps. Veremos que se adiciona a ambos códigos antes de modularse en las portadoras.
- Actualmente son varios los satélites que emiten el nuevo código C sobre la portadora L2, son los satélites del Bloque IIR(M) en la figura posterior éste no aparece. En el resto del tema obviaremos esta novedad.

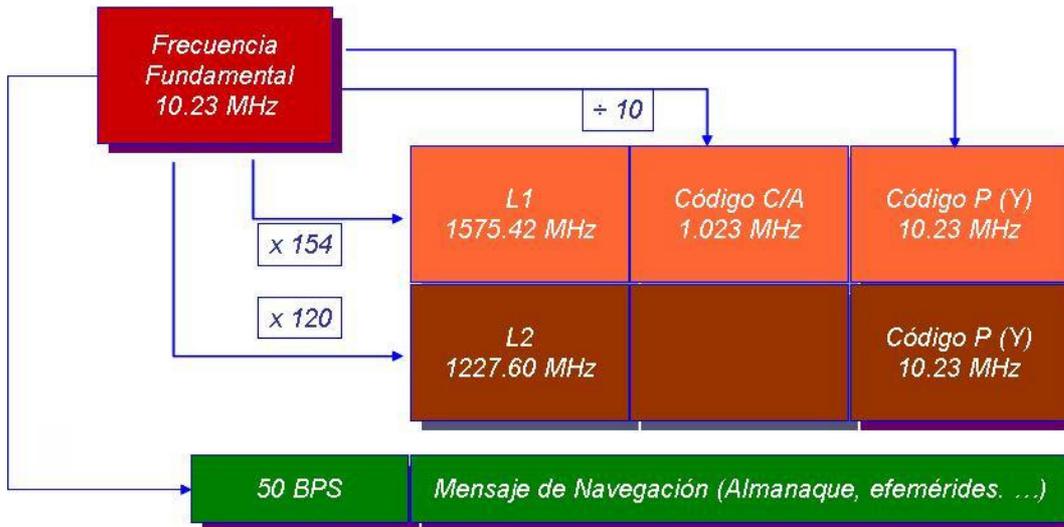


Figura 3. Señales transmitidas por el GPS

6.4.2. Variaciones de la señal

Debemos considerar que las señales emitidas por el emisor, en nuestro caso el satélite, se genera en éste mediante de los pulsos de un reloj, que el emisor y el receptor están a gran distancia, uno próximo o sobre la tierra y el otro no, y que además tanto el emisor como el receptor se desplazan con velocidades diferentes. Estos hechos causan que la frecuencia recibida por el receptor sea diferente que aquella emitida por el emisor. Principalmente debemos considerar las siguientes causas:

- Deriva del oscilador del reloj del satélite.
- Efectos relativistas.
- Efecto Doppler.

La variación de la frecuencia se puede expresar analíticamente como,

$$\frac{f_r}{f_t} = 1 + \frac{1}{c^2}(\phi_t - \phi_r) + \frac{1}{2} \left(\frac{V_r^2}{c^2} - \frac{V_t^2}{c^2} \right) + \frac{\vec{\kappa}}{c}(\vec{V}_t - \vec{V}_r) + \dots = 1 + \delta$$

siendo f_r la frecuencia recibida, f_t la frecuencia transmitida. Los términos a la derecha de la igualdad son debidos a la perturbación gravitacional, a la relatividad especial y al efecto Doppler, respectivamente. La diferencia observada entre ambas frecuencias suele ser un valor aproximadamente de $\delta \cong 447,9 \times 10^{-12}$. Las perturbaciones debidas al abultamiento ecuatorial terrestre y al Sol (el de la Luna no se considera por ser despreciable) varían entre $445,8 \times 10^{-12}$ y $450,2 \times 10^{-12}$. Debemos considerar, además, que aunque los satélites llevan patrones atómicos de frecuencia, éstos están sujetos a variaciones a lo largo del tiempo.

Debido a todas estas perturbaciones, se hace una corrección media por efectos relativistas, haciendo que el satélite emita una frecuencia más baja de la frecuencia nominal con el fin de que el receptor reciba la frecuencia nominal $f_0 = 10,23$ MHz, en $4,55 \times 10^{-12}$ MHz, de manera que la frecuencia obtenida es 10,22999999545 MHz.

Por otro lado, el efecto Doppler es debido a la diferencia de velocidades radiales entre el satélite y el observador. Para un observador estacionario en el polo ($v_r=0$), la máxima velocidad radial del satélite se produce cuando éste cruza el horizonte, así para una frecuencia de transmisión de 1,5 GHz, el desplazamiento Doppler es de 4.500 Hz.

6.4.3. Códigos de ruido pseudoaleatorios

Un código es un sistema que se genera con la finalidad de representar información. Así el código se usa junto con las reglas que lo definen para transmitir algún tipo de información. La mayoría de los códigos que se utilizan actualmente son de tipo binarios (secuencias de 0 y 1) debido a la facilidad de manejo por los ordenadores.

Toda información contiene una cantidad no deseada que se llama ruido. El ruido es aleatorio y enmascara la información que se desea transmitir. Sin embargo, un ruido pseudo-aleatorio, **PRN**, se genera con unas reglas y por lo tanto es predecible y, de hecho es el que contiene la información a transmitir. En el caso del GPS ejemplos de información a transmitir son la lectura del reloj del satélite ó la posición.

La generación de las secuencias de código PRN, está basada en unos registros llamados **registros secuenciales retroalimentados**, que están compuestos por una secuencia de celdas que pueden almacenar los estados 0, 1, de manera que para cada pulso del reloj, el contenido de cada celda pasa a la celda siguiente, y como salida se lee el contenido de la última celda. La entrada en la primera celda está definida por una suma binaria de dos o más celdas que se fijan por definición, es decir, si los valores de las celdas primera y cuarta son iguales, la próxima primera es 0, y si son diferentes, su valor es un 1. Como ejemplo vemos en la figura siguiente un registro de 4 celdas y la secuencia que se genera para cada pulso del reloj.

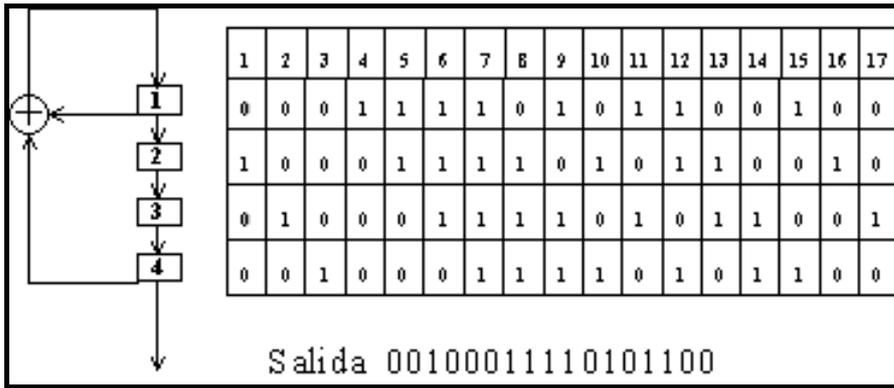


Figura 4. Generación de PRN con registros secuenciales retroalimentados a partir de valor de salida (0100)

De esta forma se va generando una secuencia con todos los estados posibles menos aquel en el que todas las celdas tienen el valor cero y, por tanto, conseguimos $2^4 - 1$ estados diferentes de una forma periódica. En general con un registro de n celdas se puede llegar a producir un código con un período $P = 2^n - 1$.

Una de las características entre las salidas es que la función de autocorrelación es casi nula para todos los desplazamientos salvo el desplazamiento nulo. Si consideramos la secuencia de salida que hemos obtenido, la función de autocorrelación de esta secuencia de código es.

$$R(i) = \frac{1}{15} \int_0^{15} s(t) \cdot s(t+i) \cdot dt \approx \frac{1}{15} \sum_0^{15} s(t) \cdot s(t+i) \cdot \Delta t$$

Para una unidad de tiempo $i = 1$, podemos ver que el producto de $s(t)s(t+1)$ es una forma desplazada de la misma secuencia, como se ve en la figura siguiente. A esta propiedad se conoce como desplazamiento y adición, o ciclo y adición. Para cualquier producto $s(t)s(t+i)$ diferente de $i=0$ o múltiplo de 2^{n-1} , la función de autocorrelación toma el valor “-1/15” y para el valor $i=0$ o múltiplo de 2^{n-1} , la función de autocorrelación toma el valor “1.”

Una vez descritos los fundamentos de los códigos de GPS, vamos a ver cuales son las características principales de éstos, el código C/A y P.

6.4.4. El código P

Una vez descritos los fundamentos de los códigos de GPS, vamos a ver cuales son las características principales de éstos, el código C/A y P.

El código de precisión P, se genera por el producto de dos códigos $X1(t)$ y $X2(t+n_iT)$, donde $X1$ tiene un periodo de 1,5 segundos o 15.345.000 bits y $X2$ tiene un periodo de 15.345.037 bits. Ambas secuencias se inicializan al principio de cada semana y en la misma época (media noche

del sábado al domingo). Estos dos códigos se alinean en fase y tienen una frecuencia ($1/T$) de 10,23 MHz, pero con un desfase de n_i intervalos del código del reloj de T segundos cada uno.

$$XP_i(t) = X_1(t) \cdot X_2(t + n_i T) \quad 0 \leq n_i \leq 36$$

Cada satélite tiene un único desplazamiento $n_i T$ que permite una subdivisión semanal, proporcionando un código P único para cada satélite. El incremento del periodo de X_2 en 37 bit permite que los valores de n_i varíen de 0 a 36 de manera que cada satélite pueda tener su propio código P .

El código $X_1(t)$ y $X_2(t)$ son generados por dos registros secuenciales de 12 celdas cada uno X_{1A} , X_{1B} para el X_1 y X_{2A} y X_{2B} para el X_2 .

El periodo de un producto de códigos es el producto de sus periodos, así que el periodo del código P es de 235.469.592.765.000 bits que corresponde a un periodo de un poco más de 38 semanas. Este periodo se ha subdividido de manera que cada satélite tiene un periodo de una semana que no se solapa con el de ninguno otro.

Por otro lado, se define la cuenta Z como el número de épocas de 1,5 segundos del código X_1 desde el principio de una semana. Estas cuentas serán las que nos permitan saber en que época se encuentra la secuencia del código P y correlarlo con nuestro receptor en un tiempo razonable.

6.4.5. El Código C/A

El código de fácil adquisición C/A tiene una longitud de 1023 bits o una duración de 1 ms y una frecuencia de 1,023 MHz. Está formado por el producto de dos códigos $G_1(t)$ y $G_2(t)$ que tienen un periodo de 1023 bits.

$$XG_i(t) = G_1(t) \cdot G_2[t + N_i(10T)]$$

donde N_i indica el desplazamiento entre los dos códigos. Como N_i puede tomar 1023 valores, existen 1023 códigos diferentes dependiendo de los desplazamientos de N_i .

Esta familia de códigos tiene la propiedad de ser casi ortogonales, por consiguiente la correlación cruzada de dos de ellos es casi nula, lo evita que haya confusión al discriminar las señales de los satélites.

6.4.6. Modulación

Una señal consistente en una onda electromagnética pura no transporta ningún tipo de información. El sistema GPS usa la modulación en fase, pues mantiene toda la información y no hace variar la frecuencia o longitud de onda aumentando sus costes. La modulación consiste en asignar a los valores binarios 0 y 1 del código de ruido pseudo-aleatorio, los estados +1 ó -1.

Se trata de multiplicar la portadora por una función de estado $P(t)$ que toma los valores +1 ó -1 en función de que el código tenga un 0 ó un 1. Ésto produce que la fase se mantenga sin alterar (0) o sufra un desplazamiento de 180° (1). Así, si X es el valor del código (0 o 1),

$$P(X) = 1 - 2 \cdot X$$

es el valor del estado. En concreto en GPS se usa la modulación bifásica binaria.

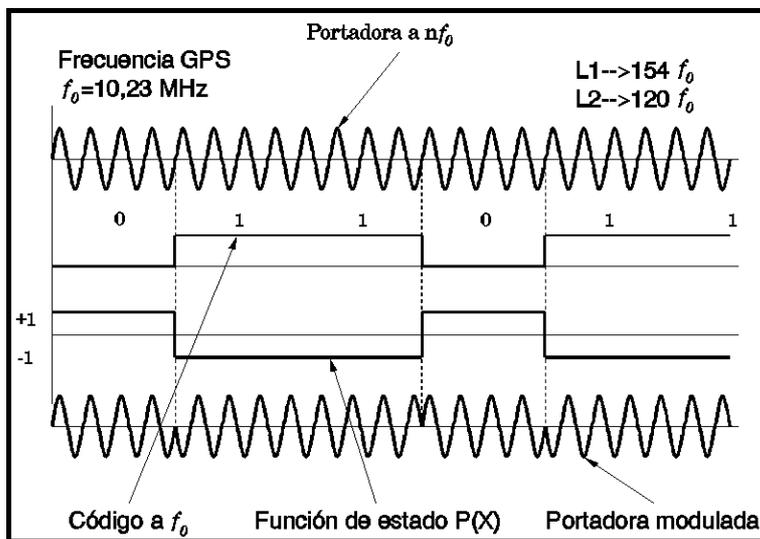


Figura 5. Modulación en GPS

El desfase de 180° se aplica al código P, mientras que en el caso del código C/A se genera un desfase de 90° . Así, se dice que el código P está en fase y el C/A en cuadratura.

6.4.7. El mensaje de navegación

Como su propio nombre indica, el mensaje de navegación contiene los datos que necesita recibir el usuario para llevar a cabo los cálculos y operaciones necesarios para la navegación, es decir, la determinación de la posición y la velocidad (si es necesaria) de dicho usuario. Dicha información se transmite a 50 bps y como hemos visto se modula sobre las dos portadoras.

Los datos que incluye el mensaje son:

- Información sobre sincronización de tiempos.
- Estado de los satélites.
- Parámetros para calcular las correcciones al reloj.

- Las efemérides (posición y velocidad) para el vehículo espacial.
- Correcciones a la señal por retardos atmosféricos.
- Almanaque de toda la constelación.
- Mensajes especiales.
- Mensajes para uso militar.

Cada página del mensaje de navegación tiene una longitud de 1500 bits y tarda 30 segundos en ser transmitida. Una página esta constituida por 5 párrafos de 300 bits cada uno, es decir, tarda 6 segundos en ser transmitido cada uno. A su vez, cada párrafo se divide en 10 palabras comenzando con una palabra de telemetría (TLM) y otra palabra de gestión del código C/A al P (HOW) (hand over word). Cada palabra tarda 0,6 segundos en ser transmitida.

El mensaje completo consta de 25 páginas, en cada una de ellas los párrafos 1, 2, y 3 son iguales y el cuarto y quintos diferentes, por lo que el mensaje completo tarda 12,5 minutos en ser transmitido de forma completa.

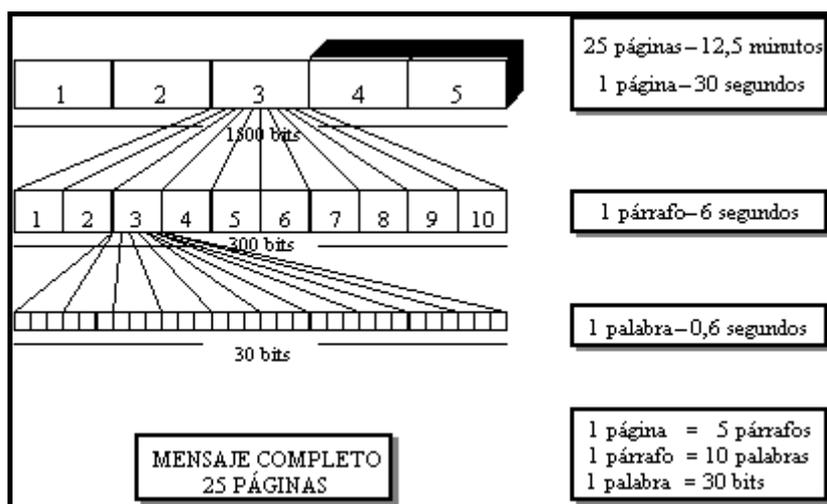


Figura 6. Esquema del mensaje en GPS.

6.4.8. Composición de la señal

Vamos a ver exactamente cuales son las señales que se generan en los satélites.

Partimos de la base de que cada una de las señales se genera con el código P o con el código C/A y una adición módulo-2 de los datos de navegación (correcciones a los relojes, efemérides de los satélites...), formando las siguientes señales, que se modulan sobre las portadoras L1 y L2. Pudiendo escribirlas de la forma $P \oplus D$ y $A/C \oplus D$.

Señal L1

La señal L1 está modulada por ambos códigos. El código P está en fase y el C/A en cuadratura. Por consiguiente el código P, para el estado 0, está alineado con un ángulo de fase 0° y para el estado 1 produce un desplazamiento de 180° en la fase. El código C/A para el estado 0 produce un aumento en la fase de 90° , y para el estado 1 la fase disminuye 90° . Puesto que cada satélite transmite un único código XP_i y XG_i . Y Las amplitudes relativas de los códigos P y C/A se controlan por medio de las constantes A_p y A_c . La portadora modulada L1 viene representada entonces, para cada satélite, por:

$$L1_i(t) = A_p XP_i(t) D_i(t) \cos(2\pi f_1 t) + A_c XG_i(t) D_i(t) \sin(2\pi f_1 t)$$

Señal L2

La señal L2 puede ser modulada bien por el código P, bien por el código C/A, según sea seleccionado por el segmento de control en tierra. Asimismo, se modulan los mismos datos de navegación de la señal L1 sobre la L2. Por tanto, la portadora L2 modulada para cada satélite i . Si B_p representa la amplitud de la señal. Actualmente algunos satélites también modulan un código C sobre la L2, pero en la expresión posterior no lo tenemos en cuenta.

$$L2_i(t) = B_p XP_i(t) D_i(t) \cos(2\pi f_2 t)$$

6.4.9. Procesamiento de la señal

El procesamiento es realizado por el segmento usuario que es el receptor GPS del usuario. La señal emitida por un satélite, en definitiva, contiene tres componentes que designamos (L1, C/A, D), (L1, P, D) y (L2, P, D), como se aprecia en las expresiones del apartado anterior. El procesamiento de la señal en el receptor GPS tiene como finalidad recuperar las componentes de la señal, incluyendo la reconstrucción de la portadora y la extracción de los códigos para obtener las lecturas del reloj de los satélites, así como el mensaje de navegación. Gráficamente:

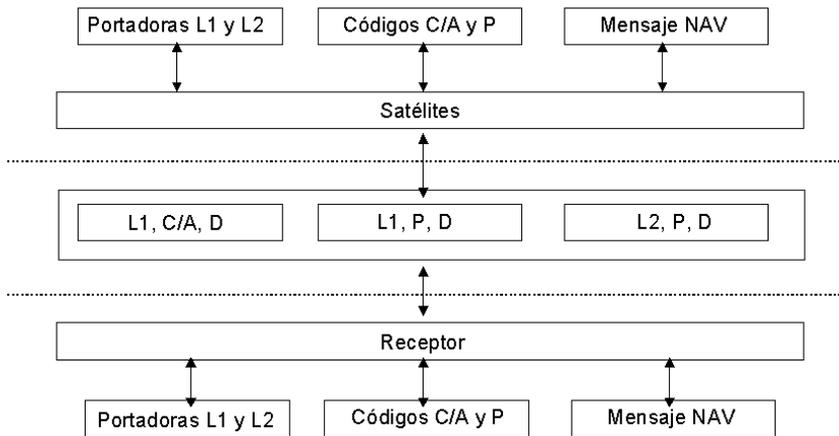


Figura 7. Flujo de códigos entre satélite y receptor.

El receptor contiene elementos para la recepción y procesamiento de la señal recibida desde los satélites. Básicamente, podemos ver el esquema de funcionamiento en el siguiente gráfico:

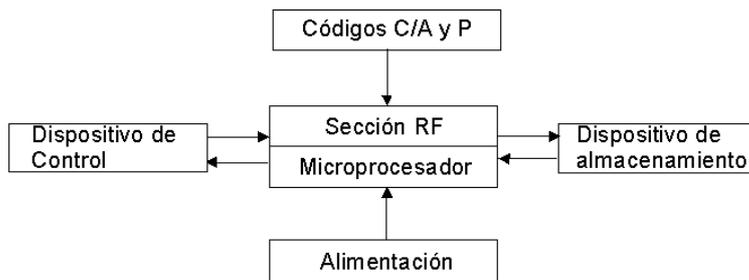


Figura 8. Elementos principales del receptor..

Una antena omnidireccional recibe la señal proveniente de todos los satélites que se encuentran por encima del horizonte del plano de tierra de la señal (Horizonte) y, tras una preamplificación de dicha señal, la transmite a la sección de radiofrecuencia. La antena puede estar diseñada para recepción simple de la portadora L1 o bien para recepción dual de las ambas portadoras L1 y L2. EL criterio de diseño más importante de la antena es la sensibilidad del *centro de fase*. El centro electrónico de la antena debe estar lo más cerca posible del centro físico de ésta y debe ser insensible a cambios por rotación o inclinación (particularmente en las aplicaciones cinemáticas en que la antena se mueve de forma continua durante la navegación).

El microprocesador controla el sistema al completo y permite la navegación en tiempo real por medio de las medidas de pseudodistancias. Este dispositivo es el que permite la comunicación interactiva con el receptor.

Se necesita también un dispositivo de almacenamiento para los observables y el mensaje de navegación de manera que estén disponibles para el procesamiento de los datos posteriormente.

La sección de radiofrecuencia (RF) es básicamente el corazón del receptor. Una vez la señal entra por la antena, se discriminan unas señales de otras, por ejemplo, por medio de los códigos C/A, únicos para cada satélite. Otra técnica consiste en controlar el desplazamiento Doppler. Las señales recibidas se asignan a canales independientes para cada satélite y a partir de su identificación, se siguen de forma continua mientras sea posible. La sección RF básicamente consta de osciladores para generar la frecuencia de referencia, filtros para eliminar frecuencias no deseadas y mezcladores. En estos últimos, se multiplican matemáticamente dos ondas y_1 e y_2 con amplitudes a_1 y a_2 y frecuencias f_1 e f_2 , que escribimos como:

$$y = y_1 \cdot y_2 = a_1 \cos(f_1 t) \cdot a_2 \cos(f_2 t) = \frac{a_1 \cdot a_2}{2} [\cos((f_1 - f_2) \cdot t) + \cos((f_1 + f_2) \cdot t)]$$

y como resultado obtenemos una oscilación que consiste en una parte de baja frecuencia y otra de alta frecuencia. Después de aplicar un filtro de paso bajo, la parte de alta frecuencia queda eliminada y se procesa la parte restante, es decir, la de baja frecuencia.

Las medidas de fase actuales se realizan en circuitos de seguimiento en los que se suelen usar dos métodos:

- Técnica de correlación del código (implica el conocimiento del código PRN)
- Técnicas independientes del código.

Con la técnica de correlación del código obtenemos todos los componentes de la señal del satélite: lectura del reloj del satélite, mensaje de navegación y la portadora remodulada. El proceso se realiza en varios pasos. Primero, el receptor genera una portadora de referencia que es modulada con una réplica del código PRN. Una vez generada, la señal resultante se correla con la señal recibida del satélite. La señal se desplaza en el tiempo hasta que se adaptan de forma óptima. Entonces, el desplazamiento en el tiempo necesario, Δt , suponiendo que no hay errores, se corresponde con el tiempo de viaje de la señal desde el satélite hasta el centro de fase de la antena del receptor. Una vez que se elimina el código PRN, la señal todavía contiene el mensaje de navegación, que se puede decodificar con un filtro de paso alto. El resultado final es la portadora desplazada (Doppler) sobre la que se realiza la medida de fase. Obviamente, éste proceso requiere el conocimiento del código PRN por lo que se suele aplicar únicamente al código C/A modulado sobre la portadora L1. Se suele usar el código C/A para adaptar la réplica a la señal recibida mediante la correlación y, una vez eliminado y obtenido el mensaje de navegación, extraer la palabra **HOW** para cada párrafo. Con la palabra HOW, el receptor es capaz de buscar la parte del código P correspondiente y comenzar el mismo proceso con el código P, lo que nos dará mayor precisión.

La técnica independiente del código se basa en la demodulación por cuadratura. La señal recibida se mezcla (multiplica) por sí misma consiguiendo eliminar todas las modulaciones de la portadora. Esto ocurre puesto que el desplazamiento de 180° en fase durante el proceso de modulación es equivalente a cambiar el signo de la señal.

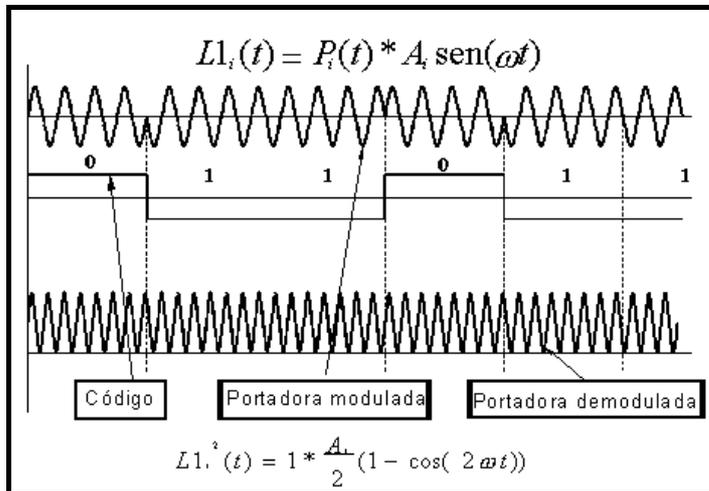


Figura 9. Flujo de códigos entre satélite y receptos.

El resultado es una portadora con el doble de frecuencia y la mitad de la longitud de onda. Esta técnica tiene la ventaja de ser independiente del código PRN pero el inconveniente de que las lecturas de reloj del satélite y la información orbital se pierden, por lo que la navegación en tiempo real no se puede realizar y el receptor debe sincronizarse de forma externa. Además, al multiplicar la señal por sí misma, la relación señal ruido se reduce puesto que el propio ruido que trae la señal recibida, también se multiplica por sí mismo.

6.5. El Segmento de control del Sistema GPS. Estaciones maestras y monitoras

Opera y monitoriza el sistema GPS. Este segmento tiene la función de realizar el seguimiento continuo de los satélites, calcular su posición precisa desde puntos de situación conocida, la transmisión de datos y la supervisión necesaria para el control diario de todos los sistemas de satélites.

Para los satélites de la constelación NAVSTAR en un principio había cinco estaciones de seguimiento: la estación maestra de Colorado Springs y otras cuatro secundarias distribuidas homogéneamente. Posteriormente se han añadido 6 estaciones más pertenecientes al NIMA además de la estación de Cañaverl.

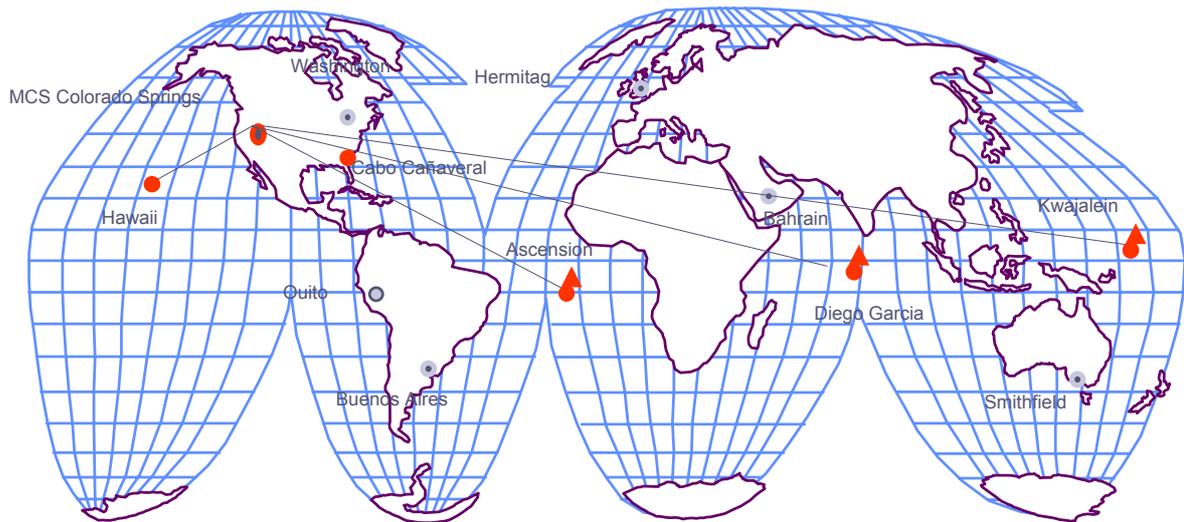


Figura 10. Puntos de control. Estaciones de Seguimiento

Estas estaciones reciben continuamente las señales de los satélites que estén sobre el horizonte, estableciendo con gran precisión las órbitas de éstos. Estos datos se envían a la estación principal o maestra donde se procesan y calculan las efemérides, los estados de relojes, y otras informaciones. Todos estos datos son transmitidos y se almacenan en la memoria del correspondiente satélite, formando parte esencial del mensaje del satélite para su radiodifusión.

La sincronización del tiempo de los satélites es una de las misiones más importantes del segmento de control. Por ello la estación principal de control está conectada directamente con el tiempo estándar del Observatorio Naval de los Estados Unidos (USNO).

Los datos que son calculados y posteriormente extrapolados al usuario son básicamente los siguientes.

- Efemérides Transmitidas (posiciones de los satélites) o parámetros orbitales de cada satélite.
- Datos Ionosféricos-troposféricos.
- Estados de los relojes de los satélites.

Una vez determinados los parámetros anteriores, la misión es emitirlos a los satélites para que éstos puedan transmitirlos a los usuarios. También prevén y realizan maniobras, lanzamientos, reprogramaciones, desactivaciones.

Por lo tanto, hay tres tipos de instalaciones: Estación Maestra de Control (MCS), Estaciones Monitoras (MS) y Antenas Terrestres.

La Estación Maestra de Control (MCS), localizada en la base aérea de Falcon en Colorado, mantiene las operaciones permanentemente, siendo responsable de todos los aspectos de control de la constelación, entre los que incluye.

- Monitorización y control de todos los satélites.
- Mantenimiento y solución de problemas en los distintos satélites.
- Monitorización y control del cumplimiento del estándar de posicionamiento GPS.
- Actualización de los *mensajes de navegación* necesarios para mantener los requisitos de precisión en el sistema.

Las Estaciones Monitoras reciben en todo momento las señales transmitidas por los satélites visibles y obtienen la información necesaria para calcular con gran precisión las órbitas de los satélites. Una vez enviados estos datos a la Estación Maestra, ésta calcula las efemérides de los satélites con un error menor de 1 m en sentido radial, 7 m en el de la trayectoria y 3 m en la perpendicular a la misma. Todo ello es incluido en el *mensaje de navegación*. Las Antenas Terrestres, que son sólo 3 se comunican con el receptor y le incorporan la nueva información.

También existen los llamados Segmentos de Control Independientes, que a partir de las observaciones en estaciones de referencia calculan mejores parámetros que los transmitidos por el mensaje de navegación los datos que calcula pueden ser de efemérides más precisas de correcciones estados de los relojes de los satélites, modelos ionosféricos, troposféricos ó parámetros de rotación terrestre. El principal Segmento de Control Independiente lo constituye la Red global del International GNSS Service (IGS). Más información en [9].

TIPOS DE ÓRBITAS	CALIDAD (M)	RETRASO EN DISPONIBILIDAD	DISPONIBLE EN
Órbitas emitidas	3 m	Tiempo Real	Mensaje de navegación
Órbitas predichas por el CODE	0.20m	Tiempo Real	CODE a través de FTP
Órbitas rápidas por el CODE	0.10m	Después de 16 horas	CODE a través de FTP
Órbitas ultrarápidas del IGS (igu)	0.15m	Después de 3 horas	Centros de Datos del IGS
Órbitas rápidas por el IGS (igr)	0.10m	Después de 24 horas	Centros de Datos del IGS
Órbitas finales por el IGS (igs)	0.05m	Después de 11 días	Centros de Datos del IGS

Tabla 4. Disponibilidad de efemérides y precisión en el IGS. Disponibles en <http://igsb.jpl.nasa.gov/products>

6.6. Órbitas de los satélites. Cálculo de efemérides

Las aplicaciones del GPS dependen en gran medida del conocimiento de las órbitas de los satélites. La determinación precisa de la órbita es esencial para conseguir el objetivo fundamental del GPS, es decir, la determinación de la posición del observador, generalmente sobre la superficie de la tierra. En el caso del posicionamiento absoluto o puntual, hay una fuerte correlación entre el error en la órbita del satélite y el error al calcular la posición del receptor. En el posicionamiento relativo, el error que cometemos en la determinación de la órbita del satélite tiende a cancelarse. Un análisis en más profundidad encontramos en [6], [7], aunque también en [3], [4], [5]. Y desde un punto de vista más práctico en [1].

6.6.1. Movimiento Kepleriano

Para comenzar el estudio del movimiento de un satélite, es necesario comenzar estudiando las leyes que gobiernan el movimiento alrededor de La Tierra. Vamos a considerar una Tierra “ideal”. Supondremos que toda la masa de La Tierra se encuentra concentrada en el origen (sólo se considera la parte central del Potencial Gravitatorio), que no existe atmósfera y que no existe ninguna otra fuerza externa al sistema Tierra-Satélite más que la de atracción de dos masas. En este caso, el movimiento de un satélite está determinado por la ley de gravitación de Newton. Consideremos dos puntos de masas m_1 y m_2 separados una distancia r , entonces el movimiento de la masa m_2 respecto de m_1 viene expresado por la ecuación diferencial homogénea de segundo grado:

$$\ddot{\vec{r}} + \frac{G(m_1 + m_2)}{r^3} \cdot \vec{r} = \vec{0}$$

Siendo,

\vec{r} vector posición relativo.

$\ddot{\vec{r}} = \frac{d^2\vec{r}}{dt^2}$ vector aceleración relativa.

G constante de gravitación universal

Si por otro lado llamamos M_E a la masa de La Tierra, el producto $\mu = G \cdot M_E$ es una constante conocida y es uno de los parámetros que definen el sistema de referencia WGS84

$$\mu = G \cdot M_E = 3986005 \cdot 10^8 \text{ m}^3 \text{ s}^{-2}$$

y si tenemos en cuenta que la masa del satélite es despreciable en comparación con la masa de la Tierra, obtenemos la ecuación diferencial,

$$\frac{d^2\vec{r}}{dt^2} = -\mu \cdot \frac{\vec{r}}{r^3}$$

La solución analítica de esta ecuación diferencial es un problema clásico de mecánica celeste que nos lleva al conocido movimiento Kepleriano, definido por los seis parámetros orbitales que se corresponden con las seis constantes de integración de ecuación diferencial de segundo orden vectorial anterior.

1ª Ley de Kleper

El movimiento de un cuerpo respecto a otro debido a la atracción de las masas se reduce a una cónica, estando uno de los dos cuerpos en el foco de la cónica. En el caso del sistema Tierra-satélite, suponiendo la Tierra “ideal” y considerando un campo gravitatorio central, el movimiento se reduce a una elipse en uno de cuyos focos se encuentra situada la Tierra.

Para situar una órbita en el espacio, vamos a partir de unas definiciones previas. Llamaremos *perigeo* a la posición, dentro de la órbita del satélite, en que éste se encuentra más próximo de la Tierra. Llamamos *apogeo* a la posición, dentro de la órbita, en que el satélite se encuentra más alejado de la Tierra. La línea que une el perigeo con el centro de masas de la Tierra recibe el nombre de *línea de ápsides*. La línea que resulta de la intersección del plano orbital con el ecuador se llama *línea nodal*, dentro de la cual hay que destacar el *nodo ascendente*, punto de la órbita en que el satélite pasa del hemisferio sur al hemisferio norte.

Consideremos un sistema de ejes coordenados con origen, O, en el centro de masas de la Tierra, el eje \vec{OX} en la dirección del equinoccio medio (punto Aries), el eje \vec{OZ} en la dirección del eje de rotación medio y el eje \vec{OY} formando un triedro trirrectángulo con orientación positiva. Gráficamente, la órbita kepleriana se describe en la siguiente figura.

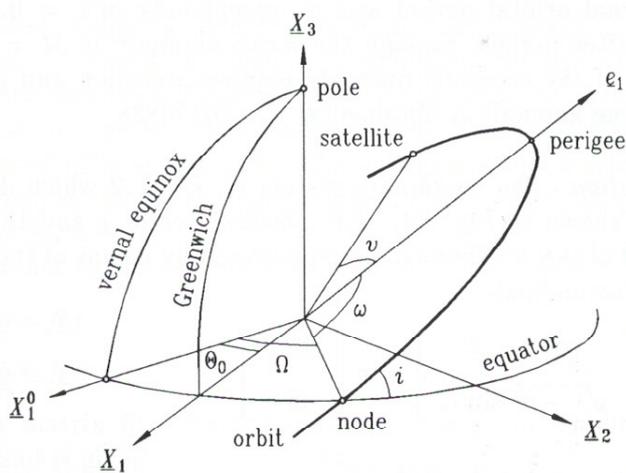


Figura 11. Órbita y sistemas de referencia

Una vez consideradas las definiciones anteriores, los seis parámetros que sitúan de forma única una órbita en el espacio (también llamados *elementos keplerianos*) son:

- Ω , *Ascensión recta del nodo ascendente*. Es el ángulo, medido en el plano ecuatorial, entre el equinoccio vernal o punto Aries y el nodo ascendente.
- i *Inclinación de la órbita*. Es el ángulo formado por el plano orbital y el plano ecuatorial.
- a , *Semieje mayor de la elipse orbital*.
- e , *Excentricidad de la elipse orbital*. Relaciona los dos semiejes de la elipse:
$$e = \frac{a^2 - b^2}{b^2}$$
- ω , *Argumento del perigeo*. Es el ángulo, medido en el plano orbital, entre el nodo ascendente de la órbita y el perigeo.
- $\nu(t)$, *Anomalía verdadera*. Es el ángulo, medido en el plano orbital, entre la línea de ápsides y la posición del satélite.

Los dos primeros parámetros orbitales, es decir, Ω e i , sitúan el plano en el que está contenida la órbita (plano orbital) en el espacio. Los parámetros orbitales tercero y cuarto, a y e , definen la forma y dimensiones de la órbita dentro del plano orbital. Por último, los parámetros orbitales quinto y sexto, ω y $\nu(t)$, sitúan para cada instante de tiempo la posición del satélite dentro de la órbita. Con los seis elementos keplerianos tenemos, por tanto, definida de forma única la posición de un satélite en el espacio para cada instante. En el caso que nos trata de movimiento no perturbado, la anomalía verdadera es el único de los parámetros keplerianos que es función del tiempo. El resto permanecen constantes a lo largo del tiempo.

2ª Ley de Kleper

La segunda ley de Kepler establece que el radio vector del satélite dentro de la órbita recorre áreas iguales en tiempos iguales.

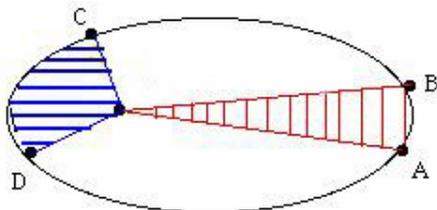


Figura 12.

La posición instantánea del satélite dentro de la órbita se describe por medio de una cantidad angular conocida como *anomalía*. Existen varios tipos de anomalías según se considere el ángulo medido desde el foco de la órbita (geocentro) o bien desde el centro de la órbita:

- $\mathbf{v}(t)$ *Anomalía verdadera*. Es el ángulo, medido en el plano orbital y desde el geocentro, entre la línea de ápsides (perigeo-geocentro-apogeo) y la posición del satélite.
- $\mathbf{E}(t)$ *Anomalía excéntrica*. Es el ángulo, medido en el plano orbital y desde el centro de la órbita, entre la línea de ápsides y la posición del satélite proyectada a una circunferencia de radio el semieje mayor de la elipse, a .
- $\mathbf{M}(t)$ *Anomalía media*.

Mientras que la anomalía verdadera y la anomalía excéntrica tienen sentido geométrico, la anomalía media es una abstracción matemática.

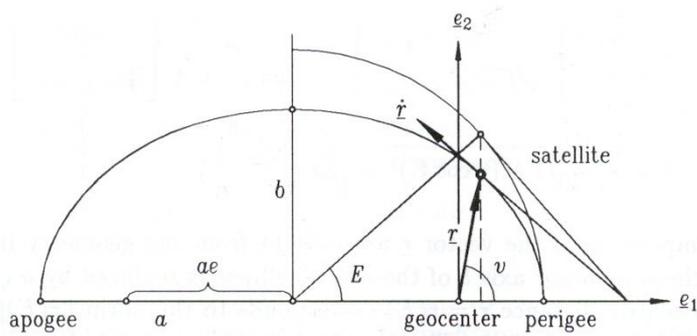


Figura 13.

3ª Ley de Kepler

Nos aporta el conocimiento del periodo orbital del satélite, es decir, el tiempo que tarda en recorrer una órbita completa alrededor de la Tierra. Establece que el cuadrado del periodo orbital es proporcional al cubo del semieje mayor de la elipse. En forma de ecuación se expresa:

$$\mu = 4\pi^2 \frac{a^3}{T^2}$$

Así, fijado el semieje mayor de una órbita para un satélite alrededor de la Tierra, conocemos su periodo orbital a través de esta tercera ley. El conocimiento de este periodo nos lleva a conocer la velocidad angular media del satélite, también llamada *movimiento medio*:

$$\mathbf{n} = \frac{2\pi}{T} = \sqrt{\frac{\mu}{a^3}}$$

y es el que va a dar sentido a la *anomalía media*. Si llamamos T_0 al tiempo de paso por el perigeo del satélite, se define la *anomalía media* para un instante t como (abstracción matemática, no tiene sentido geométrico):

$$M(t) = n \cdot (t - T_0)$$

de forma que podemos relacionar las tres diferentes anomalías mediante las igualdades:

$$M(t) = n \cdot (t - T_0)$$

$$E(t) = M(t) + e \cdot \sin E(t) \quad \rightarrow \quad \text{Conocida como Ecuación de Kepler.}$$

$$v(t) = 2 \operatorname{arctg} \left[\sqrt{\frac{1+e}{1-e}} \operatorname{tg} \frac{E(t)}{2} \right]$$

Estas igualdades, que relacionan las diferentes anomalías dentro de la órbita, nos van a permitir identificar diferentes conjuntos de elementos keplerianos para la definición de la posición de un satélite en el espacio:

- $\{\Omega, i, a, e, \omega, v(t)\}$ con la anomalía verdadera.
- $\{\Omega, i, a, e, \omega, M(t)\}$ con la anomalía media.
- $\{\Omega, i, a, e, \omega, E(t)\}$ con la anomalía excéntrica.

6.6.2. Movimiento Perturbado

La órbita kepleriana es una órbita teórica que supone una Tierra esférica cuya masa se acumula en un punto, un sistema en el que no actúa más fuerza que la de atracción entre dos masas y que no existe atmósfera. Las fuerzas o aceleraciones perturbadoras son todas aquellos factores que causan que el satélite se desvíe de su órbita kepleriana teórica, por lo que la ecuación del movimiento perturbado será la del movimiento kepleriano más la acción de las aceleraciones perturbadoras.

Debe tenerse en cuenta que, para los satélites GPS, la aceleración derivada de la fuerza atractiva central es 10^4 veces mayor que las aceleraciones perturbadoras. Las fuerzas perturbadoras que afectan a un satélite en su movimiento alrededor de la Tierra podemos dividir las en dos grandes grupos:

- *Gravitacionales*
 - No esfericidad de la Tierra
 - Atracción de mareas (efecto directo e indirecto)
- *No gravitacionales*
 - Presión por radiación solar
 - Efecto del Albedo

- Rozamiento atmosférico
- Efectos relativistas
- Viento solar, campo magnético, etc...

En el caso de los satélites GPS, las fuerzas perturbadoras principales a tener en cuenta son la no esfericidad de la Tierra, los efectos producidos por la marea del Sol y la Luna y la presión por radiación solar.

Las ecuaciones del movimiento en un sistema coordenado cartesiano podemos escribirlas como:

$$\left. \begin{aligned} \frac{d\vec{x}}{dt} &= \dot{\vec{x}} \\ \frac{d\dot{\vec{x}}}{dt} + \mu \cdot \frac{\vec{x}}{\|\dot{\vec{x}}\|^3} &= \ddot{\vec{x}}_g + \ddot{\vec{x}}_S + \ddot{\vec{x}}_L + \ddot{\vec{x}}_{PRS} \end{aligned} \right\}$$

donde el primer término es la parte central del campo gravitatorio que hemos estudiado en el caso de movimiento no perturbado. Vamos a estudiar las aceleraciones perturbadoras más importantes.

La no esfericidad de la Tierra se estudia a partir del potencial gravitatorio terrestre V que puede expresarse mediante un desarrollo en serie de armónicos esféricos en la forma:

$$V = \frac{\mu}{r} \left[1 - \sum_{n=2}^{\infty} \left(\frac{a_E}{r} \right)^n \cdot J_n \cdot P_n(\sin \varphi) - \sum_{n=2}^{\infty} \sum_{m=1}^n \left(\frac{a_E}{r} \right)^n [J_{nm} \cos m\lambda + K_{nm} \sin m\lambda] \cdot P_{nm}(\sin \varphi) \right]$$

El primer término de la parte derecha de la igualdad, $V_o = \left(\frac{\mu}{r} \right)$, representa el potencial V_o para una Tierra esférica de manera que el potencial total que genera las perturbaciones podemos escribirlo como:

$$T = V - V_o$$

por lo que la aceleración de la parte no central del campo gravitatorio será el gradiente de T:

$$\ddot{\vec{x}}_g = \overrightarrow{\text{grad}T}$$

El término más importante del desarrollo del potencial perturbador es el J_2 y representa el abultamiento ecuatorial en el campo gravitatorio. Aproximadamente es tres órdenes de magnitud (10^3) mayor que el resto de coeficientes y menor que el debido al potencial V_o en un

factor de 10^4 . Así, aceleración generada por la parte no perturbada del movimiento es de $0,57 m \cdot s^{-2}$ y la generada por el potencial perturbador es de $0,5 \times 10^{-6} m \cdot s^{-2}$.

En cuanto al efecto de marea debido a la atracción del Sol y la Luna se basa en que una masa externa al sistema Tierra-satélite ejerce una atracción sobre la Tierra y el satélite. Para ver como afecta dicha aceleración al movimiento del satélite, sólo hay que tener en cuenta la diferencia entre la atracción que dicha masa externa ejerce sobre la Tierra y la que ejerce sobre el satélite. Si tenemos en cuenta que estas atracciones son del tipo gravitatorio, cada una de las masas externas generará un potencial con un desarrollo en armónicos esféricos. De este potencial va a derivar la aceleración perturbadora que desvía al satélite de su movimiento ideal, y cada cuerpo celeste generará una atracción diferente. Las aceleraciones perturbadoras debidas a la atracción del Sol y la Luna tienen unos valores aproximados de:

$$\ddot{\vec{x}}_S \approx 2 \cdot 10^{-6} ms^{-2}$$

$$\ddot{\vec{x}}_L \approx 5 \cdot 10^{-6} ms^{-2}$$

donde se puede apreciar que el efecto de la Luna es mayor que el del Sol debido a la “proximidad” de ésta en comparación con el Sol.

La presión por radiación solar es la perturbación producida por el impacto, sobre la superficie del satélite, de los fotones procedentes del Sol. Los parámetros básicos que hay que considerar para estudiar la presión por radiación solar son:

La magnitud de la aceleración perturbadora por efecto de la presión por radiación solar es aproximadamente:

$$\ddot{\vec{x}}_{PRS} \approx 10^{-7} ms^{-2}$$

Para tener una idea de cómo afectan todas estas aceleraciones perturbadoras al movimiento del satélite, consideremos una aceleración perturbadora media constante de $10^{-9} ms^{-2}$ actuando sobre un satélite GPS. El desplazamiento asociado que sufre el satélite se obtiene integrando dos veces esta aceleración, y si dicha integración la realizamos en un periodo de 12 horas (aproximadamente es el periodo orbital), obtenemos que tras una revolución completa del satélite, éste ha sufrido un desplazamiento de 1 m.

6.6.3. Cálculo de efemérides

El cálculo de efemérides, consiste en el paso inverso al cálculo de la posición del receptor. En este caso partimos de unas coordenadas de estaciones de referencia conocidas ya de antemano en un determinado sistema de referencia. Normalmente es a partir de las ecuaciones de dobles

diferencias que se forman a partir de las observaciones de las estaciones de referencia, dejando constreñidas las coordenadas de los puntos se obtienen las coordenadas de las órbitas. En el cálculo no sólo pueden usar las observaciones GPS, sino también las observaciones de microondas a satélite *Satellite Laser Ranging* (SLR). Una red global y relativamente densa es necesaria para un cálculo suficientemente válido. Los parámetros orbitales que se calculan incluyen los 6 elementos keplerianos como condiciones iniciales de los arcos, hasta 9 parámetros del modelo empírico de presión de radiación del *Center of Orbit Determination of Europe* (CODE) y pulsos estocásticos como cambios de velocidad a determinadas épocas. La estimación orbital se hace menos precisa cuando lo que se pretende es predecirla. Hoy en día, la mayoría de los usuarios no necesita calcularlas, ya que las órbitas proporcionadas por el IGS [9] tienen la suficiente precisión, como se observa en la Tabla 2. Información adicional sobre las estrategias de cálculo de órbitas, encontramos en [1].

Bibliografía

- [1] Dach, R., Hugentobler, U., Fridez, P., Meindl, M., (2007), The Bernese GPS Software Version 5.0, Astronomical Institute, Universidad de Berna.
- [2] GPS-ICD (1993), GPS Interface Control Document, Revision C (ICD-GPS-200C, US Department of Defense (DoD), <http://www.navcen.uscg.gov/pubs/gps/icd200/>.
- [3] Hofmann-Wellenhof, B., H. Lichtenegger, and J. Collins (1992), GPS: Theory and Practice, Springer, ISBN 3- 211-82364-6.
- [4] Leick, A. (1995), GPS Satellite Surveying, Wiley, ISBN 0-471-30626-6.
- [5] Nuñez-García, A., Valbuena, J.L., Velasco, J., (1992) GPS La nueva era de la Topografía. ISBN 84-87510-31-0.
- [6] Rothacher, M. (1992), Orbits of Satellite Systems in Space Geodesy, Geodeatisch-geophysikalische Arbeiten in der Schweiz, Band 46, Schweizerische Geodetische Kommission, Institut für Geodesie und Photogrammetrie, Eidg. Technische Hochschule Zurich, Zurich
- [7] Springer, T.A. (2000), Modeling and Validating Orbits and Clocks Using the Global Positioning System, Geodeatisch-geophysikalische Arbeiten in der Schweiz, Band 60, Schweizerische Geodeatische Kommission, Institut für Geodesie und Photogrammetrie, Eidg. Technische Hochschule Zurich Zurich.
- [8] Teunissen, P. J.G., and A. Kleusberg (eds.) (1998), GPS for Geodesy, Springer, ISBN 3-540-63661-7.
- [9] IGS (2007). Página oficial del IGS. <http://igsb.jpl.nasa.gov>

Tema 7. Observables GPS. Captura de datos. Combinación de datos. Fuentes de error en GPS. Modelos matemáticos para el posicionamiento. Métodos de posicionamiento. Procesamiento de datos. Cálculo y compensación de una campaña GPS.

7.1. Introducción

Los observables del GPS pueden ser de varios tipos, de fase, de pseudodistancia de código o de lo que se llama corrimiento Doppler. De cualquier manera el primer paso en este proceso es la captura del dato por el receptor. Los diferentes observables se pueden combinar antes de ser procesados para mejorar su objetivo final que es el cálculo de la posición y el tiempo. Aun combinados o no estos observables, estos vienen afectados por diferentes fuentes de error, que aunque minimizadas por una buena observación a veces son corregibles preprocesando el observable original. Para la resolución de los parámetros que buscamos construimos el modelo matemático apropiado. Con el, normalmente, al final tendremos muchas más ecuaciones que parámetros incógnitas. Aplicando el modelo estadístico apropiado resolveremos los parámetros. Los métodos ó técnicas de medición en GPS son muy variados en función del objetivo final que casi siempre se traduce en una determinada precisión. En cualquier caso la forma en que se tratan los datos del GPS desde su observación hasta su resolución final suele ser bastante compleja por lo que es necesario establecer una serie de consideraciones previas en cuanto al procesamiento de datos para que no fracase el proyecto que tengamos entre manos.

7.2. Observables GPS

En una primera aproximación la determinación de la posición de un punto mediante GPS se basa en la medida de la distancia entre el receptor y el satélite. El concepto de observable GPS es una medida de dicha distancia derivada bien de medidas de tiempo, bien de las diferencias de fase, basadas en la comparación entre la señal recibida por el receptor procedente del satélite, y la réplica de dicha señal generada por el receptor. Actualmente los satélites del Bloque IIR(M) emiten un nuevo código C sobre la portadora L2, en el resto de explicaciones se ha ignorado esa novedad. También próximamente habrá una nueva portadora L5, tampoco lo desarrollaremos en el resto del tema. Un excelente estudio de los observables se ofrece en algunos trabajos publicados, como [4] y sobre todo [3].

7.2.1. Captura de datos

En observable de código nos sirve para asimilar el proceso de forma general. En ese caso, el observable es el retardo o tiempo ΔT que tarda la señal en viajar desde el centro de fase de la antena del satélite (instante de emisión) hasta el centro de fase de la antena del receptor (instante de recepción). El valor proporciona la distancia “aparente” entre ambos, $D=c*\Delta T$. A partir de la correlación del código (P o C/A) recibido del satélite con una réplica del mismo que genera el receptor, se obtiene el tiempo de propagación ΔT (ver figura). Hemos de tener en cuenta que en el proceso intervienen dos relojes, el del receptor y el del satélite. Por lo tanto, las distancias medidas estarán afectadas por los errores de dichos relojes y en consecuencia las llamaremos *pseudodistancias*. Como se establece en [2].

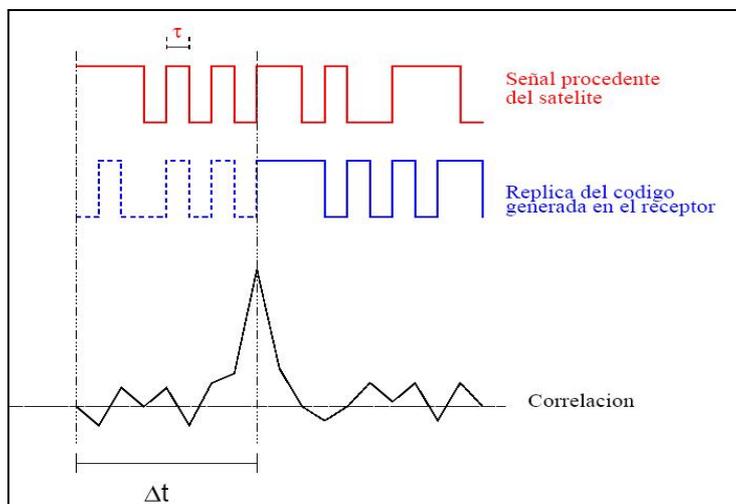


Figura 1. Correlación de señales de código de satélite y receptor

Desde una forma más general, podemos agrupar los observables GPS en dos grupos.

- Observables de tiempo:
 - Código C/A modulado sobre la portadora L1.
 - Código P modulado sobre la portadora L1.
 - Código P modulado sobre la portadora L2.
- Observables de diferencia de fase de la portadora:
 - Diferencia de fase de la portadora L1, que denotamos Φ_1 .
 - Diferencia de fase de la portadora L2, que denotamos Φ_2 .
- Observables de desplazamiento Doppler
 - Desplazamiento Doppler de la portadora L1, que denotamos D_{L1} .
 - Desplazamiento Doppler de la portadora L2, que denotamos D_{L2} .

Con ambos tipos de observables se pueden obtener medidas de la distancia receptor-satélite que nos van a permitir calcular la posición de nuestro receptor.

Pseudodistancias de código. Ecuación de pseudodistancias de código.

Para obtener las pseudodistancias el receptor genera una réplica de la señal emitida por el satélite y compara ambas mediante la correlación del ruido pseudoaleatorio (PRN). Se obtiene entonces un desplazamiento de la señal que se corresponde con el tiempo que tarda la señal en llegar del satélite al receptor. Llamemos t^S a la lectura del reloj del satélite en el momento de la emisión, transmitida vía código PRN, y t_R a la lectura del reloj del receptor en el momento de la recepción. Si ambos tiempos estuvieran referidos al mismo sistema de tiempos y no tuviesen errores, la diferencia de ambas lecturas sería el tiempo de viaje de la señal entre satélite y receptor. Ahora bien, ambos relojes tienen unos retardos o adelantos respecto al sistema de tiempos GPS. Llamemos entonces δ^S y δ_R a la diferencia entre las lecturas de los relojes del satélite y el receptor respecto al sistema de tiempo GPS, respectivamente. La diferencia entre ambas lecturas será entonces la cantidad Δt que corresponde al desfase de tiempo entre la señal del satélite y la del receptor en el proceso de correlación de código en el receptor.

$$\Delta t = t_R - t^S = [t_R(GPS) + \delta_R] - [t^S(GPS) + \delta^S]$$

De forma que si denotamos la diferencia de tiempos en la escala GPS y la diferencia de errores de los relojes como,

$$\Delta t(GPS) = t_R(GPS) - t^S(GPS)$$

$$\Delta \delta = \delta_R - \delta^S$$

obtenemos,

$$\Delta t = \Delta t(GPS) + \Delta \delta$$

Ahora bien, el error δ^S del reloj del satélite se puede modelar mediante un polinomio a partir de los coeficientes transmitidos en la primera parte del mensaje de navegación de cada satélite. Por lo tanto, si consideramos dicha corrección aplicada, la parte debida a los errores de los relojes se reduce simplemente al valor del error del reloj del receptor.

Si multiplicamos el intervalo de tiempo que tenemos Δt por la velocidad de la luz c , obtenemos la medida de pseudodistancia (R) que buscamos:

$$R = c \cdot \Delta t = c \cdot \Delta t(GPS) + c \cdot \Delta \delta = \rho + c \cdot \Delta \delta$$

La precisión de una pseudodistancia obtenida a partir de medidas de código está entorno al 1% de la longitud del chip, por lo que la precisión obtenida a partir de medidas de los códigos C/A y P son 3 m y 0,3 m, respectivamente. Sin embargo, recientemente se ha demostrado que la

precisión puede llegar a ser del orden del 0,1% de la longitud del chip (30 y 3 cm respectivamente).

Para obtener una ecuación de pseudodistancias más genérica, debemos tener en cuenta que la onda electromagnética procedente del satélite atraviesa la atmósfera y por lo tanto sufre retardos debidos principalmente a la Ionosfera y la Troposfera. Introduciendo estos términos en la ecuación obtenemos:

$$R = \rho + c \cdot \Delta\delta + D_{Ion} + D_{Trop}$$

donde el término troposférico tiene un valor aproximado de unos 10 m, siendo igual en valor absoluto pero de signo opuesto para pseudodistancias y fases y diferente para L1 y L2, y el término ionosférico de unos 50 m que es igual para todos los observables. El paso de la señal a través de la Troposfera genera un retardo en el código, lo que se traduce en un aumento en la pseudodistancia

Esta ecuación se puede expresar para cada satélite i desde cada receptor j y en cada una de las frecuencias L1 y L2, escribiremos:

$$R_{jLk}^i = \rho_j^i + c \cdot \Delta\delta_j^i + D_{Ion} + D_{Trop}$$

que se puede obtener a su vez tanto del código C/A como del código P.

Además de los errores por falta de sincronismo de los relojes del receptor y satélite y la propagación de la señal a través de la atmósfera, se tienen en cuenta efectos relativistas, retardos instrumentales, efecto multicamino, y un término final de efectos no modelados.

$$R_{jLk}^i = \rho_j^i + c \cdot \Delta\delta_j^i + D_{Ion} + D_{Trop} + rel_j^i + K1_j^i + MP_j^i + \varepsilon_j^i$$

Pseudodistancias de fase. Ecuación de pseudodistancias de fase.

El observable de fase es la diferencia entre la fase de la portadora recibida del satélite y la fase generada internamente por el oscilador del receptor. Estas medidas de fase se registran en intervalos iguales de tiempos del receptor, y no tienen en cuenta el número de ondas enteras que hay entre el receptor y el satélite. La pseudodistancia que vamos a obtener se deriva de una medida de la fase de la portadora. La distancia satélite-receptor está relacionada con el número entero de longitudes de onda y su fase:

$$D = c \cdot \Delta\varphi + \lambda \cdot N$$

Sea $\varphi^S(t)$ la fase recibida de la portadora, y sea $\varphi_R(t)$ la fase generada en el receptor en un tiempo de recepción t . La medida que calcula el receptor es una función que varía en el tiempo,

$\varphi^S(t) - \varphi_R(t)$, desplazando la fase generada en el receptor, $\varphi_R(t)$, para ir siguiendo la fase recibida $\varphi^S(t)$. Cuando la diferencia de fase aumenta o disminuye en un ciclo (2π), el observable de fase de la portadora $\varphi^S_R(t) = \varphi^S(t) - \varphi^R(t)$ también cambia en un ciclo. Si ni el receptor ni el satélite se mueven, la medida $\varphi^S(t) - \varphi_R(t)$, se mantendrá constante. Si el satélite se aleja en una longitud de onda de segundo. En ese mismo segundo, el receptor generará una onda interna extra y por lo tanto, la medida $\varphi^S(t) - \varphi_R(t)$ será igual a 1.

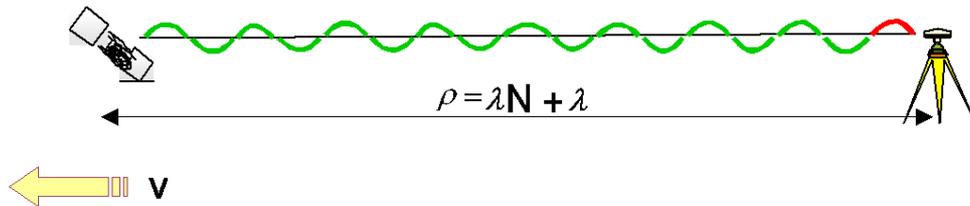


Figura 2. Gráfico de medición de pseudodistancia de fase

Por tanto, un cambio en la fase observada, refleja un cambio en la distancia receptor-satélite, y el observable $\varphi^S_R(t) = \varphi^S(t) - \varphi^R(t)$ cuando se considera a lo largo del tiempo nos da información de la posición relativa entre el satélite y el receptor.

En el instante inicial en que el receptor comienza a seguir la fase recibida del satélite, se desconoce el número entero de longitudes de onda entre el satélite y el receptor. Ésta incógnita es lo que se llama ambigüedad de ciclo inicial N .

Una vez el receptor sigue la fase recibida desde el satélite a partir de una época inicial t_0 , la medida de fase acumulada consistirá en,

$$\varphi^S_R(t) = Fr(n) + Ent(n, t_0, t) + N(t_0)$$

- la medida fraccional de la fase $Fr(n)$,
- la cuenta entera de ciclos $Ent(n, t_0, t)$ que varía con el tiempo,
- un número entero de ciclos desconocido $N(t_0)$ (ambigüedad de ciclo inicial), fijo en el tiempo, a lo largo de todo el seguimiento de la señal de ese satélite.

La diferencia de fase, $\varphi^S(t) - \varphi_R(t)$, depende tanto del estado del reloj del satélite como del estado del reloj del receptor. La fase generada en el receptor, $\varphi_R(t)$, estará referida al tiempo medido por el reloj del receptor. La fase de la señal generada y transmitida por el satélite y recibida en el receptor, $\varphi^S(t)$, estará referida al tiempo medido por reloj del satélite. Para un instante genérico t_g posterior habrá que introducir los correspondientes términos correctivos por estados de ambos relojes, δ^S y δ_R . Siendo la ecuación de ondas $\varphi = f \cdot t$ tenemos

$$\varphi_R^S(t_g) = (\varphi^S(t) - f\delta^S) - (\varphi_R(t) - f\delta_R) = \varphi_R^S(t) - (f\delta^S - f\delta_R)$$

Por otro lado, la distancia real ρ entre un satélite S y un receptor R en ese instante podemos escribirla de la forma (expresada en ciclos $\lambda = c/f$). Si n_R^S es el número entero de ciclos contados desde que el receptor sigue la fase enviada por el satélite y N_R^S es la ambigüedad de ciclo inicial.

$$\rho_R^S\left(\frac{1}{\lambda}\right) = \varphi_R^S - \left(\delta^S \cdot \frac{c}{\lambda} - \delta_R \cdot \frac{c}{\lambda}\right) + n_R^S + N_R^S$$

Pero como hemos visto, para un instante genérico t_g , la ‘lectura’ de fase en el receptor nos dará la fase φ (parte fraccionaria) más el número entero de longitudes de onda o ciclos n que ha crecido (o decrecido) esta fase desde el instante inicial en que se mantiene el seguimiento al satélite.

$$\Phi_R^S(t_g) = \varphi_R^S(t_g) + n_R^S \quad (\text{que es lo realmente medido por el receptor})$$

Sustituyendo en la ecuación de distancia anterior:

$$\rho_R^S\left(\frac{1}{\lambda}\right) = \Phi_R^S - \left(\delta^S \cdot \frac{c}{\lambda} - \delta_R \cdot \frac{c}{\lambda}\right) + N_R^S$$

Si suponemos ahora que el error del reloj del satélite lo podemos eliminar a partir de los datos de estado del reloj del mensaje de navegación y dejamos las incógnitas a la derecha de la igualdad, y generalizando la ecuación para un receptor A y un satélite j en un instante t.

$$\Phi_A^j(t) - \frac{c}{\lambda} \delta^j(t) = \frac{1}{\lambda} \rho_A^j(t) - N_A^j - \frac{c}{\lambda} \delta_A(t)$$

y si además incluimos las correcciones por retardos ionosféricos y troposféricos, expresado siempre en ciclos, obtendríamos finalmente la expresión de la ecuación de fase general como:

$$\boxed{\Phi_A^j(t) - \frac{c}{\lambda} \delta^j(t) = \frac{1}{\lambda} \rho_A^j(t) - N_A^j - \frac{c}{\lambda} \delta_A(t) - d_{Aion}^j(t) + d_{Atrop}^j(t)}$$

El término troposférico es negativo puesto que el paso de la señal a través de la Troposfera genera un retardo en la fase de la portadora recibida por el receptor. La fase puede ser medida por un receptor mejor que 0,01 ciclos, lo cual implica una precisión mejor de 2 – 3 mm.

Pseudodistancias Doppler. Ecuación de pseudodistancias de Doppler.

Algunos de los primeros modelos propuestos para el posicionamiento mediante GPS se basaban en las medidas Doppler de forma análoga a como se trabajaba con el sistema TRANSIT. Las medidas Doppler se basan en el desplazamiento que sufre la fase procedente de un emisor en

movimiento respecto del receptor. Utilizaba los desplazamientos Doppler integrados o *cuentas Doppler*, es decir, diferencias de fase que se escalaban mediante diferencias de distancias. En nuestro caso, los datos brutos Doppler son linealmente dependientes de la velocidad radial, lo que nos permite determinar en tiempo real la velocidad (muy útil para la navegación).

La ecuación de pseudodistancias para medidas Doppler es por tanto,

$$D = \lambda \cdot \frac{d\Phi}{dt} = \frac{d\rho}{dt} + c \cdot \Delta \frac{d\delta}{dt} = \dot{\rho} + c \cdot \Delta \dot{\delta}$$

En ella D es el desplazamiento Doppler observado, $\dot{\rho}$ es la velocidad radial instantánea entre el satélite y el receptor, y $\Delta \dot{\delta}$ es la derivada respecto al tiempo del error combinado de los relojes del satélite y el receptor. Una estimación de la precisión es de 0.001 Hz, lo cual corresponde a 0.3 m/s si el desplazamiento Doppler es medido en el código C/A. Los datos de desplazamiento Doppler son aplicados para determinar ambigüedades enteras en medidas cinemáticas o también como un observable adicional para posicionamiento simple.

7.2.2. Combinación de datos

Los observables principales del GPS, si tenemos en cuenta que el código C/A sólo se modula en L1 y que el código P en L1 y L2 para una época determinada, son las fases ϕ_1 y ϕ_2 , sus correspondientes desplazamientos doppler D_{L_1} y D_{L_2} , y las pseudodistancias derivadas $R_{L_1,C/A}$, $R_{L_1,P}$ y $R_{L_2,P}$, siempre que el receptor que estemos considerando sea de doble frecuencia. Existen diversas combinaciones de datos posibles en receptores de doble frecuencia que nos aportan ventajas para el procesado de los datos, así como la posibilidad de suavizado de medidas de código. No tendremos en cuenta las observaciones doppler. L1 y L2 representa los observables de fase (ya sea de las cero o las dobles diferencias) en metros. Información práctica de su posible uso en el cálculo se encuentra en [1].

Combinaciones lineales de fase .

La combinación lineal de dos fases ϕ_1 y ϕ_2 viene dada por la expresión:

$$\varphi = n_1\phi_1 + n_2\phi_2$$

con n_1 y n_2 como números arbitrarios y $\varphi_i = f_i \cdot t$ para las frecuencias L_1 y L_2 queda

$$\varphi = n_1 f_1 t + n_2 f_2 t = f \cdot t \quad \text{de forma que} \quad f = n_1 f_1 + n_2 f_2$$

es la frecuencia resultante de la combinación con una longitud de onda $\lambda = c/f$.

En algunos casos las propiedades de estas combinaciones también se cumplen para los observables de código, en las mismas condiciones que para los de fase.

Combinación libre ionosfera L3

Se llama así porque el retardo ionosférico queda virtualmente eliminado, por eso su nombre. L_1 y L_2 representa los observables de fase (ya sea de las cero o las dobles diferencias) en metros.

$$L_3 = \frac{f_1^2 L_1 - f_2^2 L_2}{f_1^2 - f_2^2}$$

La forma de la combinación tiene que ver con el hecho de que el efecto de la ionosfera depende del cuadrado de la frecuencia ($\alpha_i = 40.3 / f_i^2$).

Combinación ionosférica o de libre geometría L4

Esta combinación cancela la parte geométrica de la medida (órbitas, coordenadas), quedando únicamente el efecto de la ionosfera y las constantes instrumentales (multipath y ruido). Lógicamente, también contiene las ambigüedades iniciales de ciclo. Es útil para detectar saltos de ciclo de fase y modelar la ionosférica.

$$L_4 = L_1 - L_2$$

Combinación wide-lane o banda ancha Lw

Proporciona un observable de $\lambda = 86.2$ cm, cuatro veces superior a L_1 y L_2 , factor muy útil para localizar saltos de ciclo en la fase (esto utilizando la combinación W siguiente).

También se utiliza para cálculo de ambigüedades.

$$L_w = \frac{f_1 L_1 - f_2 L_2}{f_1 - f_2}$$

El problema principal del uso de cualquiera de las combinaciones lineales de fase es que, asumiendo un cierto nivel de ruido en la medida de la fase, el nivel de ruido en la combinación es mayor. Así, por ejemplo, si aplicamos la ley de propagación de errores considerando el mismo nivel de ruido en ambas fases, obtenemos que el ruido, tanto en $\Phi_{L_1+L_2}$ como en $\Phi_{L_1-L_2}$, es mayor en un factor $\sqrt{2}$ que el ruido de cualquiera de las dos fases simples. Existen,

no obstante, combinaciones de código y fase que evitan este problema y que se basan en el suavizado de las pseudodistancias de código por medio de las pseudodistancias de fase.

Combinación de pseudodistancias de código y fase, suavizado de código.

En las combinación de obserbables de código y fase uno de objetivos es conseguir pseudodistancias de código suavizadas por medio de pseudodistancias de fase. Actualmente las combinaciones de pseudodistancias de fase y código son una parte importante de la determinación del posicionamiento en tiempo real. La fase multiplicada por la longitud de onda es una distancia, y por tanto, las diferencias de fase (multiplicadas por la longitud de onda) son diferencias de distancias. Si no existiera retardo ionosférico, las diferencias de pseudodistancias y las diferencias de fase, serían iguales para un satélite determinado y en dos instantes de tiempo sucesivos (siempre que no haya pérdida de ciclos). Pero la ionosfera produce un retardo en las pseudodistancias y un adelanto en la fase.

El funcionamiento básico de un algoritmo de suavizado lo tenemos en el siguiente ejemplo. Si consideremos una serie de n observaciones a un determinado satélite, siendo P las pseudodistancias y C los valores de la fase en metros (nº de ciclos * longitud de onda):

Epoca	Pseudodistancia	Fase (m)
1	P ₁	C ₁
2	P ₂	C ₂
..
n	P _n	C _n

En la hipótesis del algoritmo, efectuamos las diferencias:

$$\begin{array}{lll}
 P_n - P_1 = C_n - C_1 & \Rightarrow & P_n = P_1 + C_n - C_1 \\
 P_n - P_2 = C_n - C_2 & \Rightarrow & P_n = P_2 + C_n - C_2 \\
 \dots\dots\dots & \Rightarrow & P_n = P_{n-1} + C_n - C_{n-1} \\
 P_n - P_n = C_n - C_n & \Rightarrow & P_n = P_n + C_n - C_n
 \end{array}$$

Sumando y tomando la media:

$$\bar{P}_n = \frac{P_1 + P_2 + \dots + P_n}{n} + C_n - \frac{C_1 + C_2 + \dots + C_n}{n}$$

Ordenando términos, se obtiene finalmente:
$$\bar{P}_n = \frac{P_n}{n} + (\bar{P}_{n-1} + C_n - C_{n-1}) \cdot \frac{n-1}{n}$$

siendo \bar{P}_{n-1} la pseudodistancia suavizada de la época anterior.

Este filtro de suavizado de código (pseudodistancia) reduce el efecto multipath y el ruido del receptor.

En cuanto al uso combinado de código y fase para determinar saltos de ciclo, tenemos la Combinación Melbourne-Wübbena (W). Es una combinación de código P y fase en ambas frecuencias que elimina el efecto de la ionosfera, medida geométrica (órbitas, coordenadas), estados de reloj y troposfera. Por ello se usa para determinar y reparar saltos de ciclo.

$$W = L_W - P_W = \frac{f_1 L_1 - f_2 L_2}{f_1 - f_2} - \frac{f_1 P_1 + f_2 P_2}{f_1 + f_2}$$

7.2.3. Fuentes de error en GPS

Como en toda observación geodésica o topográfica, cualquier observación con GPS está sometida a diversas fuentes de error. Con el fin de corregir o aminorar estos errores es importante conocer sus causas y así saber como minimizarlos, si no se pueden eliminar completamente. En el caso de que no se puedan anular o modelar, al menos prever cuales pueden ser sus valores máximos. Podemos agruparlas en tres causas diferentes las provenientes del satélite, las que afectan mientras se produce la propagación de la señal y las que nos afectan por el equipo de medida. En la siguiente tabla podemos ver las más importantes. Información más completa se encuentra en [3], [5] y [6].

ORIGEN	FUENTE DE ERROR
Satélite	Errores en el oscilador (reloj).
	Errores en los parámetros orbitales (efemérides transmitidas).
	Disponibilidad Selectiva (S/A).
	Anti-spoofing (AS).
	Efecto relativista.
Propagación de la señal	Refracción ionosférica.
	Refracción troposférica.
	Pérdidas de ciclos.
	Multitrayectoria.
Equipo	Errores en el oscilador (reloj).
	Variación y desfase del centro de la antena.
	Error en el estacionamiento y manipulación.
	Ruido e incertidumbre de medida.
	Retardos instrumentales.

Tabla 1. Clasificación de las fuentes de error

Algunos de estos errores pueden ser modelados e incluso eliminados utilizando combinaciones apropiadas de los observables a partir de una o dos frecuencias, o trabajando en modo diferencial, utilizando dos receptores. La modelación de los errores se realiza dejando en las ecuaciones de observación a los parámetros del modelo como incógnitas que se conocerán cuando se resuelva el sistema de ecuaciones.

Errores dependientes del satélite.

Errores en el oscilador

La causa es el desfase del reloj del satélite respecto al Tiempo GPS. Los satélites llevan relojes atómicos con osciladores de cesio o de rubidio, pero incluso estos relojes u osciladores necesitan ser corregidos de sus imperfecciones.

Los errores en los osciladores de los satélites pueden eliminarse mediante las correcciones enviadas en el mensaje de navegación que recibe cada receptor, las cuales son calculadas, enviadas y actualizadas por las estaciones de seguimiento. Se corrige la deriva de cada reloj de cada satélite con más de 10 relojes atómicos en tierra muy precisos. Para cada reloj de satélite se determina el desfase respecto a una época inicial y los coeficientes de la marcha o deriva del estado del reloj. Generalmente lo que suele ocurrir es un adelanto del reloj debido a que el satélite está situado en un campo gravitatorio más débil. Por eso se diseñan los relojes de tal

forma que en la superficie terrestre atrasen y al ponerlos en órbita funcionen bien, pero aun así existe una deriva de aproximadamente 1 ns cada tres horas.

En cualquier caso, la corrección de esta fuente de error es casi total excepto un pequeño error residual estimado en unos pocos nanosegundos que no es posible predecir. Si no se corrige estimar el error producido en 1 m. Ya que un error de 1 nanosegundo de imprecisión en el reloj del satélite produce 30 cm en la pseudodistancia.

Para modelarlo, se utiliza una ecuación de segundo grado, de la forma,

$$dt^j = a_0 + a_1 (t - t_0) + a_2 (t - t_0)^2$$

donde el parámetro a_0 representa la deriva del reloj, a_1 la evolución de la deriva y a_2 un componente de aceleración de deriva respecto a un tiempo t_0 . Estos parámetros para cada satélite los proporciona el mensaje de navegación de la señal.

Errores en los parámetros orbitales del satélite.

Las efemérides transmitidas por los satélites en el mensaje de navegación tienen asociado un error ya que en principio no es imposible predecir exactamente sus posiciones.

Los satélites se desvían de las órbitas calculadas por diferentes razones, la variación del campo gravitatorio, de la presión de radiación solar ó la fricción del satélite con moléculas libres. En el mensaje de navegación se transmiten parámetros de corrección basados en datos empíricos para reducir los errores de estas fuentes.

Además no se cumplen exactamente las leyes de Kepler para el caso ideal, la Tierra no es una esfera perfecta ni su distribución de masas es homogénea otras masas como los planetas, el Sol y la Luna ejercen atracción sobre el satélite. Por otro lado, el satélite no está viajando por un perfecto "vacío" y experimentará fricciones atmosféricas. Finalmente el impacto de fotones de luz emitidas por el sol directa e indirectamente (efecto albedo) también perturba la órbita. Esto es conocido como presión de radiación solar y será función del área efectiva del satélite la superficie de reflectividad, la luminosidad del Sol y la distancia al Sol. La mayoría de todos estos efectos son modelados, aunque es difícil en el caso de la presión de radiación solar. Las fuerzas perturbadoras sobre el satélite son cuantificadas en aceleraciones perturbadoras. Los parámetros de perturbación se expresan finalmente como:

- Corrección al movimiento medio calculado.
- Magnitud del cambio de la ascensión recta.
- Magnitud del cambio de la inclinación.
- Términos de corrección en forma de seno y coseno del argumento de la latitud.
- Términos de corrección en forma de seno y coseno del radio geocéntrico.
- Términos de corrección en forma de seno y coseno de la inclinación del plano orbital.

Si se utilizan efemérides precisas, como las del IGS, se reducen las incertidumbres significativamente. Trabajando en modo diferencial podemos eliminar casi todos los errores relativos a los satélites, ya que afectan de forma casi igual a ambos receptores.

Disponibilidad Selectiva (SA)

Con el fin de limitar las precisiones del sistema el Departamento de Defensa de los EE.UU. implementó la Disponibilidad Selectiva, *Selective Availability* (SA). Es una degradación intencionada de la señal para el usuario civil. En esta degradación de la señal se actuaba sobre los estados de los relojes (dither) y los parámetros orbitales (epsilon). "Dither" es una manipulación de la frecuencia del reloj del satélite, de tal forma que en la generación de la onda portadora y los códigos se varían las longitudes de onda de la portadora. En otras palabras, la distancia entre cada chip del código C/A puede variar, no siendo los 293 metros diseñados. La réplica generada en el receptor asume la longitud nominal y las medidas de pseudodistancia están basadas en ello. La componente "epsilon" de la SA se refiere a un error en las efemérides transmitidas, es decir, en la posición del satélite. Para posicionamiento, las coordenadas de los satélites estaban alteradas y los errores en esas coordenadas se propagaban en la posición calculada por el receptor. El 1 de Mayo de 2000, Bill Clinton anunció el fin de la degradación intencionada de la señal GPS, llamada SA. Esto significaba que los usuarios civiles podían determinar coordenadas 10 veces más precisas que antes (unos 10 metros). En GPS diferencial, el error introducido en ambos receptores era similar, con lo cual este error se eliminaba. La magnitud de los errores que se introducían eran de unos 100m en los peores casos.

Anti-Spoofing.

Consiste en encriptar el código P mediante el uso del llamado código protegido Y, mezclando P+W de tal forma que solamente usuarios autorizados tienen acceso al código P. Como solución, algunos receptores tienen desarrolladas técnicas para hacer medidas de código P con sólo añadir un poco de ruido, son las técnicas de correlación cruzada como se describe en [8]. La correlación cruzada se basa en el principio de que el código Y transmitido por un satélite es el mismo en ambas frecuencias. Por lo tanto, correlando los dos códigos Y en L1 y L2, la diferencia entre los respectivos tiempos de viaje puede ser deducida. Esta diferencia es igual al tiempo de retardo que la frecuencia sufre cuando pasa por la ionosfera. Sumando la diferencia de retrasos de tiempo a la medida de código C/A de la L1 resulta una medida de pseudorange conteniendo la misma información que la medida de código P en L2.

Efecto relativista

Esta íntimamente relacionado con los estados del reloj de los satélites. Éstos se adelantan debido a que el satélite está situado en un campo gravitatorio más débil lo que se llama relatividad general, y que la diferente velocidad entre emisor y receptor, lo que se llama relatividad especial. Se modela mediante una componente constante que depende del valor del semieje mayor de la órbita que corrige la frecuencia del oscilador del satélite y una componente periódica debida a la excentricidad de la órbita que corrige el receptor.

Errores dependientes de la propagación de la señal.

Pueden ser predecibles (por ejemplo, las causadas por la geometría) o de naturaleza aleatoria (como las turbulencias atmosféricas). Un medio en el cual la velocidad de propagación de las ondas electromagnéticas depende de la frecuencia es un medio dispersivo. En tal medio la refractividad depende de la frecuencia o la longitud de onda. Información más completa la encontramos en [7] y [8].

El efecto de dispersión es causado por interacciones electromagnéticas entre un campo cargado eléctricamente del medio y un campo externo de ondas penetrantes. Habrá una velocidad de propagación de la fase de una onda particular con longitud de onda uniforme (velocidad de fase v_{ph}) y una velocidad de propagación de un grupo de ondas, generada por la superposición de ondas diferentes de diferente longitud de onda (velocidad de grupo v_{gr}). La relación entre ambas velocidades viene dada por la ecuación de Rayleigh:

$$v_{gr} = v_{ph} - \lambda \frac{dv_{ph}}{d\lambda}$$

Ionosfera

La Ionosfera es la región de la atmósfera comprendida entre 100 y 1000 Km de altitud, donde las radiaciones solares y otras radiaciones ionizan una porción de las moléculas gaseosas liberando electrones, que interfieren en la propagación de ondas de radio. La Ionosfera es un medio dispersivo para ondas de radio, por lo tanto su índice de refracción es función de la frecuencia de la onda. Este error es negativo para la medida de fase (se produce un avance de la portadora y se miden distancias más pequeñas), y positivo para las pseudodistancias (se produce un retardo y se miden distancias más largas), pero tienen el mismo valor absoluto. En otras palabras, el código GPS es retrasado resultando las pseudodistancias de código más largas comparadas con la distancia geométrica al satélite y las de fase, más cortas.

El error es proporcional a la densidad de electrones o *Total Electron Content* (TEC) a lo largo del camino seguido por la señal y éste depende de cinco factores fundamentalmente la latitud geomagnética del receptor, la hora del día, la elevación del satélite las variaciones estacionales y diurnas y la actividad del Sol (ciclos de 11 años). Se producen grandes retardos para señales emitidas desde satélites con baja elevación, ya que viajan a lo largo de una gran sección de la ionosfera, con un pico en las horas centrales del día y bajando durante la noche debido a la radiación solar. En regiones cerca del ecuador geomagnético o de los polos, también el retardo es mucho mayor.

Para conocer el valor el TEC se pueden utilizar modelos ionosféricos como el de Klobuchar, medir el TEC como hacen algunos organismos como el CODE, explicado en [9] y [10], calculando modelos en tiempo casi real a partir de observaciones y los ponen a disposición pública (cada 2 horas), ó eliminar el efecto del TEC que es el método más eficiente y preciso. Esta eliminación se produce mediante la utilización de las dos observaciones de doble frecuencia. Es la principal razón también de que la señal GPS tenga dos portadoras L1 y L2, haciendo combinaciones lineales de observaciones de fase en estas dos frecuencias se consigue prácticamente eliminar su efecto. La combinación lineal que elimina el efecto es la llamada "libre ionosfera" o L_3 :

$$L_3 = \frac{1}{f_1^2 - f_2^2} (f_1^2 L_1 - f_2^2 L_2)$$

Su valor puede llegar a ser de hasta 100 m en pseudodistancia, normalmente unos 10 m. En modo diferencial, el retardo afectará por igual a estaciones siempre que estén cercanas. Esto no tiene porqué cumplirse en líneas base largas. La combinación libre ionosfera se recomienda aplicar en líneas base a partir de 10 km.

Troposfera

Es la capa más baja de la atmósfera, contiene vapor de agua, con lo cual el índice de refracción para un área parcial es función de su temperatura, de la presión y del contenido de vapor de agua. La Troposfera es la última capa de la atmósfera (hasta unos 80 Km) donde se produce retardo y donde las temperaturas decrecen con el incremento de altura. El espesor de la Troposfera no es el mismo en todas las zonas, es irregular y los primeros 40 Km son los que más afectan la señal. La presencia de átomos y moléculas neutros en la Troposfera afecta a las señales de propagación electromagnética. Esta atmósfera neutra es un medio no dispersivo con respecto a las ondas de radio de frecuencias superiores a 15 GHz, por lo tanto, la propagación es independiente de la frecuencia, y esto precisamente hace que el tratamiento y la eliminación del retardo troposférico no se pueda realizar como en la ionosfera, con diferentes frecuencias.

La componente húmeda varía espacialmente y temporalmente, mientras que la seca permanece más estable. La componente seca es la causante de un 90% del total del retardo y puede ser obtenida con precisión de algunos milímetros a partir de medidas de presión en superficie. La componente húmeda es función del vapor de agua a lo largo del camino de la señal, y es difícilmente evaluable. El simple uso de medidas meteorológicas en superficie no puede dar la precisión alcanzable con los radiómetros de vapor de agua. Estos instrumentos miden la radiación basal que se recibe desde el espacio en la dirección de la observación, y son capaces de medir el contenido de vapor de agua en la atmósfera. En la práctica, se introducen modelos de refractividad para la componente seca y la húmeda.

Por lo tanto se intenta conocer el efecto de la troposfera utilizando modelos atmosféricos simplificados como Hopfield, Saastamonien, Hopfield modificado, Goad y Goodman, Black, Essen y Froome, etc. Los parámetros que definen el comportamiento concreto de cada uno de estos modelos pueden ser incógnitas en el proceso de resolución del sistema de ecuaciones de observación.

El retardo troposférico causa un error de 1.9-2.5 m en la dirección cenital y se incrementa cuando decrece el ángulo, llegando a ser de 20-28 m a unos 5°. Los modelos pueden llegar a corregir el error hasta dejarlo en 1 – 5 cm.

Pérdidas de ciclo.

Las pérdidas de ciclos suponen un salto en el registro de las medidas de fase por la interrupción o pérdida de la señal enviada por el satélite a causa de obstáculos que es sin duda es la causa más frecuente, la baja calidad de la señal debido a una baja elevación del satélite, malas condiciones ionosféricas, multipath, fallos en el receptor incluso a un mal funcionamiento del oscilador del satélite que es menos probable.

Los receptores, lo que miden es la diferencia entre la fase de la portadora transmitida del satélite y la fase de la señal réplica que genera el receptor. Esta medida puede estar entre 0 y 1 ciclo (0 y 2π). Durante el registro, el contador se incrementa en una unidad cuando la fase (fraccional) cambia de 2π a 0. El número entero inicial de ciclos entre el satélite y el receptor no se conoce y tiene que ser calculado (ambigüedad). Esta cantidad o ambigüedad inicial de fase permanece siempre que no exista pérdida de señal. Una pérdida de señal causa un salto en la fase acumulada de un número entero de ciclos entre un tiempo t_1 y un tiempo t_2 . Obviamente, esto afecta únicamente a las medidas de fase.

La detección es sencilla por medio de un chequeo. Una vez detectado que hay un salto de ciclo y el tamaño de la pérdida de ciclo, la reparación se hace corrigiendo a todas las observaciones de fase siguientes para este satélite, según una cantidad fija. A veces el software interno del receptor es capaz de detectar y corregir estas pérdidas.

En realidad la formulación para la detección y reparación de saltos de ciclo es bastante amplia e incluye observaciones de código y de fase, combinaciones de fases, combinaciones de código y fase o combinaciones de fase con frecuencia Doppler para su detección. El chequeo “in situ” es muy importante porque permite detectar el salto y corregirlo mediante un software interno en el receptor para una estación.

Multipath o Multitrayectoria.

Es causado por múltiples reflexiones de la señal emitida por el satélite en superficies cercanas a la antena. La consecuencia es que las señales recorren un camino más largo y puede distorsionar la amplitud y forma de la onda. Se pueden agrupar los errores de multipath en tres clases diferentes, los de difusión proveniente de un área grande (por ejemplo, si la señal pasa a través de una tela metálica), los de reflexión especular en objetos bien definidos o superficies

reflectantes al lado de la antena y fluctuaciones de muy baja frecuencia asociados generalmente con reflexión en la superficie del agua.

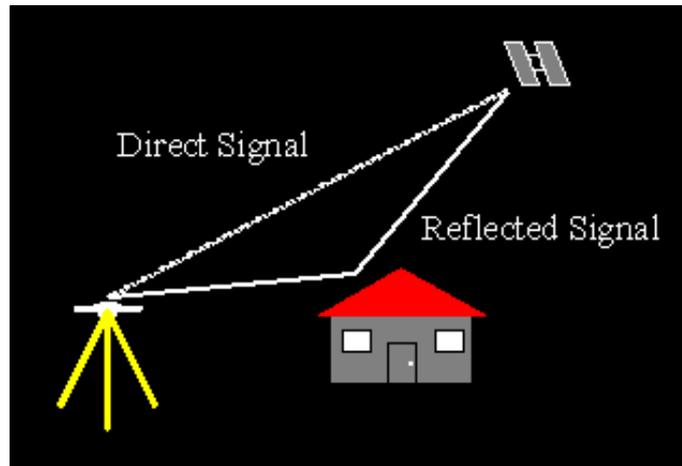


Figura 3. Multitrayectoria.

Las técnicas de GPS diferencial no eliminan los efectos de multipath, puesto que es dependiente del sitio de observación, sin embargo el equipamiento y la elección de un buen sitio de estación sí que evitan que se produzca este indeseado efecto en una observación GPS. Imponer una máscara de elevación, ya que con señales procedentes de satélites a baja altura, el efecto será mayor o más fácil que pueda darse. El diseño de la antena reduce considerablemente el efecto, mediante antenas con ‘anillos de choque’ conocidas como *choke ring* ó con antenas con ‘plano de tierra’, para evitar ondas que provengan de la parte inferior de la antena, se reducen las interferencias de señales con baja elevación o procedentes de multipath.

El efecto depende de la frecuencia y por ello las medidas de fase están menos afectadas que las de código, donde pueden llegar a 1 m. Incluso en observaciones de código pueden ser del orden de hasta 10-20 m e incluso 100 m en las cercanías de edificios. Casos extremos de multipath producen pérdidas de ciclo o señal.

Errores dependientes del receptor.

Estado de reloj de receptor

El oscilador del GPS se usa para generar la señal réplica. Cuando el receptor recibe una señal su reloj interno tendrá un desfase respecto a la escala de tiempo. Sin embargo, este error afectará por igual a todas las medidas de los satélites de los que están registrando simultáneamente. En observación de navegación ó código para determinar la posición se necesitan 3 satélites y uno más para determinar el error de reloj del receptor. Esto hará que se pueda ajustar el reloj de

nuestro receptor con una precisión como si fuera atómico. Los receptores corrigen su reloj cada segundo siguiendo una observación a un satélite, por eso a veces se emplean receptores GPS simplemente para tener una referencia precisa de tiempos. Se eliminarán trabajando con posicionamiento relativo por medidas de fase, planteando las ecuaciones de dobles diferencias. El offset del reloj del receptor se estima al mismo tiempo que sus coordenadas. En receptores geodésicos o topográficos estos errores se minimizan, debido a la mayor precisión de sus relojes.

Variación del centro de fase de la antena.

Se debe a la falta de coincidencia entre el centro radioeléctrico o punto al que realmente llega la señal y el centro mecánico o físico, generando un error residual por excentricidad que puede ser de unos milímetros. Además, este centro de fase varía en función de la altura y azimut de los satélites observados. Para receptores de doble frecuencia, habrá dos centros de fase, para L1 y L2, con sus diferentes variaciones, su valor es variable, dependiendo del tipo de antena en el mercado. Todas las antenas del mercado tienen su calibración con los correspondientes valores de variación del centro de fase. Estas tablas de calibración de cada marca y modelo de antena pueden ser relativas, respecto a una antena de referencia supuesta casi no afectada ó actualmente las calibraciones absolutas obtenidas con robots en cámaras anecoicas. Estas calibraciones absolutas pueden ser genéricas a una marca ó modelo ó individuales para cada antena específica. Al trabajar en modo diferencial, se han de orientar todas las antenas hacia el mismo punto aproximadamente ya que en fábrica se montan todas las antenas con la misma orientación en la carcasa. Información práctica sobre modelos, relativos ó absolutos, y su uso en [1].

Incertidumbre de medida.

Cualquier medida electrónica está sujeta a un error de medida aleatorio (o ruido).

El error aleatorio es considerado como la desviación con respecto a cero de las medidas hechas. Para la mayoría de los receptores de precisión, la incertidumbre en la medida de fase es de unos 2 mm o incluso 1 mm en condiciones ideales (geometría satélites, actividad atmosférica, obstáculos...) y de algún centímetro para medidas de código, aunque lógicamente dependerá de la calidad del receptor GPS.

Retardos instrumentales.

Antenas, cables y filtros utilizados en receptores y satélites los generan. Se modelan descomponiéndolos en los retardos del satélite y los del receptor. Los primeros se transmiten en

el mensaje de navegación (Total Group Delay) de cada satélite, y los segundos se modelan incluyéndolos en el offset del reloj del receptor.

7.3. Modelos matemáticos para el posicionamiento. Métodos de posicionamiento

Dependiendo del tipo de aplicación y el nivel de precisión que se quiera obtener usando el GPS, existen importantes ventajas y desventajas formando ciertas combinaciones lineales de las ecuaciones básicas de pseudodistancias y fase de la portadora. Utilizando el GPS, se puede realizar bien *posicionamiento puntual o absoluto*, bien *posicionamiento relativo* entre puntos. Debido a las incertidumbres en la posición de los satélites, el comportamiento de los relojes y los retardos en la propagación de las señales, el posicionamiento absoluto sólo alcanza precisiones próximas a algunos metros. Para la mayoría de las necesidades en aplicaciones geodésicas o geodinámicas, es necesario utilizar el GPS en posicionamiento relativo. Las fuentes de error que afectan a la señal GPS, órbitas de satélites, relojes de los satélites y receptores, errores debidos a la propagación atmosférica muestran algún tipo de correlación entre señales recibidas en varias estaciones que se encuentren recibiendo la señal de varios satélites simultáneamente. La finalidad del posicionamiento diferencial es aprovechar estas correlaciones para mejorar la precisión de las posiciones relativas.

Actualmente, los métodos del posicionamiento se han visto mejorados con su ejecución en tiempo real. Así en el caso de el código, tendremos el GPS diferencial, DGPS, que transmite correcciones relativas al código en tiempo real y en el caso de la medida de fase el cinemático en tiempo real RTK. Los fundamentos matemáticos tanto de uno como del otro, son los mismos que en el caso de tiempo diferido.

7.3.1. Posicionamiento absoluto

El posicionamiento absoluto se puede llevar a cabo utilizando los datos de pseudodistancias obtenidas con código y con fase de la portadora. A continuación vemos los tres casos.

Posicionamiento absoluto con medidas de código

Las pseudodistancias de código en una época determinada t puede escribirse como,

$$R_i^j(t) = \rho_i^j(t) + c \cdot \Delta\delta_i^j(t)$$

donde $R_i^j(t)$ es la medida de pseudodistancia de código entre el observador i y el satélite j , $\rho_i^j(t)$ es la distancia geométrica entre el satélite y el observador, y c es la velocidad de la luz.

El último término, $\Delta\delta_i^j(t)$, es el error combinado de los relojes del receptor i y el satélite j .

Si examinamos la ecuación anterior, las coordenadas del observador (nuestras incógnitas) se encuentran de forma implícita en la distancia geométrica puesto que:

$$\rho_i^j(t) = \sqrt{(X^j(t) - X_i)^2 + (Y^j(t) - Y_i)^2 + (Z^j(t) - Z_i)^2}$$

donde $X^j(t), Y^j(t), Z^j(t)$ son las componentes del vector de posición geocéntrico (coordenadas) del satélite en la época t y X_i, Y_i, Z_i son las tres incógnitas de las coordenadas del observador. Si consideramos el término $\Delta\delta_i^j(t)$ para una época, cada satélite contribuye con una incógnita en el estado del reloj. Si por un momento omitimos el error en el reloj del receptor del observador, la ecuación de pseudodistancias tiene cuatro incógnitas, las tres coordenadas del observador más el error del reloj del satélite. Cada satélite adicional genera una ecuación más con las mismas coordenadas del observador como incógnitas, más otra incógnita del error de su propio reloj, de manera que siempre obtenemos más incógnitas que ecuaciones. Incluyendo una época adicional, los nuevos errores en los relojes de los satélites se pueden modelar con su deriva. Afortunadamente, la información de los relojes de los satélites son conocidos y se transmiten vía el mensaje de navegación en forma de tres coeficientes de a_0, a_1 y a_2 respecto a una época t_0 , así $\delta^j(t) = a_0 + a_1(t - t_0) + a_2(t - t_0)^2$ que nos permite el cálculo del error del satélite para una época cualquiera t . Este polinomio elimina la mayor parte del error debido al reloj del satélite, si bien una pequeña parte siempre permanece presente.

El término del error combinado se separa en dos partes, $\Delta\delta^j(t) = \delta^j(t) - \delta_i(t)$, donde la parte relacionada con el satélite la consideramos ya conocida, y el término relacionado con el receptor sigue siendo incógnita. Substituyendo la última igualdad:

$$R_i^j(t) = \rho_i^j(t) + c\delta^j(t) - c\delta_i(t)$$

Ahora las incógnitas son cuatro, las tres coordenadas del observador y el error del reloj del receptor $\delta_i(t)$. De manera que las incógnitas se pueden calcular de forma inmediata si se observan cuatro satélites de forma simultánea.

Si llamamos ahora n_j al número de satélites, n_t al número de épocas, el producto $n_j \cdot n_t$ nos da el número de ecuaciones de observación de que se dispone. Para obtener una única solución, el número de incógnitas no puede ser mayor que el número de ecuaciones de observación. Si colocamos las incógnitas en el lado derecho de la igualdad, tenemos:

$$R_i^j(t) - c\delta^j(t) = \rho_i^j(t) - c\delta_i(t)$$

con lo que la condición se convierte en que, $n_j \cdot n_t \geq 3 + n_t$

Posicionamiento absoluto con medidas de fase de la portadora

Como ya vimos, las pseudodistancias se pueden obtener también a partir de las medidas de fase. El modelo matemático para estas medidas es,

$$\phi_i^j(t) = \frac{1}{\lambda} \rho_i^j(t) + N_i^j + f^j \Delta \delta_i^j(t)$$

Aquí, $\phi_i^j(t)$ es la fase portadora medida expresada en ciclos, λ es la longitud de onda, y $\rho_i^j(t)$ es la misma distancia que para el modelo de distancia código, es decir, la distancia geométrica entre el satélite y el receptor. La ambigüedad fase, N_i^j , independiente del tiempo es un número entero y, por lo tanto, a menudo es llamado *ambigüedad entera* o incógnita entera. El término f^j denota la frecuencia de la señal del satélite en ciclos por segundo, y $\Delta \delta_i^j(t)$ es el sesgo del receptor del reloj y el satélite combinado, que si lo sustituimos en la ecuación, queda:

$$\phi_i^j(t) = \frac{1}{\lambda} \rho_i^j(t) + N_i^j + f^j \delta^j(t) - f^j \delta_i(t)$$

ecuación en la que se vuelve a suponer conocido el error en el reloj del satélite. Así, para n_j satélites, n_t épocas y un observador i , tenemos otra vez $n_j \cdot n_t$ posibles medidas (ecuaciones). Teniendo en cuenta que la ambigüedad entera es una incógnita para cada satélite, la condición para que se obtenga una única solución se transforma ahora en, $n_j \cdot n_t \geq 3 + n_j + n_t$.

7.3.2. Posicionamiento relativo

El objetivo del posicionamiento relativo es determinar las coordenadas de un punto desconocido con respecto a otro de coordenadas conocidas, es decir, permite la determinación del vector entre dos puntos, normalmente llamado *baselínea*. Supongamos un punto de referencia A de coordenadas conocidas, un punto B de coordenadas desconocidas y el vector *baselínea* \underline{b}_{AB} . Denotemos además sus correspondientes vectores de posición \underline{X}_A y \underline{X}_B de forma que se puede escribir, $\underline{X}_B = \underline{X}_A + \underline{b}_{AB}$ siendo las componentes del vector *baselínea*.

$$\underline{b}_{AB} = \begin{bmatrix} X_B - X_A \\ Y_B - Y_A \\ Z_B - Z_A \end{bmatrix} = \begin{bmatrix} \Delta X_{AB} \\ \Delta Y_{AB} \\ \Delta Z_{AB} \end{bmatrix}$$

Los modelos matemáticos para el código y la fase de la portadora que ya hemos visto se pueden aplicar de forma análoga con la única diferencia de que ahora existen las coordenadas conocidas del punto de referencia. El posicionamiento relativo sólo es efectivo cuando se realizan observaciones simultáneas entre el punto de referencia de coordenadas conocidas y el punto a calcular. La observación simultánea en los puntos A y B a los satélites j y k nos va a permitir generar lo que se viene llamando simples diferencias, dobles diferencias y triples diferencias.

Posicionamiento relativo con medidas de código

Dada una estación de referencia de coordenadas conocidas, se calculan las correcciones a cada una de las pseudodistancias de cada satélite para las estaciones de referencia, y se aplican para cada época en la estación móvil. El resultado son unas coordenadas del punto no conocido, o móvil mejoradas.

Posicionamiento relativo con medidas de fase

Vamos a ver los modelos matemáticos para el posicionamiento relativo que se pueden obtener por medio de la combinación de las ecuaciones de observación para el posicionamiento absoluto.

- Simples diferencias.

Vamos a suponer que tenemos dos puntos observando a un mismo satélite. Llamemos a los puntos A y B y al satélite j. Podemos escribir para cada observador su ecuación de observación para la fase, de la forma,

$$\Phi_A^j(t) - f^j \delta^j(t) = \frac{1}{\lambda} \rho_A^j(t) + N_A^j - f^j \delta_A(t)$$

$$\Phi_B^j(t) - f^j \delta^j(t) = \frac{1}{\lambda} \rho_B^j(t) + N_B^j - f^j \delta_B(t)$$

y la diferencia de ambas ecuaciones nos da:

$$\Phi_B^j(t) - \Phi_A^j(t) = \frac{1}{\lambda} [\rho_B^j(t) - \rho_A^j(t)] + [N_B^j - N_A^j] - f^j [\delta_B(t) - \delta_A(t)]$$

que nos da la ecuación de simples diferencias. Si nos fijamos en las incógnitas del lado derecho de la igualdad, existe un problema respecto a la obtención de la solución. Al tener las ambigüedades y los errores de los relojes los mismos coeficientes, obtenemos una matriz deficiente de rango pues existen columnas con los mismos valores (nos lleva a un determinante nulo y por lo tanto matriz no invertible). Para evitar este problema, se incluyen ahora las cantidades relativas:

$$N_{AB}^j = N_B^j - N_A^j$$

$$\delta_{AB}^j(t) = \delta_B^j(t) - \delta_A^j(t)$$

y para el resto de la ecuación utilizamos la notación,

$$\Phi_{AB}^j(t) = \Phi_B^j(t) - \Phi_A^j(t)$$

$$\rho_{AB}^j(t) = \rho_B^j(t) - \rho_A^j(t)$$

de forma que la ecuación de simples diferencias queda como,

$$\Phi_{AB}^j(t) = \frac{1}{\lambda} \rho_{AB}^j(t) + N_{AB}^j - f^j \delta_{AB}^j(t)$$

Si nos fijamos en la ecuación, los errores del reloj del satélite han desaparecido. Las simples diferencias entre receptores eliminan o reducen los efectos asociados con los satélites como son los errores de los relojes y gran parte de los errores orbitales siempre que el vector entre los puntos sea menor que la distancia geocéntrica del satélite. La ecuación de simples diferencias se puede formar también considerando un solo punto A y dos satélites j y k , lo que nos llevaría a la ecuación:

$$\Phi_A^{jk}(t) = \frac{1}{\lambda} \rho_A^{jk}(t) + N_A^{jk} - f^j \delta^{jk}(t)$$

en la que hemos supuesto que las frecuencias de ambos satélites son iguales. La ventaja de este modelo es que se elimina el error del reloj del observador, si bien, como estamos hablando de posicionamiento relativo entre dos puntos, no tiene mucho sentido.

- Dobles diferencias.

Si ahora suponemos dos puntos A y B y dos satélites j y k , podemos formar dos simples diferencias, bien entre dos receptores, bien entre dos satélites si bien la ecuación de dobles diferencias que obtendríamos sería la misma. Supongamos entonces dos simples diferencias entre dos puntos y a dos satélites.

$$\Phi_{AB}^j(t) = \frac{1}{\lambda} \rho_{AB}^j(t) + N_{AB}^j - f^j \delta_{AB}^j(t)$$

$$\Phi_{AB}^k(t) = \frac{1}{\lambda} \rho_{AB}^k(t) + N_{AB}^k - f^k \delta_{AB}^k(t)$$

y para obtener la doble diferencia, las restamos. Si suponemos que las frecuencias de los satélites son iguales $f^j = f^k$ obtenemos:

$$\Phi_{AB}^k(t) - \Phi_{AB}^j(t) = \frac{1}{\lambda} [\rho_{AB}^k(t) - \rho_{AB}^j(t)] + [N_{AB}^k - N_{AB}^j]$$

y usando una notación análoga al caso de las simples diferencias obtenemos la forma final de la ecuación de dobles diferencias:

$$\Phi_{AB}^{jk}(t) = \frac{1}{\lambda} \rho_{AB}^{jk}(t) + N_{AB}^{jk}$$

El error del reloj del receptor se ha cancelado, y esta es la principal razón de que se usen las dobles diferencias. La eliminación de este error en el reloj del receptor es consecuencia de las hipótesis de simultaneidad en la observación y la igualdad en la frecuencia de las señales de los satélites.

La notación que hemos usado:

$$*_{AB}^{jk} = *_{AB}^k - *_{AB}^j$$

en la que podemos sustituir * por Φ , ρ o N . Hay que tener en cuenta que los términos que hemos utilizado como notación, comprenden cuatro términos, así:

$$*_{AB}^{jk} = *_{AB}^k - *_{AB}^j = *_{B}^k - *_{B}^j - *_{A}^k + *_{A}^j$$

y por lo tanto, los términos desarrollados quedan como:

$$\Phi_{AB}^{jk}(t) = \Phi_{B}^k(t) - \Phi_{B}^j(t) - \Phi_{A}^k(t) + \Phi_{A}^j(t)$$

$$\rho_{AB}^{jk}(t) = \rho_{B}^k(t) - \rho_{B}^j(t) - \rho_{A}^k(t) + \rho_{A}^j(t)$$

$$N_{AB}^{jk} = N_{B}^k - N_{B}^j - N_{A}^k + N_{A}^j$$

- Triples diferencias.

Hasta ahora sólo hemos considerado una época t . Si recordamos, las ambigüedades son constantes y por lo tanto independientes del tiempo. Consideremos entonces dos épocas t_1 y t_2 y sus correspondientes dobles diferencias:

$$\Phi_{AB}^{jk}(t_1) = \frac{1}{\lambda} \rho_{AB}^{jk}(t_1) + N_{AB}^{jk}$$

$$\Phi_{AB}^{jk}(t_2) = \frac{1}{\lambda} \rho_{AB}^{jk}(t_2) + N_{AB}^{jk}$$

y las restamos obteniendo:

$$\Phi_{AB}^{jk}(t_2) - \Phi_{AB}^{jk}(t_1) = \frac{1}{\lambda} [\rho_{AB}^{jk}(t_2) - \rho_{AB}^{jk}(t_1)]$$

que podemos escribir de forma simplificada como:

$$\Phi_{AB}^{jk}(t_{12}) = \frac{1}{\lambda} \rho_{AB}^{jk}(t_{12})$$

que es la ecuación de triples diferencias en la que hemos usado la notación:

$$*(t_{12}) = *(t_2) - *(t_1)$$

de forma que los términos de la ecuación de triples diferencias estén formados por ocho términos:

$$\Phi_{AB}^{jk}(t_{12}) = \Phi_B^k(t_2) - \Phi_B^j(t_2) - \Phi_A^k(t_2) + \Phi_A^j(t_2) - \Phi_B^k(t_1) - \Phi_B^j(t_1) - \Phi_A^k(t_1) + \Phi_A^j(t_1)$$

$$\rho_{AB}^{jk}(t_{12}) = \rho_B^k(t_2) - \rho_B^j(t_2) - \rho_A^k(t_2) + \rho_A^j(t_2) - \rho_B^k(t_1) - \rho_B^j(t_1) - \rho_A^k(t_1) + \rho_A^j(t_1)$$

Las dos principales ventajas de las triples diferencias es que el efecto de las ambigüedades desaparece y por lo tanto son independientes de los cambios en el valor de las ambigüedades, es decir, de las pérdidas de ciclo.

En general hay dos tipos de correlación, la correlación física y la correlación matemática. La correlación física se debe a que la señal que emite un satélite se recibe en varias estaciones al mismo tiempo, o bien que la misma señal emitida por un satélite se recibe en varios puntos. Normalmente no se tiene en cuenta si se calcula en modo de baselíneas aisladas, pero si se tiene en cuenta si el cálculo se realiza en modo multibaselínea o multiestación. Sin embargo, se han obtenido excelentes resultados aún obviando esta correlación. En cuanto a la corrección matemática. las ecuaciones de observación de simples diferencias están incorreladas. Las dobles diferencias están correladas. Las triples diferencias están más correladas entre si que las dobles diferencias. Las correlaciones se tienen que tener en cuenta a la hora de aplicar los pesos a las observaciones en el ajuste estadístico.

7.4. Procesamiento de datos. Cálculo y compensación de una campaña GPS

El procesamiento de datos de una campaña GPS abarca todas las fases necesarias hasta conseguir las coordenadas finales de los puntos incógnita, Éste siempre requiere la utilización de medida de fase dadas las precisiones que se exigen. Entre estas fases encontraremos primero el preprocesamiento y manipulación previa de los datos de los receptores, después el cálculo individual de baselíneas, en el caso diferencial, y finalmente un ajuste estadístico de los datos redundantes.

7.4.1. Preprocesamiento de datos

Tratamiento de los datos

El primer paso es la descarga de los datos grabados en los receptores que han participado en una campaña dada. También es necesaria la descarga de efemérides y datos adicionales para el cálculo. Ya sea directamente del mensaje de navegación, también grabado por el receptor, o de efemérides precisas, datos de rotación o estado de relojes de los satélites desde servidores públicos como el del IGS.

Estructuración de datos

En una campaña son muchos los datos a tratar, como los datos medidos, los resultados intermedios y información suplementaria necesaria (reseñas, informes de campo). Es conveniente ordenarlos y prepararlos para las fases siguientes. Una campaña GPS puede fracasar sin una adecuada estructuración previa de los datos necesarios.

Exportación de datos

Normalmente en una campaña se usan más de un tipo de receptor, o sus datos van a ser procesados en distintos programas. Es necesaria la escritura de los datos observados en un formato independiente del tipo de receptor. Así nació el actual formato de datos independientes de intercambio usado y reconocido mundialmente. Es el llamado Receiver Independent Exchange (RINEX). Para más detalle [11].

7.4.2. Detección y reparación de saltos de ciclo

Como ya se explico en el capítulo de fuentes de error en la propagación de la señal son los saltos en la medida de fase del receptor que es necesario detectar, normalmente a partir de la comparación de diferentes observables hasta detectarlos y con algoritmo adecuado corregir o reparar este corte de la medición. Una vez reparado en el preproceso evitamos que se introduzca una nueva incógnita de ambigüedad por salto en el ajuste de observaciones.

7.4.3. Resolución de ambigüedades

La ambigüedad N depende del receptor y del satélite y es independiente del tiempo, siempre que no se pierda la recepción inicial. Debe ser un valor entero, y su resolución previa generalmente mejora el cálculo posterior de la base línea.

$$\Phi_A^j(t) = \frac{1}{\lambda} \rho_A^j(t) + N_A^j + f\Delta\delta_A(t) - d_{Aion}^j(t)$$

Son numerosas las técnicas de resolución de ambigüedades, unas para la resolución con una sola frecuencia y otras para la resolución con dos frecuencias.

En el caso de una sola frecuencia o base líneas de menos de 10 Km. La resolución puede consistir en, utilizando la expresión anterior, realizar su linealización. Con las ecuaciones resultado calcular los parámetros que declaremos incógnitas (coordenadas, estados de relojes, y siempre la ambigüedad). La ambigüedad debe ser un número entero. Hay resoluciones que se conforman con la solución en número real, pero otras realizan ajustes sucesivos hasta encontrar el entero más probable para cada N .

En el caso de dos frecuencias las posibilidades se multiplican por la posibilidad de realizar combinaciones con las mediciones de fase de las dos portadoras e incluso con el código, como vimos en el capítulo de combinación de datos. Algoritmos como el de Wide-Lane ó Wübená ya fueron expuestos anteriormente.

Cuando se consigue después del ajuste la matriz covarianza de la solución de ambigüedades, como número real, es posible utilizarla para encontrar el valor de ambigüedad, como número entero, de los posibles cercanos probando con cada uno de ellos ó buscando. Por eso se llaman técnicas de búsqueda. Son muchas las técnicas de búsqueda desarrolladas como el de fijar de forma secuencial las ambigüedades, método de Frei, el de malla de cubo, ó el método de la función ambigüedad.

7.4.4. Ajuste del modelo matemático

Dada la superabundancia de observaciones, es necesario escoger el modelo estadístico de ajuste más adecuado. La técnicas de ajuste de mínimos cuadrados suele ser la escogida en que su modelo estocástico establece que la media de los residuos de los observables es cero y su varianza es igual a un valor a priori dado $N(0, \sigma^2)$, y su modelo matemático requiere ecuaciones de observación lineales. Tanto para la resolución en posicionamiento absoluto ó relativo ya sea

con ecuaciones de código o de fase o combinadas, es precisa la linealización previa de sus ecuaciones ya que es un requisito previo para la resolución por mínimos cuadrados.

En la resolución de una red de puntos se plantean dos estrategias diferentes. La primera es llamada solución de simple base línea. Básicamente todas las base líneas son calculadas una a una para posteriormente realizar un ajuste posterior de estas base líneas. Este proceso, se puede realizar con dobles ó triples diferencias, viene afectado por la correlación matemática de que hablamos en el capítulo de modelos matemáticos. Existen técnicas de recorrelación, pero en el caso de triples diferencias es un esfuerzo inútil dado el ruido que conllevan. Los presupuestos de la resolución de simple base línea sobre la no correlación de las base líneas ajustadas hacen que no sea un planteamiento totalmente correcto. Ya que base líneas observadas simultáneamente si que están correladas. Existe una segunda técnica de ajuste de red, llamada solución multipunto, en contraste con la anterior, todos los puntos de la red son considerados a la vez. Las correlaciones entre base líneas si se tienen en cuenta.

Bibliografía

- [1] Dach, R., Hugentobler, U., Fridez, P., Meindl, M., (2007), The Bernese GPS Software Version 5.0, Astronomical Institute, Universidad de Berna.
- [2] GPS-ICD (1993), GPS Interface Control Document, Revision C (ICD-GPS-200C, US Departement of Defense (DoD), <http://www.navcen.uscg.gov/pubs/gps/icd200/>
- [3] Hofmann-Wellenhof, B., H. Lichtenegger, and J. Collins (1992), GPS: Theory and Practice, Springer, ISBN 3-211-82364-6.
- [4] Leick, A. (1995), GPS Satellite Surveying, Wiley, ISBN 0-471-30626-6.
- [5] Nuñez-García, A., Valbuena, J.L., Velasco, J., (1992) GPS La nueva era de la Topografía. ISBN 84-87510-31-0
- [6] Teunissen, P. J.G., and A. Kleusberg (eds.) (1998), GPS for Geodesy, Springer, ISBN 3-540-63661-7.
- [7] Niell, A. E. (1996), Global Mapping Functions for the Atmosphere Delay at Radio Wavelengths, Journal of Geophysical Research, 101(B2), pp. 3227–3246.
- [8] Wild, U. (1994), Ionosphere and Geodetic Satellite Systems, Permanent GPS Tracking Data for Modelling and Monitoring, Geodatisch-geophysikalische Arbeiten in der Schweiz, Band 48.
- [9] Schaer, S. (1999), Mapping and Predicting the Earth's Ionosphere Using the Global Positioning System, Vol. 59 of Geodatisch-geophysikalische Arbeiten in der Schweiz,

Schweizerische Geodetische Kommission, Institut für Geodasie und Photogrammetrie, Eidg. Technische Hochschule Zurich, Suiza.

- [10] Schaer, S., G. Beutler, M. Rothacher, and T.A. Springer (1996), Daily Global Ionosphere Maps Based on GPS Carrier Phase Data Routinely Produced by the CODE Analysis Center, in Proceedings of the IGS Analysis Center Workshop, Silver Spring, Maryland, USA, edited by R.E. Neilan et al., pp. 181–192, IGS Central Bureau, JPL, Pasadena, California, USA, Marzo 1996.
- [11] Gurtner, W. (1994), RINEX: The Receiver-Independent Exchange Format, GPS World, 5(7), pp. 48–52, July 1994, format specifications available at <ftp://igscb.jpl.nasa.gov/igscb/data/format/rinex2.txt>.

Tema 8. Problemas directo e inverso de la Geodesia. Fundamentos de los tres tipos de soluciones: Integración del sistema de ecuaciones diferenciales de la geodésica, transferencia del triángulo polar elipsoidal a una esfera concéntrica y empleo de la esfera osculatriz de Gauss. Convergencia de meridianos.

8.1. Introducción

Estudiaremos a continuación algunos de los métodos clásicos para resolver los denominados *problemas geodésicos principales*. Como sabemos, en Geodesia esferoidal las coordenadas geodésicas de los vértices de la red están referidas a la superficie del elipsoide ($h = 0$), si bien cada par de datos en longitud y latitud van acompañados de la altitud ortométrica H (sobre el geoido) a la que se encuentra la correspondiente señal del terreno. Además de ello, cada uno de los lados de la red geodésica está caracterizado por su distancia geodésica y por los acimutes geodésicos de la dirección en ambos extremos, lo que configura la distribución y posición de la red sobre el elipsoide, así como su orientación respecto a la rejilla de meridianos y paralelos.

Los problemas geodésicos principales se refieren en concreto al tema del *transporte de coordenadas* de los puntos sobre el elipsoide, es decir, al cálculo de las coordenadas de un punto a partir de las coordenadas conocidas de otro vértice, o bien al cálculo de la distancia y de los acimutes correspondientes entre dos puntos de coordenadas conocidas. En el primer caso el problema geodésico se conoce como *problema directo* y se enuncia como sigue:

Dadas las coordenadas geodésicas φ y λ de un punto A sobre el elipsoide, el acimut geodésico \hat{A}_{AB} de la dirección a otro vértice B y la distancia geodésica s entre A y B, se pretenden obtener la longitud y latitud geodésicas de B así como el acimut inverso \hat{A}_{BA} .

Por su parte, el segundo caso constituye el *problema inverso* cuyo enunciado resulta ser:

Conocidas las coordenadas geodésicas de dos puntos A y B sobre la superficie del elipsoide, se han de calcular la distancia geodésica s entre ambos y los acimutes de la dirección en los extremos de la línea geodésica que los une.

Estos dos problemas, que se complementan con el cálculo de la convergencia de meridianos entre los puntos A y B, son la base para el establecimiento posterior de la red geodésica. Ambos se

pueden resolver por distintos procedimientos desde el punto de vista de la Geodesia clásica, si bien cada uno de los métodos se caracteriza por el margen de distancias en que puede utilizarse al objeto de obtener resultados aceptables en precisión. Atendiendo a la magnitud de la distancia de separación entre los puntos A y B, clasificaremos los procedimientos existentes según la siguiente subdivisión:

- Métodos para distancias cortas. Empleo hasta 30 ó 40 kilómetros como máximo.
- Métodos para medias distancias. Utilizables hasta los 250 ó 300 kilómetros.
- Métodos para distancias grandes. Pueden emplearse hasta los 5000 kilómetros.
- Métodos para distancias muy grandes. Útiles hasta 20000 kilómetros y más.

La clasificación expuesta es típica de la Geodesia esférica. Sin embargo, con la aparición de las tecnologías que dieron lugar a la Geodesia espacial, el estudio de los métodos clásicos diseñados para transporte de coordenadas sobre el elipsoide tiene más interés teórico que práctico. Por tal razón prescindiremos de tratar aquí los procedimientos elaborados para distancias muy grandes, limitando nuestro estudio al de aquellos métodos aplicables hasta distancias medias o grandes que aún pudieran ser de utilidad bajo diversas circunstancias.

No obstante, antes de entrar en los distintos procedimientos se deducen las ecuaciones diferenciales a las que atiende la línea geodésica, repasando a su vez el concepto de convergencia de meridianos entre dos puntos situados sobre la superficie del elipsoide.

8.2. Ecuaciones de la línea geodésica. Fórmulas de Laplace y Clairaut

Siguiendo lo expuesto en el tema 9 de la referencia [1], sean los puntos A y B representados en la figura 1, situados sobre la superficie del elipsoide y separados por el arco elemental de curva geodésica ds . Dibujemos los correspondientes meridianos geodésicos PA y PB, así como el ángulo de acimut \hat{A} en el punto A. Llevemos este acimut a la prolongación del arco de geodésica en B, y sea BP' el otro lado de dicho ángulo, el cual forma con el meridiano de B el ángulo elemental de convergencia de meridianos dA . Tomemos también un arco diferencial del paralelo que pasa por el punto A, y situémoslo en B perpendicularmente a las direcciones BP' y PA. Llamando BC a este arco, la distancia AC medida sobre el meridiano de A será el elemento diferencial de arco de meridiano $d\beta$ entre los puntos A y B.

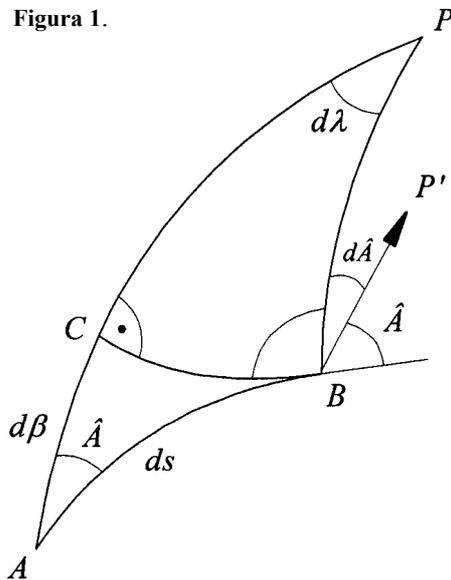
Dado el carácter elemental del triángulo ABC, consideramos iguales en A y en B tanto los radios principales de curvatura M y N como el radio r del paralelo. De este modo, teniendo en cuenta por un lado el radio de curvatura M de la elipse meridiana y el diferencial de latitud geo-

désica, y por otro el ángulo de acimut y el diferencial ds de arco de línea geodésica, se tienen las dos relaciones siguientes:

(1)

$$d\beta = M d\varphi; \quad d\beta = ds \cos \hat{A}$$

Figura 1.



Asimismo, para el lado BC se verifica:

(2)

$$r d\lambda = ds \sin \hat{A} \Rightarrow N \cos \varphi d\lambda = ds \sin \hat{A}$$

Por otra parte en el triángulo PCB , rectángulo en C , el ángulo en B tiene por valor $90^\circ - dA$, y el ángulo en P vale $d\lambda$. Podemos escribir entonces:

$$\begin{aligned} \cos PB &= \cot P \cot B \Rightarrow \\ &\Rightarrow \cos(90^\circ - \varphi) = \cot(d\lambda) \cot(90^\circ - dA) \end{aligned}$$

De aquí se obtiene:

$$\begin{aligned} \text{sen } \varphi &= \cot(d\lambda) \tan(dA) \Rightarrow \\ &\Rightarrow \tan(dA) = \tan(d\lambda) \text{sen}(\varphi) \end{aligned}$$

Y como dA y $d\lambda$ son elementos diferenciales, resulta finalmente:

(3)

$$\boxed{dA = d\lambda \text{sen } \varphi}$$

Esta ecuación se conoce como fórmula de Laplace, y proporciona la variación del acimut de una geodésica con relación a dos meridianos infinitamente próximos.

** A continuación obtendremos las derivadas con respecto del arco de los ángulos de latitud, longitud y acimut geodésicos en un punto de la curva de mínima distancia.

La primera resulta de forma inmediata considerando las dos igualdades (1); se tiene así:

(4)

$$M d\varphi = ds \cos \hat{A} \quad \Rightarrow \quad \frac{d\varphi}{ds} = \frac{\cos \hat{A}}{M}$$

La segunda se encuentra a partir de la fórmula (2):

(5)

$$N \cos \varphi d\lambda = ds \sin \hat{A} \quad \Rightarrow \quad d\lambda = \frac{\sin \hat{A}}{N \cos \varphi} ds \quad \Rightarrow \quad \frac{d\lambda}{ds} = \frac{\sin \hat{A}}{N} \sec \varphi$$

Y la tercera se obtiene al sustituir en (3) el elemento $d\lambda$ que acabamos de hallar:

(6)

$$dA = \frac{\sin \hat{A} \sin \varphi}{N \cos \varphi} ds \quad \Rightarrow \quad \frac{dA}{ds} = \frac{\sin \hat{A}}{N} \operatorname{tg} \varphi$$

Las fórmulas (4), (5) y (6) constituyen las ecuaciones diferenciales de las líneas geodésicas sobre la superficie del elipsoide. Dichas fórmulas pueden ser escritas en el parámetro c del elipsoide y en la función V cuyas expresiones son ([1]):

$$c = \frac{a^2}{b}; \quad V = \sqrt{1 + e'^2 \cos^2 \varphi}; \quad M = \frac{c}{V^3}; \quad N = \frac{c}{V}$$

Resultan así estas otras ecuaciones diferenciales de la geodésica, equivalentes a las anteriores:

(7)

$$\frac{d\varphi}{ds} = \frac{V^3}{c} \cos \hat{A}; \quad \frac{d\lambda}{ds} = \frac{V}{c} \sec \varphi \sin \hat{A}; \quad \frac{dA}{ds} = \frac{V}{c} \operatorname{tg} \varphi \sin \hat{A}$$

** El conjunto de relaciones deducidas nos van a permitir obtener la ecuación fundamental de la línea geodésica. Con esta finalidad representamos en la figura 2 la sección meridiana en el punto A de la figura 1, donde hemos trazado la recta tangente en C. Evidentemente, si r es el radio del paralelo de A, al punto C le corresponderá el valor $r + dr$, siendo:

$$dr = -AC \sin \varphi = -d\beta \sin \varphi = -M \sin \varphi d\varphi$$

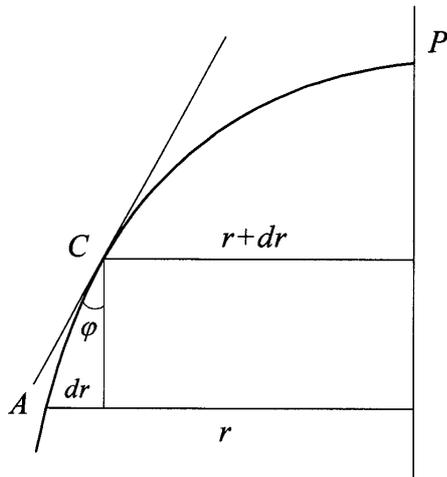


Figura 2.

Tomemos ahora las expresiones del coseno y seno del acimut \hat{A} respectivamente de las derivadas (4) y (5):

$$\cos \hat{A} = M \frac{d\varphi}{ds}, \quad \text{sen } \hat{A} = r \frac{d\lambda}{ds}$$

Multiplicando la primera por $r dA$ y la segunda por dr nos queda:

$$r \cos \hat{A} dA = Mr \frac{d\varphi}{ds} dA$$

$$\text{sen } \hat{A} dr = r \frac{d\lambda}{ds} dr = -r \frac{d\lambda}{ds} M \text{sen } \varphi d\varphi$$

Y sumando ambas miembro a miembro:

$$r \cos \hat{A} dA + \text{sen } \hat{A} dr = Mr \frac{d\varphi}{ds} dA - Mr \frac{d\lambda}{ds} \text{sen } \varphi d\varphi$$

A continuación extraemos factor común el producto Mr por la derivada $d\varphi/ds$:

$$N \cos \varphi d\lambda = ds \text{sen } \hat{A} \Rightarrow d\lambda \frac{\text{sen } \hat{A}}{N \cos \varphi} ds \Rightarrow \frac{d\lambda}{ds} = \frac{\text{sen } \hat{A}}{N} \sec \varphi$$

La diferencia entre paréntesis es nula por la fórmula de Laplace (3); se tiene entonces:

$$r \cos \hat{A} dA + \operatorname{sen} \hat{A} dr = 0$$

Esta ecuación diferencial se resuelve fácilmente considerando que su primer miembro es la diferencial total siguiente, función de \hat{A} y r :

$$d(r \operatorname{sen} \hat{A}) = 0$$

En consecuencia la ecuación de la línea geodésica resulta ser:

(8)

$$\boxed{r \operatorname{sen} \hat{A} = Cte. \quad \Leftrightarrow \quad r_1 \operatorname{sen} \hat{A}_1 = r_2 \operatorname{sen} \hat{A}_2 = \dots = r_n \operatorname{sen} \hat{A}_n = Cte.}$$

La igualdad obtenida se conoce como *teorema o fórmula de Clairaut*, y nos permite escribir la siguiente definición de la línea geodésica: las líneas geodésicas sobre la superficie del elipsoide de revolución son aquellas en las cuales el producto del radio del paralelo por el seno del acimut de la curva en cada uno de sus puntos es una cantidad constante.

Además, si en lugar del radio del paralelo se utiliza como variable el ángulo de latitud reducida según la igualdad: $r = a \cos u$, al ser a constante encontramos que el teorema de Clairaut adquiere la expresión equivalente:

(9)

$$\boxed{\cos u \operatorname{sen} \hat{A} = Cte. \quad \Leftrightarrow \quad \cos u_1 \operatorname{sen} \hat{A}_1 = \cos u_2 \operatorname{sen} \hat{A}_2 = \dots = \cos u_n \operatorname{sen} \hat{A}_n = Cte.}$$

8.3. Integración numérica del sistema de ecuaciones de la geodésica

El primer procedimiento que estudiaremos para la resolución de los problemas geodésicos principales consiste en la integración de las ecuaciones diferenciales de la geodésica utilizando métodos numéricos paso a paso. El margen de distancias para el cual el procedimiento es válido depende por lo tanto del método de integración, y en particular del orden de éste, por lo que teóricamente puede extenderse a grandes distancias si resulta necesario.

Puesto que se están tratando los problemas geodésicos desde el punto de vista de la Geodesia clásica, en este apartado aplicaremos uno de los métodos numéricos más tradicionales como es

el método de Runge – Kutta ([2] y [3]), el cual se particulariza para orden 4 por motivos didácticos. Bajo esta precisión, el procedimiento resolutivo que se expone se estima válido para distancias hasta los 400 ó 500 kilómetros.

8.3.1. Problema directo

Escribamos las ecuaciones diferenciales (4), (5) y (6) mediante las siguientes funciones al objeto de abreviar la notación:

$$f(\varphi, \hat{A}) = \frac{d\varphi}{ds} = \frac{\cos \hat{A}}{M}; \quad g(\varphi, \hat{A}) = \frac{d\lambda}{ds} = \frac{\text{sen } \hat{A}}{N \cos \varphi}; \quad h(\varphi, \hat{A}) = \frac{dA}{ds} = \frac{\text{sen } \hat{A}}{N} \text{tg } \varphi$$

Y supongamos conocidas las coordenadas geodésicas de un vértice A, así como el acimut \hat{A} y la distancia geodésica s a un segundo vértice B. Básicamente, el proceso de resolución del problema directo consiste en dividir la distancia s en n tramos sobre la línea geodésica que une A y B, e ir obteniendo los incrementos de latitud, longitud y acimut del punto final de cada tramo a partir de los datos correspondientes del punto anterior. De esta forma, tomando como valores iniciales los datos en A, se obtienen los del punto P_1 extremo del primer tramo, con éstos los de P_2 extremo del segundo tramo, y así sucesivamente hasta encontrar los de B, el cual es el punto extremo de tramo enésimo y último.

Las expresiones mediante las que resultan los incrementos en coordenadas y acimut del punto P_{i+1} a partir de los obtenidos en P_i , son las que siguen:

$$\Delta \varphi = \frac{k_1 + 2k_2 + 2k_3 + k_4}{6}, \quad \Delta \lambda = \frac{l_1 + 2l_2 + 2l_3 + l_4}{6}, \quad \Delta \hat{A} = \frac{a_1 + 2a_2 + 2a_3 + a_4}{6} \quad (10)$$

Relaciones donde los valores k , l , y a , se determinan sucesivamente a través de las igualdades:

$$\begin{aligned} k_1 &= \delta f(\varphi_i, \hat{A}_i), & l_1 &= \delta g(\varphi_i, \hat{A}_i), & a_1 &= \delta h(\varphi_i, \hat{A}_i) \\ k_2 &= \delta f(\varphi_i + k_1/2, \hat{A}_i + a_1/2), & l_2 &= \delta g(\varphi_i + k_1/2, \hat{A}_i + a_1/2), & a_2 &= \delta h(\varphi_i + k_1/2, \hat{A}_i + a_1/2) \\ k_3 &= \delta f(\varphi_i + k_2/2, \hat{A}_i + a_2/2), & l_3 &= \delta g(\varphi_i + k_2/2, \hat{A}_i + a_2/2), & a_3 &= \delta h(\varphi_i + k_2/2, \hat{A}_i + a_2/2) \\ k_4 &= \delta f(\varphi_i + k_3, \hat{A}_i + a_3), & l_4 &= \delta g(\varphi_i + k_3, \hat{A}_i + a_3), & a_4 &= \delta h(\varphi_i + k_3, \hat{A}_i + a_3) \end{aligned} \quad (11)$$

$$\delta = \frac{s}{n}, \quad i = 0, 1, \dots, n-1, \quad (i = 0 \rightarrow A, \quad i + 1 = n \rightarrow B)$$

Basta pues con escoger un valor concreto de n para obtener una solución al problema. Y evidentemente, mientras menor sea el valor δ del *paso de integración*, más precisa será la solución obtenida dentro de los márgenes de aproximación del método.

Una vez calculadas las coordenadas del vértice B y el acimut \hat{A}_B de la geodésica en este punto, el acimut recíproco \hat{A}_{BA} y la convergencia de meridianos en B con respecto de A se obtienen sencillamente mediante las relaciones siguientes:

$$\hat{A}_{BA} = \hat{A}_B \pm \pi, \quad \Delta\hat{A} = \hat{A}_B - \hat{A} \quad (12)$$

8.3.2. Problema inverso

Conocidas las coordenadas geodésicas de dos vértices A y B, para encontrar la distancia geodésica s que los separa y los acimutes de la línea geodésica en ambos puntos, se comienza calculando un valor promedio R para el radio medio de curvatura en ambos puntos. Atendiendo a lo expuesto en [1], dicho valor lo obtendremos en función de los radios principales de curvatura M y N en cada extremo de la geodésica mediante la expresión:

$$R = \frac{1}{2} \left(\sqrt{N_A M_A} + \sqrt{N_B M_B} \right) \quad (13)$$

A continuación, sobre la esfera cuyo radio es R se resuelve el triángulo esférico aproximado definido por los puntos A, B y el polo P de la esfera. Considerando el ángulo y los lados conocidos de este triángulo (P, a, b), se obtienen mediante trigonometría esférica los valores del lado p y del ángulo en el vértice A, por ejemplo a través de las ecuaciones:

$$p = \text{Arccos} (\cos b \cos a + \text{sen } b \text{sen } a \cos P) \quad (14)$$

$$A = \text{Arcsen} \left[\frac{\text{sen } P \text{sen } a}{\text{sen } p} \right]$$

Como consecuencia, unos primeros valores aproximados para la distancia geodésica y el acimut geodésico en A son los siguientes:

$$s_0 = pR, \quad \hat{A}_0 = A \quad (15)$$

Donde hemos nombrado los puntos de manera que la longitud de A es menor que la de B.

A partir de aquí las expresiones del problema directo permitirán encontrar la latitud y la longitud del punto B así como el acimut geodésico en el mismo, si bien de manera igualmente aproximada. Supongamos que una vez finalizado este cálculo se han obtenido los valores: φ_b , λ_b y \hat{A}_b . Con ellos se inicia entonces un proceso de aproximaciones sucesivas con objeto de encontrar las correcciones que superponen el punto b sobre B. Dicha superposición se consigue proyectando la diferencia en latitud y longitud entre b y B sobre las tangentes al meridiano y paralelo que pasan por B, para lo cual deberán calcularse inicialmente los incrementos longitudinales y el valor angular de giro siguientes:

$$\Delta s_\varphi = (\varphi_b - \varphi_B) M_b, \quad \Delta s_\lambda = (\lambda_b - \lambda_B) N_b \cos \varphi_b, \quad \alpha = \frac{\pi}{2} - \hat{A}_b \quad (16)$$

De esta manera, los nuevos valores para las variables (15) resultan ser ([2]):

$$\begin{pmatrix} \Delta s \\ \Delta \hat{A}s \end{pmatrix} = \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} \Delta s_\lambda \\ \Delta s_\varphi \end{pmatrix} \Rightarrow \begin{cases} s_1 = s_0 - \Delta s \\ \hat{A}_1 = \hat{A}_0 + \Delta \hat{A}s / s_0 \end{cases} \quad (17)$$

Mediante las coordenadas del vértice A y los resultados (17) podemos reiterar el problema directo para calcular unos nuevos valores aproximados del punto b y con ellos unos valores más precisos de la distancia geodésica y del acimut en A. Este proceso puede repetirse por tanto hasta que la diferencia de posición entre b y B esté por debajo de una tolerancia prefijada.

8.4. Los problemas geodésicos mediante la esfera oscultriz de Gauss

8.4.1. Fundamento

El segundo procedimiento que estudiamos para resolver los problemas geodésicos principales consiste en aplicar la *teoría de Gauss* para representar un elipsoide de poco aplanamiento, como es el terrestre, sobre la superficie de una esfera.

Como es sobradamente conocido, no puede conseguirse una proyección del elipsoide, ni sobre un plano ni sobre una esfera, que conserve las distancias en todas las direcciones, y mucho menos que simultáneamente sea conforme, o sea, que también mantenga las magnitudes angulares.

Sin embargo, el ingenio de Gauss le llevó a diseñar un procedimiento matemático por el cual podemos obtener la representación conforme de una faja del elipsoide sobre una esfera, de modo que las deformaciones lineales en el caso de un elipsoide de poco aplanamiento resulten prácticamente nulas en la zona objeto de estudio. Siendo así, resulta evidente que una vez dispuestas las expresiones de la proyección y calculados los parámetros necesarios, quedarán establecidas las relaciones entre los puntos del elipsoide y sus correspondientes en la esfera. Con ello podemos trasladar a la esfera los datos correspondientes los distintos problemas planteados sobre el elipsoide, resolver estos problemas mediante las expresiones clásicas de la trigonometría esférica, y finalmente transformar en sentido inverso los resultados al elipsoide.

El conjunto de las ecuaciones y desarrollos que siguen con sus correspondientes deducciones, los encontrará el lector con todo detalle en la referencia [1].

8.4.2. Representación conforme del elipsoide sobre la esfera

Una proyección del elipsoide sobre la esfera se denomina *conforme* cuando un contorno infinitamente pequeño de la superficie elipsoidal se representa en la esfera por un contorno semejante. Ello implica la conservación de los ángulos que las aristas de las distintas figuras forman entre sí.

Por otra parte, y como indica la Cartografía matemática ([1]-tema 14), en cualquier proyección del elipsoide, ya sea sobre la esfera o sobre el plano, existen siempre dos direcciones perpendiculares que se mantienen asimismo perpendiculares en la proyección. Estas dos direcciones se conocen como *direcciones principales*, y se caracterizan porque la escala de la representación adquiere sobre ellas sus valores extremos. Como consecuencia, para que una proyección sea conforme es necesario y suficiente que, salvo en las posibles singularidades, los factores de escala en cada punto tomados en las direcciones principales sean iguales.

Para obtener las ecuaciones de la transformación conforme del elipsoide sobre la esfera, designemos con λ y φ las coordenadas de un punto del elipsoide y con λ' y φ' las de su correspondiente en la esfera. Supondremos como primera condición la inexistencia de puntos impropios, por lo que a cada pareja de valores reales de λ y φ le corresponderá un par de valores reales de λ' y φ' . Además, y como segunda condición, restringiremos la proyección que buscamos imponiendo que los meridianos y paralelos del elipsoide se transformen respectivamente en meridianos y paralelos sobre la esfera. En tal caso las ecuaciones serán de la forma:

$$\lambda' = F_1(\lambda), \quad \varphi = F_2(\varphi) \quad (18)$$

Para determinar las funciones F_1 y F_2 denominaremos m y n a los factores de escala de la proyección en meridiano y paralelo en cada punto (ver [1]). Considerando los radios principales de curvatura M y N del elipsoide en dicho punto y el radio R de la esfera, dichos factores, para un elemento de arco en meridiano y otro en paralelo, resultan dados mediante las razones:

$$m = \frac{R d\varphi'}{M d\varphi}; \quad n = \frac{R \cos \varphi' d\lambda'}{N \cos \varphi d\lambda} \quad (19)$$

Pero siendo que la proyección buscada ha de ser conforme, deberá cumplirse que ambos factores de escala sean iguales en todo punto. Haciendo pues $m = n$ resulta:

$$\frac{d\varphi'}{\cos \varphi'} = \frac{M}{N} \cdot \frac{d\varphi}{\cos \varphi} \cdot \frac{d\lambda'}{d\lambda} \quad (20)$$

Esta relación, al objeto de que la latitud sobre la esfera sólo dependa de la latitud en el elipsoide, implica que la fracción $d\lambda' / d\lambda$ ha de ser necesariamente una magnitud constante K , por lo que la desglosamos en las dos ecuaciones diferenciales siguientes:

$$\frac{d\lambda'}{d\lambda} = K, \quad \frac{d\varphi'}{\cos \varphi'} = K \frac{M}{N} \cdot \frac{d\varphi}{\cos \varphi} \quad (21)$$

* Ecuación en longitud. Integrando la primera igualdad resulta la siguiente expresión para la proyección en longitud, siendo K una constante a determinar:

$$\boxed{\lambda' = K \lambda} \quad (22)$$

Lo que supone la elección de un mismo meridiano origen de longitudes tanto en el elipsoide como en la esfera (constante de integración igual a cero).

* Ecuación en latitud. Por su parte la integración de la ecuación en latitud, algo más complicada, proporciona el siguiente resultado:

$$\tan\left(\frac{\varphi'}{2} + \frac{\pi}{4}\right) = \frac{1}{k} \left(\frac{1 - e \operatorname{sen} \varphi}{1 + e \operatorname{sen} \varphi}\right)^{\frac{K e}{2}} \tan^K \left(\frac{\varphi}{2} + \frac{\pi}{4}\right) \quad (23)$$

Expresión donde k es una segunda constante a calcular.

* Factor de escala en meridiano y paralelo. Las relaciones (22) y (23) constituyen la ley de representación conforme del elipsoide sobre la esfera, y como se ha visto las hemos obtenido imponiendo la igualdad en cada punto entre los factores de escala en el meridiano y el paralelo. Por tanto dicho factor será único en cada uno de los puntos de la zona proyectada, pudiendo expresarse su valor concreto sin más que tener en cuenta (22) en la segunda de las igualdades (19):

$$m = n = K \frac{R \cos \varphi'}{N \cos \varphi} \quad (24)$$

Evidentemente no será posible calcular este parámetro en un punto dado en tanto no se defina el radio R de la esfera y se obtengan las constantes K y k de la proyección. Sin embargo, ya podemos ver que el factor de escala, al no depender de la longitud, será igual en todos los puntos de un mismo paralelo. Ello nos induce a pensar que las deformaciones lineales en la proyección de Gauss han de estudiarse según nos desplazamos en latitud. De hecho es esta coordenada la que determinará, en la forma que veremos más adelante, el margen de distancias hasta el cual puede extenderse la proyección con objeto de obtener resultados precisos.

8.4.3. Determinación de las constantes de la proyección

Como el factor de escala sólo varía en sentido de la latitud, determinaremos las constantes de la proyección de modo que resulten satisfechas las siguientes condiciones:

1. Que dicho factor sea igual a la unidad a lo largo de la proyección sobre la esfera de un paralelo previamente designado por su latitud φ_0 . Este paralelo, que se denomina paralelo de latitud normal, será por tanto línea automecoica de la transformación, y se escogerá en cada caso en función del problema concreto que se pretende resolver.
2. Que la variación de la escala según nos distanciamos del paralelo de latitud normal en uno u otro sentido sea lo más lenta posible.

Con estas condiciones, la teoría de Gauss trata de encontrar una proyección conforme, donde además la escala pueda considerarse prácticamente igual a la unidad hasta una cierta distancia a ambos lados del paralelo de latitud normal. En consecuencia las correcciones para pasar del elipsoide a la esfera y viceversa serán pequeñas y las deformaciones resultarán mínimas.

Siendo m_0 la escala de representación sobre el paralelo de latitud normal, la primera de las citadas condiciones se expresa en la forma:

$$m_0 = n_0 = K \frac{R \cos \varphi'_0}{N_0 \cos \varphi_0} = 1 \quad (25)$$

En cuanto a la segunda condición, el cálculo de las dos primeras derivadas de m respecto del ángulo de latitud y su posterior igualdad a cero al objeto de que la variación de m con este ángulo sea lo mínima posible, proporcionan sucesivamente las siguientes relaciones (ver [1]):

$$\boxed{\sin \varphi'_0 = \frac{1}{K} \sin \varphi_0, \quad K^2 = 1 + \frac{e^2 \cos^4 \varphi_0}{1 - e^2}} \quad (26)$$

Por tanto, una vez definida la latitud normal φ_0 podemos obtener la respectiva constante K y también la correspondiente latitud normal sobre la proyección esférica.

A partir de aquí, la segunda constante se obtiene sustituyendo el paralelo normal en la ecuación en latitud y despejando después el parámetro k . Resulta así:

$$k = \frac{\tan^K \left(\frac{\varphi_0 + \pi}{2} + \frac{\pi}{4} \right)}{\tan^K \left(\frac{\varphi'_0 + \pi}{2} + \frac{\pi}{4} \right)} \left(\frac{1 - e \sin \varphi_0}{1 + e \sin \varphi_0} \right)^{\frac{Ke}{2}} \quad (27)$$

Finalmente, el valor del radio R de la superficie esférica depende, al igual que las constantes K y k , del paralelo de latitud normal. Este valor puede calcularse fácilmente a partir de la condición (25); sin embargo, y como se demuestra en [1], R coincide precisamente con el radio medio de curvatura en los puntos del elipsoide de latitud φ_0 , es decir, que es igual a la media geométrica de los radios principales de curvatura calculados para la mencionada latitud. Esta es la razón por la cual la esfera de curvatura media se denomina también esfera de Gauss. Como consecuencia de lo dicho, las siguientes expresiones de R son equivalentes:

$$R = \frac{N_0 \cos \varphi_0}{K \cos \varphi'_0} \quad \Leftrightarrow \quad R = \sqrt{N_0 M_0} \quad (28)$$

8.4.4. Extensión de la proyección. Correcciones

Para estudiar el margen en distancia a ambos lados del paralelo de latitud normal donde la proyección puede considerarse automecoica además de conforme, se hace necesario obtener la expresión del factor de escala m en cada punto como función de la diferencia $\varphi - \varphi_0$. Ello no es difícil, pues con los valores encontrados para K , k y R tras anular las dos primeras derivadas de m en el paralelo de latitud normal, el citado factor responde al desarrollo de Taylor del factor m en torno al valor de φ_0 cuyos dos primeros términos son los siguientes:

$$m = 1 + \frac{1}{6} \left[\frac{d^3 m}{d\varphi^3} \right]_0 (\varphi - \varphi_0)^3$$

Calculando entonces esta derivada, según [1] se obtiene:

$$m = 1 - \frac{2}{3} \frac{e^2(1-e^2)\text{sen}\varphi_0 \cos\varphi_0}{(1-e^2\text{sen}^2\varphi_0)^2} (\varphi - \varphi_0)^3 \quad (29)$$

Además, puesto que m es una función que tiende a la unidad según nos acercamos al paralelo de latitud normal, su logaritmo tenderá a $m - 1$ en tales circunstancias. Por lo tanto, también podemos escribir:

$$Lm = -\frac{2}{3} \frac{e^2(1-e^2)\text{sen}\varphi_0 \cos\varphi_0}{(1-e^2\text{sen}^2\varphi_0)^2} (\varphi - \varphi_0)^3 \quad (30)$$

Atendiendo a la última expresión obtenida y dando distintos valores a φ , es posible establecer los límites en distancia entre los cuales la proyección conforme de Gauss puede entenderse también automecoica para una determinada latitud normal. En latitudes medias, de unos 40° como es el caso de la Península Ibérica, dentro del intervalo $[\varphi_0 - 1^\circ, \varphi_0 + 1^\circ]$ e incluso algo más allá la escala de la proyección es prácticamente 1, ya que el valor de Lm se mide en términos de 10^{-7} . Por tanto, si tenemos en cuenta que 1° de latitud supone casi 120 kilómetros sobre la superficie terrestre, podemos aceptar en general que la proyección conforme de Gauss del elipsoide sobre la esfera también conserva las distancias en una franja de 250 kilómetros de anchura cuyo paralelo central es el de latitud normal de la representación.

Esta consecuencia viene a demostrar la idoneidad de elegir una esfera para proyectar la superficie del elipsoide terrestre. Resulta así que para resolver los problemas geodésicos en una zona de la red geodésica situada hasta 120 kilómetros al norte o al sur del paralelo normal, y sin limi-

tación en longitud geográfica, se pueden trasladar los elementos de la triangulación del elipsoide a la esfera sin sufrir deformaciones apreciables: ni angulares por ser la proyección conforme, ni lineales debido al valor casi unidad de m en todos los puntos de la zona proyectada.

* Correcciones a los acimutes y distancias. Dada la conformidad de la proyección de Gauss, toda línea geodésica sobre el elipsoide se representa en la esfera por una curva cuyo acimut en cada punto es igual al respectivo de la geodésica. Sin embargo puede demostrarse que dicha curva no es exactamente un arco de círculo máximo en la esfera de curvatura media, razón por la cual se hace necesario introducir una corrección en los acimutes si se quieren utilizar las fórmulas de la trigonometría esférica. Con ello los triángulos elipsoidales cuyos lados son líneas geodésicas se representarán en la esfera por triángulos formados por arcos de círculo máximo.

Siendo A y B dos puntos sobre el elipsoide, \hat{A}_{AB} y \hat{A}_{BA} los acimutes geodésicos directo y recíproco entre ambos puntos, y s la distancia geodésica, de acuerdo con [3] los correspondientes ángulos a considerar en la superficie de la esfera resultan determinados por las siguientes expresiones:

$$\hat{A}'_{AB} = \hat{A}_{AB} - \left(\frac{2c_A + c_B}{3} \right) s \text{sen} \hat{A}_{AB}, \quad \hat{A}'_{BA} = \hat{A}_{BA} - \left(\frac{c_A + 2c_B}{3} \right) s \text{sen} \hat{A}_{AB} \quad (31)$$

$$c_i = \frac{e^2 \sqrt{1 - e^2} \text{sen} \varphi_0 \cos \varphi_0}{(1 - e^2 \text{sen} \varphi_0)^{3/2}} (\varphi_i - \varphi_0)^2 \quad \{i = A, B\}$$

No obstante, como el aplanamiento terrestre es muy pequeño, estas correcciones angulares sólo se tienen en cuenta en aquellas direcciones cuyos extremos se encuentran más allá de los 60 ó 70 kilómetros del paralelo de latitud normal. En tal caso deberemos considerar las indicadas fórmulas tanto en el paso del elipsoide a la esfera como en el proceso inverso.

Finalmente remarcar que en rigor m sólo es igual a 1 en el paralelo de latitud normal. En consecuencia, si alguno de los puntos del problema a resolver se encuentra próximo a los márgenes de la citada franja, en concreto a más de 1° del paralelo de latitud normal, han de tenerse en cuenta las distorsiones lineales de la proyección. Para ello se atiende a las siguientes igualdades que relacionan las distancias entre el elipsoide (s) y la esfera (s'):

$$m = \frac{ds'}{ds} \Rightarrow s' \int m ds \Rightarrow s' = m_m s \cong \sqrt{m_A m_B} s \quad (32)$$

De este modo podemos corregir la distancia que convenga tomando un valor medio de m apropiado, como puede ser por ejemplo la media geométrica de los valores de este factor en los extremos de la dirección considerada.

8.4.5. Resolución de los problemas geodésicos principales

Estudiada la proyección de Gauss, el camino general para resolver los problemas geodésicos principales, así como otros que se nos pueden plantear sobre la superficie del elipsoide, consiste en trasladar a la esfera según la ley de representación conforme el conjunto de datos geodésicos iniciales, resolver el problema sobre la esfera mediante las fórmulas de la trigonometría y, con la misma ley de representación pero aplicada en el sentido inverso, efectuar el paso al elipsoide de los resultados obtenidos en la esfera.

Aunque el procedimiento es bastante simple, la precisión de los resultados depende notablemente de la elección del paralelo de latitud normal ya que, según hemos visto, las fórmulas de la proyección y el margen de validez de la misma son función de la latitud φ_0 tomada como referencia. Por lo tanto, una acertada elección del paralelo de latitud normal proporcionará resultados muy buenos siempre que el problema no sobrepase los límites prefijados en la teoría. Como norma general, el valor de φ_0 se define procurando que el paralelo normal resulte lo más centrado posible en el conjunto de los puntos que intervienen en el problema.

Cuando se trata de resolver el problema directo, la latitud normal se determina en función de los datos de acimut y distancia geodésicos. Así por ejemplo, un acimut próximo a 90° ó 270° nos permite tomar por latitud de referencia la correspondiente al punto de coordenadas conocidas aunque la distancia geodésica sea grande, puesto que el vértice incógnita quedará bastante centrado en la zona de proyección. Sin embargo, a medida que el acimut señala direcciones más cercanas al meridiano, la influencia de la distancia geodésica se hace más determinante, pudiendo incluso prohibir la utilización del método. En un caso general del problema directo, el establecimiento del paralelo φ_0 se realiza evaluando previamente las coordenadas del vértice incógnita de forma aproximada; lo que podemos hacer con los datos del problema en el supuesto de Tierra esférica. Entonces tomaremos como latitud normal el valor promedio de las latitudes de los dos puntos que intervienen en el problema.

En el caso del problema geodésico inverso, al disponerse como dato de las coordenadas de los dos vértices de la dirección, será común elegir como paralelo de latitud normal el correspondiente a la latitud media.

Una vez determinado el parámetro φ_0 se procede a construir la ley conforme de representación según se ha explicado en la teoría; es decir, hemos de obtener los valores de K , φ_0' , k y R . Dispuestas así las ecuaciones, realizaremos el traslado de los datos a la esfera teniendo en cuenta la conservación de los acimutes, y la necesidad de convertir las distancias en magnitudes angulares mediante el valor del radio R . Asimismo, introduciremos las correcciones necesarias en aquellos acimutes y distancias que lo requieran.

Establecido el problema sobre la esfera, se calculan a continuación los valores de las incógnitas del problema mediante la trigonometría. Para ello será común utilizar el triángulo cuyo tercer vértice es uno de los polos de la esfera.

El último paso consiste en el traslado de los resultados al elipsoide; lo que haremos aplicando la ley de proyección (22) y (23) en sentido inverso. De ello resulta:

$$\lambda = \frac{\lambda'}{K}, \quad \tan\left(\frac{\varphi}{2} + \frac{\pi}{4}\right) = \left[k \tan\left(\frac{\varphi'}{2} + \frac{\pi}{4}\right) \right]^{\frac{1}{K}} \left(\frac{1 - e \operatorname{sen} \varphi}{1 + e \operatorname{sen} \varphi} \right)^{-\frac{e}{2}} \quad (33)$$

La segunda de estas ecuaciones requiere de aproximaciones sucesivas. Dicho proceso, que suele ser de rápida convergencia, se realiza considerando en el denominador del segundo miembro el valor de φ' como aproximación inicial de φ .

En cuestión de distancias, una vez transformadas por R a partir de los ángulos respectivos, éstas se toman idénticas sobre el elipsoide, aunque si alguna de ellas se separa en demasía del paralelo normal deberemos aplicar la corrección (32) en sentido inverso. Asimismo, también los valores obtenidos en la esfera para los acimutes se mantienen invariables sobre el elipsoide, salvo que hubiese que deshacer la corrección de paso al círculo máximo. En tal caso, los acimutes geodésicos de la dirección en cuestión se calculan mediante las siguientes fórmulas equivalentes a las (31) y cuyos parámetros c_A y c_B son los allí indicados ([1]):

$$\hat{A}_{AB} = \hat{A}'_{AB} + \left(\frac{2c_A + c_B}{3} \right) s \operatorname{sen} \hat{A}'_{AB}, \quad \hat{A}_{BA} = \hat{A}'_{BA} - \left(\frac{c_A + 2c_B}{3} \right) s \operatorname{sen} \hat{A}'_{AB} \quad (34)$$

Finalmente, cuando sea necesario obtener el ángulo de convergencia de meridianos, éste puede calcularse sobre el elipsoide mediante la segunda de las expresiones (12).

8. 5. Método de transferencia del triángulo elipsoidal a la esfera de Jacobi

8. 5.1. Fundamento

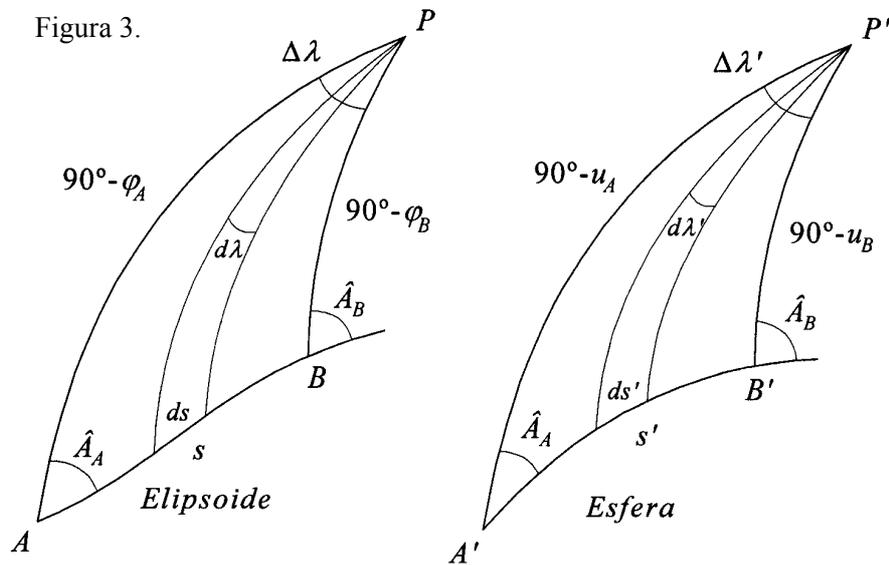
El tercer procedimiento que estudiaremos aquí para la resolución de los problemas geodésicos principales y que extraemos de las referencias [1] y [3], puede aplicarse cuando la distancia entre los vértices se extiende en cualquier dirección hasta los 800 ó 1000 kilómetros, considerando las fórmulas que se exponen.

Análogamente al caso anterior, el método se fundamenta en establecer un enlace analítico del triángulo elipsoidal que se quiere resolver con un triángulo esférico; lo que es factible, como indica la teoría de Gauss, debido a la pequeñez del aplanamiento terrestre. El triángulo elipsoidal será aquel cuyos vértices son los dos puntos A y B que intervienen en el problema, junto con el polo P del elipsoide, siendo sus lados los dos arcos de meridiano AP y BP y el arco AB de línea geodésica. En cuanto a la esfera a considerar, en este caso se elige la de Jacobi, también llamada *esfera de los acimutes constantes*, porque en ella se conservan los acimutes entre elipsoide y esfera como consecuencia de las fórmulas de Laplace y Clairaut (ver [1]). De este modo y al igual que sucede en la teoría de Gauss, la resolución de los problemas geodésicos principales supone trasladar los datos elipsoidales a la esfera, resolver allí el triángulo esférico obtenido en función de sus elementos conocidos, y trasladar en sentido inverso los resultados al elipsoide.

La esfera de Jacobi es aquella cuyo radio es igual al semieje mayor del elipsoide, que se mantiene tangente a éste en el ecuador. En ella, dados dos puntos A y B sobre la superficie del elipsoide, existen A' y B' cuyas latitudes son respectivamente iguales a las latitudes reducidas de A y B, y tales que la línea geodésica AB se corresponde con el arco de círculo máximo A'B' de la esfera, sobre el cual el acimut en cada uno de sus puntos es igual al acimut geodésico en el punto correspondiente de AB ubicado sobre el elipsoide.

Esta importante propiedad de la esfera de Jacobi, que se resume en la conservación de los acimutes, implica como consecuencia la determinación unívoca de todos los elementos del triángulo esférico relativo a dos puntos dados sobre la superficie del elipsoide. Para ver esto observemos el triángulo elipsoidal ABP de la figura 3 y el correspondiente triángulo esférico A'B'P'. Por una parte las magnitudes de los lados A'P' y B'P' quedan determinadas a partir de las latitudes geodésicas de A y B por la siguiente expresión:

$$\operatorname{tg} u = \frac{b}{a} \operatorname{tg} \varphi \quad \Leftrightarrow \quad \operatorname{tg} u = \sqrt{1-e^2} \operatorname{tg} \varphi \quad (35)$$



Además, los ángulos del triángulo elipsoidal en los vértices A y B son iguales que los del triángulo esférico en A' y B' por la propiedad enunciada anteriormente. Por otra parte, si consideramos la expresión (9) de la línea geodésica en función de las latitudes reducidas, se tiene:

$$\frac{\operatorname{sen} \hat{A}_A}{\operatorname{cos} u_B} = \frac{\operatorname{sen} \hat{A}_B}{\operatorname{cos} u_A} \quad \Leftrightarrow \quad \frac{\operatorname{sen} \hat{A}_A}{\operatorname{sen}(90^\circ - u_B)} = \frac{\operatorname{sen}(180^\circ - \hat{A}_B)}{\operatorname{sen}(90^\circ - u_A)} \quad (36)$$

Pero esta relación es la que resulta al aplicar el teorema del seno al triángulo esférico A'B'P' de la figura. Por tanto, la ecuación de la línea geodésica sobre el elipsoide constituye a la vez una dependencia exacta de la trigonometría esférica establecida en el respectivo triángulo sobre la esfera de Jacobi. Se deduce de esta manera que los valores de los seis elementos del triángulo esférico son únicos para dos puntos dados sobre el elipsoide, si bien la posición en longitud del triángulo esférico queda indeterminada en tanto no se conozca la relación entre longitudes elipsoidales y esféricas y se fije un meridiano de referencia.

También conviene señalar que, al ser iguales las latitudes reducidas y también los acimutes de la geodésica AB y del círculo máximo A'B' en los puntos correspondientes, la ecuación de Clairaut para ambas curvas es la misma. Ello implica que el acimut \hat{A}_0 con el cual el círculo máximo corta al ecuador es igual al de la geodésica a su paso por dicho plano; y asimismo, que los puntos de latitud reducida u_1 donde el acimut de las dos curvas es 90° , son correspondientes. Considerando estas consecuencias, el valor común de la constante de la línea será:

$$C = \cos u_A \operatorname{sen} \hat{A}_A = \cos u_B \operatorname{sen} \hat{A}_B = \operatorname{sen} \hat{A}_0 = \cos u_1 \quad (37)$$

Vistas las dependencias existentes entre los triángulos en correspondencia sobre el elipsoide y la esfera, este método se ocupa de establecer a partir de ellas las relaciones que ligan las diferencias de longitud $\Delta\lambda$ e $\Delta\lambda'$, así como las distancias s y s' de la línea geodésica AB y del arco de círculo máximo A'B' respectivamente.

8. 5.2. Relaciones diferenciales entre los elementos elipsoidales y jacobianos

En la figura 3 consideremos un elemento diferencial de arco ds sobre la línea geodésica que une los puntos A y B en el elipsoide, y el correspondiente elemento ds' de círculo máximo situado en la esfera. Asimismo sean $d\lambda$ y $d\lambda'$ los respectivos diferenciales de longitud abarcados por dichos elementos. A continuación obtendremos las ecuaciones diferenciales que relacionan entre sí cada una de estas parejas de elementos, pues ni las longitudes ni las distancias se conservan al pasar del elipsoide a la esfera de Jacobi.

Comencemos trayendo aquí las ecuaciones (4) y (5) que caracterizan la línea geodésica sobre el elipsoide, y por las cuales los elementos de latitud y longitud están relacionados con el diferencial de arco mediante las igualdades:

$$M d\varphi = ds \cos \hat{A}, \quad N \cos \varphi d\lambda = ds \operatorname{sen} \hat{A} \quad (38)$$

En la esfera de Jacobi (elipsoide con excentricidad nula y radio a), donde latitudes geodésicas y reducidas son una misma cosa y se conservan los acimutes del elipsoide, estas igualdades adquieren las siguientes formas respectivas:

$$a du = ds' \cos \hat{A}, \quad a \cos u d\lambda' = ds' \operatorname{sen} \hat{A} \quad (39)$$

* Ecuación diferencial entre los elementos de arco. Dividiendo miembro a miembro las dos primeras ecuaciones (38) y (39) obtenemos:

$$\frac{ds}{ds'} = \frac{M}{a} \frac{d\varphi}{du}$$

Tengamos en cuenta la siguiente expresión de M que figura en el tema 7 de [1]:

$$M = \frac{b}{W} \frac{du}{d\varphi} \text{ (donde } W = \sqrt{1 - e^2 \sin^2 \varphi}) \Rightarrow \frac{d\varphi}{du} = \frac{b}{WM} \Rightarrow \frac{ds}{ds'} = \frac{b}{aW}$$

Y asimismo, la relación de los semiejes a y b del elipsoide con la excentricidad, y la relación entre las funciones W y V (tema 2, referencia [1]); se tiene entonces:

$$ds = \frac{\sqrt{1 - e^2}}{W} ds' \quad \Leftrightarrow \quad ds = \frac{1}{V} ds'$$

Para escribir esta ecuación en función de la latitud reducida u , basta considerar la definición de la función V según la latitud reducida:

$$V = \frac{1}{\sqrt{1 - e^2 \cos^2 u}}$$

Resulta así la ecuación diferencial:

$$\boxed{ds = \sqrt{1 - e^2 \cos^2 u} ds'} \quad (40)$$

* Ecuación diferencial entre los elementos de longitud. La relación entre los diferenciales de longitud se establece dividiendo miembro a miembro las dos segundas ecuaciones (38) y (39):

$$\frac{N \cos \varphi}{a \cos u} \frac{d\lambda}{d\lambda'} = \frac{ds}{ds'} \quad \Rightarrow \quad \frac{d\lambda}{d\lambda'} = \frac{a \cos u}{N \cos \varphi} \frac{ds}{ds'}$$

Consideremos aquí la siguiente ligadura entre los cosenos de los ángulos de latitud, dada por la fórmula [2-33] de la referencia [1]:

$$\frac{\cos u}{\cos \varphi} = \frac{1}{W} \quad \Rightarrow \quad \frac{d\lambda}{d\lambda'} = \frac{a}{NW} \frac{ds}{ds'}$$

Y teniendo en cuenta que $N = a / W$, obtenemos que la proporción entre los diferenciales de longitud es la misma que la existente entre los diferenciales de arco:

$$\frac{d\lambda}{d\lambda'} = \frac{ds}{ds'}$$

Con ello, al sustituir el segundo miembro según las mismas relaciones anteriores, se tiene:

$$d\lambda = \frac{\sqrt{1-e^2}}{W} d\lambda' \quad \Leftrightarrow \quad d\lambda = \frac{1}{V} d\lambda'$$

Y por tanto la ecuación diferencial de longitudes en función de u es la siguiente:

$$\boxed{d\lambda = \sqrt{1-e^2 \cos^2 u} d\lambda'} \quad (41)$$

8.5.3. Fórmulas integrales en distancias e incrementos de longitud

Las respectivas integraciones de las ecuaciones (40) y (41) no son sencillas, además de que son distintas debido a que ha de expresarse la variable u como función de s' en el primer caso, y como función de λ' en el segundo. No obstante, el lector interesado las encontrará realizadas con detalle en [1], por lo que nos limitaremos aquí a exponer los resultados.

* Relación entre distancias. Considerando como referencia los valores de acimut y latitud reducida en el vértice A, se introducen los parámetros m y n definidos y relacionados mediante las igualdades siguientes:

$$\operatorname{tg} n = \frac{\operatorname{tg} u_A}{\cos \hat{A}_A}, \quad \operatorname{sen} m = \operatorname{sen} \hat{A}_A \cos u_A \quad (42)$$

$$\operatorname{cotg} n = \frac{\operatorname{sen} m \operatorname{cotg} \hat{A}_A}{\operatorname{sen} u_A}, \quad \operatorname{cos} m = \frac{\operatorname{sen} u_A}{\operatorname{sen} n} = \frac{\cos u_A \cos \hat{A}_A}{\operatorname{cos} n} \quad (43)$$

Asimismo, con el semieje mayor del elipsoide se define la magnitud angular σ , teniendo lugar entonces la relación siguiente para un punto cualquiera de la geodésica:

$$k^2 = e'^2 \cos^2 m, \quad K_1 = 1 + \frac{k^2}{4} - \frac{3k^4}{64}, \quad K_2 = -\frac{k^2}{4} + \frac{k^4}{16}, \quad K_3 = -\frac{k^4}{128} \quad (44)$$

Y con la segunda excentricidad se consideran también los siguientes factores:

$$k^2 = e'^2 \cos^2 m, \quad K_1 = 1 + \frac{k^2}{4} - \frac{3k^4}{64}, \quad K_2 = -\frac{k^2}{4} + \frac{k^4}{16}, \quad K_3 = -\frac{k^4}{128} \quad (45)$$

La integral solución, donde b es el semieje menor del elipsoide, resulta ser:

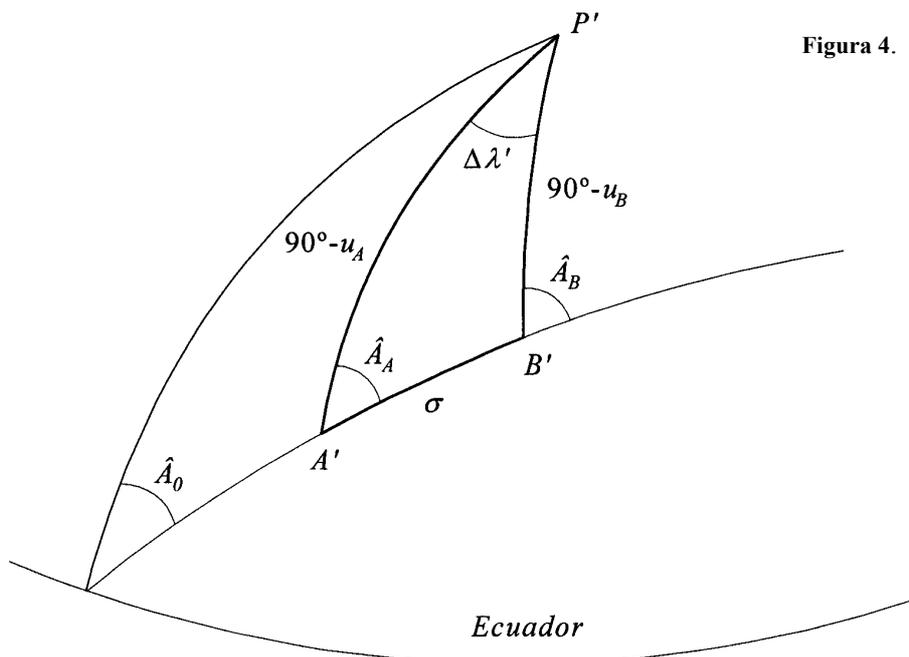
$$s = b[K_1\sigma + K_2\text{sen}\sigma \cos(2n + \sigma) + K_3\text{sen}2\sigma \cos(4n + 2\sigma)] \quad (46)$$

* Relación entre los incrementos de longitud. En la integral entre los incrementos de longitud en el elipsoide y la esfera intervienen, análogamente al caso anterior, los parámetros definidos en (42), (43) y (44). En este caso, siendo e la primera excentricidad del elipsoide, la integral solución de la ecuación diferencial (41) es la que sigue:

$$\Delta\lambda = \Delta\lambda' - \frac{e^2 \text{sen}m}{2} \left(1 + \frac{e^2}{4} - \frac{e^2}{8} \cos^2 m \right) \sigma - \frac{e^4 \text{sen}m \cos^2 m}{16} \text{sen}\sigma \cos(2n + \sigma) \quad (47)$$

8. 5.4. Acimut geodésico en el ecuador. Acimutes geodésicos en A y en B

A continuación se deduce una expresión para el cálculo del acimut con el que corta al ecuador la geodésica que pasa por los puntos A y B del elipsoide. Dicho acimut es igual al correspondiente del círculo máximo que une los puntos A' y B' sobre la esfera de Jacobi.



Consideremos el triángulo esférico de la figura 4, donde se ha prolongado el círculo máximo A'B' hasta su intersección con el ecuador. Sea \hat{A}_0 el acimut en este punto y σ el valor angular del arco A'B'. Por los teoremas del seno y el coseno, en este triángulo se verifican las igualdades:

$$\frac{\text{sen } \sigma}{\text{sen } \Delta\lambda'} = \frac{\text{cos } u_B}{\text{sen } \hat{A}_0}, \quad \text{cos } \sigma = \text{sen } u_A \text{ sen } u_B + \text{cos } u_A \text{ cos } u_B \text{ cos } \Delta\lambda' \quad (48)$$

Atendiendo a la primera expresión aplicaremos la ecuación de Clairaut en el punto A', considerando que esta ecuación es común tanto para la geodésica AB como para el círculo máximo A'B':

$$\frac{\text{sen } \sigma}{\text{sen } \Delta\lambda'} = \frac{\text{cos } u_A \text{ cos } u_B}{\text{sen } \hat{A}_0} \quad \Rightarrow \quad \text{sen } \hat{A}_0 = \text{sen } \Delta\lambda' \frac{\text{cos } u_A \text{ cos } u_B}{\text{sen } \sigma}$$

Elevando al cuadrado la relación obtenida, se tiene:

$$\text{sen}^2 \hat{A}_0 = \frac{\text{sen}^2 \Delta\lambda' \text{cos}^2 u_A \text{cos}^2 u_B}{\text{sen}^2 \sigma} \quad \Rightarrow \quad \text{cos}^2 \hat{A}_0 = \frac{\text{sen}^2 \sigma - \text{sen}^2 \Delta\lambda' \text{cos}^2 u_A \text{cos}^2 u_B}{\text{sen}^2 \sigma}$$

Y por tanto:

$$\text{ctg}^2 \hat{A}_0 = \frac{1 - \text{cos}^2 \sigma - \text{sen}^2 \Delta\lambda' \text{cos}^2 u_A \text{cos}^2 u_B}{\text{sen}^2 \Delta\lambda' \text{cos}^2 u_A \text{cos}^2 u_B}$$

Sustituyamos ahora el coseno de σ según la segunda de las expresiones (48):

$$\text{cot}^2 \hat{A}_0 = \frac{1 - (\text{sen } u_A \text{ sen } u_B + \text{cos } u_A \text{ cos } u_B \text{ cos } \Delta\lambda')^2 - \text{sen}^2 \Delta\lambda' \text{cos}^2 u_A \text{cos}^2 u_B}{\text{sen}^2 \Delta\lambda' \text{cos}^2 u_A \text{cos}^2 u_B}$$

A partir de aquí, tras algunas operaciones se obtiene (ver [1]):

$$\boxed{\text{ctg}^2 \hat{A}_0 = \frac{\text{tg}^2 u_A + \text{tg}^2 u_B - 2 \text{tg } u_A \text{tg } u_B \text{ cos } \Delta\lambda'}{\text{sen}^2 \Delta\lambda'}} \quad (49)$$

Calculando el valor de \hat{A}_0 por esta fórmula, la ecuación de Clairaut nos permite obtener de forma inmediata los acimutes geodésicos en los puntos A y B.

8. 5.5. Cálculo de la convergencia de meridianos

Aunque el cálculo de la convergencia de meridianos sobre el elipsoide es cuestión de solución sencilla una vez resueltos los problemas geodésicos principales, atendiendo a la conservación de los acimutes sobre la esfera de Jacobi vamos a deducir una expresión que nos permite obtener directa y rigurosamente el valor de $\Delta\hat{A}$ sin necesidad de conocer previamente los acimutes de la línea geodésica en sus dos puntos extremos. Para ello, volviendo sobre el triángulo esférico de la figura 4, sean las siguientes notaciones:

$$\Delta u = u_B - u_A, \quad u_m = \frac{u_A + u_B}{2}, \quad \Delta\hat{A} = \hat{A}_B - \hat{A}_A, \quad \Delta\lambda' = \lambda'_B - \lambda'_A \quad (50)$$

En dicho triángulo las analogías de Neper permiten escribir la igualdad:

$$\tan\left(90^\circ + \frac{\hat{A}_A - \hat{A}_B}{2}\right) = \cos\left(\frac{u_A - u_B}{2}\right) \sec\left(90^\circ - \frac{u_A - u_B}{2}\right) \cot\frac{\Delta\lambda'}{2}$$

En el primer miembro de esta ecuación observamos que aparece la convergencia de meridianos; por lo que se tiene:

$$\cot\frac{\Delta\hat{A}}{2} = \cos\left(\frac{u_A - u_B}{2}\right) \sec\left(90^\circ - \frac{u_A - u_B}{2}\right) \cot\frac{\Delta\lambda'}{2}$$

O bien, utilizando las notaciones (50):

$$\operatorname{ctg}\frac{\Delta\hat{A}}{2} = \frac{\cos\frac{\Delta u}{2}}{\operatorname{sen}u_m} \operatorname{ctg}\frac{\Delta\lambda'}{2}$$

Invirtiendo ambos miembros resulta la siguiente expresión para determinar la convergencia de meridianos entre B' y A'; valor igual al medido en el elipsoide entre B y A:

$$\boxed{\operatorname{tg} \frac{\Delta \hat{A}}{2} = \frac{\operatorname{sen} u_m}{\cos \frac{\Delta u}{2}} \operatorname{tg} \frac{\Delta \lambda'}{2}} \quad (51)$$

Mediante las analogías de Neper, y de manera análoga a la seguida para alcanzar la fórmula anterior, se demuestra esta otra expresión:

$$\operatorname{tg} \frac{\hat{A}_A + \hat{A}_B}{2} = -\frac{\cos u_m}{\operatorname{sen} \frac{\Delta u}{2}} \operatorname{tg} \frac{\Delta \lambda'}{2} \quad (52)$$

Como consecuencia, si eliminamos el incremento de longitud esférica entre las dos últimas relaciones, encontramos la dependencia de la convergencia de meridianos con las latitudes reducidas de dos puntos y la suma de los acimutes geodésicos:

$$\boxed{\operatorname{tg} \frac{\Delta \hat{A}}{2} = -\operatorname{tg} u_m \operatorname{tg} \frac{\Delta u}{2} \operatorname{tg} \frac{\hat{A}_A + \hat{A}_B}{2}} \quad (53)$$

* Teorema de Dalby para pequeñas distancias geodésicas. En la referencia [1] se demuestra que en el caso de pequeñas distancias, tales como las propias de los lados de la triangulación geodésica clásica, la fórmula (53) puede utilizarse directamente en términos de las coordenadas elipsoidales (λ, φ) de los puntos A y B, facilitándose con ello la obtención inmediata de la convergencia de meridianos. Dicha afirmación es consecuencia del teorema de Dalby, cuya expresión y enunciado son los siguientes:

$$\Delta \hat{A} = \Delta \lambda \operatorname{sen} \varphi_m \left(1 + \frac{\Delta \varphi^2}{8} + \frac{\Delta \lambda^2}{12} \cos^2 \varphi_m \right) \quad (54)$$

Para un lado normal de la triangulación geodésica, la convergencia de meridianos sobre el elipsoide es la misma, hasta los términos de tercer orden inclusive, que la obtenida considerando las coordenadas geográficas λ y φ referidas a la esfera de Jacobi.

8. 5.6. Problema directo

Para resolver el problema directo, donde los datos son las coordenadas geodésicas λ y φ de un punto A, así como el acimut y la distancia geodésica a un segundo punto B, los pasos a seguir son los siguientes:

1. Cálculo de la latitud reducida de A mediante (35).
2. Cálculo de los coeficientes m y n por las expresiones (42) y (43), utilizando para ello el acimut conocido \hat{A}_{AB} y la latitud u_A .
3. Obtención del valor de k y de los coeficientes K_i por (45), para seguidamente calcular σ por aproximaciones sucesivas en la fórmula (46) a partir de la distancia dada s .
4. Resolución del triángulo esférico A'B'P' donde se conocen los lados A'P' y A'B' así como el ángulo en el punto A. Se obtiene entonces el incremento de longitud esférica $\Delta\lambda'$, el ángulo en el vértice B y la latitud reducida u_B de este punto.
5. Cálculo del incremento $\Delta\lambda$ elipsoidal a partir del correspondiente esférico por (47).
6. Obtención de la latitud geodésica de B por (35), así como de su longitud a partir de λ_A e $\Delta\lambda$.
7. Cálculo del acimut inverso \hat{A}_{BA} sobre la base del ángulo obtenido en B, y determinación del valor de convergencia de meridianos por ejemplo mediante (51).

8. 5.7. Problema inverso

Para resolver el problema inverso, supuestas conocidas las coordenadas geodésicas de dos puntos A y B, operamos de la siguiente forma:

1. Obtención de las latitudes reducidas de ambos puntos por (35).
2. Cálculo de los valores de \hat{A}_A , m , n , y σ por aproximaciones sucesivas. Para ello se supone inicialmente un valor del incremento de longitud esférica $\Delta\lambda'$ igual al de $\Delta\lambda$ sobre el elipsoide, con el que se obtiene por (49) una primera aproximación de \hat{A}_0 . Con dicha magnitud, y mediante la ecuación de Clairaut (9), se busca el acimut geodésico correspondiente en el punto A. Este valor

\hat{A}_A nos servirá para encontrar la primera aproximación de m y n en las expresiones (42) y (43). Con dichos parámetros, en las fórmulas (44) (referidas al punto B) se busca un primer valor para σ . Aplicando entonces la relación (47) se calcula la segunda aproximación de $\Delta\lambda'$ con la que se repite el proceso. Continuaremos esta serie de operaciones hasta conseguir la convergencia del conjunto de valores.

3. Determinación de la distancia s de línea geodésica mediante (45) y (46), utilizando para ello las magnitudes finales halladas en el punto anterior.

4. Cálculo de la convergencia de meridianos y del acimut recíproco a través de (37) ó (51).

8. 6. Fórmulas de Cotichia – Surace

Cuando las distancias son pequeñas, del orden de los 30 kilómetros, la resolución de los problemas geodésicos principales puede realizarse de una forma sencilla mediante las siguientes fórmulas extraídas de la referencia [1] y debidas a los geodesias italianos Cotichia y Surace.

8. 6.1. Problema directo

Datos de entrada: coordenadas geodésicas de un punto A , distancia s y acimut geodésico \hat{A} a otro vértice B .

(55)

$$\varphi_B = \varphi_A + \frac{s \cos \hat{A}}{M_A} \left[1 - \frac{s^2 \text{sen}^2 \hat{A}}{6M_A^2} (1 + 3 \tan^2 \varphi_A) \right] - \frac{s^2 \tan \varphi_A}{2M_A N_A} (\text{sen}^2 \hat{A} + 3e'^2 \cos^2 \hat{A} \cos^2 \varphi_A)$$

(56)

$$\lambda_B = \lambda_A + \frac{s \text{sen} \hat{A}}{N_A \cos \varphi_A} + \frac{s^2}{N_A^2 \cos \varphi_A} \left[\frac{\text{sen}^2 \hat{A} \tan \varphi_A}{2} + \frac{s}{6} \left(\frac{\cos \hat{A} \text{sen} 2\hat{A}}{M_A} + \frac{2 \tan^2 \varphi_A \text{sen} 3\hat{A}}{N_A} \right) \right]$$

8. 6.2. Problema inverso

Las sucesivas operaciones a realizar para resolver este problema a partir de las coordenadas geodésicas de los puntos A y B son las que siguen:

1. Cálculo de los parámetros iniciales:

$$\varphi = \frac{\varphi_A + \varphi_B}{2}, \quad \eta^2 = e'^2 \cos^2 \varphi, \quad \nu^2 = 1 + \eta^2, \quad \varepsilon = \frac{\varphi_B - \varphi_A}{2\nu^2}, \quad \Delta\lambda = \frac{\lambda_B - \lambda_A}{2}$$

2. Coordenadas auxiliares:

$$x' = \operatorname{sen} \varepsilon \cos \Delta\lambda, \quad y' = \operatorname{sen} \Delta\lambda \cos \varphi, \quad z' = \frac{\tan(\nu \Delta\lambda) \operatorname{sen} \varphi}{\cos(\nu^2 \varepsilon)}$$

3. Transformación de coordenadas:

$$x = x' \left(1 - \frac{\eta^2 y'^2}{3} + \frac{\eta^2 x'^2}{3} \left(\nu^2 - \frac{5 - 4\nu^2}{\cot^2 \varphi} \right) \right)$$

$$y = y' \left[1 + \frac{\eta^2 x'^2}{6} \left(1 - \frac{2\nu^2 + 7}{\cot^2 \varphi} \right) \right], \quad z = z' \left(1 + \frac{\eta^2 x'^2}{3} \right)$$

4. Cálculo de los parámetros finales:

$$\gamma = \frac{\operatorname{Arctg} z}{\nu}, \quad S = \sqrt{x^2 + y^2}, \quad \alpha = \operatorname{Arctg} \frac{y}{x},$$

5. Distancia y acimutes geodésicos:

$$s = \frac{2a^2}{\nu b} \operatorname{Arcsen} S, \quad \hat{A}_{AB} = 180^\circ + \alpha - \gamma, \quad \hat{A}_{BA} = \alpha + \gamma$$

Bibliografía

- [1] Mena, J.B., 2008. *Geodesia Superior*. Texto en dos tomos. En proceso de publicación por el Instituto Geográfico Nacional. Madrid.
- [2] Múgica, F. 1971. *El ordenador y el análisis numérico*. Boletín del Servicio Geográfico del Ejército, núm.15, 1971.
- [3] Zakatov, P.S., 1997. *Curso de Geodesia Superior*. Ed. Mir. Moscú.

Tema 9. Compensación de una red geodésica por variación de coordenadas: Descripción general del método. Fórmulas diferenciales sobre el plano y sobre el elipsoide. Precisión requerida en las coordenadas aproximadas. Relaciones de observación: Direcciones, distancias y acimutes Laplace. Pesos de las observaciones. Elipse de error.

9.1. Compensación de una red geodésica por variación de coordenadas: Descripción general del método

Una red geodésica es un conjunto de puntos perfectamente localizados en el terreno mediante monumentaciones adecuadas entre los que se han efectuado observaciones de tipo geodésico con el fin de obtener sus coordenadas, precisión y fiabilidad en términos relativos y absolutos, respecto de un sistema de referencia establecido.

Una vez obtenidas las coordenadas de los puntos de una red geodésica por observación de direcciones, distancias y distancias zenitales, y reducidos estos datos al sistema de referencia o superficie de representación, tenemos que proceder a la compensación de la red geodésica para obtener un resultado homogéneo, único y fiable de la geometría de la red.

En la compensación de una red, primero se hace un estudio del modelo matemático a utilizar, es decir, se eligen las variables, las relaciones entre ellas y la técnica estadística con la que se va a trabajar.

Existen dos métodos utilizados en la compensación de redes geodésicas:

- Método de direcciones (o de ecuaciones de condición): donde los datos son las propias medidas realizadas, las incógnitas son las correcciones a esas medidas y el modelo matemático de ajuste el de ecuaciones de condición.
- Método de variación de las coordenadas (o de ecuaciones de observación): donde con los datos (medidas de ángulos o distancias) se obtienen unos valores aproximados de las coordenadas de cada vértice y las incógnitas son las correcciones a las coordenadas aproximadas.

Aunque ambos métodos tienen similares características matemáticas, el método elegido en la compensación de redes es el de variación de coordenadas, entre otras razones porque es

necesario conocer las coordenadas aproximadas de los puntos y las correcciones aplicadas tras la compensación es sobre estas coordenadas aproximadas y no sobre las observaciones de campo, como el método de ecuaciones de condición.

El método de variación de coordenadas puede aplicarse sobre puntos aislados pero donde mayor interés muestra este método es en el cálculo de un gran número de puntos, formando un cálculo en bloque y permitiendo añadir al cálculo observaciones con facilidad.

En el método de variación de coordenadas, cada observación de campo va a dar lugar a una expresión matemática, que relaciona los valores aproximados, las correcciones buscadas y los valores medidos. La expresión que se forma se denomina relación de observación:

$$(1) \text{ VALOR APROXIMADO} + \text{CORRECCIÓN} - \text{VALOR OBSERVADO} = \text{RESIDUO}$$

El valor aproximado se calcula a partir de las coordenadas aproximadas y el valor observado son las medidas de campo. La corrección es el valor que buscamos para que se cumpla la relación de observación y el residuo es el resto que quedará siempre, distinto de cero debido a los errores que se comenten en las medidas.

Si las observaciones fueran perfectas (Residuo = 0), entonces el valor aproximado más la corrección sería igual al valor observado.

Cuando contamos con una serie de observaciones podemos establecer un sistema de ecuaciones de la siguiente forma:

$$\begin{aligned} F_1(x_1, \dots, x_n; l_1, \dots, l_m) &= 0 \\ &\dots \\ (2) \quad &\dots \\ &\dots \\ F_m(x_1, \dots, x_n; l_1, \dots, l_m) &= 0 \end{aligned}$$

donde (x_1, \dots, x_n) son los parámetros incógnitas (las correcciones), y (l_1, \dots, l_m) son las observaciones. Los valores aproximados los representamos por (x_1^0, \dots, x_n^0) y desarrollando en serie de Taylor las funciones F_1, \dots, F_m en el punto x_0 tenemos que:

$$\begin{aligned} &F_j(x_1, \dots, x_n; l_1, \dots, l_m) = \\ (3) \quad &= F_j(x_1^0, \dots, x_n^0) + \left(\frac{\partial F_j}{\partial x_1} \right)_0 (x_1 - x_1^0) + \dots + \left(\frac{\partial F_j}{\partial x_n} \right)_0 (x_n - x_n^0) = \\ &= F_j(x_1^0, \dots, x_n^0; l_1, \dots, l_m) + \sum_{i=1}^n \left(\frac{\partial F_j}{\partial x_i} \right) dx_i \end{aligned}$$

El valor $dx_i = x_i - x_i^0$ corresponde a la variación de coordenadas, es decir, la desviación entre el valor verdadero y el aproximado, y es el valor incógnita que se calcula.

Si definimos el valor de constantes de observación como:

$$(4) F_j(x_1^0, \dots, x_n^0; l_1, \dots, l_m) = -t_j$$

Obtenemos:

$$(5) \begin{aligned} \delta_{11}(x_1 - x_1^0) + \dots + \delta_{1n}(x_n - x_n^0) &= t_1 \\ \dots & \\ \delta_{m1}(x_1 - x_1^0) + \dots + \delta_{mn}(x_n - x_n^0) &= t_m \end{aligned}$$

que es un sistema de ecuaciones linealizado formado por relaciones de observación y puede expresarse en forma matricial de la siguiente forma:

$$(6) A dx = t$$

De la solución de este sistema lineal extraemos las diferencias de coordenadas a aplicar a las coordenadas aproximadas para obtener las coordenadas finales compensadas. Este es un sistema, en general, con m ecuaciones y n incógnitas, donde $m > n$, es decir, con observaciones sobreabundantes y con puntos conocidos que entran como fijos en el cálculo, son los denominados constreñimientos de la red geodésica.

Las ecuaciones de observación las podemos estudiar en función de la superficie de referencia utilizada para realizar la compensación:

- Compensación sobre el plano
- Compensación sobre el elipsoide

de esta forma se plantean diferentes formulaciones.

9.2. Fórmulas diferenciales sobre el plano

Para este caso, y en función de la magnitud a observar, vamos a estudiar las fórmulas para los siguientes casos:

- Relación de observación para direcciones
- Relación de observación de distancias

Partimos de la consideración de dos puntos P_i y P_j conocidos o desconocidos, de coordenadas x_i, y_i y x_j, y_j ([1], Pag. 63):

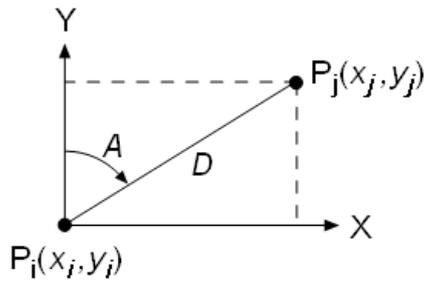


Figura 1. Relaciones de observación entre dos puntos en el plano.

Se sabe que el acimut A en P_i en la recta P_i-P_j viene dado por:

$$(7) \quad A = \text{actg} \left(\frac{x_j - x_i}{y_j - y_i} \right)$$

Diferenciando esta expresión:

$$(8) \quad dA = d \left(\text{actg} \frac{x_j - x_i}{y_j - y_i} \right) = \frac{1}{1 + \left(\frac{x_j - x_i}{y_j - y_i} \right)^2} \frac{d(x_j - x_i)(y_j - y_i) - d(y_j - y_i)(x_j - x_i)}{(y_j - y_i)^2}$$

y haciendo operaciones llegamos a la expresión diferencial para el acimut:

$$(9) \quad dA = a_1 dx_i + a_2 dx_j + a_3 dy_i + a_4 dy_j$$

donde D_0 , es la distancia calculada con las coordenadas aproximadas. Los coeficientes a_i son los siguientes:

$$(10) \quad \begin{aligned} a_1 &= -\frac{(y_j - y_i)}{D_0^2} \\ a_2 &= \frac{(y_j - y_i)}{D_0^2} \\ a_3 &= \frac{(x_j - x_i)}{D_0^2} \\ a_4 &= -\frac{(x_j - x_i)}{D_0^2} \end{aligned}$$

Si nos fijamos de nuevo en la Figura 1, sabemos que la distancia entre P_i y P_j es:

$$(11) \quad D = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$$

Diferenciando esta expresión:

$$(12) \quad 2DdD = 2(x_j - x_i)d(x_j - x_i) + 2(y_j - y_i)d(y_j - y_i)$$

y despejando convenientemente, llegamos a la expresión diferencial para las distancias:

$$(13) dD = b_1 dx_i + b_2 dy_i + b_3 dx_j + b_4 dy_j$$

donde D_0 , es la distancia calculada con las coordenadas aproximadas. Los coeficientes b_i son los siguientes:

$$(14) \begin{aligned} b_1 &= -\frac{(x_j - x_i)}{D_0} \\ b_2 &= -\frac{(y_j - y_i)}{D_0} \\ b_3 &= \frac{(x_j - x_i)}{D_0} \\ b_4 &= \frac{(y_j - y_i)}{D_0} \end{aligned}$$

En ambas expresiones: (9) y (10), si el punto P_i , por ejemplo, fuera conocido entonces las expresiones se simplifican ya que $x_i = y_i = 0$.

9.3. Fórmulas diferenciales sobre el elipsoide

Si las coordenadas φ y λ de los puntos son las incógnitas, entonces tenemos dos formas de formular el modelo matemático que relaciona éstas con los observables: el modelo se formula bien en el espacio tridimensional, o directamente en el elipsoide de referencia

En el primer caso, las observaciones se usan tal y como fueron observadas en la superficie de la Tierra (corregidas por efectos instrumentales y refracción), en este caso se considera el sistema local de observación terrestre y su relación con el sistema cartesiano global.

La formulación, en el segundo caso (coordenadas φ y λ sobre el elipsoide), emplea las observaciones reducidas al elipsoide de referencia.

En este segundo método se necesita únicamente en el modelo, altitudes geodésicas h aproximadas de los puntos de la red. Esta es la principal diferencia con las redes tridimensionales.

El modelo matemático de una red horizontal en el elipsoide se basa en el uso de la línea geodésica entre cualesquiera dos puntos del elipsoide. Entonces las observaciones de acimut, dirección y distancia deben ser reducidas, para que ellas se refieran a la geodésica.

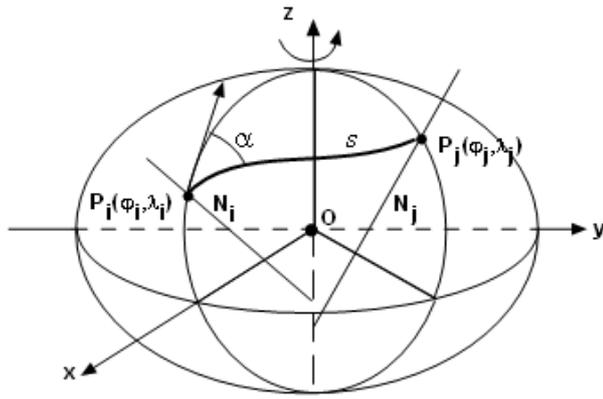


Figura 2. Observables sobre el elipsoide

El acimut elipsoidal entre P_i y P_j se puede escribir como:

$$(15) \quad \alpha_{ij} = \alpha(\varphi_i, \lambda_i, \varphi_j, \lambda_j)$$

Aproximando por la parte lineal del desarrollo de Taylor tenemos que:

$$(16) \quad \alpha_{ij} = \alpha(\varphi_i^{(0)}, \lambda_i^{(0)}, \varphi_j^{(0)}, \lambda_j^{(0)}) + \delta\alpha = \alpha^{(0)} + \delta\alpha$$

donde $\alpha^{(0)}$, es el valor calculado del acimut elipsoidal con las coordenadas aproximadas $\varphi_i^{(0)}, \lambda_i^{(0)}, \varphi_j^{(0)}, \lambda_j^{(0)}$.

La fórmula diferencial para el acimut geodésico $\delta\alpha$ es igual a:

$$(17) \quad \delta\alpha = \frac{\delta\alpha}{\delta\varphi_i} d\varphi_i + \frac{\delta\alpha}{\delta\lambda_i} d\lambda_i + \frac{\delta\alpha}{\delta\varphi_j} d\varphi_j + \frac{\delta\alpha}{\delta\lambda_j} d\lambda_j = a_1 d\varphi_i + a_2 d\lambda_i + a_3 d\varphi_j + a_4 d\lambda_j$$

donde los coeficientes a son:

$$(18) \quad \begin{aligned} a_1 &= \frac{\rho_i}{s} \operatorname{sen} \alpha_{ij} \\ a_2 &= \frac{N_j}{s} \cos \alpha_{ji} \cos \varphi_j \\ a_3 &= \frac{\rho_j}{s} \operatorname{sen} \alpha_{ji} \\ a_4 &= \frac{-N_j}{s} \cos \alpha_{ji} \cos \varphi_j \end{aligned}$$

siendo ρ y N los radios principales de curvatura.

De la misma forma, la distancia elipsoidal entre P_i y P_j se puede escribir como:

$$(19) \quad s_{ij} = s(\varphi_i, \lambda_i, \varphi_j, \lambda_j)$$

y aproximando por la parte lineal del desarrollo de Taylor tenemos que:

$$(20) s_{ij} = s(\varphi_i^{(0)}, \lambda_i^{(0)}, \varphi_j^{(0)}, \lambda_j^{(0)}) + \delta s = s^{(0)} + \delta s$$

donde $s^{(0)}$, es el valor calculado de la distancia elipsoidal con las coordenadas aproximadas $\varphi_i^{(0)}, \lambda_i^{(0)}, \varphi_j^{(0)}, \lambda_j^{(0)}$.

La fórmula diferencial para la distancia δs es igual a:

$$(21) \delta s = \frac{\delta s}{\delta \varphi_i} d\varphi_i + \frac{\delta s}{\delta \lambda_i} d\lambda_i + \frac{\delta s}{\delta \varphi_j} d\varphi_j + \frac{\delta s}{\delta \lambda_j} d\lambda_j = b_1 d\varphi_i + b_2 d\lambda_i + b_3 d\varphi_j + b_4 d\lambda_j$$

donde los coeficientes b son:

$$(22) \begin{aligned} b_1 &= -\rho_i \cos \alpha_{ij} \\ b_2 &= N_j \operatorname{sen} \alpha_{ji} \cos \varphi_j \\ b_3 &= -\rho_j \cos \alpha_{ij} \\ b_4 &= -N_j \operatorname{sen} \alpha_{ji} \cos \varphi_j \end{aligned}$$

9.4. Precisión requerida en las coordenadas aproximadas

Al montar un sistema de ecuaciones con observaciones e incógnitas para realizar un ajuste por mínimo cuadrados, la convergencia del ajuste depende de los datos de entrada. El método de compensación por mínimos cuadrados lo que hace es distribuir el error de entrada y cuando más fiables y precisas sean las observaciones introducidas el sistema será más capaz de detectar posibles errores en las observaciones.

Las coordenadas aproximadas de los puntos visados se suelen calcular con las propias observaciones, aunque a veces se obtienen, por ejemplo, directamente de una cartografía. La convergencia en el cálculo es visible cuando estas coordenadas aproximadas son buenas, hay casos en que el sistema diverge por un mal cálculo de las aproximadas pero el caso más general es que con un cálculo obtengamos un resultado correcto.

En ocasiones, es necesario o aconsejable iterar más de una vez el cálculo del sistema de ecuaciones para comprobar que el resultado obtenido es el mejor posible. Esto se realiza introduciendo la solución de las incógnitas en el nuevo cálculo de la siguiente forma:

- Partimos de unas coordenadas aproximadas del punto visado $P_j(X_j^0, Y_j^0)$, al montar el sistema de ecuaciones lineales el objetivo es calcular la variación de coordenadas, es decir, calcular la variación del valor real respecto del aproximado utilizado para el cálculo en base a las observaciones realizadas: $dX_j = X_j - X_j^0, dY_j = Y_j - Y_j^0$.

- La solución del sistema nos da: dX_{j1}, dY_{j1} , que utilizamos para calcular:

$$X_j = X_j^0 + dX_{j1}, Y_j = Y_j^0 + dY_{j1}.$$

- Entonces montamos una nueva iteración para el sistema, considerando:

$$X_j^{00} = X_j; Y_j^{00} = Y_j, \text{ entonces los nuevos valores a calcular } dX_{j2}, dY_{j2}, \text{ nos generan unas}$$

$$\text{nuevas soluciones: } X_j = X_j^{00} + dX_{j2}, Y_j = Y_j^{00} + dY_{j2}$$

- El sistema converge cuando $dX_{jn} \Rightarrow dX_{j(n-1)}$. En general, con unas coordenadas aproximadas calculadas a partir de las observaciones de campo, con una iteración es suficiente para obtener unos resultados buenos.

Hay formas de determinar los errores máximos permitidos en las observaciones para la obtención de unas coordenadas aproximadas buenas ([1], Págs. 145-146), expresando el error relativo en las coordenadas aproximadas en función de una distancia y acimut entre dos puntos próximos (el punto verdadero y el aproximado para el cálculo). Un error de 1 parte en 4000 en la distancia aproximada o de un minuto en su acimut, son siempre aceptables.

9.5. Relaciones de observación: Direcciones, distancias y acimutes Laplace

Recordando la expresión (1), vamos a determinar las ecuaciones de observación para los diferentes casos citados.

Para el caso de direcciones partimos de la definición de la relación de observación para un acimut y la aplicamos de la siguiente forma:

$$(23) \alpha^{(0)} + d\alpha - \alpha_{obs} = v_i$$

donde $d\alpha$ es la corrección que hemos de aplicar al acimut calculado $\alpha^{(0)}$ y v_i es el residuo.

El valor de α_{obs} es el valor observado en campo que se compone de la lectura angular realizada y de una desorientación:

$$(24) \alpha_{obs} = l_A + O = l_A + O_0 + dO$$

donde l_A , es la lectura angular realizada y O_0, dO son la desorientación calculada y la corrección a esa desorientación que entrará también como incógnita en el cálculo.

De esta manera, sustituyendo la expresión (25) en la expresión (24):

$$(25) \alpha^{(0)} + d\alpha - (l_A + O_0 + dO) = v_i$$

y ordenando convenientemente, llegamos a la relación de observación para direcciones:

$$(26) \alpha^{(0)} + (d\alpha - dO) - (l_A + O_0) = v_i$$

Para el caso de distancias, formamos la relación de observación de la siguiente forma:

$$(27) s^{(0)} + ds - s_{obs} = v_i$$

donde $s^{(0)}$ es la distancia calculada con las coordenadas aproximadas, ds es la corrección a aplicar a la distancia calculada (incógnita a calcular), s_{obs} es el valor de distancia observado y v_i es el residuo.

Por último, para el caso de acimutes Laplace, tenemos que la ecuación de Laplace es:

$$(28) A - \alpha = \eta \operatorname{tg} \varphi - [\xi \operatorname{sen} A - \eta \cos A] \operatorname{ctg} v$$

donde A es el acimut astronómico y α es el acimut geodésico, ξ, η son las componentes de la desviación de la vertical y v es el ángulo cenital observado. Esta expresión se puede simplificar a¹:

$$(29) A_{ij} - \alpha_{ij} = (\lambda_i^A - \lambda_i^G) \operatorname{sen} \varphi_i$$

siendo λ_i^A la longitud astronómica, φ_i, λ_i^G la latitud y la longitud geodésicas en P_i .

El acimut Laplace es el acimut astronómico corregido en la cantidad $(\lambda_i^A - \lambda_i^G) \operatorname{sen} \varphi_i$. La relación de observación se puede tratar de la misma forma que para el caso de un acimut teniendo en cuenta que aquí el acimut observado es el acimut astronómico, es decir, aplicamos la corrección $(\lambda_i^A - \lambda_i^G) \operatorname{sen} \varphi_i$ para obtener el acimut geodésico, donde la longitud geodésica λ_G puede diferir del valor final a calcular. Entonces la relación de observación que se plantea para el acimut Laplace α_L es la siguiente:

$$(30) \alpha_L^{(0)} + d\alpha_L - \alpha_{L_obs} = v_i$$

donde la expresión para $d\alpha_L^2$ (ver coeficientes a_i en expresiones (15) y (16)) es:

$$(31) a_1 d\varphi_i + (a_2 - \operatorname{sen} \varphi_i) d\lambda_i + a_3 d\varphi_j + a_4 d\lambda_j$$

$\alpha_L^{(0)}$, es el valor calculado con las coordenadas aproximadas y el valor α_{L_obs} es el acimut Laplace *provisional*, que equivale al valor observado, teniendo en cuenta que en su cálculo entra $\lambda_i^{(0)G}$ la longitud geodésica en P_i que pueden que no sea la final. Este valor α_{L_obs} es igual a:

$$(32) \alpha_{L_obs} = A_{ij} - (\lambda_i^A - \lambda_i^{(0)G}) \operatorname{sen} \varphi_i$$

¹ Martín-Asín, 1990, págs.. 215-218.

² Bomford, 1971, pág. 149.

9.6. Pesos de las observaciones

La media aritmética de una serie de valores nos da el valor medio. Si a estos valores les introducimos unos pesos en función de la importancia del valor asociado obtenemos una media ponderada desviando el valor medio en función de aquellos valores con mayor importancia, por ejemplo, una serie de tres valores v_i con sus respectivos pesos p_i :

$$(33) \text{ mediaPonderada} = \frac{p_1 v_1 + p_2 v_2 + p_3 v_3}{p_1 + p_2 + p_3}$$

Al formar un sistema de ecuaciones linealizado a partir de las ecuaciones de observación, cada valor observado debe ser multiplicado por la raíz cuadrada de su peso. Idealmente, el peso se define como el inverso del cuadrado del error estándar. El error estándar proviene de las precisiones con que se han hechos las medidas. Los pesos generalmente, son los mismos para cada tipo de medida, distancia o dirección, es decir, las observaciones realizadas con el mismo instrumento llevan asociado el mismo valor de peso.

Según la definición de peso:

$$(34) p = \frac{\sigma_0^2}{\sigma^2}$$

donde σ_0^2 es la varianza a priori de la observación de peso unidad o variación de referencia.

Para este coeficiente se suele adoptar un valor de 1, con lo que:

$$(35) p = \frac{1}{\sigma^2}$$

teniendo σ^2 como el cuadrado del error, también denominado como varianza.

A priori se definen los pesos como el producto de la varianza a priori por la matriz cofactor: $\sigma_0^2 Q$. Esta matriz cofactor es igual a la inversa de la matriz de pesos, es decir: $\sigma_0^2 Q = \sigma_0^2 P^{-1}$.

Entonces definimos la matriz de pesos P como:

$$(36) P = \begin{vmatrix} p_1 & \dots & \dots \\ \dots & \dots & \dots \\ \dots & \dots & p_n \end{vmatrix} = \sigma_0^2 \begin{vmatrix} 1/\sigma_1^2 & \dots & \dots \\ \dots & \dots & \dots \\ \dots & \dots & 1/\sigma_n^2 \end{vmatrix} = \sigma_0^2 Q \rightarrow P = \sigma_0^2 \begin{vmatrix} 1/\sigma_1^2 & \dots & \dots \\ \dots & \dots & \dots \\ \dots & \dots & 1/\sigma_n^2 \end{vmatrix}$$

Una vez definidos los pesos para cada ecuación de observación se monta el sistema de ecuaciones normales cuya solución estará afectada por la configuración de dichos pesos.

$$(37) \xrightarrow{\text{sistema lineal}} Ax = t \xrightarrow{\text{ec.Observación}} Ax - t = v \xrightarrow{\text{resolver}} x = (A^t PA)^{-1} A^t Pt$$

donde A es la matriz de coeficientes de las fórmulas diferenciales descritas en los apartados anteriores, x es el vector de incógnitas (diferenciales de las coordenadas), t el vector de datos (valores observados menos los aproximados) y v el vector de residuos.

9.7. Elipse de error

La matriz de varianzas-covarianzas que obtenemos después de la compensación de la red geodésica incluye las precisiones del ajuste:

$$(38) Q_{\hat{x}\hat{x}} = N^{-1} = (A^t PA)^{-1}$$

En esta matriz tenemos los elementos de la diagonal principal denominados varianzas y el resto son las covarianzas:

$$(39) Q_{\hat{x}\hat{x}} = \begin{vmatrix} Q_{xx} & Q_{xy} \\ Q_{yx} & Q_{yy} \end{vmatrix}$$

La elipse de error nos da una indicación gráfica y relativa respecto de otros puntos, del error en un punto de la red en función de las componentes del ajuste. En la siguiente figura vemos los parámetros que definen esta elipse de error:

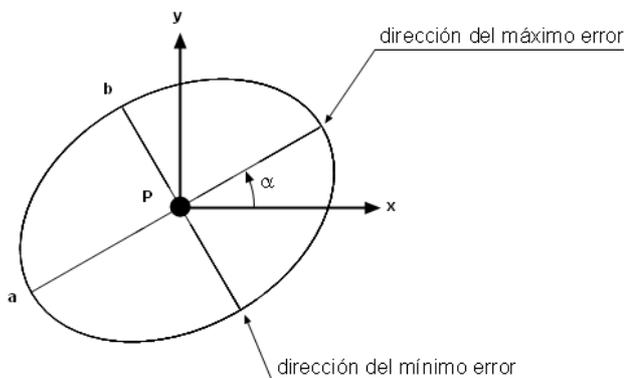


Figura 3. Elipse de error

Para los cálculos de estos parámetros de la elipse de error utilizamos la matriz de varianzas covarianzas:

$$(40) \operatorname{tg} 2\alpha = \frac{2Q_{xy}}{Q_{xx} - Q_{yy}} \xrightarrow{4 \text{ valores posibles}} \begin{aligned} \rightarrow \cos 2\alpha &= \frac{Q_{xx} - Q_{yy}}{\pm W} \\ \rightarrow \operatorname{sen} 2\alpha &= \frac{2Q_{xy}}{\pm W} \end{aligned}$$

Estos dos valores expuestos pueden tener el denominador + ó -, en estos casos, estaremos representando la dirección del semieje mayor (a) o la dirección del semieje menor (b) respectivamente. El valor de W es: $W = \sqrt{(Q_{xx} - Q_{yy})^2 + 4Q_{xy}^2}$.

Los valores de los errores máximo y mínimo representados por las longitudes de los semiejes mayor y menor (a) y (b) vienen dados por:

$$(41) \quad \begin{aligned} a &= \hat{\sigma}_0 \sqrt{Q_{uu}} \xrightarrow{\text{donde}} Q_{uu} = \frac{1}{2}(Q_{xx} + Q_{yy} + W) \\ b &= \hat{\sigma}_0 \sqrt{Q_{vv}} \xrightarrow{\text{donde}} Q_{vv} = \frac{1}{2}(Q_{xx} + Q_{yy} - W) \end{aligned}$$

Bibliografía

- [1] Bomford, G. *Geodesy*. Third edition. Oxford University Press, 1971.
- [2] Martín Asín, F. *Geodesia y cartografía matemática*. 3ª Edición, Editorial Paraninfo. Madrid 1990
- [3] Sevilla, M.J., Linkwitz, H. y Henneberg, H. *III Curso de Geodesia Superior*. Cursos y Seminarios del Instituto Geográfico Nacional. Madrid, 1985.
- [4] Torge, W. *Geodesy*. 2nd Edition. Walter de Gruyter, Berlin – New York, 1991. Pags. 44-49.

Tema 10. Aplicación del método de mínimos cuadrados a la compensación de redes geodésicas. Análisis estadístico de los resultados. Concepto de precisión y fiabilidad. Incertidumbre de medida.

10.1. Aplicación del método de mínimos cuadrados a la compensación de redes geodésicas

Una red geodésica es un conjunto de puntos perfectamente localizados en el terreno mediante monumentaciones adecuadas entre los que se han efectuado observaciones de tipo geodésico con el fin de obtener sus coordenadas, precisión y fiabilidad en términos relativos y absolutos, respecto de un sistema de referencia establecido.

Una vez realizadas las observaciones en una red geodésica es necesario efectuar un tratamiento de los datos para obtener resultados homogéneos, únicos y compatibles con las observaciones.

La red geodésica tiene por objeto la estimación de unos parámetros que describan la geometría de la superficie terrestre o de una parte de ella a partir de valores observados.

Esto plantea dos cuestiones importantes: por una parte habrá que establecer las relaciones matemáticas entre los parámetros y las observaciones (basadas en las propiedades geométricas o físicas del problema), esto nos lleva al denominado MODELO FUNCIONAL. Por otra parte, habrá que describir el comportamiento estadístico de las variables aleatorias que intervengan, lo que nos lleva al llamado MODELO ESTOCÁSTICO, este modelo viene representado por el valor esperado de dichas variables y su matriz de varianza-covarianza.

Ambos modelos, el funcional y el estocástico, constituyen el MODELO MATEMÁTICO en la compensación de redes geodésicas.

En toda red geodésica se disponen de más datos de los necesarios para la resolución del problema, es decir, tenemos redundancia de observaciones por ello es necesario recurrir a los métodos de compensación o ajuste que tienen por objeto obtener unos resultados que mejor resultan el problema.

El objetivo del modelo matemático será conseguir la mejor solución o aproximación para las coordenadas finales de los puntos de la red geodésica, solución que está relacionada con el concepto de mínima distancia. La solución de mínimos cuadrados aplicada a este problema satisface la búsqueda de mejor solución o aproximación ya que minimiza la suma al cuadrado de los residuos en el ajuste de la red, es decir, que es la que mejor aproxima los valores aproximados a la solución verdadera.

Los diferentes modelos funcionales, una vez linealizados, conducen a la resolución de sistemas de ecuaciones lineales del tipo:

$$(1) Ax = y$$

donde x es el vector de parámetros, y el vector de observaciones y A es una matriz que los relaciona denominada matriz de diseño.

Las cuestiones sobre la existencia y unicidad de solución del sistema (1) pueden analizarse con las técnicas habituales del álgebra lineal. El problema se reduce a estudiar la matriz de diseño A . Supongamos que la matriz A tiene m filas (ecuaciones) y n columnas (incógnitas), entonces podemos tener:

1.- A de rango completo: $\text{rango}(A)=\min(m,n)$, entonces nos podemos encontrar con los siguientes casos:

- $\text{rg}=m=n$. Este caso tiene solución única ya que A es regular y existe su inversa de manera que: $x = A^{-1}y$. Este sistema tiene tantas ecuaciones como incógnitas y se denomina *modelo determinado*.
- $\text{rg}=m < n$. El modelo tiene infinitas soluciones exactas. Es un modelo con más incógnitas que ecuaciones denominado *modelo indeterminado*.
- $\text{rg}=n < m$. El modelo en este caso no tiene solución exacta, tiene más ecuaciones que incógnitas pero existen infinitas soluciones aproximadas entre la que habrá que elegir la mejor.

Los modelos con $r=m \leq n$ con solución exacta se denominan *modelos consistentes*. Los modelos que no tienen solución exacta se denominan *modelos inconsistentes*.

2.- A deficiente de rango: $\text{rango}(A)=r < \min(m,n)$, en este caso podemos encontrarnos las siguientes situaciones:

- $r < m < n$. Este modelo tiene infinitas soluciones exactas, es un modelo consistente pero con la matriz A deficiente de rango.
- $R < n < m$. En este caso el modelo no tiene solución exacta, es un modelo inconsistente con infinitas soluciones aproximadas y con A deficiente de rango.

De estos casos nos vamos a quedar con los modelos deterministas de infinitas soluciones aproximadas en los que hay que buscar la mejor aproximación, que es aquella relacionada con el concepto de mínima distancia.

La solución mínimos cuadrados aplicada al modelo lineal (1) es:

$$(2) x = \min(\|y - Ax\|)$$

x es la solución que hace mínima la forma cuadrática:

$$(3) (y - Ax)^T P (y - Ax) \approx \sum v_i^2$$

Donde v_i son los residuos para cada ecuación de observación.

Desde el punto estocástico, se considera el modelo lineal:

$$(4) \quad Ax - t = v$$

donde t y v son magnitudes aleatorias con las siguientes características estocásticas que se expresan por la esperanza matemática de la siguiente forma:

$$(5) \quad E(t) = Ax \text{ o } E(v) = 0$$

y por la matriz varianza-covarianza:

$$(6) \quad E\{(t - Ax)(t - Ax)^T\} = \sum_{vv} = \sigma_0^2 Q = \sigma_0^2 P^{-1}$$

donde Q es la matriz cofactor, P la matriz de pesos y σ_0^2 la varianza de la unidad de pesos.

Esto constituye el denominado modelo de Gauss-Markov que se representa abreviadamente por (t, Ax, Σ) .

A partir de las observaciones t y conociendo su matriz de varianza-covarianza Σ se trata de obtener unos valores de los parámetros x que haga que la suma de los cuadrados de los residuos sea mínima.

Diremos que \hat{x} es una estimación insesgada de x cuando:

$$(7) \quad E(\hat{x}) = x$$

Entonces, se dice que \hat{x} es una estimación mínima varianza de x cuando:

$$(8) \quad E\{(\hat{x} - E(x))^2\} = \text{mínimo}$$

La aplicación del método de mínimos cuadrados en la compensación de redes geodésicas se define a través de modelo matemático, que recordamos, que se compone de un modelo funcional y un modelo estocástico. Este modelo lo vamos a desarrollar para el ajuste de observaciones indirectas conocido como método de ecuaciones de observación y para el ajuste de observaciones condicionadas conocido como método de ecuaciones de condición.

Método de ecuaciones de observación

1. Modelo Funcional

Tenemos:

$$(9) \quad L = F(X)$$

donde X es el vector de parámetros ajustados de dimensión $(n, 1)$ y L el vector de observaciones ajustadas de dimensión $(m, 1)$. Siendo m el número de ecuaciones y n el de incógnitas.

El vector de parámetros incógnita x de dimensión $(n, 1)$ es:

$$(10) \quad x = X - X_0$$

Donde X_0 es el vector de valores aproximados de los parámetros X de dimensión $(n, 1)$.

El vector de observaciones calculadas L_0 se determina a partir de los valores aproximados X_0 :

$$(11) L_0 = F(X_0)$$

Por otra parte tenemos l como vector de observaciones reales de dimensión $(m, 1)$ y t , el vector de constantes de observación de dimensión $(m, 1)$, entonces:

$$(12) t = l - L_0$$

es decir, las constantes de observación son iguales a las observaciones realizadas en campo menos las observaciones calculadas a partir de los valores aproximados.

El vector de errores residuales v de dimensión $(m, 1)$ es igual a:

$$(13) v = L - l$$

En este sistema de ecuaciones tenemos que: $r = m - n$, que nos da la redundancia o grados de libertad en el ajuste.

La matriz de pesos P de las observaciones l tiene dimensión (m, m) y rango m . La matriz cofactor es igual a la inversa de la matriz de pesos $Q = P^{-1}$. La varianza a priori de la unidad de peso viene designada por σ_0^2 . La matriz varianza-covarianza Σ a priori de las observaciones l tiene dimensión (m, m) y rango m y es igual a:

$$(14) \Sigma = \sigma_0^2 Q = \sigma_0^2 P^{-1}$$

Por último, la matriz A de coeficientes de las ecuaciones de observación, también denominada matriz de diseño, con dimensión (m, n) que relaciona las observaciones con los parámetros incógnita es igual a:

$$(15) A = \left. \frac{\partial F}{\partial X} \right|_{X=X_0}$$

2. Ecuaciones de observación. Linealización del sistema de ecuaciones

La linealización del modelo (8) por el método de Taylor en X_0 :

$$(16) L = F(X_0) + \left. \frac{\partial F}{\partial X} \right|_{X=X_0} (X - X_0)$$

que con los valores (10), (14) y (9), lo escribimos como:

$$(17) L = L_0 + Ax$$

Utilizando $v = L - l \rightarrow L = l + v$ y teniendo en cuenta la expresión $t = l - L_0$ obtenemos:

$$(18) Ax - t = v$$

Este es el denominado *sistema de ecuaciones de observación* con m ecuaciones y n incógnitas, teniendo que $m > n$ por lo que es un sistema con infinitas soluciones aproximadas de las que hay que seleccionar la mejor aplicando la solución de mínimos cuadrados.

3. Modelo estocástico

Viene definido por:

$$(19) \quad E(t) = Ax \text{ o bien } E(v) = 0$$

4. Condición de mínimos cuadrados. Ecuaciones normales

Los parámetros x se determinan bajo la condición de mínimo en el espacio de observaciones:

$$(20) \quad v^T P v = \text{mínimo}$$

donde sólo se requiere el conocimiento a priori de la matriz de pesos.

Desarrollando la expresión cuadrática (20) en función de los parámetros, utilizando la expresión (18), resulta:

$$(21) \quad T(x) = (Ax - t)^T P (Ax - t) = \dots = x^T A^T P A x - 2t^T P A x + t^T P t$$

El mínimo de esta expresión se obtiene igualando a cero las derivadas parciales con respecto a x :

$$(22) \quad \frac{\delta T(x)}{\delta x} = 2\hat{x}^T A^T P A - 2t^T P A = 0 \xrightarrow{\text{matrices traspuestas}} A^T P A \hat{x} - A^T P t$$

Escribiendo esto según $N = A^T P A$ y $d = A^T P t$ con dimensión para N (n, n) y para d ($n, 1$), resulta el sistema de ecuaciones normales:

$$(23) \quad N\hat{x} - d = 0$$

donde \hat{x} es la estimación mínimos cuadrados de x . La solución de \hat{x} es:

$$(24) \quad \hat{x} = N^{-1} d = (A^T P A)^{-1} A^T P t$$

Con los valores estimados de x , podemos obtener:

- $\hat{X} = X_0 + \hat{x} \rightarrow$ los valores estimados de los parámetros ajustados
- $\hat{v} = A\hat{x} - t \Rightarrow$ los valores estimados de los residuales
- $\hat{L} = l + \hat{v} \rightarrow$ y los valores estimados de las observaciones

5. Estimación de la precisión a posteriori y matrices cofactor y covarianzas

Se trata de calcular una estimación $\hat{\sigma}_0^2$ de la varianza a priori σ_0^2 de la unidad de peso. Esta estimación se llama varianza a posteriori de la unidad de peso.

Esta estimación se calcula¹ a partir de la esperanza matemática:

$$(25) \quad E\{\hat{v}^T P \hat{v}\} \longrightarrow \hat{\sigma}_0^2 = \frac{\hat{v}^T P \hat{v}}{m - n}$$

En el caso de que la matriz A sea deficiente de rango ($r < n$) entonces:

$$(26) \quad \hat{\sigma}_0^2 = \frac{\hat{v}^T P \hat{v}}{m - r}$$

¹ Para más detalle consultar [1] Sevilla Págs. 19-21

Para obtener las matrices cofactor a posteriori de los parámetros \hat{x} aplicamos la ley de propagación de las covarianzas a la expresión (24):

$$(27) Q_{\hat{x}\hat{x}} = N^{-1} A^T P Q_{tt} P A N^{-1} = N^{-1}$$

donde Q_{tt} es la matriz cofactor a priori de las constantes de observación $Q_{tt} = Q = P^{-1}$.

Y la matriz de varianza-covarianza:

$$(28) \Sigma_{\hat{x}\hat{x}} = \hat{\sigma}_0^2 N^{-1}$$

Método de ecuaciones de condición

Aunque el método de ecuaciones de observación es el más habitual de los empleados para la compensación de redes geodésicas, este método es también un modelo que aparece en Geodesia frecuentemente sobre todo en problemas de nivelación y gravimetría. En este método no aparecen los parámetros x , en este caso las observaciones tienen que cumplir condiciones y obtenemos al final del ajuste una estimación de las observaciones.

1. Modelo funcional

Las observaciones tienen que cumplir la siguiente condición:

$$(29) F(L) = 0$$

donde L es el vector de valores ajustados de las observaciones, con dimensión $(m, 1)$.

El vector de errores residuales v se define como:

$$(30) v = L - l$$

teniendo l como vector de observaciones reales de dimensión $(m, 1)$.

Definido c como el número de condiciones (ecuaciones), el número de grados de libertad r del ajuste coincide con el número de ecuaciones independientes, entonces $r=c$.

La matriz de pesos, la matriz cofactor, varianza a priori y la matriz varianza-covarianza a priori, quedan definidos de la misma manera que en el caso anterior, respectivamente:

$$(31) P = P^T, Q = P^{-1}, \sigma_0^2, \Sigma = \sigma_0^2 P^{-1}$$

La matriz B de diseño, con dimensión (c, m) en este caso es:

$$(32) B = \left. \frac{\partial F}{\partial L} \right|_{L=L_0}$$

El vector de constantes de observación t (términos independientes):

$$(33) t = -[F(L_0) + Bl - BL_0]$$

Por último, se define A como el vector de multiplicadores de Lagrange de dimensión $(c, 1)$.

2. Linealización. Ecuaciones de condición

La Linealización de la ecuación (29) por el desarrollo en serie de Taylor:

$$(34) F(L_0) + \left. \frac{\delta F}{\delta L} \right|_{L=L_0} (L - L_0) = 0$$

Teniendo en cuenta (30), (32) y (33):

$$(35) Bv - t = 0$$

Este es el sistema de ecuaciones de condición con c ecuaciones y m incógnitas.

3. Modelo estocástico

Viene definido por:

$$(36) E(v) = 0 \text{ o bien } E(t) = BE(v) = 0$$

4. Condición de mínimos cuadrados. Ecuaciones correlativas y ecuaciones normales

Los valores más probables o estimaciones de los errores residuales v , se determinan de forma que verifiquen las ecuaciones de condición (35) y a la vez minimicen la forma cuadrática:

$$(37) v^T P v = \text{mínimo}$$

Se trata, por tanto, de encontrar un mínimo para la función matricial:

$$(38) T(v, \Lambda) = v^T P v - 2\Lambda^T (Bv - t)$$

El mínimo de esta ecuación se obtiene igualando a 0 las derivadas parciales respecto de cada variable:

$$(39) \begin{aligned} \frac{\delta T(v, \Lambda)}{\delta v} &= 2\hat{v}^T P - 2\Lambda^T B = 0 \\ \frac{\delta T(v, \Lambda)}{\delta \Lambda} &= -2(B\hat{v} - t) = 0 \end{aligned}$$

Trasponiendo la primera ecuación, simplificando y expresando de forma matricial, tenemos:

$$(40) \begin{bmatrix} -P & B^T \\ B & 0 \end{bmatrix} \begin{bmatrix} \hat{v} \\ \Lambda \end{bmatrix} = \begin{bmatrix} 0 \\ t \end{bmatrix}$$

Este sistema con m+c ecuaciones y m+c incógnitas se llama de *ecuaciones correlativas*.

De la primera ecuación obtenemos:

$$(41) \hat{v} = P^{-1} B^T \Lambda = 0$$

y sustituyendo en la segunda:

$$(42) B P^{-1} B^T \Lambda - t = 0$$

Y utilizando $M = B P^{-1} B^T$ la notación llegamos a:

$$(43) M \Lambda - t = 0$$

Este es un sistema de c ecuaciones con c incógnitas y se llama sistema de *ecuaciones normales*.

El sistema se resuelve directamente por inversión de la matriz M :

$$(44) \Lambda = M^{-1} t$$

y sustituyendo en (41) obtenemos la solución única para \hat{v} :

$$(45) \hat{v} = P^{-1} B^T M^{-1} t = P^{-1} B^T (B P^{-1} B^T)^{-1} t$$

Con los valores estimados de v , obtenemos:

$$(46) \hat{L} = l + \hat{v} \rightarrow \text{los valores estimados de las observaciones}$$

5. Estimación de la precisión a posteriori y matrices cofactor y covarianzas

Tratamos de calcular una estimación $\hat{\sigma}_0^2$ de la varianza a priori σ_0^2 de la unidad de peso. Esta estimación se llama varianza a posteriori de la unidad de peso.

Esta estimación se calcula² a partir de la esperanza matemática:

$$(47) E\{\hat{v}^T P \hat{v}\} \rightarrow \hat{\sigma}_0^2 = \frac{\hat{v}^T P \hat{v}}{c}$$

Para las observaciones l y las constantes t , el modelo estocástico define las matrices:

$$(48) \begin{aligned} Q_{ll} &= Q = P^{-1} \xrightarrow{\text{varianza}} \Sigma_{ll} = \sigma_0^2 Q = \sigma_0^2 P^{-1} \\ Q_{tt} &= BQB^T = M \xrightarrow{\text{varianza}} \Sigma_{tt} = B\Sigma B^T = \sigma_0^2 M \end{aligned}$$

Para obtener las matrices cofactor a posteriori de los multiplicadores $\hat{\Lambda}$ y de los errores residuales \hat{v} , aplicamos la ley de propagación de las covarianzas a las expresiones (44) y (45), y obtenemos:

$$(49) \begin{aligned} Q_{\Lambda\Lambda} &= M^{-1} Q_{tt} (M^{-1})^T = M^{-1} M M^{-1} = M^{-1} \\ Q_{\hat{v}\hat{v}} &= P^{-1} B^T M^{-1} Q_{tt} (P^{-1} B^T M^{-1})^T = \dots = P^{-1} B^T M^{-1} B P^{-1} \end{aligned}$$

Y las matrices varianza:

$$(50) \begin{aligned} \Sigma_{\Lambda\Lambda} &= \hat{\sigma}_0^2 M^{-1} \\ \Sigma_{\hat{v}\hat{v}} &= \hat{\sigma}_0^2 P^{-1} B^T M B P^{-1} \end{aligned}$$

Existe un tercer modelo matemático, menos utilizado, denominado modelo mixto en el que se incluyen ecuaciones de observación junto con ecuaciones de condición. El sistema de ecuaciones es el siguiente:

$$(51) Ax - Bv - t = 0$$

Este es un sistema con c ecuaciones (número de ecuaciones de condición) y $m+n$ incógnitas (m residuales y n parámetros) en el que $m+n > c$, se denomina sistema de ecuaciones de condición, es indeterminado y tiene infinitas soluciones.

La condición de mínimos cuadrados se consigue haciendo mínimo:

$$(52) T(v, \Lambda, x) = \text{mínimo} \rightarrow v^T P v - 2\Lambda^T (Ax + Bv - t) = \text{mínimo}$$

Derivando respecto de cada variable se llega a las ecuaciones correlativas y al sistema de ecuaciones normales:

$$(53) N\hat{x} - d = 0$$

Este es un sistema con n ecuaciones y n incógnitas.

La solución a este sistema viene dada por:

$$(54) \hat{x} = N^{-1}d$$

$$(55) \hat{x} = (A^T M^{-1} A)^{-1} A^T M^{-1} t$$

Y la precisión a priori:

$$(56) \hat{\sigma}_0^2 = \frac{\hat{v}^T P \hat{v}}{r}$$

donde $r=c-n$ es el número de grados de libertad del ajuste.

10.2. Análisis estadístico de los resultados

En el apartado anterior se ha planteado el modelo de Gauss-Markov:

$$\begin{aligned} A\hat{x} - t &= v \\ E(t) &= A\hat{x} \text{ o bien } E(v) = 0 \\ (57) \text{cov}(\hat{v}) &= E(\hat{v}\hat{v}^T) = \sigma_0^2 Q = \Sigma \\ v &\approx N(0, \sigma_0^2 Q) \end{aligned}$$

Vemos que v sigue una distribución normal m -dimensional de media 0 y matriz $\sigma_0^2 Q$ de varianza-covarianza con σ_0^2 de varianza a priori de peso unidad.

En este modelo se ha supuesto algunas hipótesis que hay que contrastar:

1. El vector de residuos sigue una distribución normal
2. $E(v)=0$, lo que equivale a decir que no existen errores sistemáticos ni groseros
3. Las precisiones a priori de las observaciones determinadas en la matriz $\sigma_0^2 Q$ han de ser consistentes con los resultados a posteriori.

Estas hipótesis se pueden contrastar con test estadísticos de contraste de hipótesis.

El análisis estadístico de los resultados del ajuste se hace con el fin de:

- chequear las hipótesis del modelo Gauss-Markov
- detectar errores groseros y posibles sistematismos
- estudiar la fiabilidad de la red
- encontrar el modelo óptimo

Se aplican los siguientes test estadísticos:

1. Test de varianza de peso unidad

En este test se contrasta la hipótesis nula: $H_0 : \sigma_0^2 = \hat{\sigma}_0^2$

Se define el estadístico:

² Para más detalle consultar [1] Sevilla. Págs. 34-37

$$(58) Y = \text{máx} \left[\frac{\sigma_0^2}{\hat{\sigma}_0^2}, \frac{\hat{\sigma}_0^2}{\sigma_0^2} \right]$$

La hipótesis nula H_0 se rechaza si $Y > F_{f, \infty, \alpha}$, donde f y ∞ son los grados de libertad y F es el valor de la F de Snedecor a un nivel de significancia de α .

2. Test de normalidad de los residuales

Para investigar la normalidad de la distribución de los residuos se puede utilizar el test X^2 . La hipótesis a contrastar es si los residuales tipificados siguen una distribución normal $N(0,1)$.

Los residuales tipificados son $\frac{\hat{v}_i}{\sigma_{\hat{v}_i}}$.

La hipótesis se acepta si:

$$(59) \sum_{i=1}^p \frac{(n_i - mp_i)^2}{mp_i}$$

es menor que el valor de la distribución $X^2_{(p-1), \alpha}$, siendo p el número de clases, n_i la frecuencia absoluta de cada clase, m el número de observaciones y p_i la probabilidad teórica de cada clase, es decir, probabilidad teórica de que un resultado esté en el intervalo i -ésimo.

La no aceptación de este test puede ser debida a dos causas:

- existencia de sistematismos en el ajuste o de errores groseros en las observaciones
- este test tiene sentido riguroso cuando los residuos son variables incorreladas y esto no es cierto en la práctica.

3. Análisis de posibles sistematismos

Para averiguar si existen sistematismos es aconsejable utilizar el t-test.

Se define el estadístico:

$$(60) Y = \frac{\bar{w}_m}{\sigma_{\bar{w}_m}}$$

en donde \bar{w}_m es el valor medio de los residuos tipificados y $\sigma_{\bar{w}_m}$ su desviación típica. La hipótesis nula se plantea como la no existencia de sistematismos:

$$(61) H_0 : \bar{w} = 0$$

La hipótesis se acepta si $Y < t_{(m-1), \alpha}$ donde, $t_{(m-1)}$ es el valor de la distribución t de Student con $m-1$ grados de libertad y α es el nivel de significación. (m es el número de observaciones)

4. Detección de errores groseros

Se denomina fiabilidad de una red geodésica a su resistencia a los errores groseros, es decir, a la facilidad de la red para la detección de los errores groseros.

Cuando más precisas y homogéneas sean las observaciones y óptima la configuración de la red, un error será más fácilmente detectable.

Los test más utilizados en geodesia para estos fines son el t-test de Pope y el test de Baarda.

Test Tau de Pope

La hipótesis sobre cada observación se plantea de la siguiente forma:

H_0 : se acepta la observación i -ésima

H_1 : existe un error grosero en la observación i -ésima y se rechaza

El test se aplica a los residuos tipificados $\frac{\hat{v}_i}{\sigma_{\hat{v}_i}}$.

Se acepta si $\left| \frac{\hat{v}_i}{\sigma_{\hat{v}_i}} \right| < \Gamma$, donde

$$(62) \Gamma = \frac{\sqrt{r} t_{r-1, 1-\alpha}}{\sqrt{r-1 + t_{r-1, 1-\alpha}^2}}$$

Este es el valor crítico que depende del nivel de confianza $(1-\alpha)$ y del número de grados de libertad r . t es la variable t-student.

Se rechaza H_0 , es decir, se acepta H_1 si $\left| \frac{\hat{v}_i}{\sigma_{\hat{v}_i}} \right| \geq \Gamma$.

El valor de la precisión de los residuos tipificados es:

$$(63) \hat{\sigma}_{\hat{v}_i}^2 = \hat{\sigma}_0^2 (P^{-1} - AN^{-1}A^T)_{ii}, \text{ donde } N = A^T P A$$

Test de Baarda

Una forma de analizar una red es realizar el examen de los residuales \hat{v} que produce el ajuste mínimos cuadrados.

Si el vector \hat{v} estuviera formado por variables incorreladas, entonces la distribución de los residuos tipificados $\frac{\hat{v}_i}{\sigma_{\hat{v}_i}}$ será $N(0, 1)$, aunque esto en la práctica esto no sucede.

Baarda(1968) demuestra que un test es más sensible con respecto a la detección de errores groseros si se contrasta la cantidad $P\hat{v}$, siendo P la matriz de pesos de las observaciones.

Cada observación puede analizarse por separado para lo cual se introduce el vector:

$$(64) \mathbf{c} = \begin{pmatrix} 0 \\ \dots \\ 1_{(i)} \\ \dots \\ 0 \end{pmatrix}$$

si se quiere testear la observación i -ésima.

Se define la cantidad:

$$(65) \mathbf{f} = -\mathbf{c}^T \mathbf{P}\hat{\mathbf{v}}$$

y se calcula su varianza aplicando la ley de propagación de las varianzas:

$$(66) \sigma_f^2 = \mathbf{c}^T \mathbf{P} \sum_{\hat{\mathbf{v}}} \mathbf{P}\mathbf{c} = \sigma_0^2 \mathbf{c}^T \mathbf{P}\mathbf{Q}_{\hat{\mathbf{v}}}\mathbf{P}\mathbf{c}$$

donde:

$$(67) \mathbf{Q}_{\hat{\mathbf{v}}} = \mathbf{P}^{-1} - \mathbf{A}(\mathbf{A}^T \mathbf{P}\mathbf{A})^{-1} \mathbf{A}^T = \mathbf{Q} - \mathbf{A}\mathbf{Q}_{\hat{\mathbf{x}}}\mathbf{A}^T$$

Se define la cantidad σ_f :

$$(68) \sigma_f = \sigma_0 \sqrt{\mathbf{c}^T \mathbf{P}\mathbf{Q}_{\hat{\mathbf{v}}}\mathbf{P}\mathbf{c}}$$

Consideramos w_i :

$$(69) w_i = \frac{\mathbf{f}}{\sigma_f}$$

que sigue una distribución $N(0,1)$ aproximadamente, debido a la correlación de las variables. El w -test de Baarda plantea la normalidad de los residuales w_i . Dado el nivel de significación α , el valor crítico w_0 se obtiene por:

$$(70) w_0 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{w_0} e^{-t^2/2} dt = 1 - \alpha$$

Si $|w_i| \geq w_0$ existe una probabilidad $(1-\alpha)$ de que la observación i -ésima sea errónea. Distintas elecciones del vector \mathbf{c} para $i=1, \dots, m$ van definiendo las cantidades w_i y se testean los residuales individualmente.

10.3. Concepto de precisión y fiabilidad

Generalmente se considera que la calidad de una red geodésica depende de tres factores: el económico, la precisión y la fiabilidad. Es evidente que a más precisión y fiabilidad el coste es mayor, tanto por plazos de tiempo a invertir como por el instrumental a utilizar.

La precisión es la medida de las características de la red en el aspecto de propagación de errores aleatorios y la fiabilidad describe la habilidad de las observaciones redundantes para detectar errores en la red. La precisión nos indica la capacidad de un sistema de medida de dar el mismo

resultado en mediciones diferentes realizadas en las mismas condiciones. Este concepto está asociado al parámetro dispersión en el sentido que el resultado de sucesivas medidas de un mismo mensurando es más preciso, cuanto menor dispersión tenga el conjunto de medidas realizadas para llegar a ese resultado.

El concepto de precisión aplicado a redes geodésicas está íntimamente ligado al sistema de referencia elegido. En una red bidimensional son necesarios cuatro parámetros para fijar el sistema de referencia, las coordenadas de un punto, una orientación y una escala.

Con estos cuatro parámetros compensamos la red y determinamos las coordenadas y las precisiones del resto de los puntos respecto a ese datum elegido. La elección de un datum determinado permite obtener un conjunto infinito de soluciones (coordenadas de los puntos de la red) y matrices varianza-covarianza de estas coordenadas que dan la precisión de la red para cada solución.

A la hora de dar un parámetro de precisión en la compensación de redes geodésicas

Si suponemos una red que ha sido observada dos veces y obtenemos las matrices de covarianzas a posteriori $Q_{\hat{x}_1}$ y $Q_{\hat{x}_2}$, referidas al mismo sistema de referencia, podríamos determinar qué medida es la mejor comparando estas dos matrices pero esta metodología no es práctica ya que suelen ser matrices muy grandes. Entonces, el conjunto de la información debe ser comprimido en una o varias medidas representativas que puedan ser fácilmente comprensibles.

Para conocer las precisiones de cada una de las coordenadas calculadas, partimos de la matriz de varianzas-covarianzas a posteriori:

$$(71) \Sigma_{\hat{x}\hat{x}} = \hat{\sigma}_0^2 Q_{\hat{x}\hat{x}}$$

donde para el caso de red bidimensional en un sistema coordenado x-y, podemos calcular la precisión de cada componente a partir de la desviación estándar de las coordenadas.

Si la matriz de covarianzas es:

$$(72) Q_{\hat{x}\hat{x}} = \begin{vmatrix} q_{xx} & q_{xy} \\ q_{yx} & q_{yy} \end{vmatrix}$$

y conocida la estimación de la precisión a posteriori $\hat{\sigma}_0 = \sqrt{\frac{\hat{v}^T P \hat{v}}{gl}}$, con gl grados de libertad

en el ajuste, tenemos que la desviación estándar de las coordenadas viene dada por:

$$(73) \hat{\sigma}_x = \hat{\sigma}_0 \sqrt{q_{xx}} \quad ; \quad \hat{\sigma}_y = \hat{\sigma}_0 \sqrt{q_{yy}}$$

Estos valores se suelen llamar errores medios cuadráticos de las coordenadas y están asociados con el valor de máxima dispersión o agrupación de la medida y por lo tanto a su precisión.

En el modelo de Gauss-Markov se plantea que $E\{\hat{\sigma}_0^2\} = \sigma_0^2$, pero resulta que la varianza a posteriori es un mal estimador de la varianza a priori cuando el número de repeticiones de la misma medida es pequeño, es decir, hay una dependencia con la redundancia y es recomendable

sustituir las ecuaciones (73) en sus respectivos intervalos de confianza. En este caso estaríamos hablando de incertidumbre de la medida que se ve en el apartado siguiente.

En cuanto a la fiabilidad de la red primero hay que citar las condiciones de partida requeridas para que se pueda determinar la fiabilidad en las operaciones geodésicas

- técnicos bien entrenados, responsables y cuidadosos
- instrumentación precisa y calibrada
- emplear procesos de medida que contemplen comprobaciones
- diseños de red en los que todas las observaciones sean realizadas por más de un técnico

Partiendo de estas condiciones, la red estaría preparada para detectar errores sistemáticos y groseros asumidos por el modelo Gauss-Markov aplicado en la compensación de la red

La fiabilidad en las redes geodésicas puede ser clasificada en:

- Fiabilidad interna: se refiere a la capacidad de la propia red para detectar errores sistemáticos y localizar errores groseros.
- Fiabilidad externa: es la capacidad para determinar el efecto del error mínimo detectable sobre la geometría de la red.

En ambos casos existen medidas globales y locales de la fiabilidad.

La medida global de la fiabilidad interna expresa la probabilidad de detección de desviaciones del modelo, especialmente de errores groseros y sistemáticos. La medida local de la fiabilidad interna expresa la probabilidad de detección de *outlier*.

En cuanto a la fiabilidad externa, la medida global indica que un modelo con alta fiabilidad externa responde de manera no significativa a errores no detectados. La medida local indica que si se pierden observaciones buenas el sistema no se vería afectado indicando una buena fiabilidad externa local.

10.4. Incertidumbre de medida

La incertidumbre de una medida es un parámetro asociado al resultado de una medición que caracteriza el intervalo de valores que puede ser razonablemente atribuido a la magnitud a medir, es decir, el concepto de incertidumbre de una medida refleja la duda que existe sobre la veracidad del resultado obtenido una vez evaluadas las posibles fuentes de error.

En general, el resultado de una medida es únicamente una aproximación o estimación del valor de la cantidad objeto de medida y entonces, el resultado solo es completo si va acompañado del valor cuantitativo de la incertidumbre. Por lo tanto, la incertidumbre de la medida nos da una idea de la calidad del resultado ya que muestra el intervalo alrededor del valor estimado dentro del cual se encuentra el valor verdadero.

Para el caso de los resultados de la compensación de una red geodésica, partiendo de las ecuaciones (73) que representan el grado de dispersión de la medida, vamos a formar los intervalos de confianza para determinar la incertidumbre en la medida de las coordenadas.

El intervalo de confianza se deriva de la distribución t de Student conocida la varianza a posteriori del ajuste $\hat{\sigma}_0^2$:

$$(74) (x - \hat{x}) / \hat{\sigma}_x \approx t(n - r)$$

donde $n-r$ son los grados de libertad, entonces la probabilidad se determina por:

$$(75) P\left\{\hat{x} - t_{\alpha/2} \hat{\sigma}_x < x < \hat{x} + t_{\alpha/2} \hat{\sigma}_x\right\} = 1 - \alpha$$

y de forma análoga para y , donde el valor $t_{\alpha/2}$ depende del nivel de confianza $1-\alpha$ y de la redundancia del modelo Gauss-Markov.

Bibliografía

- [1] Caspary, W.F. *Concepts of network and deformation analysis*. Edited by J.M. Rüeger. The University of New South Wales. Kensington, Australia. August 1988.
- [2] Gil, A.J. *Apuntes de Redes Geodésicas y Cartografía Matemática*. Universidad de Jaén.
- [3] Grafarend, E.W. and Sansó, F. *Optimization and Design of Geodetic Networks*. Springer-Verlag, Berlin, 1985.
- [4] ISO/TAG 4. Published by ISO, 1993 (corrected and reprinted, 1995) in the name of the BIPM, IEC, IFCC, ISO, IUPAC, IUPAP and OIML. *Guide to the Expression of Uncertainty in Measurement*.
- [5] Núñez, A. *Concepto de precisión y fiabilidad en redes geodésicas*. Boletín del Servicio Geográfico del Ejército, nº68. Madrid 1989.
- [6] Sevilla, M.J., Linkwitz, H. y Henneberg, H. *III Curso de Geodesia Superior*. Cursos y Seminarios del Instituto Geográfico Nacional. Madrid, 1985.

Tema 11. Elasticidad. Ley de Hooke. Ecuaciones del movimiento en un medio elástico, homogéneo e isótropo. Funciones potencial de los desplazamientos y fuerzas. Función de Green.

11.1. Introducción

El estudio analítico de los movimientos sísmicos engloba al menos dos componentes principales: la descripción matemática de la fuente sísmica y las ecuaciones que rigen la propagación de los movimientos iniciados en la fuente.

Excepto en la cercanía inmediata de la fuente la mayoría de los movimientos producidos en el terreno son efímeros. Es decir, el terreno regresa a su posición inicial una vez estos movimientos han sucedido. Por este motivo, para desarrollar las ecuaciones que describen la propagación de los movimientos sísmicos, trabajaremos bajo la hipótesis de que la Tierra se comporta como un medio elástico.

De este modo, comenzaremos desarrollando aquellos principios de la teoría de la elasticidad que tienen aplicación en la sismología, alcanzando a enunciar la ecuación del movimiento. A continuación describiremos las denominadas funciones potenciales, útiles para resolver la citada ecuación, y por último aplicaremos la ecuación del movimiento al caso concreto de una fuerza impulsiva, obteniendo la Función de Green, con la cual caracterizaremos el medio de propagación.

11.2. Elasticidad

En aras de una mayor sencillez es común considerar que la Tierra es un medio *continuo* y *elástico*.

Un *medio continuo* es un medio ideal donde la distancia entre dos puntos contiguos puede hacerse infinitamente pequeña. De este modo usaremos el término “partícula” para describir un punto sin dimensiones. Asimismo, la densidad y las propiedades mecánicas serán consideradas como funciones continuas en el espacio, así como sus derivadas en las coordenadas espaciales.

Antes de explicar qué entendemos por un medio *elástico*, definiremos dos importantes conceptos: *deformación* y *esfuerzo*.

11.2.1. Deformación

Generalmente en la propagación de las ondas sísmicas tienen lugar pequeñas deformaciones del material (cambios relativos de longitud de $\sim 10^{-6}$) durante cortos periodos de tiempo (< 3600 s), por lo que de ahora en adelante nos referiremos a la teoría de *deformaciones infinitesimales*. Esto sin embargo no se cumple en el entorno inmediato de la fuente sísmica ni al considerar deformaciones de las fallas como las que se estudian en geología estructural.

Para describir el movimiento en cada punto del medio continuo utilizaremos el campo vectorial de desplazamientos $\mathbf{u}(\mathbf{x},t)$. Este vector representa la distancia de una partícula en un tiempo t , de la posición \mathbf{x} que ocupaba en un tiempo de referencia t_0 .

En este medio pueden tener lugar dos tipos de movimiento: i) translaciones y/o rotaciones del medio en su conjunto y ii) deformaciones internas.

Las *translaciones* pueden ser descritas por medio de un único vector desplazamiento común a todos los puntos del medio.

En cambio, las *deformaciones internas*, de las que trataremos ahora, van asociadas a variaciones espaciales y temporales del campo de desplazamientos $\mathbf{u}(\mathbf{x},t)$.

Consideremos dos partículas O y P del medio en las posiciones \mathbf{x}_O y \mathbf{x}_P separadas por una distancia infinitesimal ds en el instante t_1 . Al aplicar una fuerza al medio estas partículas pasan a la posición \mathbf{x}_O' y \mathbf{x}_P' separadas ahora por una distancia ds' en el instante t_2 -las posiciones que ocupan O y P en t_2 son las definidas por sus vectores desplazamiento $\mathbf{u}(\mathbf{x}_O,t)$, $\mathbf{u}(\mathbf{x}_P,t)$ -.

Para describir la deformación del medio tendremos que tener en cuenta tanto la variación de longitud de ds a ds' (*deformaciones longitudinales*) como la rotación de la línea ds para pasar a ds' (*deformaciones de cizalla*) relativa a un sistema fijo de coordenadas (ver *Figura 1*).

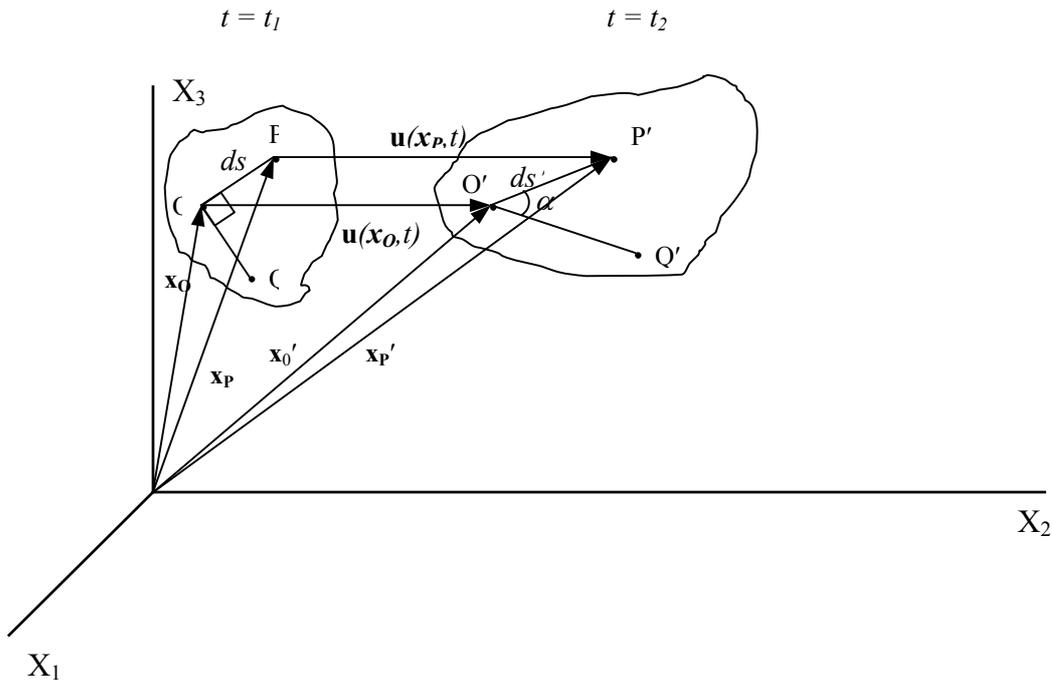


Figura 1. Deformación de un medio continuo.

En concreto, definimos deformaciones longitudinales como:

$$(1) e_l = \left(\frac{ds' - ds}{ds} \right)$$

Y deformaciones de cizalla como:

$$(2) e_s = \frac{1}{2} \left(\frac{\pi}{2} - \alpha \right)$$

Donde α es el ángulo que pasan a formar dos líneas que en el instante inicial eran perpendiculares.

Con el fin de definir la deformación en función del campo de desplazamientos $\mathbf{u}(\mathbf{x},t)$ utilizaremos un sistema de coordenadas cartesianas con ejes X_1 , X_2 y X_3 . En este sistema el vector desplazamiento vendrá dado por sus componentes $\mathbf{u}(\mathbf{x},t) = (u_1(\mathbf{x},t), u_2(\mathbf{x},t), u_3(\mathbf{x},t))$. Y como debemos describir todas las deformaciones en cada punto del medio con respecto a cada uno de los ejes coordenados, tendremos que utilizar nueve términos: tres deformaciones longitudinales e_{11} , e_{22} , e_{33} caracterizando las variaciones relativas en longitud de segmentos lineales orientados en la dirección de los ejes X_1 , X_2 y X_3 y seis deformaciones de cizalla e_{12} , e_{13} , e_{21} , e_{23} , e_{31} , e_{32} de cada una de las tres direcciones con respecto a las otras dos.

De este modo definimos el *tensor de deformaciones* como:

$$(3) \quad e_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) = \frac{1}{2} (u_{i,j} + u_{j,i}) \quad ^1$$

Esta definición supone que el incremento en los desplazamientos ($\partial \mathbf{u}$) es menor que las variaciones en la posición ($\partial \mathbf{x}$).

Este tensor es por definición simétrico ($e_{ij} = e_{ji}$) y, por lo tanto, sólo seis de sus nueve componentes son diferentes.

Como podemos observar la deformación no depende del valor absoluto de los desplazamientos y no tiene unidades.

Las deformaciones longitudinales implican variaciones de volumen, por lo que la traza del tensor de deformaciones representa variaciones de volumen por unidad de volumen y se denomina *dilatación cúbica*, θ .

$$(4) \quad \theta = \frac{\Delta V}{V} = e_{ii} = u_{1,1} + u_{2,2} + u_{3,3} = \nabla \cdot \mathbf{u} \quad ^2$$

Por último, para describir las *rotaciones* del medio en su conjunto, utilizamos el *tensor de rotación*, ω_{ij} , que corresponde a la rotación de un sólido rígido sin deformación.

$$(5) \quad \omega_{ij} = \frac{1}{2} (u_{i,j} - u_{j,i})$$

Este tensor es antisimétrico ($\omega_{ij} = -\omega_{ji}$), por lo tanto sus términos diagonales son nulos y sólo tiene tres componentes distintas. Teniendo esto en cuenta podemos definir un vector ω con sus componentes:

$$(6) \quad \omega_i = \varepsilon_{ijk} \omega_{jk} = \varepsilon_{ijk} u_{k,j} = \nabla \times \mathbf{u} \quad ^3$$

Usando los tensores deformación y rotación las derivadas parciales de los desplazamientos están dadas por:

$$(7) \quad u_{i,j} = e_{ij} + \omega_{ij}$$

11.2.2. Esfuerzo

Consideramos ahora fuerzas actuando en un pequeño volumen V rodeado por una superficie S , en el interior de un medio continuo mayor. El material en el interior de V está afectado por fuerzas internas actuando en todo V y fuerzas de superficie o de contacto actuando en S .

¹ Este tensor de deformación se deduce a partir de las definiciones de deformación longitudinal y deformación de cizalla dadas en (1) y (2), (ver Lay y Wallace, 1995).

² Utilizamos la notación de Einstein, donde un índice repetido dos veces implica una suma sobre ese índice ($e_{ii} = e_{11} + e_{22} + e_{33}$; $a_i \cdot b_i = a_1 \cdot b_1 + a_2 \cdot b_2 + a_3 \cdot b_3$).

³ ε_{ijk} : Tensor permutación. (=0 cuando algún índice está repetido, =1 cuando sus índices están ordenados y =-1 cuando sus índices están desordenados).

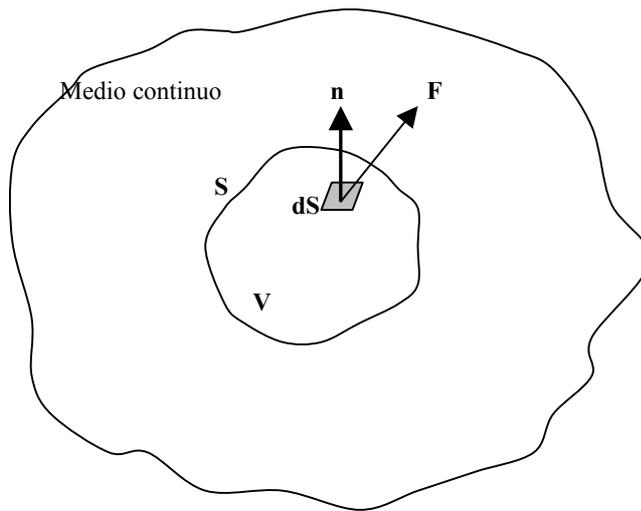


Figura 2. Esfuerzo actuando sobre una superficie dS.

Los esfuerzos representan las fuerzas de contacto que actúan por unidad de área sobre la superficie S (en el interior de un medio continuo) y cuantifican la fuerza (por unidad de área) que las partículas de un lado de la superficie ejercen sobre las del otro lado.

Para un determinado punto en la superficie S , definimos el *vector esfuerzo* \mathbf{T} como el cociente de la fuerza \mathbf{F} que actúa en ese punto por unidad de superficie, cuya orientación es \mathbf{n} , cuando la superficie tiende a cero (en cuyo caso podemos considerarla como un plano) (Figura 2).

$$(8) \quad \mathbf{T}(\mathbf{n}) = \lim_{\Delta S \rightarrow 0} \frac{\mathbf{F}}{\Delta S}$$

Por tanto en la definición están incluidas la dirección de la fuerza y de la normal a la superficie, \mathbf{n} . El vector $\mathbf{T}(\mathbf{n})$ tiene la misma dirección que la fuerza \mathbf{F} y su valor depende de la orientación de la superficie.

Los esfuerzos se deben a la acción de una parte del medio sobre la otra por lo que sólo existen si hay material en los dos lados de la superficie.

Los esfuerzos pueden también ser representados por un tensor de segundo orden τ_{ij} con nueve elementos, de modo que con este tensor podemos definir el estado de esfuerzo en cualquier punto sobre un plano de dirección arbitraria. El primer índice de τ_{ij} corresponde a la dirección de la normal al plano sobre el que actúa la fuerza y el segundo índice indica la dirección de la fuerza. Por lo tanto, los términos de la diagonal principal corresponden a esfuerzos que actúan perpendicularmente a la superficie, *esfuerzos normales*, y los términos de fuera de la diagonal a *esfuerzos tangenciales o de cizalla*. Los esfuerzos normales positivos producen un aumento de volumen, por lo que corresponden a *tensiones* y los negativos producen una reducción del volumen y se denominan *compresiones*. En general, en el interior de la Tierra, los esfuerzos

normales son negativos, debido al peso del material situado por encima, por lo que suele hablarse de esfuerzos compresivos máximos y mínimos.

Para un plano cuya normal viene dada por el vector \mathbf{n} , la relación de Cauchy entre el vector \mathbf{T} y el tensor $\boldsymbol{\tau}$ es:

$$(9) \quad T_i = \tau_{ji} n_j$$

Es posible demostrar que en ausencia de momentos externos $\boldsymbol{\tau}$ es simétrico ($\tau_{ij} = \tau_{ji}$)⁴.

Las dimensiones físicas de $\mathbf{T}(\mathbf{n})$ son fuerza por unidad de área. En geofísica suele utilizarse el bar como unidad para el esfuerzo. Sus equivalencias con los sistemas de unidades cgs y S.I. son respectivamente: $1 \text{ bar} = 10^6 \text{ dyn/cm}^2$ y $1 \text{ bar} = 10^5 \text{ Pa}$. La presión atmosférica al nivel del mar es aproximadamente 1 bar y, en la corteza, a 3-4 km de profundidad, los esfuerzos de compresión son del orden de 1 kbar.

Hemos visto que tanto el tensor de deformaciones como el de esfuerzos son simétricos y por lo tanto diagonalizables. Es decir, que para cualquier punto en el interior de V existen tres planos ortogonales sobre los que sólo actúan esfuerzos normales (las normales a estos planos constituyen los *ejes principales de esfuerzo*) y tres ejes ortogonales sobre los que sólo existen deformaciones longitudinales (*ejes principales de deformación*). Así mismo, las trazas de estos dos tensores son invariantes ante un cambio de sistema de coordenadas.

11.3. Ley de Hooke

El comportamiento mecánico de un material continuo está definido por la relación existente entre esfuerzos y deformaciones (*ecuación constitutiva*). Se dice que un *medio es elástico* si posee un estado natural al cual regresa una vez dejan de aplicársele fuerzas, y bajo la aplicación de cargas cumple la *Ley de Hooke*, es decir, que la deformación es proporcional al esfuerzo.

$$(10) \quad \tau_{ij} = C_{ijkl} e_{kl}$$

Siendo ésta la ecuación básica de la teoría de la *elasticidad lineal*.

A partir de ahora supondremos que los materiales de la Tierra cumplen la ley de Hooke. Esto no es válido para pequeños esfuerzos aplicados durante millones de años o grandes esfuerzos repentinos. Sin embargo, en los pequeños esfuerzos aplicados durante poco tiempo que tienen interés en la propagación de ondas sísmicas, casi todos los materiales terrestres presentan proporcionalidad entre esfuerzo y deformación.

El tensor de cuarto orden de *coeficientes o módulos elásticos* C_{ijkl} describe las propiedades del material y tiene 81 componentes. No obstante, debido a la simetría de los tensores deformación

⁴ Ver Lay y Wallace (1995).

y esfuerzo sólo 36 son diferentes, pudiendo escribirse la ecuación (10) en función de los desplazamientos como:

$$(11) \tau_{ij} = C_{ijkl} u_{k,l}$$

Además, para un medio *perfectamente elástico* (no existe disipación de energía cuando los esfuerzos producen deformaciones), se deduce que $C_{ijkl} = C_{klij}$ por lo que los términos diferentes se reducen a 21.

Por último, para un medio isótropo, en el que sus propiedades son iguales en todas las direcciones, los 21 términos de C_{ijkl} se reducen a 2 coeficientes independientes, λ y μ , denominados *coeficientes de Lamè*.

$$(12) C_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \mu (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}) \quad ^5$$

Sustituyendo la expresión (12) en (10) y (11) obtenemos:

$$(13) \tau_{ij} = C_{ijkl} e_{kl} = \lambda \delta_{ij} e_{kk} + 2\mu e_{ij}$$

$$(14) \tau_{ij} = C_{ijkl} u_{k,l} = \lambda \delta_{ij} u_{k,k} + \mu (u_{i,j} + u_{j,i})$$

El coeficiente μ se denomina *módulo de cizalla o rigidez*, tiene unidades de esfuerzo y representa una medida de la resistencia al esfuerzo de cizalla ($\tau_{12} = 2\mu e_{12}$; $\tau_{13} = 2\mu e_{13}$, etc.). En un fluido $\mu = 0$ y cuanto más elevado sea μ menos se deforma el cuerpo con un esfuerzo dado.

Otros coeficientes obtenidos a partir de λ y μ son:

El *coeficiente volumétrico o incompresibilidad, K*:

$$(15) K = \frac{-dP}{\theta} = \frac{\tau_{kk}}{3e_{kk}} = \lambda + \frac{2}{3}\mu$$

Representa la resistencia del material a variaciones de volumen (razón entre la presión aplicada y la variación relativa de volumen que produce).

El *módulo de Young, E*:

$$(16) E = \frac{\tau_{11}}{e_{11}} = \mu \frac{3\lambda + 2\mu}{\lambda + \mu}$$

Relaciona el esfuerzo longitudinal con la deformación en la misma dirección.

El *coeficiente de Poisson, σ* , es la relación entre la deformación en una dirección perpendicular (e_{22} o e_{33}) y la deformación en la misma dirección (e_{11}) que un esfuerzo normal aplicado ($\tau_{11} \neq 0$, $\tau_{22} = \tau_{33} = 0$).

$$(17) \sigma = -\frac{e_{22}}{e_{11}} = \frac{\lambda}{2(\lambda + \mu)}$$

El coeficiente de Poisson presenta valores entre 0 y 0.5. Cuando $\lambda = \mu$ este coeficiente toma el valor $\sigma = 0.25$. Es lo que se conoce como condición de Poisson. Esta condición se cumple aproximadamente en muchos materiales del interior de la Tierra. En ese caso, el

⁵ δ_{ij} : Tensor delta de Kroneker ($\delta_{ij} = 1$ si $i=j$ y $\delta_{ij} = 0$ si $i \neq j$).

comportamiento elástico del material esta definido por un solo parámetro, simplificándose así muchos de los problemas sismológicos.

Al igual que μ , los coeficientes λ , K y E tienen dimensiones de esfuerzo.

11.4. Ecuaciones del movimiento en un medio elástico, homogéneo e isótropo

El movimiento en cada punto del interior de un volumen está determinado por las fuerzas y esfuerzos que actúan respectivamente en el interior del volumen y en su superficie exterior.

La *Segunda Ley de Newton* aplicada a un cuerpo continuo de volumen V rodeado por una superficie S , puede escribirse como:

$$(18) \int_V F_i dV + \int_S T_i dS = \frac{d}{dt} \int_V \rho v_i dV$$

Donde F_i son las *fuerzas internas* o fuerzas por unidad de volumen que actúan sobre los elementos de volumen dV , T_i son los esfuerzos que actúan en los elementos de superficie dS y v_i es la velocidad en cada punto del volumen.

Poniendo T_i en función del tensor de esfuerzos y aplicando el teorema de Gauss, para transformar la integral de superficie en integral de volumen, obtenemos:

$$(19) \int_V \left(F_i + \frac{\partial \tau_{ij}}{\partial x_j} \right) dV = \frac{d}{dt} \int_V \rho v_i dV$$

Si la densidad es constante en el tiempo y para un medio infinito, podemos escribir:

$$(20) \tau_{ij,j} + F_i = \rho \ddot{u}_i \quad ^6$$

Particularizando ahora para un medio *elástico y homogéneo*, donde se cumple la *Ley de Hooke* y los coeficientes de elasticidad son constantes:

$$(21) C_{ijkl} u_{k,lj} + F_i = \rho \ddot{u}_i$$

Si además el material es *isótropo*, se cumple la ecuación (14), por lo que:

$$(22) \left[\lambda \delta_{ij} u_{k,k} + \mu (u_{i,j} + u_{j,i}) \right]_{,j} + F_i = \rho \ddot{u}_i$$

Y operando llegamos a⁷:

$$(23) (\lambda + \mu) u_{k,ki} + \mu u_{i,jj} + F_i = \rho \ddot{u}_i$$

Que en notación vectorial toma la forma:

$$(24) (\lambda + \mu) \nabla(\nabla \mathbf{u}) + \mu \nabla^2 \mathbf{u} + \mathbf{F} = \rho \ddot{\mathbf{u}}$$

⁶ Aquí hemos supuesto que las deformaciones son infinitesimales y que las velocidades y las variaciones espaciales de las velocidades son pequeñas, por lo que $\ddot{u}_i = \frac{\partial^2 u_i}{\partial t^2}$.

⁷ $(\delta_{ij} u_{k,k})_{,j} = \delta_{ij} u_{k,kj} = u_{k,ki}$; $(u_{j,i})_{,j} = u_{j,ij} = u_{j,ji} = u_{k,ki}$.

Y representa la ecuación del movimiento en función del desplazamiento para un *medio continuo, elástico, infinito, homogéneo e isótropo*.

Esta ecuación tiene gran importancia en sismología, pues podemos suponer que un terremoto esta generado por procesos que pueden representarse por un sistema de fuerzas internas que actúan en una cierta región focal y que el campo de desplazamientos fuera de la región focal debido a ese sistema de fuerzas es la solución de la ecuación (24).

La ecuación del movimiento (24) también puede expresarse en términos de la dilatación cúbica θ y del vector rotación $\boldsymbol{\omega}$ dados por las ecuaciones (4) y (6).

$$(25) (\lambda + 2\mu)\nabla(\theta) - \mu\nabla \times \boldsymbol{\omega} + \mathbf{F} = \rho\mathbf{u} \quad ^9$$

Si definimos las constantes:

$$(26) \alpha^2 = \frac{\lambda + 2\mu}{\rho}$$

$$(27) \beta^2 = \frac{\mu}{\rho}$$

Expresamos la ecuación (25) como:

$$(28) \alpha^2\nabla(\theta) - \beta^2\nabla \times \boldsymbol{\omega} + \frac{\mathbf{F}}{\rho} = \mathbf{u}$$

Si las fuerzas internas son nulas ($\mathbf{F}=0$), obtenemos la ecuación *homogénea* del movimiento, también conocida como *ecuación de Navier*:

$$(29) (\lambda + \mu)\nabla(\nabla\mathbf{u}) + \mu\nabla^2\mathbf{u} = \rho\mathbf{u}$$

$$(30) \alpha^2\nabla(\theta) - \beta^2\nabla \times \boldsymbol{\omega} = \mathbf{u}$$

Cuyas soluciones son los desplazamientos elásticos en un medio donde no hay fuerzas actuando.

11.5. Funciones potencial de los desplazamientos y fuerzas

Los desplazamientos $\mathbf{u}(\mathbf{x},t)$ en un medio elástico constituyen un campo vectorial. Por lo tanto podemos aplicar el teorema de Helmholtz para expresar el campo \mathbf{u} en función de un potencial escalar, ϕ , y un potencial vectorial, $\boldsymbol{\psi}$, tal que $\nabla\boldsymbol{\psi} = 0$:

$$(31) \mathbf{u} = \nabla\phi + \nabla \times \boldsymbol{\psi}$$

Aplicando la divergencia y el rotacional a ambos miembros de la ecuación (31) y usando las expresiones (4) y (6) podemos deducir que:

$$(32) \theta = \nabla^2\phi$$

⁹ Para deducir esta ecuación se ha tenido en cuenta la relación: $\nabla^2\mathbf{a} = \nabla(\nabla\mathbf{a}) - \nabla \times \nabla \times \mathbf{a}$.

$$(33) \boldsymbol{\omega} = -\nabla^2 \boldsymbol{\Psi}$$

De un modo análogo podemos representar las fuerzas internas, \mathbf{F} , por el potencial escalar Φ y el potencial vectorial $\boldsymbol{\Psi}$:

$$(34) \mathbf{F} = \nabla \Phi + \nabla \times \boldsymbol{\Psi}$$

De esa forma se puede expresar la ecuación del movimiento en función de estos potenciales al sustituir \mathbf{u} y \mathbf{F} de la ecuación (28) por las relaciones dadas en (31) y (34), quedando la ecuación (28) dividida en una parte escalar y una parte vectorial:

$$(35) \alpha^2 \nabla^2 \phi + \frac{\Phi}{\rho} = \ddot{\phi}$$

$$(36) \beta^2 \nabla^2 \boldsymbol{\Psi} + \frac{\boldsymbol{\Psi}}{\rho} = \ddot{\boldsymbol{\Psi}}$$

Hay que tener en cuenta que los dos potenciales vectoriales deben satisfacer la condición de que sus divergencias son nulas. Esta forma de expresar la ecuación del movimiento tiene gran importancia puesto que simplifica enormemente su solución.

11.6. Función de Green

Como hemos mencionado anteriormente las fuerzas internas de la ecuación del movimiento pueden ser utilizadas para representar los procesos que generan los terremotos. En general estas fuerzas serán diferentes para cada terremoto y están definidas sólo dentro de un cierto volumen (la región focal).

Un tipo de fuerzas de gran interés por su sencillez y utilidad en sismología son las *fuerzas impulsivas unidireccionales* en el espacio y en el tiempo, que podemos expresar matemáticamente por medio de la Delta de Dirac como:

$$(37) F_i(\mathbf{x}, t) = \delta(\mathbf{x} - \boldsymbol{\xi}) \delta(t - \tau) \delta_{in}$$

Esta fuerza está aplicada en un punto de coordenadas $\boldsymbol{\xi}_i$ y en un instante τ , su módulo en $(\boldsymbol{\xi}, \tau)$ es la unidad -y cero para todo $(\mathbf{x}, t) \neq (\boldsymbol{\xi}, \tau)$ - y tiene la dirección \mathbf{n} .

Para un medio continuo y elástico, el campo de desplazamientos producido por esta fuerza es un tensor de segundo orden que se denomina *Función de Green* de la elastodinámica, $G_{ni}(\mathbf{x}, t; \boldsymbol{\xi}, \tau)$, el cual representa la respuesta mecánica del medio a una excitación impulsiva. Estos desplazamientos dependen del tiempo y de la posición del punto donde los midamos (\mathbf{x}, t) ; de las coordenadas y el instante de aplicación de la fuerza, $(\boldsymbol{\xi}, \tau)$; y de la dirección de la fuerza, \mathbf{n} .

Si utilizamos una fuerza impulsiva en la ecuación del movimiento para un medio continuo y elástico tenemos que:

$$(38) \int_V \delta(\mathbf{x}-\xi)\delta(t-\tau)\delta_{,in}dV + \int_S C_{ijkl}G_{nk,l}n_jdS = \int_V \rho\ddot{G}_{ni}dV \quad 10$$

La Función de Green caracteriza el medio, puesto que depende de sus coeficientes elásticos, su densidad, la forma de V y las condiciones de contorno de su superficie exterior S .

11.6.1. Teoremas de reciprocidad y representación

La utilidad de la Función de Green se pone de manifiesto por medio del teorema de representación:

Consideremos de nuevo un medio continuo y elástico de volumen V rodeado por una superficie S . Suponemos que existe en ese medio un sistema de fuerzas internas \mathbf{f} que actúan en cada elemento de volumen dV y de esfuerzos \mathbf{T}^u que actúan en cada elemento de superficie dS produciendo un campo de desplazamientos \mathbf{u} . De igual modo en el mismo medio actúa un sistema de fuerzas \mathbf{g} y de esfuerzos \mathbf{T}^w produciendo unos desplazamientos \mathbf{w} .

En cada caso se cumplirá la ecuación (18):

$$(39) \int_V (f_i - \rho u_{i,t})dV + \int_S T_i^u dS = 0$$

$$(40) \int_V (g_i - \rho w_{i,t})dV + \int_S T_i^w dS = 0$$

Multiplicando escalarmente la primera ecuación por \mathbf{w} y la segunda por \mathbf{u} y restándolas obtenemos:

$$(41) \int_V (f_i - \rho u_{i,t})w_i dV + \int_S T_i^u w_i dS = \int_V (g_i - \rho w_{i,t})u_i dV + \int_S T_i^w u_i dS$$

Expresión conocida como *teorema de reciprocidad de Betti*.

Integrando para todo el tiempo obtenemos:

$$(42) \int_{-\infty}^{\infty} dt \int_V \rho(u_i \dot{w}_i - \dot{u}_i w_i) dV = \int_{-\infty}^{\infty} dt \int_V (u_i g_i - w_i f_i) dV + \int_{-\infty}^{\infty} dt \int_S (T_i^w u_i - T_i^u w_i) dS$$

Generalización del teorema de reciprocidad, conocida como la fórmula de Green-Volterra.

El teorema de reciprocidad sigue siendo válido si evaluamos las funciones \mathbf{f} , \mathbf{T}^u , \mathbf{u} y $\dot{\mathbf{u}}$ en un instante t_1 y \mathbf{g} , \mathbf{T}^w , \mathbf{w} y $\dot{\mathbf{w}}$ en un instante diferente t_2 . Si elegimos $t_1 = t$ y $t_2 = \tau - t$, podemos escribir la integral del primer miembro de (42), cambiando el orden de integración como:

$$(43) \int_V dV \int_{-\infty}^{\infty} \rho [u_i(t) \dot{w}_i(\tau - t) - \dot{u}_i(t) w_i(\tau - t)] dt$$

Si imponemos que los desplazamientos y velocidades son nulos antes de un tiempo dado y elegimos ese instante en $t=0$, es decir, que $u_i = \dot{u}_i = w_i = \dot{w}_i = 0$ para $t \leq 0$ (condiciones de causalidad), es posible reducir los límites de integración en el tiempo de 0 a τ , y escribir la integral como:

¹⁰ A partir de la ecuación del movimiento (18), donde F_i viene dada por (37), T_i por (9) y para un medio elástico, donde se cumple (11).

$$(44) \int_V dV \int_0^\tau \rho \frac{d}{dt} [-\dot{u}_i(t)w_i(\tau-t) - u_i(t)\dot{w}_i(\tau-t)] dt$$

E integrando (44) obtenemos que la integral es igual a cero. Por lo que la fórmula de Green-Volterra queda como:

$$(45) \int_{-\infty}^{\infty} dt \int_V (u_i g_i - w_i f_i) dV = \int_{-\infty}^{\infty} dt \int_S (T_i^u w_i - T_i^w u_i) dS$$

Donde se pone de manifiesto que es posible representar los desplazamientos debidos a un sistema de fuerzas por medio de los producidos por otro sistema diferente, siempre que se cumplan las condiciones de causalidad. Por lo tanto podemos representar los desplazamientos de un complicado sistema de fuerzas por medio de los producidos por otro más simple. En concreto usaremos como sistema de fuerzas complementario, \mathbf{g} , una fuerza impulsiva del tipo de la mostrada en (37) por lo que su desplazamiento, \mathbf{w} , asociado es la función de Green:

$$(46) \int_{-\infty}^{\infty} dt \int_V (u_i \delta(\mathbf{x} - \boldsymbol{\xi}) \delta(t - \tau) \delta_{in} - G_{ni} f_i) dV = \int_{-\infty}^{\infty} dt \int_S (T_i^u G_{ni} - u_i C_{ijkl} G_{nk,l} n_j) dS$$

Y utilizando la definición de la función delta e intercambiando las variables \mathbf{x} , t , $\boldsymbol{\xi}$ y τ despejamos el campo de desplazamientos u_i :

$$(47) u_i(\mathbf{x}, t) = \int_{-\infty}^{\infty} d\tau \int_V f_i G_{ni} dV + \int_{-\infty}^{\infty} d\tau \int_S T_i G_{ni} dS - \int_{-\infty}^{\infty} d\tau \int_S u_i C_{ijkl} G_{nk,l} n_j dS$$

Obteniendo la expresión conocida como *teorema de representación*.

Esta fórmula nos proporciona el valor de los desplazamientos elásticos en cualquier punto del volumen V debidos a un sistema de fuerzas definidas en la región focal y un sistema de esfuerzos y desplazamientos en su superficie por medio de la función de Green.

Para poder utilizar esta ecuación debemos determinar con anterioridad la función de Green para el medio en el que queramos conocer los desplazamientos. En general, este proceso no es sencillo y la dificultad aumenta con la complejidad del medio. El caso más sencillo corresponde a un medio elástico, infinito, homogéneo e isótropo.

La ventaja es que una vez hemos calculado la función de Green para el medio en cuestión, podemos determinar los desplazamientos elásticos para cualquier sistema de fuerzas definidas en un volumen y de esfuerzos y desplazamientos definidos en la superficie que rodea ese volumen, sin más que resolver la ecuación (47).

Bibliografía

- [1] Lay T. y Wallace, T. C. Modern Global Seismology. Academic Press, 1995. pág. 35-54.
- [2] Udías, A. Principles of Seismology. Cambridge University Press, 1999. pág. 10-28.

Tema 12. Ondas: La ecuación de ondas. Ondas sísmicas internas. Propagación en un medio elástico, homogéneo e isótropo. Reflexión y refracción en medios estratificados, planos y esféricos.

12.1. La ecuación de ondas

En un medio elástico, isótropo, homogéneo e infinito, en el que las fuerzas internas son nulas, habíamos visto que la ecuación de movimiento viene dada por la ecuación de Navier en función de la dilatación cúbica θ y el vector rotación ω

$$\alpha^2 \nabla \theta - \beta^2 \nabla \times \omega = \ddot{u}$$

Si aplicamos en esta ecuación el operador divergencia y considerando que la divergencia del gradiente de θ es su Laplaciana, que la divergencia del rotacional de ω es nulo y que la divergencia del desplazamiento u es la dilatación cúbica θ , obtenemos

$$(1) \quad \nabla^2 \theta = \frac{1}{\alpha^2} \frac{\partial^2 \theta}{\partial t^2}$$

Tomando de nuevo la ecuación de Navier, le aplicamos ahora el operador rotacional. Sabemos que el rotacional del gradiente de una función escalar es cero y que el rotacional del desplazamiento u es el vector rotación ω . Además, el rotacional del rotacional de ω es igual al gradiente de la divergencia, que es nulo, menos la Laplaciana. Por tanto quedará:

$$(2) \quad \nabla^2 \omega = \frac{1}{\beta^2} \frac{\partial^2 \omega}{\partial t^2}$$

Las ecuaciones (1) y (2) tienen la forma de ecuaciones de onda para la función escalar θ y la función vectorial ω , respectivamente. Las soluciones de ambas ecuaciones representan ondas que se propagan en un medio elástico, siendo α y β sus velocidades, que son función de los coeficientes elásticos λ y μ y de la densidad ρ

$$\alpha = \sqrt{\frac{\lambda + 2\mu}{\rho}}$$

$$\beta = \sqrt{\frac{\mu}{\rho}}$$

12.2. Ondas sísmicas internas

Dado que la dilatación cubica θ representa cambios en el volumen sin cambios en la forma, las soluciones a la ecuación (1) corresponden a movimientos de compresión y dilatación, o también ondas longitudinales. Cuando estas ondas se propagan, el material elástico se expande y se contrae manteniendo la misma forma. Este tipo de ondas son las llamadas ondas P, ya que son las primeras que se observan en un sismograma ($\alpha > \beta$). Las soluciones a la ecuación (2) representan ondas de cizalla que se propagan con velocidad β . En este caso el medio cambia de forma pero no de volumen y las ondas se propagan con movimiento perpendicular al del frente de ondas. Estas ondas se denominan ondas S debido a que son el segundo tipo de ondas importantes que aparecen en el sismograma. Podemos concluir que un medio elástico infinito, homogéneo e isótropo solo existen estos dos tipos de ondas, que se denominan conjuntamente ondas de volumen u ondas internas.

La velocidad de las ondas P aumenta con la profundidad a pesar de aumentar también ρ con la profundidad, siendo esto debido a que λ y μ aumentan más rápidamente que ρ con la profundidad. Asimismo, las ondas S aumentan también de velocidad con la profundidad.

12.2.1. Ecuación de Helmholtz

Habíamos visto en el tema anterior que los desplazamientos \mathbf{u} en un medio elástico forman un campo vectorial y que se podían representar en términos de dos funciones potenciales, un potencial escalar ϕ y un potencial vector ψ

$$(3) \quad \mathbf{u} = \nabla \phi + \nabla \times \psi$$

Si hacemos nulos los potenciales Φ y Ψ de las fuerzas internas, las ecuaciones de los potenciales del desplazamiento quedarían:

$$(4) \quad \nabla^2 \phi = \frac{1}{\alpha^2} \frac{\partial^2 \phi}{\partial t^2}$$

$$(5) \quad \nabla^2 \psi = \frac{1}{\beta^2} \frac{\partial^2 \psi}{\partial t^2}$$

Por tanto, en las condiciones de ausencia de fuerzas internas, los potenciales ϕ y ψ son también soluciones de la ecuación de ondas. Debido a que α y β son las velocidades de las ondas P y S, entonces ϕ es el potencial de las ondas P y ψ el de las ondas S. El desplazamiento elástico total \mathbf{u} es la suma de los desplazamientos de las ondas P y S y puede ser escrito como

$$u = u^P + u^S$$

Y de acuerdo con la ecuación (3)

$$u^P = \nabla \phi$$

$$u^S = \nabla \times \psi$$

Por lo que los desplazamientos de las ondas P y S pueden ser deducidos de los potenciales ϕ y ψ respectivamente.

Si suponemos que tanto ϕ como ψ tienen dependencia armónica del tiempo

$$\phi(x, t) = \phi(x) e^{-i\omega t}$$

Y sustituyendo en las ecuaciones (4) y (5) se obtiene:

$$(\nabla^2 + k_\alpha^2) \phi = 0$$

$$(\nabla^2 + k_\beta^2) \psi_i = 0$$

donde $k_\alpha = \omega/\alpha$ y $k_\beta = \omega/\beta$ son el número de onda de las ondas P y S respectivamente. Estas expresiones se conocen como las ecuaciones de Helmholtz en las cuales ha sido eliminada la dependencia del tiempo.

12.2.2. Ondas planas

En el caso más simple de una sola dimensión, podemos escribir la ecuación de ondas de la forma:

$$\frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2}$$

donde c es la velocidad de propagación de la onda. La solución más general puede ser escrita como:

$$y = f(x - ct) + f(x + ct)$$

siendo $c = \omega/k$, con ω la frecuencia angular y k el número de onda $k = \frac{2\pi}{\lambda}$.

Dependiendo del signo, la onda se propagará en la dirección de x positiva o negativa. Una solución particular puede ser:

$$y = f(x, t) = A e^{i(kx - \omega t)} + B e^{i(kx + \omega t)}$$

Para el caso de que el foco emisor este muy alejado, sucede que el frente de onda es plano y las ondas se propagan en la dirección positiva de x . Una solución será, considerando solo la parte real:

$$y = f(x, t) = A \cos [k(x - ct) + \varepsilon] = A \cos [kx - \omega t + \varepsilon]$$

ésta ecuación representa una onda cosenoidal que se propaga con una velocidad c , y los parámetros que aparecen son: A la amplitud de la onda, el número de onda k ya definido, ω la frecuencia angular $\omega = 2\pi/T$, donde T es el periodo y ε la fase en el origen. Este tipo de soluciones es el más sencillo y corresponde a ondas monocromáticas ($\omega = \text{constante}$) planas.

12.3. Propagación en un medio elástico, homogéneo e isótropo

Consideremos un medio elástico, infinito e isótropo en el que se propagan los dos tipos de ondas P y S. Las ondas P viajan con velocidad α y las componentes de sus desplazamientos cumplen:

$$u_i^\alpha \wedge v_i = 0$$

v_i dirección propagación de la onda

u_i^α componentes de la onda P.

La onda P tiene un movimiento en el sentido de la propagación del frente de ondas. El segundo tipo, las ondas S, se propagan con velocidad β y su movimiento está en el plano normal a la dirección de propagación del frente de ondas, por lo cual sus desplazamientos cumplen la condición:

$$u_i^\alpha \cdot v_i = 0$$

En general, las componentes de los desplazamientos de las ondas P y S se acostumbra a referir a un sistema de ejes geográficos (x_1, x_2, x_3) coincidentes con las direcciones norte, oeste y cenit. Los vectores que representan los desplazamientos de las ondas P y S tienen cada uno tres componentes:

$$u_i^\alpha = (u_1^\alpha, u_2^\alpha, u_3^\alpha) \quad \text{onda P}$$

$$u_i^\beta = (u_1^\beta, u_2^\beta, u_3^\beta) \quad \text{onda S}$$

Definimos un nuevo sistema de ejes (r, SH, SV) (figura 1) donde r es la dirección de propagación del frente de ondas, o bien v_j , y SH y SV están en el plano normal a r . SV es la intersección del plano normal con el plano " $x_3 r$ " y SH es la intersección del plano normal con el plano " $x_1 x_2$ ". SV representa la componente vertical de la onda S y SH representa la componente horizontal de la onda S. En este nuevo sistema de coordenadas:

$$u_i^\alpha = (u_r^\alpha, 0, 0)$$

$$u_i^\beta = (0, u_{SH}^\beta, u_{SV}^\beta)$$

El ángulo que forma r con el eje vertical x_3 es el ángulo de incidencia i_0 . El ángulo que forma según las agujas del reloj el norte con la proyección de r sobre el plano horizontal $x_1 x_2$ es el acimut a_z . El ángulo que forman S con SV es el ángulo de polarización ε .

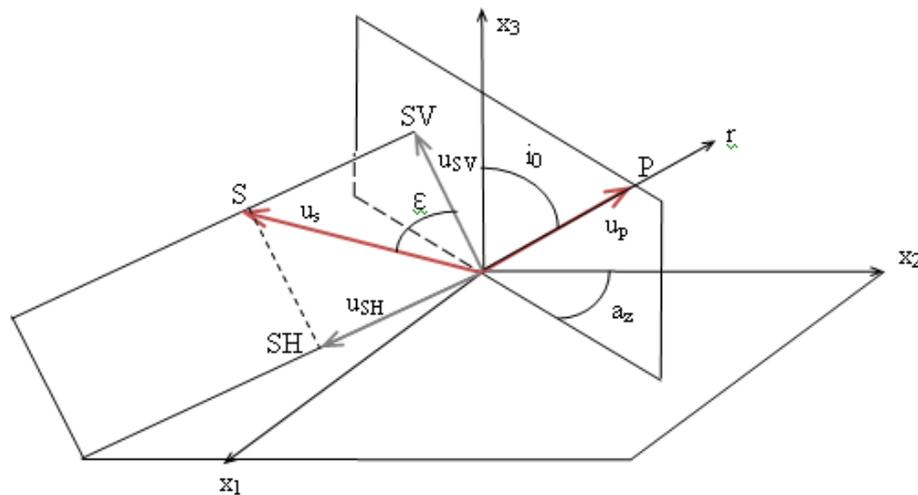


Figura 1. Geometría de los desplazamientos de las ondas P y S referidos a un sistema de coordenadas geográficas.

$$\operatorname{tg} \varepsilon = \frac{u_{SH}}{u_{SV}}$$

Como la dirección de propagación del rayo viene dada por el vector

$$\vec{e}_r = (v_1, v_2, v_3)$$

sus cosenos directores estarán relacionados con el ángulo de incidencia i_0 y el acimut a_z mediante las relaciones:

$$v_1 = \operatorname{sen} i_0 \operatorname{sen} a_z$$

$$v_2 = \operatorname{sen} i_0 \operatorname{cos} a_z$$

$$v_3 = \operatorname{cos} i_0$$

12.4. Reflexión y refracción en medios estratificados, planos y esféricos

Hasta ahora hemos estado considerando un medio de propagación homogéneo, sin embargo la Tierra no lo es, existiendo discontinuidades que separan medios de diferentes propiedades. Cuando una onda incide desde un medio sobre una superficie de separación con otro medio, parte de la energía es reflejada y parte es transmitida o refractada hacia el segundo medio.

En el caso de la propagación de las ondas internas P y S, sus trayectorias y tiempos de llegada desde el foco al punto de observación se pueden estudiar a partir de la dirección del rayo o normal

al frente de onda. La teoría de rayos está basada en el principio de Fermat de tiempo mínimo en el camino recorrido, que tiene dos importantes consecuencias en la refracción y reflexión de ondas sobre una superficie plana que separa dos medios. La primera es que los rayos incidentes, reflejados y refractados están situados en el mismo plano, que a su vez es normal al plano de separación de ambos medios. La segunda consecuencia es que las trayectorias de los rayos incidentes, reflejados y refractados siguen la ley de Snell.

12.5. Propagación en una Tierra plana

Consideremos una tierra plana constituida por capas homogéneas de espesor cualquiera y con velocidad constante en cada capa:

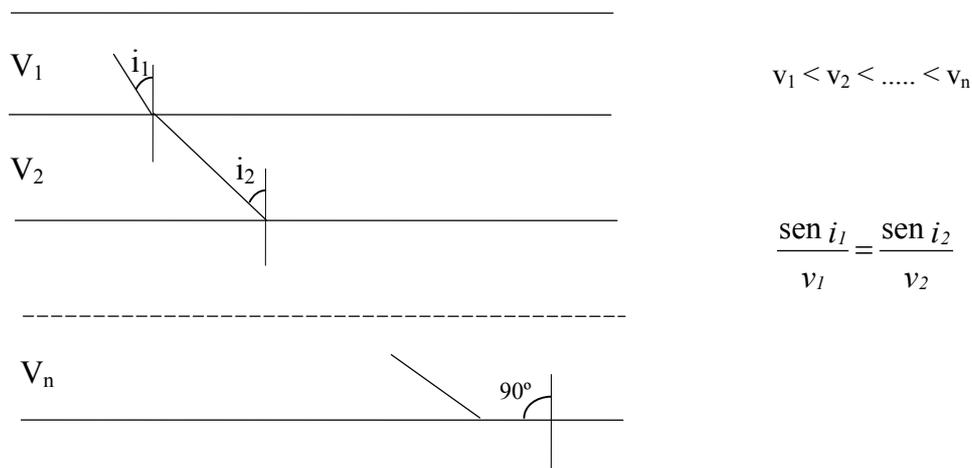


Figura 2. Trayectorias de los rayos para medios estratificados

Como la velocidad va aumentando con la profundidad, la dirección del rayo se va alejando de la normal y llegará a una capa en que la incidencia será de 90°

$$\frac{\text{sen } i_1}{v_1} = \frac{\text{sen } i_2}{v_2} = \dots = \frac{\text{sen } i_{n-1}}{v_{n-1}} = \frac{\text{sen } 90^\circ}{v_n} = \frac{l}{v_n} = p$$

donde p es una constante a lo largo de cada rayo y se le denomina parámetro del rayo. En general, la ley de Snell se puede expresar:

$$\frac{\text{sen } i}{v} = p$$

Si v es constante, i será constante, y si v cambia, también deberá cambiar i , por lo que el rayo se curvará, bien hacia arriba o hacia abajo. Si comprimimos las capas, con un aumento de la velocidad con la profundidad, podemos aproximar a una curva la dirección del rayo, tal y como se aprecia en la figura 3.

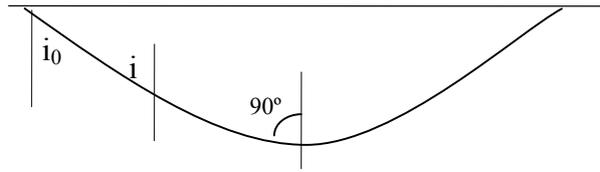


Figura 3. Trayectoria de rayos en un medio con velocidad aumentando con la profundidad.

Consideremos un frente de onda en dos posiciones t y $t + \Delta t$ (figura 4)

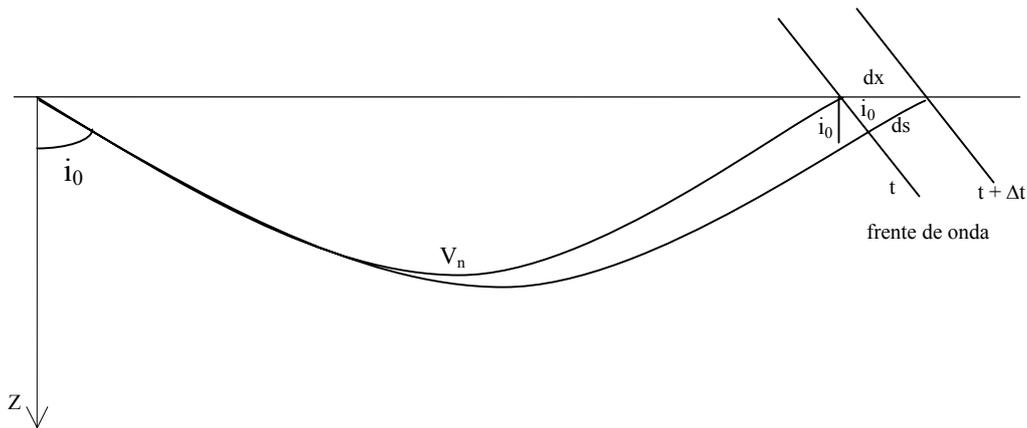


Figura 4. Trayectoria de rayos que determinan la velocidad aparente.

De la figura se deduce que

$$\text{sen } i_0 = \frac{ds}{dx} = \frac{ds/dt}{dx/dt} = \frac{v_0}{v_a}$$

donde v_0 es la velocidad del rayo y v_a es la velocidad aparente que tiene el rayo al medirlo sobre la superficie, luego:

$$\frac{\text{sen } i_0}{v_0} = \frac{1}{v_a}$$

$$\frac{\text{sen } i_0}{v_0} = \dots = \frac{\text{sen } 90^\circ}{v_n} = \frac{1}{v_n} = p$$

Por tanto $v_a = v_n$ y el parámetro del rayo es también la inversa de la velocidad aparente.

12.5.1. Trayectorias y tiempos de recorrido

Consideremos un rayo general de parámetro p , tal y como se expresa en la figura 5

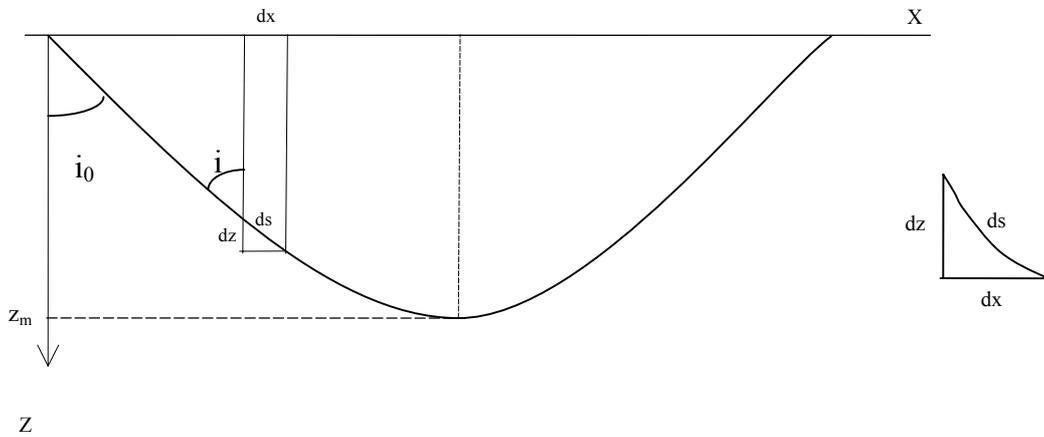


Figura 5. Trayectoria de un rayo con velocidad aumentando con la profundidad

Consideremos un elemento diferencial de trayectoria ds y el triángulo rectángulo que conforma. La relación con el ángulo de incidencia sería:

$$\operatorname{tg} i = \frac{dx}{dz} \Rightarrow dx = \operatorname{tg} i \, dz$$

Integrando para todo el rayo, al que suponemos simétrico:

$$X = 2 \int_0^{z_m} \operatorname{tg} i \, dz = 2 \int_0^{z_m} \frac{\operatorname{sen} i}{\operatorname{cos} i} \, dz$$

Como:

$$\operatorname{sen} i = p v$$

$$\operatorname{cos} i = \sqrt{1 - p^2 v^2}$$

Sustituyendo en la integral anterior

$$(6) \quad X = 2 \int_0^{z_m} \frac{p v}{\sqrt{1 - p^2 v^2}} \, dz$$

Introduciendo un parámetro inverso de la velocidad llamado *lentitud* $\eta = \frac{1}{v}$, obtenemos la distancia total

$$(7) \quad X = 2 \int_0^{z_m} \frac{p}{\sqrt{\eta^2 - p^2}} dz$$

Veamos el tiempo de recorrido T a lo largo de la trayectoria de todo el rayo:

$$T = 2 \int_0^{z_m} \frac{ds}{v}$$

y sabiendo que: $\cos i = \frac{dz}{ds} \Rightarrow ds = \frac{dz}{\cos i}$

el tiempo de recorrido será:

$$(8) \quad T = 2 \int_0^{z_m} \frac{dz}{v \cos i} = 2 \int_0^{z_m} \frac{dz}{v \sqrt{1 - p^2 v^2}} = 2 \int_0^{z_m} \frac{dz}{v^2 \sqrt{\frac{1}{v^2} - p^2}} = 2 \int_0^{z_m} \frac{\eta^2}{\sqrt{\eta^2 - p^2}} dz$$

las dos ecuaciones (7) y (8), obtenidas de una forma general son también las expresiones de las distancias y tiempos de recorrido para el caso de una variación continua de la velocidad con la profundidad desde 0 hasta Z_m .

Como caso particular, podemos determinar la distancia y el tiempo de recorrido para un medio estratificado de n capas con velocidad constante en cada capa, $v_j = \text{cte}$. (figura 6). Por debajo de la corteza (30 km) esta situación ya no se podría considerar.

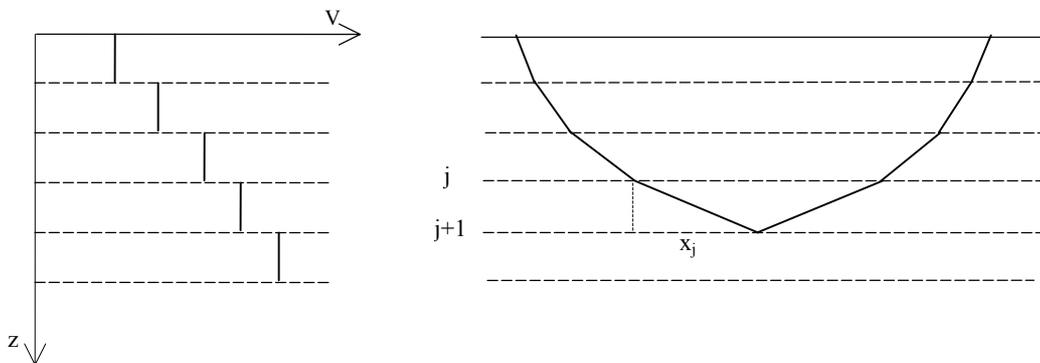


Figura 6. Modelo de velocidades en capas estratificadas de velocidad constante

Aplicamos la expresión (6) de la distancia x para la capa $j \leftrightarrow j + 1$

$$x_j = 2 \int_{z_j}^{z_{j+1}} \frac{p v}{\sqrt{1 - p^2 v^2}} dz$$

La distancia total X será:

$$X = 2 \sum_{j=0}^n \int_{z_j}^{z_{j+1}} \frac{p v}{\sqrt{1 - p^2 v^2}} dz$$

Dentro de cada capa, el parámetro p y la velocidad v son constantes, quedando

$$X = 2 \sum_{j=0}^n \frac{p v}{\sqrt{1 - p^2 v^2}} (z_{j+1} - z_j)$$

Análogamente, el tiempo de recorrido se puede obtener de la ecuación (8)

$$T = 2 \sum_{j=0}^n \frac{\eta^2}{\sqrt{\eta^2 - p^2}} (z_{j+1} - z_j)$$

12.5.2. Variación de la distancia respecto al parámetro del rayo

Para entender el comportamiento de las trayectorias del rayo en medios donde existen diferentes distribuciones de velocidad, es importante considerar los cambios de distancia respecto de los cambios del parámetro del rayo p , es decir la relación dx/dp . Recordemos la ecuación (7) y consideremos h como la máxima profundidad:

$$x = 2 \int_0^h \frac{p dz}{\sqrt{\eta^2 - p^2}}$$

Derivando con respecto a p

$$\frac{dx}{dp} = 2 \int_0^h \frac{dz}{\sqrt{\eta^2 - p^2}} + 2 p \frac{d}{dp} \int_0^h \frac{dz}{\sqrt{\eta^2 - p^2}}$$

Realizando operaciones e introduciendo una nueva variable ξ , que definimos como el gradiente de la velocidad por unidad de velocidad:

$$\xi = \frac{1}{v} \frac{dv}{dz} = -v \frac{d\eta}{dz}$$

llegamos a:

$$(9) \quad \frac{dx}{dp} = - \frac{2}{\xi_0 \sqrt{\eta_0^2 - p^2}} + 2 \int_0^h \frac{1}{\xi^2 \sqrt{\eta^2 - p^2}} \frac{d\xi}{dz} dz$$

Esta ecuación representa el cambio en la distancia horizontal cuando se pasa de un rayo con parámetro p a otro de parámetro $p + dp$ y resulta de interés para estudiar el comportamiento de los rayos en medios que tienen una cierta distribución de la velocidad con la profundidad, $v(z)$.

12.5.3. Distribuciones de la velocidad con la profundidad

Vamos a realizar un análisis del comportamiento de los rayos para las diferentes distribuciones de velocidad con la profundidad, mediante el estudio de las ecuaciones que definen las trayectorias y los tiempos de llegada.

Velocidad aumentando linealmente con la profundidad

La distribución de la velocidad aumentando linealmente con la profundidad es aplicable a muchos casos dentro de la corteza terrestre y el manto superior:

$$v(z) = v_0 + b z$$

siendo $b = dv/dz$ el gradiente de la velocidad, que en la mayor parte de los casos tiene un valor de 0,2 (figura 7).

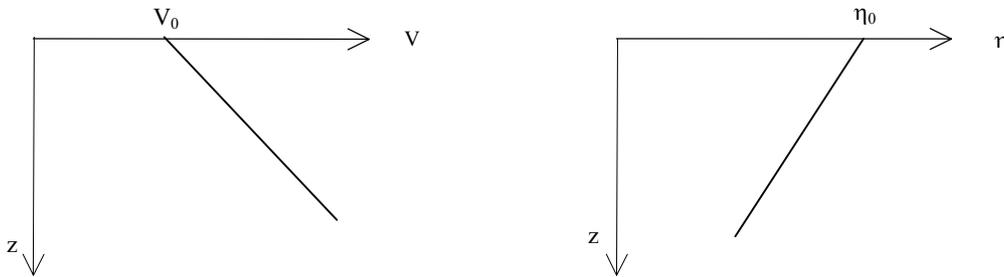


Figura 7. Aumento lineal de la velocidad con la profundidad

El valor de la variable $\xi = \frac{1}{v} \frac{dv}{dz}$ es pequeño y positivo, y su derivada respecto a z

$$\frac{d\xi}{dz} = -\frac{1}{v^2} \left(\frac{dv}{dz} \right)^2 + \frac{1}{v} \frac{d^2v}{dz^2}$$

donde el segundo sumando vale cero y el primero es muy pequeño, por lo que podemos despreciar el termino integral en la ecuación (10) que define la relación dx/dp y quedaría:

$$\frac{dx}{dp} = -2 \frac{1}{\xi_0 \sqrt{\eta_0^2 - p^2}}$$

cuyo valor es negativo.

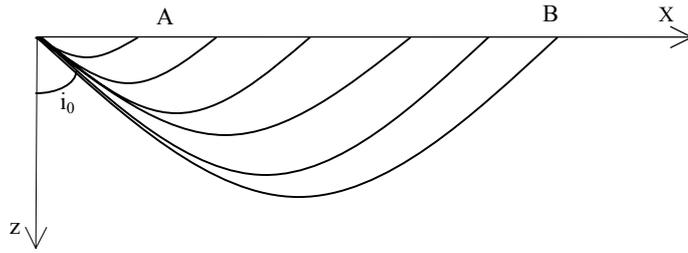


Figura 8. Trayectoria de los rayos en un medio con velocidad aumentando con la profundidad.

Cuando los rayos dejan el foco con un gran i_0 (figura 8)

$$p = \frac{\text{sen } i_0}{v_0} \quad ; \quad i_0 \cong 90^\circ \quad ; \quad \frac{\text{sen } i_0}{v_0} = \frac{1}{v_0} = p_{\text{max}}$$

p va decreciendo, ya que los rayos van dejando el foco más rápidamente o también i_0 va decreciendo, entonces $\eta_0^2 - p^2$ va aumentando desde cero y $\frac{dx}{dp}$ va decreciendo (figura 9).

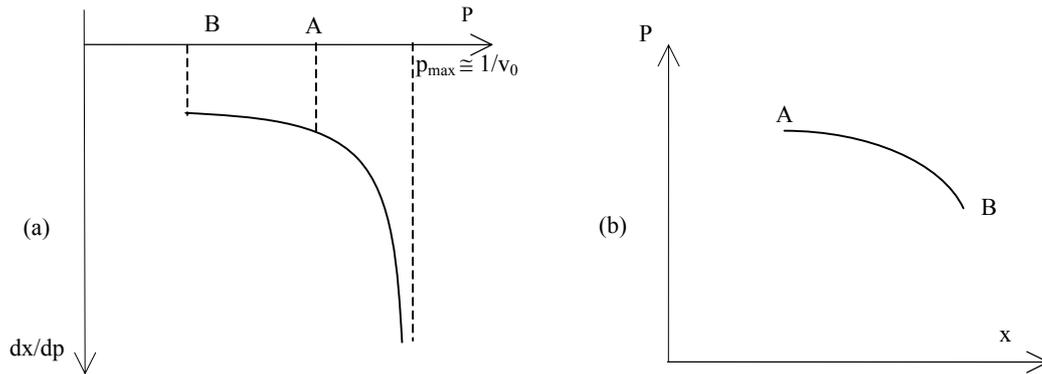


Figura 9. (a) Variación de la distancia con el parámetro del rayo. (b) Variación del parámetro del rayo con la distancia.

Como $\frac{dT}{dx} = p$; $\frac{dT}{dx}$ versus x será una función decreciente y la gráfica (T,x) llamada dromocrónica, tendrá la forma expresada en la figura 10.

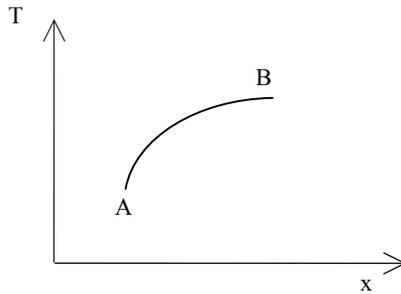


Figura 10. Dromocronica.

Velocidad aumentando de forma continúa y no regular con la profundidad

Consideremos un incremento de la velocidad con la profundidad hasta una cierta distancia. Después un incremento más rápido sobre una cierta capa intermedia y después otra vez un incremento normal de la velocidad (figura 11).

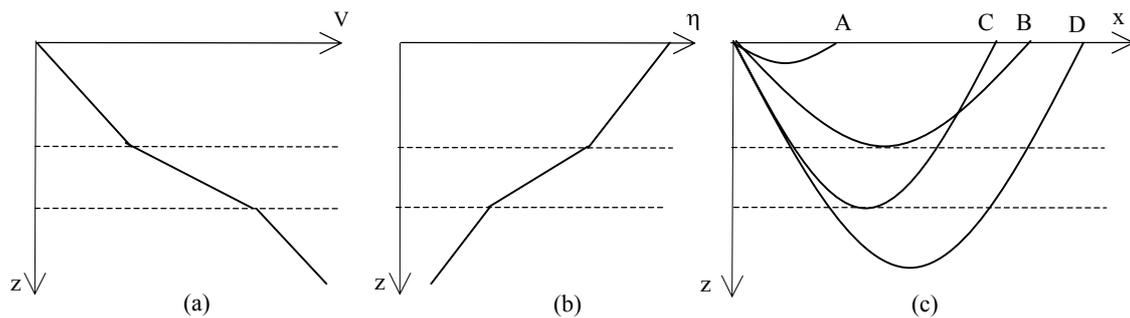


Figura 11. (a) Aumento no regular de la velocidad con la profundidad. (b) Variación de la lentitud con la profundidad. (c) Trayectorias de los rayos.

La relación $\frac{dx}{dp}$ se comportará como en el caso anterior solo hasta la 1ª discontinuidad. Para la capa de fuerte incremento de velocidad, la variable $\xi = \frac{1}{v} \frac{dv}{dz}$ llega a tener un cierto valor y $\frac{d\xi}{dz}$ ya no va a ser despreciable, por lo que los dos términos de $\frac{dx}{dp}$ van a ser significativos, de hecho el término integral en la ecuación (10) va a ser mayor que el otro para ciertos valores de p y por lo tanto $\frac{dx}{dp} > 0$ (figura 12).

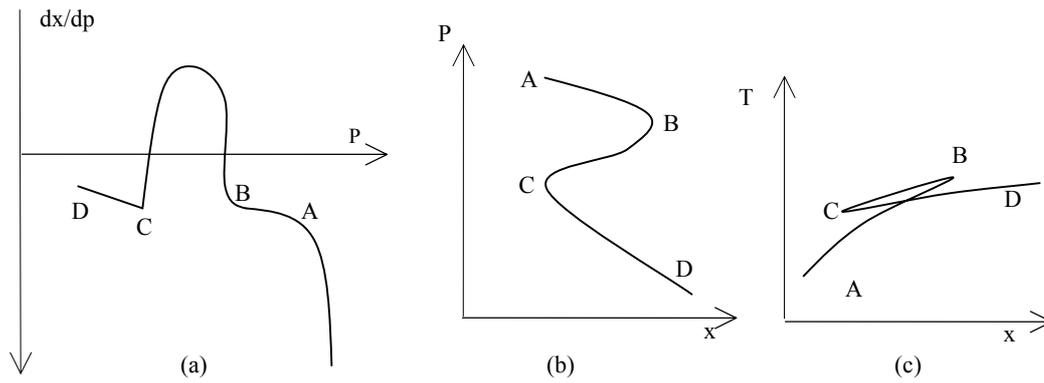


Figura 12. (a) Variación de la distancia con el parámetro del rayo. (b) Variación del parámetro del rayo con la distancia. (c) Dromocronica.

En la dromocrónica (figura 12, (c)) vemos que hay un segmento B a C que está invertido. En este intervalo de distancias se reciben tres llegadas para una misma distancia. El segmento de A a B es el correspondiente a la más baja velocidad y el tramo de C a D es el de mayor velocidad. Entre estos dos, el B a C corresponde al de mayor incremento en la velocidad.

Velocidad aumentando de forma discontinua con la profundidad

Consideremos un incremento lineal de la velocidad hasta una profundidad z_1 donde existe una discontinuidad, aumentando posteriormente de nuevo de forma lineal (figura 13).

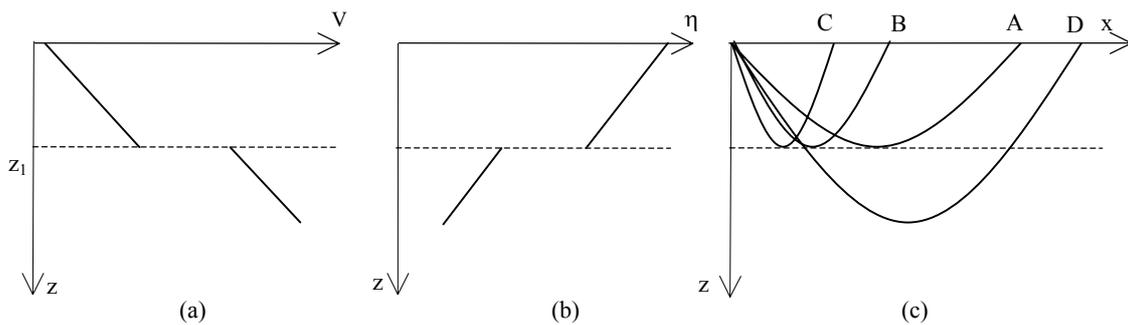


Figura 13. (a) Aumento discontinuo de la velocidad con la profundidad. (b) Variación de la lentitud con la profundidad. (c) Trayectorias de los rayos.

Sea p el parámetro del último rayo que viaja por la primera capa:

$$p = \frac{\text{sen } i_0}{v_0} = \frac{l}{v} \quad \text{con } v \text{ la velocidad cuando el rayo viaja horizontalmente.}$$

Si disminuimos un poco el ángulo i_0 , tendremos un i_1 y el parámetro de este rayo será:

$$p' = \frac{\text{sen } i_1}{v_0} = \frac{l}{v'} \quad \text{con } v' \text{ la velocidad cuando viaja horizontalmente por la 2ª capa.}$$

Pero puede ocurrir que esta v' no entre en el valor posible para poder viajar por la capa de más abajo, por lo tanto se refleja. Podemos afirmar que a partir de un ángulo el rayo se refleja y a partir de otro ya si puede penetrar en el otro medio (figuras 14 y 15).

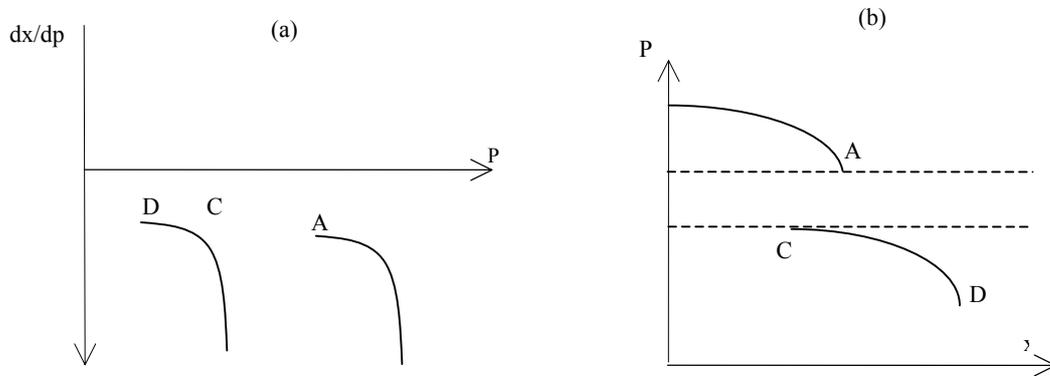


Figura 14. a) Variación de la distancia con el parámetro del rayo. (b) Variación del parámetro del rayo con la distancia.

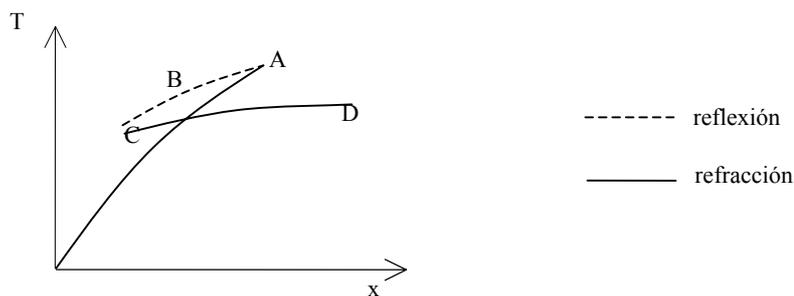


Figura 15. Dromocrónica.

Capa de baja velocidad

Consideremos una capa con decrecimiento en velocidad, llamada capa de baja velocidad, situada entre dos capas con incremento lineal de velocidad. Este tipo de capas es importante para la

tectónica de placas ya que podría indicar que se trata de un material en semi-fusión sobre el cual flotarían las placas (figura 16).

Por la ley de Snell, al disminuir la velocidad el rayo se acerca a la normal y no podrá viajar horizontalmente hasta que no se iguale la velocidad que tenía cuando empezó a disminuir la velocidad.

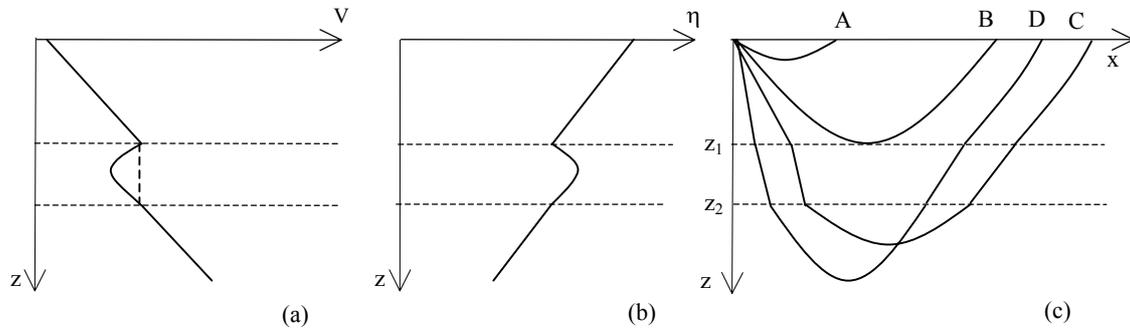


Figura 16. (a) Capa de baja velocidad, dentro de un aumento continuo de la velocidad con la profundidad. (b) Variación de la lentitud con la profundidad. (c) Trayectorias de los rayos.

Entre los puntos B y D no va a haber ningún rayo (figura 17) y para el rayo B:

$$p = \eta_a = \frac{1}{v_a}$$

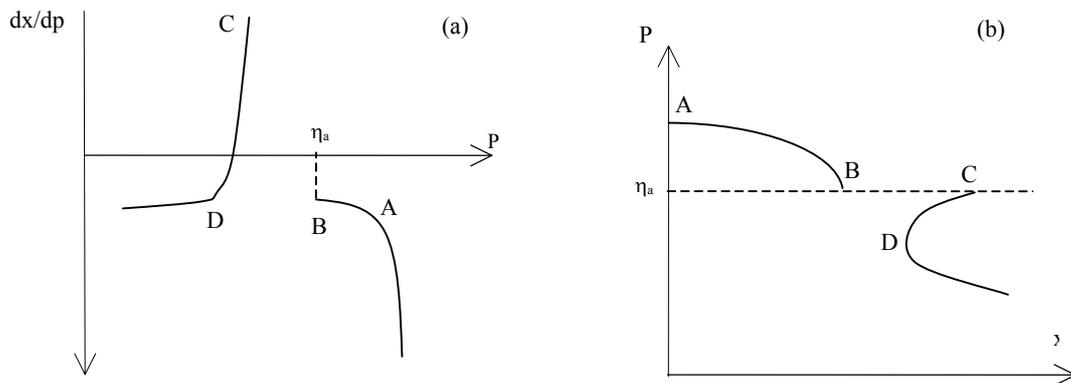


Figura 17.a) Variación de la distancia con el parámetro del rayo. (b) Variación del parámetro del rayo con la distancia.

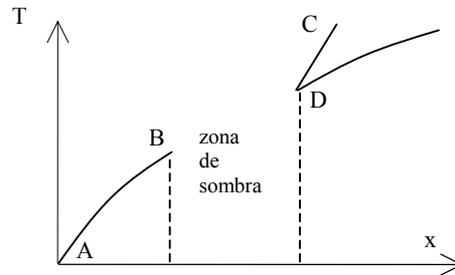


Figura 18. Dromocrónica.

En la dromocrónica (figura 18) existe una discontinuidad entre B y D, donde no existen rayos recibidos en la superficie, esta región es conocida como zona de sombra. Una característica de las copas de baja velocidad es que en el punto D existe una acumulación de energía.

12.6. Propagación en una Tierra esférica

En el estudio del comportamiento de las ondas sísmicas para grandes distancias ($x > 1000$ km) la aproximación de Tierra plana no es válida y se debe considerar con forma esférica. Consideremos una esfera de radio r_0 y velocidad dependiente del radio $v(r)$ y un rayo que se propaga desde un foco en la superficie F a un punto P también en la superficie (figura 19). Las distancias entre dos puntos de la superficie (F y P) se representan por el ángulo Δ o distancia angular (sobre la superficie terrestre la distancia sería $r_0\Delta$). El ángulo de incidencia i se mide desde la trayectoria del rayo al radio. La distancia radial desde el centro de la esfera al punto donde el rayo vuelve a ascender es r_p , que se corresponde con una velocidad v_p .

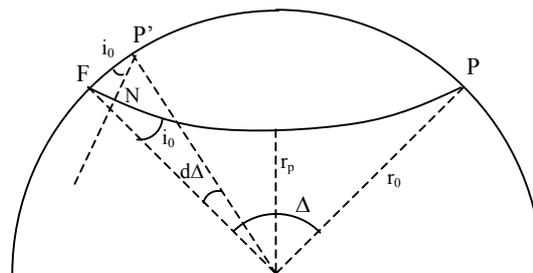


Figura 19. Trayectoria de un rayo en un medio esférico de velocidad aumentando de forma continua con la profundidad.

Si se considera el caso de capas concéntricas de velocidad constante (figura 20), de acuerdo con la ley de Snell

$$\frac{\text{sen } i_1}{v_1} = \frac{\text{sen } f}{v_2}$$

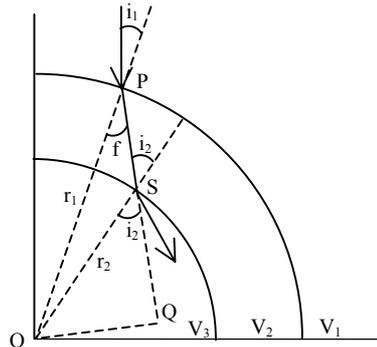


Figura 20. Trayectoria de un rayo en regiones esféricas de velocidad constante ($V_1 < V_2 < V_3$).

Si se consideran los triángulos PQO y SQO se tiene que:

$$r_2 \text{ sen } i_2 = r_1 \text{ sen } f$$

y sustituyendo en la expresión anterior:

$$\frac{r_1 \text{ sen } i_1}{v_1} = \frac{r_2 \text{ sen } i_2}{v_2}$$

Por tanto, la ley de Snell se puede expresar, en general cuando la velocidad varía de forma continua con el radio, para medios esféricos en la forma:

$$\frac{r \text{ sen } i}{v} = p$$

donde r , i y v están referidos a la misma capa y p es el parámetro del rayo.

Continuando con el caso de variación continua de la velocidad con el radio, consideremos el rayo que va de F a P (figura 19) correspondiente a una distancia angular Δ . El elemento de arco FP' será:

$$FP' = r_0 d\Delta$$

El segmento NP' es normal al rayo y representa un frente de onda que tarda en llegar de F a N un tiempo dt , si la velocidad en la superficie es $v(r_0) = v_0$:

$$FN = v_0 dt$$

y de acuerdo al triangulo FP'N:

$$\text{sen } i_0 = \frac{FN}{FP'}$$

y sustituyendo sus valores quedará:

$$\text{sen } i_0 = \frac{v_0 dt}{r_0 d\Delta}$$

que de acuerdo con la ley de Snell, para cada rayo:

$$\frac{r \text{ sen } i}{v} = p = \frac{dt}{d\Delta}$$

Luego el parámetro del rayo viene dado por la pendiente de la curva (t, Δ) o dromocronica. La relación anterior, llamada de Benndorf, es muy útil para obtener la distribución de velocidades a partir de la curva $t(\Delta)$.

12.6.1. Esfera de velocidad constante

Consideremos una esfera de radio R y velocidad constante V , cuyos rayos se propagan en línea recta y cuyo tiempo de viaje viene dado por (figura 21, (a)):

$$t = \frac{2R}{V} \text{sen } \frac{\Delta}{2}$$

La dromocronica no es una línea recta, sino una función seno (figura 21, (b)). El parámetro p del rayo varía con la distancia Δ en la forma:

$$p = \frac{dt}{d\Delta} = \frac{R}{V} \cos \frac{\Delta}{2}$$

Finalmente, podemos obtener la variación de la distancia angular respecto al parámetro p :

$$\frac{d\Delta}{dp} = \frac{-2}{(\eta^2 - p^2)^{1/2}}$$

donde $\eta = R/V$.

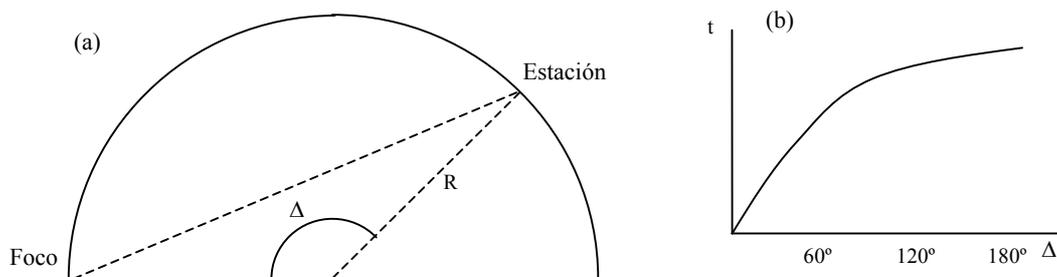


Figura 21.(a) Rayo en un medio esférico de velocidad constante. (b) Dromocrónica correspondiente.

12.6.2. Esfera de velocidad variable con el radio

Consideremos la trayectoria del rayo representado en la figura 22.

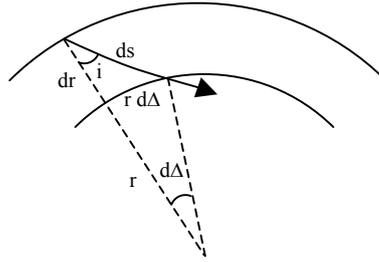


Figura 22. Elementos de la trayectoria de un rayo en un medio esférico de velocidad variable.

Se pueden escribir las siguientes relaciones:

$$(10) \quad ds^2 = dr^2 + (r d\Delta)^2$$

$$\text{sen } i = r \frac{d\Delta}{ds}$$

Sustituyendo el valor de $\text{sen } i$ en la ley de Snell, se obtiene:

$$\frac{r^2}{v} \frac{d\Delta}{ds} = p$$

Dividiendo en la ecuación (10) primero por ds^2 y después por dr^2 , y teniendo en cuenta además que $\eta = r/v$ y $d\Delta/ds = pv/r^2$ se obtienen las ecuaciones:

$$\frac{ds}{dr} = \frac{\eta}{\sqrt{\eta^2 - p^2}}$$

$$\frac{d\Delta}{dr} = \frac{p}{r} \frac{1}{\sqrt{\eta^2 - p^2}}$$

Integrando estas expresiones a lo largo del rayo desde la superficie (r_0) hasta el punto más profundo (r_p), se llega a:

$$\Delta = 2 \int_{r_p}^{r_0} \frac{p}{r} \frac{dr}{\sqrt{\eta^2 - p^2}}$$

$$S = 2 \int_{r_p}^{r_0} \frac{\eta \, dr}{\sqrt{\eta^2 - p^2}}$$

$$t = 2 \int_{r_p}^{r_0} \frac{\eta \, dr}{v \sqrt{\eta^2 - p^2}}$$

donde Δ es la distancia angular, S la distancia recorrida medida sobre el rayo y t el tiempo de recorrido.

Bibliografía

- [1] Udias, A. Principles of Seismology. Ed. Cambridge University Press. 1999. Cap. 3, 5, 6, 7 y 8.
- [2] Bullen, K.E. and Bolt, B.A. An introduction to the theory of seismology. Ed. Cambridge University Press. 1985. Cap. 7.
- [3] Udias, A. y Mezcuá, J. Fundamentos de Geofísica. Ed. Alhambra. 1986. Cap. 8.

Tema 13. Ondas sísmicas superficiales. Ondas superficiales en un medio semiinfinito. Ondas superficiales en una capa. Ondas Love. Dispersión de ondas. Velocidad de fase y de grupo.

13.1. Introducción

Las ondas P y S proporcionan una solución completa de la ecuación de movimiento en un medio continuo, elástico, isótropo e infinito, pero la presencia de una superficie libre en un medio elástico introduce una serie de fenómenos como son la reflexión de las ondas internas, la generación de ondas inhomogéneas y la aparición de ondas superficiales.

Este último fenómeno será el que trataremos en este tema, donde analizamos por separado los dos tipos de ondas superficiales: las ondas Rayleigh y las ondas Love, y una de las características principales de estas ondas: la dispersión.

13.2. Ondas sísmicas superficiales

La superficie de la Tierra delimita un medio elástico -el interior de la Tierra- de un medio que se aproxima al vacío. Por tanto, esta superficie puede considerarse una *superficie libre*, donde los esfuerzos sobre la misma son nulos.

Las *ondas superficiales* se definen como aquellas que se producen en un medio con una superficie libre, se propagan en una dirección paralela a esta superficie y su amplitud decrece con la distancia a la misma. Estas ondas se generan a partir de la energía de las ondas internas, P y S, que inciden en la superficie libre.

Además veremos como la presencia de otras superficies de discontinuidad entre capas con distintas propiedades elásticas, altera las características de las ondas superficiales.

Debido a que la energía de las ondas superficiales se propaga de forma paralela y en una región próxima a la superficie libre, su frente de ondas no es esférico, como en el caso de las ondas internas propagándose en un medio infinito, sino que es más bien cilíndrico. Es decir, que la propagación se produce fundamentalmente en dos dimensiones en lugar de en tres. Por lo tanto la energía disminuye con la distancia a la fuente, r , aproximadamente como r^{-1} , en lugar de r^{-2} , como ocurre cuando el frente de ondas es esférico. Por ello, a grandes distancias de la fuente, las ondas superficiales son las que predominan en los sismogramas, debido a su menor dispersión geométrica.

Estas ondas suelen presentar periodos comprendidos entre 10 y 300 s y aportan la información predominante de la parte de largo periodo del espectro de la fuente sísmica.

Por otra parte, el estudio de estas ondas es de gran valor para el conocimiento de la estructura interna de la corteza y la parte superior del manto.

13.3. Ondas superficiales en un medio semiinfinito

Consideremos un sistema de coordenadas cartesianas con ejes X_1 , X_2 y X_3 (X_3 en la dirección vertical y positivo hacia arriba).

Dado un semiespacio infinito, homogéneo e isótropo en $x_3 < 0$, limitado por una superficie libre coincidente con el plano $x_3 = 0$, vamos a suponer, siguiendo la definición de ondas superficiales, unas ondas que se propagan en la dirección del eje X_1 , con velocidad c y cuya amplitud decrece con $-x_3$, a las que denominaremos *ondas Rayleigh*.

Los desplazamientos de las ondas internas pueden escribirse en función de los potenciales, ϕ y ψ , como $\mathbf{u} = \nabla\phi + \nabla \times \boldsymbol{\psi} = \mathbf{u}^P + \mathbf{u}^S$. Si consideramos, sin perder generalidad, que las ondas inciden en el plano formado por X_1 y X_3 , las funciones potencial dependerán únicamente de X_1 y X_3 , por lo que podemos expresar los desplazamientos en las componentes X_1 y X_3 , u_1 y u_3 , en función de los potenciales escalares ϕ y $\psi = \psi_2$, dejando u_2 aparte:

$$\begin{aligned} u_1 &= \phi_{,1} - \psi_{,3} \\ (1) \quad u_2 &= u_2 \\ u_3 &= \phi_{,3} + \psi_{,1} \end{aligned}$$

Estamos buscando ondas que se propaguen en la dirección de X_1 y cuya amplitud dependa del valor de x_3 . Por lo tanto para una frecuencia y velocidad arbitrarias ω y c (número de onda k) los potenciales ϕ y ψ y el desplazamiento transversal u_2 serán:

$$\begin{aligned} \phi &= f(x_3)\exp[ik(x_1 - ct)] \\ (2) \quad \psi &= g(x_3)\exp[ik(x_1 - ct)] \\ u_2 &= h(x_3)\exp[ik(x_1 - ct)] \end{aligned}$$

Si sustituimos en las ecuaciones de ondas para ϕ , ψ y u_2 ¹ las expresiones dadas en la ecuación (2) obtenemos:

$$\begin{aligned} f''(x_3) + kr^2 f(x_3) &= 0 \\ (3) \quad g''(x_3) + ks^2 g(x_3) &= 0 \\ h''(x_3) + ks^2 h(x_3) &= 0 \end{aligned}$$

Donde:

¹ En efecto, u_2 cumple la ecuación de ondas, como podemos comprobar sin más que sustituir los desplazamientos dados en las expresiones (1) con los valores de (2) en la ecuación de Navier.

$$(4) \quad \begin{aligned} r &= \left(\frac{c^2}{\alpha^2} - 1 \right)^{1/2} \\ s &= \left(\frac{c^2}{\beta^2} - 1 \right)^{1/2} \end{aligned}$$

Las soluciones a las ecuaciones (3) son de la forma:

$$(5) \quad \begin{aligned} f(x_3) &= A' \exp(ikrx_3) + A \exp(-ikrx_3) \\ g(x_3) &= B' \exp(iksx_3) + B \exp(-iksx_3) \\ h(x_3) &= C' \exp(iksx_3) + C \exp(-iksx_3) \end{aligned}$$

Pero las ondas superficiales deben decrecer con la profundidad ($-x_3$). El único modo de que $f(x_3)$, $g(x_3)$ y $h(x_3)$ disminuyan con $-x_3$ es que r y s sean imaginarios y positivos y $A'=B'=C'=0$, obteniendo:

$$(6) \quad \begin{aligned} \phi &= A \exp[-ikrx_3 + ik(x_1 - ct)] \\ \psi &= B \exp[-iksx_3 + ik(x_1 - ct)] \\ u_2 &= C \exp[-iksx_3 + ik(x_1 - ct)] \end{aligned}$$

Para que r y s sean imaginarias se debe de cumplir que $c < \beta < \alpha$, es decir, que la velocidad de las ondas superficiales sea menor que la correspondiente a las ondas S.

Para evaluar las constantes arbitrarias A , B y C aplicamos las condiciones de contorno en la superficie libre, donde los esfuerzos han de ser nulos. De este modo las condiciones de contorno en $x_3 = 0$ son $\tau_{31} = \tau_{32} = \tau_{33} = 0$. Para un medio isótropo (donde se cumple: $\tau_{ij} = \lambda \delta_{ij} u_{k,k} + \mu (u_{i,j} + u_{j,i})$) tenemos que:

$$(7) \quad \begin{aligned} \tau_{31} &= \mu (u_{3,1} + u_{1,3}) \\ \tau_{32} &= \mu (u_{3,2} + u_{2,3}) \\ \tau_{33} &= (\lambda + 2\mu) u_{3,3} + \lambda (u_{1,1} + u_{2,2}) \end{aligned}$$

Poniendo en las ecuaciones (7) las derivadas de los desplazamientos en función de los potenciales, según las expresiones (1), tenemos que para $x_3 = 0$ las ecuaciones de contorno son:

$$(8) \quad \begin{aligned} 2\phi_{,31} + \psi_{,11} - \psi_{,33} &= 0 \\ u_{2,3} &= 0 \\ (\lambda + 2\mu)\phi_{,33} + \lambda\phi_{,11} + 2\mu\psi_{,13} &= \rho\alpha^2\phi_{,33} + \rho(\alpha^2 - 2\beta^2)\phi_{,11} + \rho^2\beta^2\psi_{,13} = 0 \end{aligned}$$

De la segunda ecuación de (8) se deduce que $C = 0$. Es decir, que las ondas superficiales en un semiespacio no tienen componente transversal del desplazamiento.

Sustituyendo en la primera y tercera ecuación de (8) las expresiones para los potenciales descritas en (6) obtenemos que para $x_3 = 0$:

$$(9) \quad \begin{aligned} 2rA - (1 - s^2)B &= 0 \\ [\alpha^2(r^2 + 1) - 2\beta^2]A - 2\beta^2sB &= 0 \end{aligned}$$

Estas condiciones de contorno forman un sistema homogéneo de ecuaciones con dos incógnitas: A y B . Por lo tanto la condición para que este sistema tenga solución distinta a la trivial es que el determinante del sistema sea nulo, lo cual nos da el siguiente resultado:

$$(10) \quad [\alpha^2(r^2 + 1) - 2\beta^2](1 - s^2) - 4rs\beta^2 = 0$$

Reemplazando los valores de r y s de la expresión (4) en la ecuación (10) tenemos que:

$$(11) \quad \left(2 - \frac{c^2}{\beta^2}\right)^2 = 4\left(1 - \frac{c^2}{\alpha^2}\right)^{1/2} \left(1 - \frac{c^2}{\beta^2}\right)^{1/2}$$

En esta ecuación, conocida como *ecuación de Rayleigh*, se observa que la velocidad, c , no depende de la frecuencia ω .

De la expresión (11) deducimos, haciendo el cambio de variables $\xi = (c/\beta)^2$, que la velocidad de las ondas superficiales, c , depende únicamente del coeficiente $q = (\beta^2/\alpha^2)$. Este valor está

relacionado con el coeficiente de Poisson $\sigma = \frac{\lambda}{2(\lambda + \mu)}$, obteniéndose distintos valores de c

para distintos σ . Por ejemplo, para un medio líquido ($\sigma = 0.5$) se obtiene $c = 0$, es decir, que las ondas Rayleigh no existen en un semiespacio líquido. Para $\sigma = 0.25$, como ocurre aproximadamente en los materiales de la Tierra, $c_R = 0.92\beta$.

Los desplazamientos de las ondas superficiales se obtienen a partir de los potenciales ϕ y ψ . Suponiendo $\sigma = 0.25$ tenemos $c_R = 0.92\beta$, $r = 0.85i$ y $s = 0.39i$, con lo que sustituyendo en una de las ecuaciones de (9) obtenemos $B = -1.47iA$. Tomando únicamente la parte real obtenemos:

$$(12) \quad \begin{aligned} u_1 &= -Ak(e^{0.85kx_3} - 0.58e^{0.39kx_3})\sin[k(x_1 - c_R t)] \\ u_3 &= -Ak(-0.85e^{0.85kx_3} + 1.47e^{0.39kx_3})\cos[k(x_1 - c_R t)] \end{aligned}$$

El hecho de no tener desplazamientos horizontales transversales a la dirección de propagación ($u_2 = 0$; sus desplazamientos están contenidos en el plano de incidencia), refleja que las ondas Rayleigh están generadas por la interacción de ondas P y SV con la superficie libre. Para puntos de esta superficie ($x_3 = 0$) los desplazamientos son:

$$(13) \quad \begin{aligned} u_1 &= 0.42a \sin[k(x_1 - c_R t)] \\ u_3 &= 0.62a \cos[k(x_1 - c_R t)] \end{aligned}$$

Donde $a = -Ak$.

Dado que las ondas Rayleigh no tienen componente transversal, están polarizadas en el plano de incidencia $x_2 = 0$. Como vemos en las ecuaciones (12), las componentes horizontal y vertical están desfasadas $\pi/2$ por lo que el movimiento de la partícula es elíptico con el semieje mayor en la dirección vertical (ver *Figura 1*). En $x_3 = 0$ el movimiento es retrógrado y conforme vamos aumentando la profundidad disminuye el valor de u_1 hasta ser nulo y posteriormente

cambiar de signo, variando también el sentido de giro de la partícula. El valor de u_3 en cambio nunca se hace cero, y siempre es mayor que el de u_1 , aunque ambas componentes del desplazamiento decrecen exponencialmente con la profundidad.

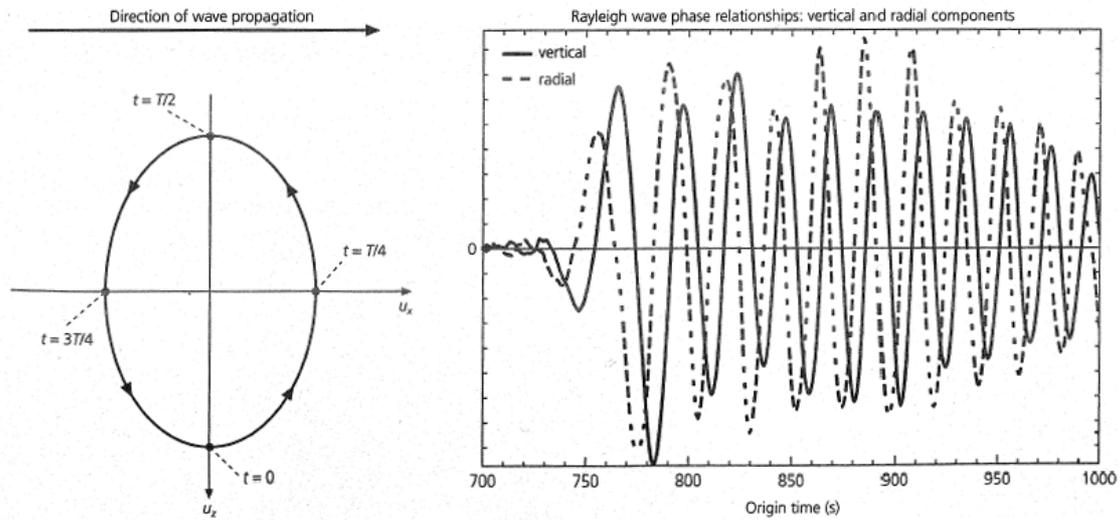


Figura 1. Izquierda: Movimiento de la partícula en un punto de la superficie libre. Derecha: Comparación de las componentes vertical y radial en un registro de ondas Rayleigh. (Stein y Wysession, 2003).

Dado que las amplitudes de los desplazamientos de las ondas Rayleigh tienen una dependencia con la profundidad del tipo $e^{kx_3} = e^{(2\pi/\lambda)x_3}$, las ondas Rayleigh con grandes longitudes de onda, λ , presentan mayores desplazamientos en profundidad que las de pequeñas longitudes de onda.

13.4. Ondas superficiales en una capa

La existencia de una o más capas de espesor finito sobre el semiespacio modifica los resultados obtenidos anteriormente. Para estudiar este efecto podemos analizar el caso más sencillo: una capa líquida de espesor H sobre un semiespacio rígido (donde no se propagan ondas elásticas). En este caso la velocidad de las ondas superficiales que se propagan en el medio líquido cumplen la ecuación:

$$(14) \quad kH \left(\frac{c^2}{\alpha^2} - 1 \right)^{1/2} = \left(n + \frac{1}{2} \right) \pi$$

Donde n es cualquier número entero positivo.

En esta ecuación observamos que la velocidad, c , a diferencia del caso anterior, depende del número de onda k , y por lo tanto de la frecuencia de la onda, ω . Este fenómeno se conoce como *dispersión* y la ecuación que relaciona c con k , *ecuación de dispersión*.

Además según sea el valor de n existe un número infinito de ecuaciones de dispersión. Es decir que existe un número infinito de tipos de propagación y cada uno de ellos se denomina *modo*, siendo el de $n = 0$ el *modo fundamental*.

Estas dos propiedades son consecuencia directa de la propagación de las ondas superficiales en una capa finita.

13.5. Ondas Love

Si situamos una capa elástica por encima de un semiespacio elástico infinito, la solución general de este problema indica la existencia de dos tipos de ondas: uno formado por ondas Rayleigh, con desplazamientos en el plano de incidencia y otro por ondas con desplazamientos en la dirección horizontal transversal, denominadas *ondas Love*.

Vamos a analizar el segundo tipo, por lo que consideraremos ondas que sólo poseen la componente transversal del desplazamiento, u_2 , que se propagan en la dirección x_1 y con velocidad c en una capa de espesor H , densidad ρ' , y velocidad de las ondas S, β' , que se sitúa sobre un semiespacio con densidad ρ y velocidad β , siendo $\beta > \beta'$.

Para describir el campo de desplazamientos tendremos en cuenta que las ondas pueden viajar en las direcciones positiva y negativa del eje X_3 dentro de la capa y sólo hacia abajo en el semiespacio, quedando:

$$\begin{aligned} (15) \quad u_2' &= (A' e^{iks'x_3} + B' e^{-iks'x_3}) e^{ik(x_1 - ct)} \\ u_2 &= B e^{-iksx_3 + ik(x_1 - ct)} \end{aligned}$$

Donde s viene dado en la ecuación (4) y $s = \tan(e)$, siendo e el ángulo de emergencia, y s' es el equivalente a s utilizando β' en lugar de β .

De acuerdo con la definición de las ondas superficiales las amplitudes de las ondas en el semiespacio (u_2) deben decrecer con la profundidad ($-x_3$), por lo tanto s debe ser positiva e imaginaria, con lo que $c < \beta$.

Por otro lado las condiciones de contorno son la ausencia de esfuerzos en la superficie libre ($x_3 = H$) y la continuidad de los desplazamientos y esfuerzos en la superficie de contacto entre la capa y el semiespacio ($x_3 = 0$). Es decir:

$$\begin{aligned} x_3 = H : \quad \tau_{32}' &= 0 \\ (16) \quad x_3 = 0 : \quad \tau_{32}' &= \tau_{32} \\ u_2' &= u_2 \end{aligned}$$

Estas condiciones en función de los desplazamientos son:

$$\begin{aligned} x_3 = H : \quad \mu' u_{2,3}' &= 0 \\ (17) \quad x_3 = 0 : \quad \mu' u_{2,3}' &= \mu u_{2,3} \\ u_2' &= u_2 \end{aligned}$$

Sustituyendo (15) en (17) obtenemos el siguiente sistema de ecuaciones:

$$\begin{aligned} A'e^{iks'H} - B'e^{-iks'H} &= 0 \\ (18) \quad A'\mu's' - B'\mu's' + Bs\mu &= 0 \\ A' + B' - B &= 0 \end{aligned}$$

Para que exista solución no trivial a este sistema de ecuaciones homogéneo, el determinante del sistema debe ser nulo, por lo que:

$$(19) \quad \frac{\mu s}{\mu's'} = \frac{e^{iks'H} - e^{-iks'H}}{e^{iks'H} + e^{-iks'H}}$$

Si s' es imaginario haciendo las sustituciones $s' = i\bar{s}'$ y $s = i\bar{s}$ tenemos:

$$(20) \quad \frac{\mu\bar{s}}{\mu'\bar{s}'} = -\tanh(k\bar{s}'H)$$

Ecuación imposible de cumplir dado que, tanto el primer miembro como la tangente hiperbólica, son siempre positivos. Por lo tanto s' tiene que ser real, con lo que $\beta > c > \beta'$. Es decir, que la velocidad de las ondas Love se sitúa entre las velocidades de las ondas S de los dos medios. Y la ecuación (19) queda:

$$(21) \quad \frac{\mu\left(1 - \frac{c^2}{\beta^2}\right)^{1/2}}{\mu'\left(\frac{c^2}{\beta'^2} - 1\right)^{1/2}} = \tan\left[kH\left(\frac{c^2}{\beta'^2} - 1\right)^{1/2}\right]$$

Y teniendo en cuenta que $k = \omega/c$:

$$(22) \quad \frac{\mu\left(\frac{1}{c^2} - \frac{1}{\beta^2}\right)^{1/2}}{\mu'\left(\frac{1}{\beta'^2} - \frac{1}{c^2}\right)^{1/2}} = \tan\left[\omega H\left(\frac{1}{\beta'^2} - \frac{1}{c^2}\right)^{1/2}\right]$$

Siendo esta la *ecuación de dispersión de las ondas Love*.

La función tangente da como resultado valores positivos entre cero e infinito para infinitos intervalos (entre 0 y $\frac{\pi}{2}$, π y $\frac{3\pi}{2}$, ...). Correspondiendo cada uno de ellos a un modo de propagación. De manera que:

$$(23) \quad n\pi \leq kH\left(\frac{c^2}{\beta'^2} - 1\right)^{1/2} \leq \frac{2n+1}{2}\pi$$

Donde n puede ser cualquier número entero positivo. El modo fundamental viene dado por $n = 0$ y los modos superiores por $n \geq 1$.

En la *Figura 2* se muestran las curvas de dispersión definidas por la ecuación (21), donde aparecen los valores que toma la velocidad de fase de las ondas Love, c , en función de la

frecuencia para cada modo de propagación. La forma de estas curvas, como vemos en la expresión (21) depende de los parámetros del modelo: β , β' y H .

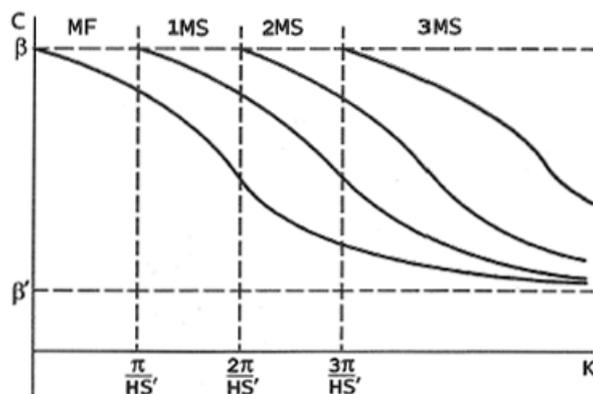


Figura 2. Curvas de dispersión de la velocidad de fase de las ondas Love en una capa sobre un semiespacio (Udías, 1999).

En el modo fundamental tenemos una velocidad de fase para cada una de las frecuencias, de cero a infinito, y los valores de la velocidad, c , toman valores de β a β' . La influencia del espesor de la capa, H , está condicionada por su relación con la longitud de onda, λ ($k = 2\pi/\lambda$). Cuando $\lambda < H$, la propagación está influenciada por las características elásticas de la capa y c se aproxima al valor de β' . Sin embargo cuando $\lambda > H$, la propagación está influenciada por el semiespacio y c se aproxima al valor de β .

Para los modos superiores existe un límite para las bajas frecuencias, denominado *frecuencia de corte*. Este valor puede deducirse de la expresión (23) y, para cada orden, n , viene dado por:

$$(24) k_n = \frac{n\pi}{HS'}$$

En los modos superiores las curvas de dispersión comienzan en su frecuencia de corte con $c = \beta$ y a medida que aumenta la frecuencia el valor de c va aproximándose al de β' .

Como vemos en la *Figura 2* para una velocidad se dan distintos modos de propagación a distintas frecuencias, de forma que ondas de bajas frecuencias del modo fundamental pueden ser registradas en el mismo instante que ondas de altas frecuencias de modos superiores.

Otra forma de obtener las ondas Love es a partir del resultado de la interferencia constructiva de las ondas dentro de la capa, las cuales se reflejan en la superficie de separación entre la capa y el semiespacio infinito, con un ángulo de incidencia, i , superior al crítico (reflexión total²). Se trata por tanto de ondas canalizadas en la capa en contacto con la superficie libre.

Por otro lado, dado que las ondas Love sólo presentan desplazamientos en la componente horizontal transversal, u_2 , deducimos que están generadas únicamente por ondas SH.

² Para que se produzca este fenómeno es imprescindible que $\beta > \beta'$.

De este modo, podemos concluir que las ondas Love están formadas por la interferencia constructiva de ondas SH dentro de la capa elástica que inciden de forma supercrítica en la superficie de discontinuidad entre la capa y el medio y de ondas inhomogéneas en el semiespacio.

Los *desplazamientos* de las ondas Love para la capa y el semiespacio se calculan a partir de las expresiones (15) con los valores de A y B obtenidos a partir de la solución del sistema de ecuaciones (18). De este modo obtenemos, quedándonos únicamente con la parte real:

$$\begin{aligned} (25) \quad u_2' &= 2A' \cos \left[ks'H \left(1 - \frac{x_3}{H} \right) \right] \cos [k(s'H + x_1 - ct)] \\ u_2 &= 2A' \cos(ks'H) e^{iksx_3} \cos [k(s'H + x_1 - ct)] \end{aligned}$$

A la vista de (25) y dado que s es positiva e imaginaria y s' real, observamos que en el semiespacio los desplazamientos decrecen exponencialmente con la profundidad y dentro de la capa, éstos varían con la profundidad dependiendo de $ks'H$, por lo tanto dependiendo del modo de vibración. En la superficie libre ($x_3 = H$) la amplitud del desplazamiento es $2A'$, para todas las frecuencias y todos los modos de vibración. En el modo fundamental la amplitud del desplazamiento disminuye con la profundidad hasta alcanzar un valor entre 0 y $2A'$ en la base de la capa ($x_3 = 0$), según sea su frecuencia, pero en ningún caso el valor de la amplitud es nulo. En los modos superiores, en cambio, el número de nodos es igual al orden del modo.

13.5.1. Ondas Rayleigh en una capa sobre un semiespacio

El problema de la propagación de ondas Rayleigh en una capa y un semiespacio infinito es más complicado que el de las ondas Love, al ser dos las condiciones de contorno en la superficie libre (ausencia de esfuerzos) y cuatro en la superficie de separación entre la capa y el medio (continuidad de esfuerzos y desplazamientos).

En este caso también se obtiene la dependencia de la velocidad c con la frecuencia³ y la presencia de modos de vibración.

La forma de la curva de dispersión (*Figura 3*) es parecida a la de las ondas Love, pero ahora las velocidades máximas y mínimas para el modo fundamental no son β y β' sino c_R y c_R' (velocidades de la onda Rayleigh en un semiespacio infinito con características elásticas iguales a las del medio, c_R , y a las de la capa, c_R'). Sin embargo para los modos superiores las velocidades límite son β y β' .

³ La no dependencia de la velocidad de fase de las ondas Rayleigh con la frecuencia sólo se da cuando se considera una superficie libre sobre un semiespacio infinito y homogéneo. En medios estratificados o con heterogeneidad vertical, estas ondas presentan dispersión. Las ondas Love en cambio, siempre son dispersivas.

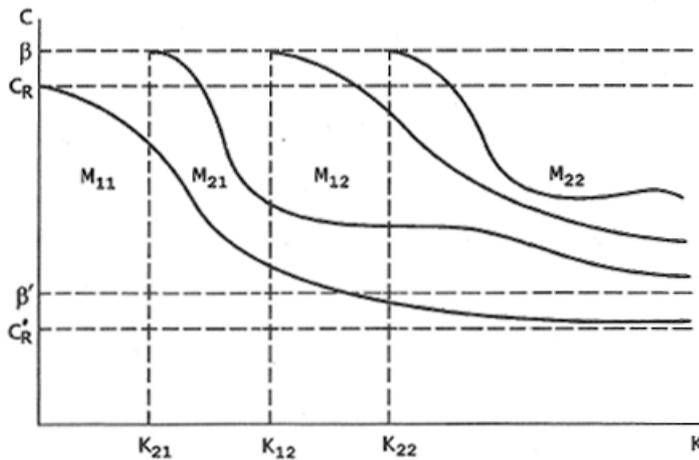


Figura 3. Curvas de dispersión de la velocidad de fase de las ondas Rayleigh en una capa elástica sobre un semiespacio (Udías, 1999).

Del mismo modo que en las ondas Love, estas ondas Rayleigh pueden obtenerse como interferencia constructiva de ondas P y SV en la capa reflejadas de forma supercrítica en la superficie de contacto.

En general, las ondas superficiales se observan en sismogramas producidos por terremotos lejanos y poco profundos como largos trenes de ondas con grandes amplitudes. Como ya hemos mencionado, las ondas Rayleigh (LR) producen desplazamientos en el plano de incidencia (componente vertical y radial) y las ondas Love (LQ) solo en la dirección transversal. Por lo tanto, en un sismograma, las ondas Love solo aparecerán en las componentes horizontales y en cambio las ondas Rayleigh en las componentes horizontales y vertical.

Las ondas Love y Rayleigh de pequeños periodos en trayectorias continentales están canalizadas en la corteza superior y se conocen como ondas Lg y Rg, en cambio, para periodos entre 60 y 300 s estas ondas viajan fundamentalmente por el manto, por lo que se denominan “ondas del manto”. En los terremotos de gran magnitud, se observan ondas superficiales que viajan alrededor de la Tierra rodeándola más de una vez. Estas ondas se designan como G_1, G_2, G_3, \dots para las ondas Love y como R_1, R_2, R_3, \dots para las ondas Rayleigh. El subíndice 1 hace referencia a ondas directas, el subíndice 2 a ondas que van del epicentro a la estación en el sentido opuesto, el subíndice 3 a ondas que viajan en el sentido de G_1 y R_1 dado una vuelta completa a la Tierra y llegando a la estación, el subíndice 4 a ondas que dan una vuelta a la Tierra en el sentido de G_2 y R_2 , y así sucesivamente (ver *Figura 4*)

Si en lugar de una capa sobre un semiespacio infinito tenemos dos semiespacios infinitos de características diferentes en contacto, se producen unas ondas de similares características, denominadas *ondas Stoneley*, cuyas amplitudes disminuyen conforme nos alejamos de la superficie de contacto y cuya velocidad no depende de la frecuencia, al igual que para las ondas Rayleigh en un semiespacio.

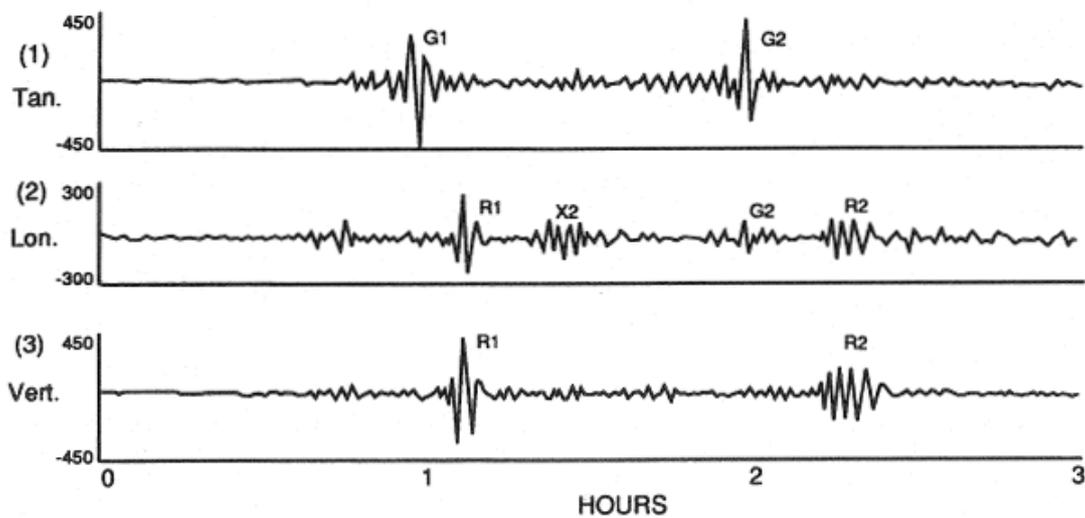


Figura 4. Registro de ondas superficiales de grandes periodos. La llegada identificada como X_2 corresponde a un modo superior de R_2 (su velocidad es mayor que para el modo fundamental). (Lay y Wallace, 1995).

13.6. Dispersión de ondas. Velocidad de fase y de grupo

Hemos visto que tanto en las ondas Love como en las ondas Rayleigh al introducir una capa de espesor finito existe dependencia entre la velocidad de propagación de las ondas y la frecuencia o periodo. Debido a esta dependencia, las ondas de distinto periodo viajan con distinta velocidad, por lo que la señal sísmica se extiende en el tiempo al aumentar la distancia recorrida en lo que denominamos *tren de ondas*. Es decir, que para una función impulsiva en la fuente, las ondas superficiales a una cierta distancia están formadas por trenes de ondas.

La velocidad de propagación a la que nos hemos referido hasta ahora es la *velocidad de fase*, o velocidad con la que se propaga la fase o el frente de ondas de cada componente armónico.

Sin embargo, las ondas sísmicas están compuestas por la suma de distintas ondas armónicas cada una de ellas con una frecuencia diferente. Si la velocidad de fase de cada una de ellas es la misma (c no depende de ω), como ocurre con las ondas internas en un medio no dispersivo, esta velocidad resulta ser igual a la velocidad de transporte de la energía. En cambio, si la velocidad de fase no es la misma para todos los componentes armónicos (c es función de la frecuencia, $c(\omega)$), la velocidad de transporte de la energía ya no es la velocidad de fase, sino la denominada *velocidad de grupo*.

Para tener una idea intuitiva de lo que representa la velocidad de grupo, vamos a analizar un caso muy sencillo: una onda formada por dos componentes monocromáticas con amplitudes iguales y frecuencias muy parecidas, ω_1 y ω_2 , tal que:

$$(26) \begin{aligned} \omega_1 &= \omega_0 + \Delta\omega, & \omega_2 &= \omega_0 - \Delta\omega, & \omega_0 &\gg \Delta\omega \\ k_1 &= k_0 + \Delta k, & k_2 &= k_0 - \Delta k, & k_0 &\gg \Delta k \end{aligned}$$

El desplazamiento de una onda sinusoidal de frecuencia ω y número de onda k viene dado por:

$$(27) u(x, t) = A \cos[(kx - \omega t) + \varepsilon]$$

Donde la velocidad de fase es $c = \frac{\omega}{k}$. Por lo tanto el desplazamiento total, considerando una fase inicial nula ($\varepsilon = 0$), vendrá dado por:

$$(28) u = A \cos(k_1 x - \omega_1 t) + A \cos(k_2 x - \omega_2 t)$$

Sustituyendo las expresiones de (26) en (28) tenemos:

$$(29) u = A \cos(k_0 x - \omega_0 t + (x\Delta k - t\Delta\omega)) + A \cos(k_0 x - \omega_0 t - (x\Delta k - t\Delta\omega))$$

Y utilizando, $2 \cos a \cos b = \cos(a + b) + \cos(a - b)$, expresamos la ecuación (29) como:

$$(30) u = 2 \cos(k_0 x - \omega_0 t) \cos(x\Delta k - t\Delta\omega)$$

Esto es el producto de dos cosenos, el segundo de los cuales varía mucho más lentamente que el primero, lo que representa una onda de frecuencia ω_0 modulada por una onda de frecuencia $\Delta\omega$ (Figura 5). Por lo tanto, está formada por paquetes, que tienen una longitud de onda $2\pi/\Delta\omega$ y se propagan con una velocidad $U = \Delta\omega/\Delta k$. Estos paquetes a su vez están formados por ondas de frecuencia ω_0 y velocidad de fase $c = \omega_0/k_0$.

Dado que estos paquetes de ondas modelan los máximos de la amplitud, la velocidad U es la velocidad de grupo o velocidad de propagación de la energía.

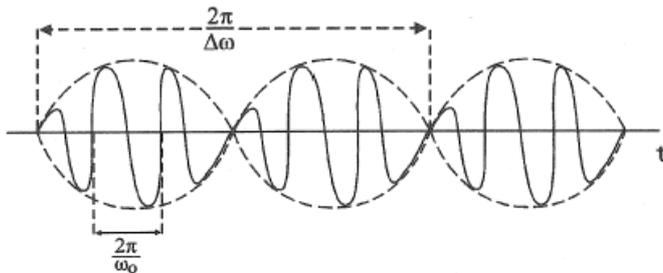


Figura 5. Suma de de dos ondas monocromáticas con frecuencias $\omega_0 + \Delta\omega$ y $\omega_0 - \Delta\omega$ (Udías, 1999).

En general, la *velocidad de grupo* se define como:

$$(31) U = \frac{d\omega}{dk}$$

Si derivamos ω respecto a k tenemos:

$$(32) U = c + k \frac{dc}{dk}$$

Donde podemos ver que si c no depende de la frecuencia, U y c coinciden.

Si aplicamos la expresión (32) a la función de la velocidad de fase para las ondas Love dada en la expresión (21) obtenemos la función $U(k)$ para las ondas Love que se muestra en la Figura 6.

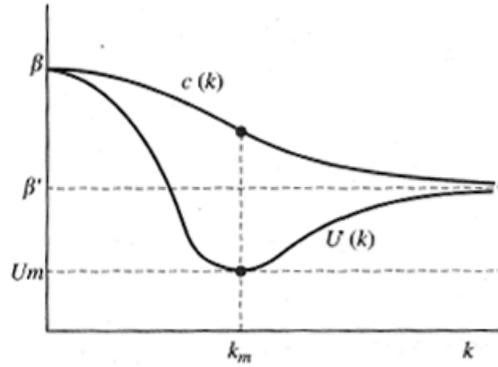


Figura 6. Curva de dispersión de la velocidad de fase y velocidad de grupo del modo fundamental de las ondas Love en una capa sobre un medio semiinfinito (Udías y Mezcuca, 1996).

13.6.1. Principio de la fase estacionaria

Una onda formada por una distribución continua de frecuencias en un ancho espectro puede escribirse como:

$$(33) u(x, t) = \int_0^{\infty} A(\omega) \cos \left[\omega \left(\frac{x}{c(\omega)} - t \right) + \varepsilon(\omega) \right] d\omega$$

Si ponemos la ecuación (33) en función de k , suponemos la fase inicial nula $\varepsilon(\omega) = 0$ y designamos la fase como $\Phi(k) = kx - \omega t$:

$$(34) u(x, t) = \int_0^{\infty} A(k) \cos[\Phi(k)] dk$$

Si la amplitud $A(k)$ varía lentamente en función de k en comparación con la fase $\Phi(k)$, la integral (34) solo tendrá valores significativos para frecuencias en las que la fase $\Phi(k)$ sea estacionaria, es decir $\frac{d\Phi}{dk} = 0$. Para variaciones rápidas de $\Phi(k)$, la función coseno cambia continuamente de signo haciendo que la integral sea nula, excepto para aquellos valores de $k = k_0$, donde se cumple la condición anterior, siendo k_0 el valor del número de onda para el cual la fase es estacionaria. Debido a esto la integral (34) se reduce a:

$$(35) u(x, t) = \int_{k_0 - \Delta k}^{k_0 + \Delta k} A(k) \cos[\Phi(k)] dk$$

Si desarrollamos la derivada de Φ y la igualamos a cero tenemos que:

$$(36) \frac{d\Phi}{dk} = x - \frac{d\omega}{dk} t = 0$$

Y dado que esta igualdad se cumple para $k = k_0$, tenemos que:

$$(37) \frac{x}{t} = U(k_0)$$

Es decir, que para cada distancia x y un tiempo t , la energía esta contenida en ondas con número de onda k_0 (o frecuencia ω_0) que llegan a la velocidad correspondiente a la velocidad de grupo para k_0 . Para cada par (x, t) k_0 toma un valor diferente.

Las amplitudes de la fase estacionaria se pueden obtener al expresar $\Phi(k)$ como un desarrollo de Taylor y después de tomar sólo los términos de orden dos, resolver la integral (34) bajo esta aproximación.

De este modo, y tomando solo la parte real, se llega a la expresión:

$$(38) u(x, t) = A(k_0) \left(\frac{2\pi}{\frac{x}{U} \frac{dU}{dk}} \right)^{1/2} \cos \left(k_0 x - \omega_0 t \pm \frac{\pi}{4} \right)$$

Por lo que las mayores amplitudes corresponden al caso $\frac{dU}{dk} = 0$ o *fase de Airy*⁴.

Así, en un sismograma, las frecuencias observadas de las ondas superficiales que presentan dispersión son, en gran medida, aquellas que corresponden a la *fase estacionaria* y sus tiempos de llegada dependerán de la *velocidad de grupo* en esas frecuencias y no de la velocidad de fase. En concreto para las ondas Love y siguiendo la *Figura 6*, las primeras ondas en llegar, para un tiempo $t = x/\beta$, serán aquellas con menores frecuencias. A continuación, cuando $t = x/\beta'$, superpuestas a las ondas de bajas frecuencias llegaran las ondas de frecuencias más altas. Por último, en $t = x/U_m$ (siendo U_m la velocidad de grupo mínima que corresponde a la *fase de Airy*), llegarán las ondas de mayor amplitud con número de onda k_m .

A una distancia fija, para cada tiempo, t , la energía llega en una onda de frecuencia ω_0 distinta cada vez y denominada *frecuencia instantánea*⁵. Si consideramos ahora un incremento de fase $\delta\Phi$, relacionado con cambios en la distancia y en el tiempo, y también en la frecuencia y en el número de onda:

$$(39) \delta(kx - \omega t) = k\delta x + x\delta k - t\delta\omega - \omega\delta t$$

Para frecuencias y números de onda correspondientes a la fase estacionaria por la ecuación (36) tenemos que $x\delta k - \omega\delta t = 0$, y por lo tanto:

$$(40) \delta\Phi = k_0\delta x - \omega_0\delta t$$

Aplicando esta ecuación, si observamos la misma fase ($\delta\Phi=0$) en una onda de la misma frecuencia ω_0 , en dos puntos separados por una distancia δx , con un intervalo de tiempo δt :

$$(41) \frac{\delta x}{\delta t} = \frac{\omega_0}{k_0} = c$$

⁴ En este caso el término de orden dos del desarrollo de Taylor antes citado es nulo, por lo que tenemos que utilizar hasta el término de orden 3, en cuyo caso la expresión (37) ya no tiene la derivada de U con respecto a k en el denominador, sino su segunda derivada.

⁵ En el ejemplo anterior, en el intervalo desde $t = x/\beta$ a $t = x/U_m$ existen dos frecuencias instantáneas correspondientes a la misma velocidad de grupo y por lo tanto al mismo tiempo.

Usando las ecuaciones (37) y (41) podemos determinar tanto la velocidad de grupo como la de fase por medio de observaciones de trenes de ondas dispersadas. Estas ideas son la base de algunos métodos para calcular las curvas de dispersión de las velocidades de fase y de grupo a partir de los sismogramas, las cuales nos aportan una valiosa información sobre la estructura interna de la corteza y la parte superior del manto.

Bibliografía

- [1] Lay T. y Wallace, T. C. Modern Global Seismology. Academic Press, 1995. Pág. 116-147.
- [2] Udías, A. Principles of Seismology. Cambridge University Press, 1999. Pág. 183-204, 217-238.
- [3] Udías, A. y Mézcua, J. Fundamentos de Geofísica. Editorial Alianza, 1996. pág . 183-197.

Tema 14. Resolución del problema inverso en sismología. Inversión de dromocrónicas: Fórmula de Herglotz-Wiechert. Inversión de ondas superficiales. Inversa generalizada.

14.1. Resolución del problema inverso en sismología

Consideremos un proceso físico que puede ser descrito por medio de un modelo matemático $G(\mathbf{x})$; el cual, para un conjunto de parámetros recogidos por el vector \mathbf{m} , da lugar a un conjunto de datos expresados por el vector \mathbf{d} . Es decir, los datos son el resultado de una función u operador, G , actuando sobre los parámetros del modelo:

$$(1) \quad \mathbf{d} = G(\mathbf{m})$$

El denominado *problema directo* consiste en predecir los datos, \mathbf{d} , a partir de un modelo y unos parámetros conocidos.

En contraposición, el *problema inverso*, para un modelo determinado, trata de encontrar los parámetros del modelo, \mathbf{m} , a partir de las observaciones o datos expresados por \mathbf{d} .

Para resolver el problema directo únicamente debemos comprender el modelo físico que tiene lugar y aplicarle unos determinados parámetros. En cambio el problema inverso presenta mayores dificultades. Para resolverlo pueden utilizarse dos acercamientos distintos: las técnicas matemáticas de inversión, donde se halla \mathbf{m} directamente a partir de \mathbf{d} , o bien técnicas de prueba-error en las que se lleva a cabo el problema directo repetidamente hasta encontrar la solución que más se ajusta a nuestros datos. Cada uno de estos dos enfoques tiene ventajas e inconvenientes, por lo que habrá que analizar cuál es el más adecuado en cada caso.

Aunque los problemas directos, en general, nos proporcionan una solución única, los problemas inversos normalmente no tienen una solución única, exacta o correcta. Esto se debe a que es común tener datos inconsistentes debido a errores de observación unido a que nuestros modelos suelen simplificar la realidad, por lo que no existe un modelo que describa exactamente los datos. Por otro lado, es posible que distintos conjuntos de parámetros se ajusten a los datos de forma similar para un modelo dado, debiendo elegir uno de entre todos ellos. Además, en muchas ocasiones, los datos pueden ser insuficientes para resolver determinados aspectos del modelo.

Una consecuencia de estas limitaciones nos lleva a que muchas veces haya que aceptar una solución de compromiso entre la resolución del modelo (cuan detallado es éste) y su estabilidad o robustez.

La sismología esta dirigida en gran medida a la resolución de *problemas inversos* relacionados con la fuente de los terremotos y con la estructura interna de la Tierra. En general, partimos del

resultado final, los *sismogramas*, y a partir de ellos tratamos de caracterizar los terremotos que han generado las ondas sísmicas y el medio a través del cual estas ondas se han propagado. Esta doble dependencia de la fuente sísmica y el medio en los sismogramas registrados trae consigo una dificultad asociada, pues suele ser difícil separar ambos efectos, ninguno de los cuales conocemos perfectamente, y por lo tanto trabajar con un conjunto de datos, \mathbf{d} , que sea el resultado de un único fenómeno físico.

El problema inverso en la sismología se da en situaciones muy variadas: inversión de tiempos de llegada de las fases sísmicas para estimar la localización hipocentral del terremoto y de las amplitudes de los sismogramas para estimar el tamaño del mismo; inversión de los tiempos de llegada, amplitudes y curvas de dispersión para caracterizar la estructura de la Tierra; inversión de polaridades, formas de onda y datos geodésicos para estudiar el mecanismo focal de los terremotos; inversión de los mecanismos focales para estudiar la tectónica regional; etc.

A continuación analizaremos algunos de los problemas inversos más importantes en sismología, como son la inversión de dromocrónicas y la inversión de ondas superficiales. Finalmente, expondremos la teoría de la inversa generalizada aplicable a cualquier problema inverso.

14.2. Inversión de dromocrónicas: Fórmula de Herglotz-Wiechert

Las curvas de los tiempos de llegada frente a la distancia epicentral o *dromocrónicas* dependen de la distribución de la velocidad de las ondas sísmicas en el interior de la Tierra. Por lo tanto, a partir de las dromocrónicas observadas experimentalmente podremos obtener la distribución de velocidades existente en el medio, siendo éste uno de los problemas inversos clásicos en sismología. En este caso, el problema directo siempre es resoluble de una manera más o menos complicada. Por el contrario, el problema inverso entraña una gran dificultad y puede que no tenga una solución única o exacta.

En este apartado analizaremos ondas internas que se propagan en el interior de la Tierra con una velocidad que solo depende de la profundidad o de la distancia al centro de la Tierra, suposición aproximadamente válida en el manto.

Vamos a considerar una Tierra esférica donde la velocidad de las ondas sísmicas varía de forma suave en función únicamente de la profundidad, $v = v(r)$. En este caso podemos expresar la distancia epicentral y el tiempo de llegada de las ondas sísmicas como:

$$(2) \quad \Delta = 2p \int_{r_p}^{r_0} \frac{dr}{r \sqrt{\eta^2 - p^2}}$$

$$(3) \quad t = 2 \int_{r_p}^{r_0} \frac{\eta^2 dr}{r \sqrt{\eta^2 - p^2}}$$

Donde $\eta = r/v$, r_p es la distancia desde el centro de la Tierra al punto donde el rayo alcanza su profundidad máxima o radio del punto de retorno, p es el parámetro de rayo y r_0 es el radio de la Tierra. A partir de la ley de Snell y la geometría del problema puede deducirse que:

$$(4) \quad p = \frac{r \operatorname{sen} i}{v(r)} = \frac{r_p}{v(r_p)}$$

Así como la ecuación conocida como ley de Benndorf:

$$(5) \quad p = \frac{dt(\Delta)}{d\Delta}$$

Si conocemos la curva dromocrona $t(\Delta)$ y esta es continua, podremos calcular su derivada para cada punto y, por medio de la expresión (5), hallar una curva $p(\Delta)$.

Hay que tener en cuenta que este proceso implica un suavizado y promediado de los datos, los cuales nunca dan lugar a una curva dromocrónica derivable en todos sus puntos debido a errores de medida y las heterogeneidades laterales presentes en un medio real. Los criterios utilizados para ajustar a una curva $t(\Delta)$ los valores individuales afectará a la forma de la curva $p(\Delta)$ y finalmente a los valores del modelo $v(r)$ que queremos determinar.

Para obtener la expresión matemática con la cual podamos obtener los valores de velocidad del modelo a partir de la curva $p(\Delta)$ procedemos del siguiente modo:

En primer lugar realizamos un cambio de la variable de integración en la expresión que define la distancia epicentral (2):

$$(6) \quad \Delta = 2p \int_{\eta_p=p}^{\eta_0} \frac{1}{r \sqrt{\eta^2 - p^2}} \left(\frac{dr}{d\eta} \right) d\eta$$

Donde $\eta_p = r_p/v(r_p) = p$ y $\eta_0 = r_0/v(r_0)$.

La expresión (6) representa la integral a lo largo de un único rayo con parámetro p . Para resolverla aplicamos a esta ecuación el operador integral¹:

$$(7) \quad \int_{p=\eta_1}^{p=\eta_0} \frac{dp}{\sqrt{p^2 - \eta_1^2}}$$

Obteniendo:

$$(8) \quad \int_{\eta_1}^{\eta_0} \frac{\Delta dp}{\sqrt{p^2 - \eta_1^2}} = \int_{\eta_1}^{\eta_0} dp \int_p^{\eta_0} \frac{2p}{r \sqrt{\eta^2 - p^2} \sqrt{p^2 - \eta_1^2}} \left(\frac{dr}{d\eta} \right) d\eta$$

Donde $\eta_1 = r_1/v(r_1) = p_1$ es el parámetro de rayo para un rayo que alcanza su máxima profundidad en r_1 . Por lo tanto este operador integra sobre todos los rayos comprendidos entre aquel donde $p = \eta_0$, y por tanto $\Delta = 0$, y el que tiene como parámetro p_1 .

La integral del miembro izquierdo podemos resolverla por medio de la integración por partes²:

¹ Por similitud con el problema conocido como problema de Abel.

$$(9) \int_{\eta_1}^{\eta_0} \frac{\Delta dp}{\sqrt{p^2 - \eta_1^2}} = \Delta \cosh^{-1} \left(\frac{p}{\eta_1} \right) \Big|_{\eta_1}^{\eta_0} - \int_{\eta_1}^{\eta_0} \cosh^{-1} \left(\frac{p}{\eta_1} \right) \left(\frac{d\Delta}{dp} \right) dp$$

El primer término de esta integral es igual a cero puesto que $\Delta(\eta_0) = 0$ por definición y dado que, $\cosh^{-1}(1) = 0$. El segundo término de (9), incluido el signo, puede escribirse poniendo los límites de integración en función de Δ como:

$$(10) \int_0^{\Delta_1} \cosh^{-1} \left(\frac{p}{\eta_1} \right) d\Delta$$

Volviendo de nuevo a la ecuación (8), realizamos un cambio en el orden de integración en su segundo miembro:

$$(11) \int_{\eta_1}^{\eta_0} dp \int_p^{\eta_0} \frac{2p}{r \sqrt{\eta^2 - p^2} \sqrt{p^2 - \eta_1^2}} \left(\frac{dr}{d\eta} \right) d\eta = \int_{\eta_1}^{\eta_0} \left(\frac{dr}{d\eta} \right) \frac{d\eta}{r} \int_{\eta_1}^{\eta} \frac{2p}{\sqrt{\eta^2 - p^2} \sqrt{p^2 - \eta_1^2}} dp$$

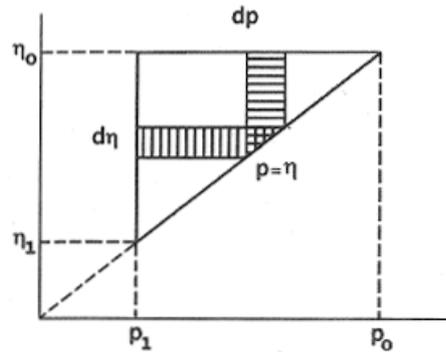


Figura 1. Superficie de integración en el plano (p, η) (Udías, 1999).

Los límites de integración de las integrales del segundo miembro de (11), pueden deducirse analizando la *Figura 1*, de modo que se integre sobre la misma superficie en los dos miembros de (11). En el primer miembro primero integramos sobre η desde $\eta = p$ hasta $\eta = \eta_0$, es decir una línea vertical que va desde algún punto de la recta $p = \eta$ hasta la ordenada η_0 y luego integramos sobre p desde p_1 hasta p_0 , es decir que extendemos la línea anterior horizontalmente desde p_1 hasta p_0 recorriendo el área del triángulo con vértices (p_0, η_0) , (p_1, η_0) y (p_1, η_1) . En el segundo miembro podemos cubrir la misma área de integración integrando sobre p desde $\eta = \eta_1$ hasta $p = \eta$ (líneas horizontales desde p_1 hasta la recta $p = \eta$) y luego integrando sobre η desde η_1 hasta η_0 .

2 Tener en cuenta que $\int \frac{dx}{\sqrt{x^2 - a^2}} = \cosh^{-1} \left(\frac{x}{a} \right), (a > 0)$

La integral respecto a p del segundo miembro de (11) puede resolverse realizando el cambio de variable $p^2 = \eta_1^2 \cos^2 \theta + \eta^2 \sin^2 \theta$, ($2pdp = 2(\eta^2 - \eta_1^2) \cos \theta \sin \theta$) obteniendo:

$$(12) \int_{\eta_1}^{\eta} \frac{2p}{\sqrt{\eta^2 - p^2} \sqrt{p^2 - \eta_1^2}} dp = \int_0^{\pi/2} 2d\theta = \pi$$

Por lo tanto el segundo miembro de (11) queda:

$$(13) \int_{\eta_1}^{\eta_0} \left(\frac{dr}{d\eta} \right) \frac{d\eta}{r} \int_{\eta_1}^{\eta} \frac{2p}{\sqrt{\eta^2 - p^2} \sqrt{p^2 - \eta_1^2}} dp = \pi \int_{\eta_1}^{\eta_0} \left(\frac{dr}{d\eta} \right) \frac{d\eta}{r} = \pi \int_{r_1}^{r_0} \frac{dr}{r} = \pi \ln \left(\frac{r_0}{r_1} \right)$$

Y volviendo por fin a la expresión (8) y teniendo en cuenta (10) y (13), tenemos que:

$$(14) \ln \left(\frac{r_0}{r_1} \right) = \frac{1}{\pi} \int_0^{\Delta_1} \cosh^{-1} \left(\frac{p}{\eta_1} \right) d\Delta$$

Ecuación conocida como la *fórmula de Herglotz-Wiechert* para un medio esférico. Utilizando esta expresión podemos calcular la distribución de velocidades con la profundidad a partir de las curvas dromocrónicas. Para ello se siguen los siguientes pasos:

1. Obtención de la curva $p(\Delta)$ a partir de las curvas dromocrónicas y la expresión (5).
2. Para un Δ_1 cualquiera calcular $\eta_1 = p(\Delta_1)$ mediante la curva obtenida anteriormente.
3. Resolver la integral (14) (normalmente de forma numérica) utilizando la función $p(\Delta)$ obtenida en el primer paso y los valores de η_1 y Δ_1 hallando así el valor de r_1 (profundidad máxima que alcanza el rayo con parámetro η_1 y que llega a una distancia epicentral Δ_1).
4. Cálculo de $v(r_1)$ por medio de $\eta_1 = r_1/v(r_1)$.
5. Repetir el proceso para distintos Δ_n desde el paso 2, obteniendo la curva $v = v(r)$.

El método es válido siempre que $dt/d\Delta$ sea una función continua y monótonamente decreciente con Δ ($r/v(r)$ aumenta de forma monótona con r). Lo que limita su aplicación a distribuciones normales de velocidad. Es decir, que el método no puede ser aplicado directamente si hay discontinuidades en la velocidad, zonas de crecimiento brusco de la velocidad o capas de baja velocidad. Sin embargo existen modificaciones del método que permiten su aplicación a ese tipo de distribuciones más o menos anómalas.

Este procedimiento fue usado extensivamente en el desarrollo de los primeros modelos de velocidad de las ondas P y S para la estructura profunda de la Tierra, donde pueden suponerse las condiciones anteriormente expuestas.

14.2.1. Tomografía sísmica

La Tierra puede representarse aproximadamente por medio de un modelo en una dimensión, donde la velocidad de las ondas sísmicas varía con la profundidad. De este modo tan solo cometemos errores en los tiempos de llegada de ondas P telesísmicas menores del 1%. Sin

embargo se sabe que existe cierta heterogeneidad lateral, que aunque no supera el 10 % para la velocidad de cizalla (parámetro elástico con mayor variabilidad), tiene gran importancia para explicar los fenómenos dinámicos que ocurren en el interior de la Tierra. Para la creación de modelos de velocidad en dos y tres dimensiones un método muy utilizado es la *tomografía sísmica*. Este método está basado en el estudio de la diferencia de los tiempos de llegada observados respecto a los esperados para un modelo inicial (residuos en los tiempos de llegada), los cuales se suponen causados por estructuras anómalas de velocidad con respecto a dicho modelo inicial. Como modelo inicial puede tomarse en primer lugar un modelo en una dimensión obtenido por la fórmula de Herglotz-Wiechert. A continuación se describe el método brevemente:

Una fase en particular está caracterizada por un tiempo de llegada, T , dado por una integral a lo largo de su trayectoria:

$$(15) T = \int_s \frac{ds}{v(s)} = \int_s u(s) ds$$

donde $u(s)$ es la *lentitud*, igual al inverso de la velocidad.

Normalmente se suele subdividir el medio en regiones donde la velocidad puede considerarse constante, y a cada una de estas regiones se les atribuirá perturbaciones en la lentitud del modelo inicial de modo que los tiempos de llegada teóricos se ajusten mejor a los observados. Es decir, la integral a través de todas las perturbaciones en la lentitud debe ser igual a los residuos en los tiempos de llegada.

$$(16) \int_s \Delta u(s) ds = \Delta T$$

siendo $\Delta u(s)$ la incógnita a determinar.

Si el medio se subdivide en p bloques y un determinado rayo cruza q bloques ($q < p$), determinando la longitud recorrida l_j en el bloque j y discretizando la fórmula anterior, tenemos:

$$(17) \Delta T = \sum_j^q l_j \Delta u_j$$

Pero en general debemos considerar varias trayectorias, de modo que cada uno de los bloques será cruzado por más de un rayo. Para N pares fuente-estación, cada uno de ellos con un residuo ΔT_i , tendremos el sistema de ecuaciones:

$$(18) \Delta T_i = \sum_j l_{ij} \Delta u_j$$

donde i toma valores de 1 a N (número de ecuaciones) y j de 1 a p (número de incógnitas), ver *Figura 2*. En la práctica para poder resolver este sistema debe estar sobredeterminado, por lo que $N > p$. La expresión (18) forma un sistema de ecuaciones lineales que puede ser resuelto por medio de la inversa generalizada como veremos más adelante.

Una vez conocidas las perturbaciones en lentitud para cada bloque Δu_j y a partir del modelo inicial, podremos deducir fácilmente la velocidad de cada uno de los bloques considerados.

Dado que la localización de los terremotos considerados depende de la estructura de velocidades podemos localizar los terremotos por medio del modelo inicial, resolver el problema de la estructura y proseguir de forma iterativa resolviendo los problemas de localización y estructura o formular el problema de modo que se resuelvan simultáneamente ambos aspectos. Sin embargo, cuanto mayor sea el número de parámetros libres en nuestro problema mayores serán los requerimientos en cuanto a cantidad y calidad de los datos, disposición geométrica de los hipocentros y estaciones, modelo inicial, etc.

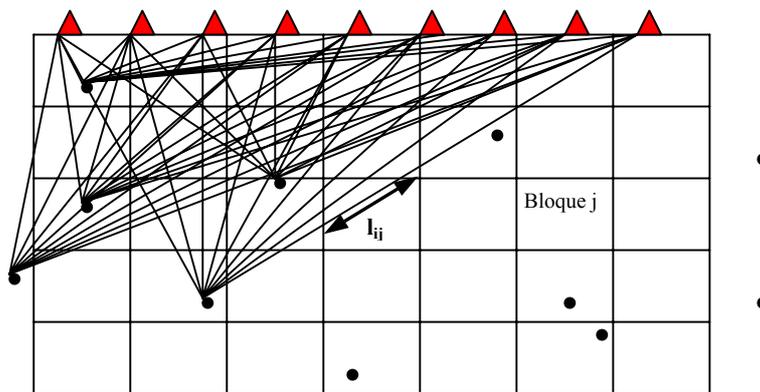


Figura 2. Representación de algunas trayectorias hipocentro-estación junto con la disposición de los bloques.

14.3. Inversión de ondas superficiales

La mayor parte de la energía de grandes periodos ($T > 20$) generada por los terremotos y registrada a distancias telesísmicas corresponde a las ondas superficiales. Éstas pueden observarse en los sismogramas correspondientes a terremotos lejanos y poco profundos como trenes de ondas dispersadas con grandes amplitudes.

En sismos de tamaño moderado, solo las ondas superficiales directas que viajan a lo largo del círculo máximo entre el epicentro y la estación tienen una relación señal-ruido significativa. En cambio para terremotos de magnitud entorno a 7 o mayor se registran durante horas paquetes de ondas superficiales que pueden dar varias vueltas a la Tierra en la dirección de dicho círculo máximo, tanto en sentido directo como inverso.

Los registros de ondas superficiales se han utilizado y siguen utilizándose ampliamente para obtener información tanto de las características de la fuente sísmica como de la estructura de la Tierra a lo largo del camino recorrido. En concreto sus velocidades de fase, velocidades de grupo y atenuación de sus amplitudes han sido de gran utilidad para un mejor conocimiento de la estructura de la corteza y el manto superior en muchas regiones de la Tierra. Por otra parte el espectro de la fuente sísmica deducido a partir de estas ondas, la modelización de su forma de

onda y las características de su patrón de radiación son utilizados para la determinación del tensor momento sísmico y la profundidad focal de grandes y moderados terremotos.

Precisamente la separación de estos dos factores (estructura y fuente sísmica), implícitos en los registros obtenidos, es un importante reto al que se enfrenta el sismólogo. Una forma de eliminar los efectos de la fuente consiste en considerar la propagación entre dos o más estaciones alineadas con el epicentro dentro del mismo círculo máximo, o para periodos altos, la observación consecutiva en la misma estación de las ondas que circundan la Tierra. En otras ocasiones, se supone que se conocen los efectos de la propagación y estos son suprimidos de las amplitudes de los registros, para así poder estimar el mecanismo de la fuente sísmica.

A partir de ahora nos centraremos en la inversión de ondas superficiales para conocer la estructura.

Las ondas superficiales viajan a lo largo de la superficie de la Tierra y su penetración depende de su longitud de onda. En periodos menores de 60 s están afectadas por la corteza y en periodos entre 60 y 300 s principalmente por el manto. Por otro lado una de las características más importantes de este tipo de ondas es que, en general, su velocidad varía con la frecuencia. El hecho de que la velocidad de las ondas superficiales varíe en función de la profundidad muestreada por cada periodo hace que estas ondas sean muy útiles para estudiar la estructura de la Tierra.

En concreto la forma de las *curvas de dispersión* de las velocidades de fase y de grupo de las ondas Rayleigh y Love dependen de las características del medio (p. ej: en un medio estratificado dependen del grosor de las capas y de las velocidades de las ondas internas en estas capas). Teniendo esto en cuenta la forma de proceder para conocer la estructura consiste en determinar las curvas de dispersión a partir de un sismograma. Una vez que contamos con una curva de dispersión observada podemos compararla con curvas teóricas de distintos modelos de estructura, con el fin de determinar cuál se aproxima más a la realidad o bien proceder a la inversión directa de esta curva experimental.

Por otro lado también existen métodos que trabajan en el dominio del tiempo, en los cuales se calculan sismogramas sintéticos basados en un modelo de referencia y este se modifica hasta que se adapta a los datos observados. En este caso la inversión se produce en un solo paso dando lugar a un modelo tridimensional de la estructura directamente a partir de los sismogramas.

Para determinar la estructura por medio de las ondas superficiales pueden utilizarse tanto las curvas de dispersión de la velocidad de grupo como las de la velocidad de fase, aunque las velocidades de grupo son más fáciles de medir puesto que son las velocidades a las cuales viaja

un tren de ondas visible en un sismograma. Los procedimientos más sencillos para determinar estas curvas son:

1. Velocidad de grupo.

Teniendo en cuenta el principio de la fase estacionaria sabemos que para cada distancia epicentro-estación, x , y tiempo, t , la energía está contenida en una oscilación con frecuencia ω_0 , o periodo T_0 , que llegan a la velocidad correspondiente a la velocidad de grupo para ese periodo,

$$\frac{x}{t} = U(T_0).$$

Por lo tanto en los métodos más sencillos la velocidad de grupo se obtiene midiendo los tiempos de llegada de los picos y valles del tren de ondas; a partir de estas medidas se calcula el periodo asociado a cada tiempo (tiempo entre dos picos sucesivos o el doble de tiempo entre un pico y el siguiente valle); por medio de la ecuación $\frac{x}{t} = U(T_0)$ la velocidad de grupo correspondiente a cada periodo y de aquí las curvas de dispersión.

En la *Figura 3* podemos ver la curva de dispersión obtenida a partir de las medidas realizadas en un sismograma junto con el ajuste de estas medidas a una curva de dispersión teórica.

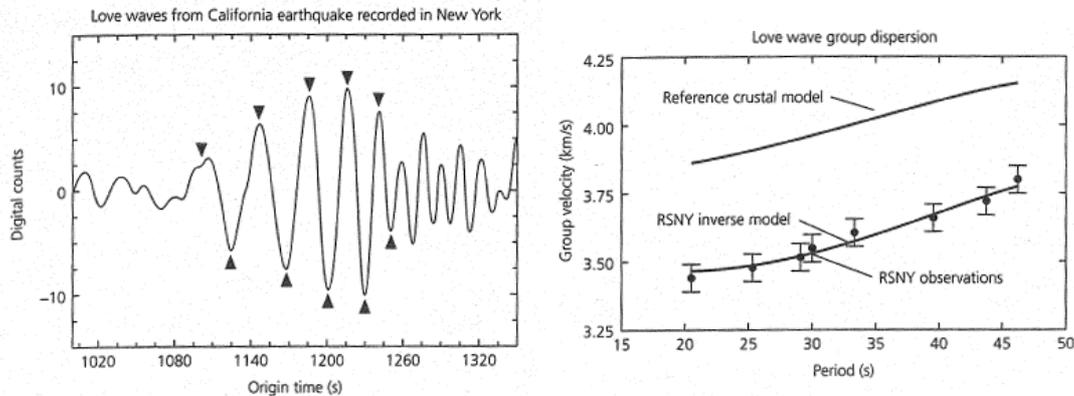


Figura 3. Ondas Love debido a un terremoto cercano a la costa de California, registrado en la componente transversal de la estación RSNY de Nueva York, con $x = 4200$ km. La curva de dispersión teórica superior corresponde a un modelo formado por una capa sobre un semiespacio con valores: $\beta = 3.9$ km/s, $\beta = 4.6$ km/s, $H = 40$ km. La curva de dispersión que se ajusta a los datos corresponde a la misma estructura pero con velocidades: $\beta = 3.6$ km/s, $\beta = 4.4$ km/s (Stein y Wysession, 2003).

Otro modo de determinar las velocidades de grupo correspondientes a cada periodo se basa en el filtrado paso-banda de la señal en estrechos rangos de frecuencia centrados en la frecuencia para la cual queremos calcular la velocidad de grupo (ver *Figura 4*). Siendo este procedimiento más estable que el anterior.

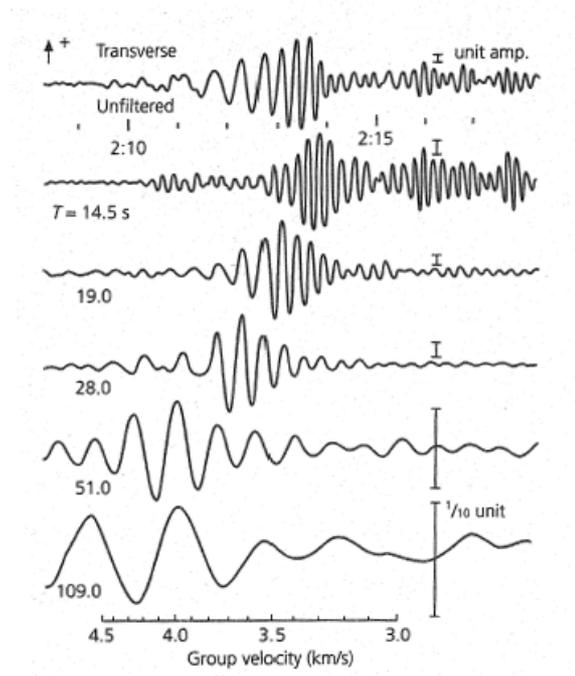


Figura 4. Ondas Love de un terremoto de Mongolia registrado en Japón. Los datos están filtrados entorno a cinco periodos, cuyo valor se muestra a la izquierda de la señal. Se observa como la energía de periodo mayor llega con velocidad de grupo mayor (Stein y Wysession, 2003).

Otra manera más sofisticada de calcular estas curvas es por medio de los *diagramas energéticos* en los que se calcula la energía que llega para cada periodo y velocidad de grupo, dado que la curva de dispersión de la velocidad de grupo coincide con el máximo de energía para cada periodo.

2. Velocidad de fase

La medida de la velocidad de fase es más complicada, pues está relacionada con ondas armónicas de una única frecuencia.

La transformada de Fourier de un sismograma determina la fase para cada frecuencia angular $\Phi(\omega)$, cuya forma es³:

$$(19) \Phi(\omega) = [\omega t - k(\omega)x] + \phi_i + 2\pi N = \left[\omega t - \frac{\omega x}{c(\omega)} \right] + \phi_i + 2\pi N$$

Donde el término ϕ_i representa la fase inicial en la fuente y la variación de la fase introducida por el sismómetro y el término $2\pi N$, refleja el número de ciclos completos.

El problema de tener que conocer el mecanismo focal para estimar ϕ_i puede evitarse usando dos estaciones alineadas con el epicentro. Si estas estaciones se sitúan en el mismo círculo máximo

³ A la expresión de $\phi(\omega)$ deben hacerse correcciones adicionales debido a los efectos de la anelasticidad y pasadas por el polo (añadir $\pi/2$ cada vez que la onda pasa por el epicentro o por sus antípodas).

junto con el epicentro a distancias x_1 y x_2 de éste, de modo que las ondas llegan en tiempos t_1 y t_2 , realizando la transformada de Fourier de ambos sismogramas, tenemos:

$$(20) \quad \begin{aligned} \Phi_1(\omega) &= \left[\omega t_1 - \frac{\omega x_1}{c(\omega)} \right] + \phi_i + 2\pi N \\ \Phi_2(\omega) &= \left[\omega t_2 - \frac{\omega x_2}{c(\omega)} \right] + \phi_i + 2\pi M \end{aligned}$$

Restando ambas fases, $\Phi_{21}(\omega) = \Phi_2(\omega) - \Phi_1(\omega)$, y despejando la velocidad de fase, llegamos a⁴:

$$(21) \quad c(\omega) = \frac{\omega(x_2 - x_1)}{\omega(t_2 - t_1) + 2(M - N)\pi - \Phi_{21}(\omega)}$$

El término $2(M - N)\pi$ se determina por prueba y error de modo que la velocidad de fase obtenida para los periodos altos converja con valores globales promedio de $c(\omega)$, los cuales son lo suficientemente estables.

De este modo es posible determinar la velocidad de fase para distintas frecuencias y así construir la curva de dispersión, a partir de la cual podremos conocer la estructura de velocidades promedio a lo largo de la trayectoria entre las dos estaciones.

Tanto las ondas Rayleigh como las Love presentan modos superiores, cada uno de los cuales tiene sus propias curvas de dispersión. Para periodos muy grandes, a partir de 200-250 s, el modo fundamental se solapa con modos superiores por lo que son necesarias técnicas de filtrado para aislar los distintos modos de propagación. Los modos superiores son útiles para estudiar zonas más profundas que las muestreadas por el modo fundamental ya que su sensibilidad es mayor en estas profundidades.

Uno de los primeros resultados del análisis de las curvas de dispersión fue la diferencia entre la corteza continental y el manto superior bajo océanos y continentes (ver Figura 5). Las trayectorias oceánicas muestran que las velocidades de grupo de ondas Rayleigh con periodos entre 20 y 100 s tienen un valor constante entorno a 4 km/s y en periodos de 10 a 20 s una caída brusca debido a la capa de agua. En cambio en las trayectorias continentales queda reflejado que la corteza tiene mayor grosor y tiene una capa superior sedimentaria con velocidades menores. A partir de periodos cercanos a unos 150 s las curvas de dispersión en océanos y continentes es similar.

⁴ Suponemos que ambos sismogramas están registrados por sismómetros con la misma respuesta.

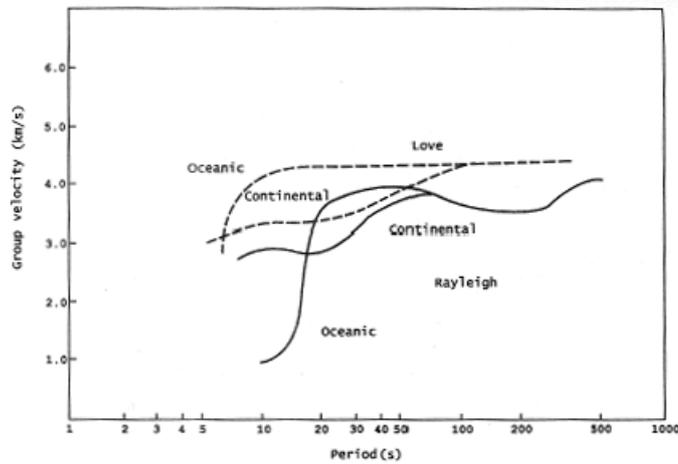


Figura 5. Curvas de dispersión de la velocidad de grupo para ondas Rayleigh y Love en trayectorias oceánicas y continentales (Udías, 1999).

Las curvas de dispersión así determinadas representan un promedio de las características a lo largo de toda la trayectoria epicentro-estación. Para poder distinguir entre distintos tramos de esta trayectoria es decir, determinar cambios laterales en la velocidad, es necesario considerar varios pares epicentro-estación y aplicar las técnicas de tomografía sísmica.

14.4. Inversa generalizada

Vamos ahora a tratar de dar una solución general al problema inverso.

Volviendo a la ecuación (1), observamos que nuestro objetivo es encontrar un operador G^{-1} , tal que al aplicarlo al vector \mathbf{d} , se obtenga \mathbf{m} como resultado.

En el caso de que la expresión (1) represente una *ecuación lineal*, podemos escribirla como:

$$(22) \mathbf{d} = \mathbf{Gm}$$

Esta ecuación relaciona una matriz de datos, \mathbf{d} , de dimensión $n \times 1$, siendo n el número de observaciones con una matriz de parámetros del modelo, \mathbf{m} , de dimensión $m \times 1$, donde m es el número de parámetros del modelo. Al ser una ecuación lineal \mathbf{G} constituye una matriz de términos constantes de dimensión $n \times m$ y \mathbf{Gm} un producto entre matrices.

Si $n = m$ podemos simplemente multiplicar ambos miembros de la ecuación (22) por la matriz inversa de \mathbf{G} , \mathbf{G}^{-1} , suponiendo que esta exista, obteniendo así:

$$(23) \mathbf{G}^{-1}\mathbf{d} = \mathbf{m}$$

Con lo que resolvemos el problema.

Sin embargo, este caso nunca se da en sismología. Los datos que estamos utilizando contienen errores de observación y por otro lado, la ecuación (22) supone que podemos predecir los datos de forma exacta. Es decir, que nuestro modelo es una fiel representación de la realidad. Por lo

tanto la expresión (22) constituye un sistema de ecuaciones inconsistentes. Para solucionar este problema se construye un sistema de ecuaciones *sobredeterminado* ($n > m$). En este caso en la ecuación (22) \mathbf{G} es una matriz de dimensión $n \times m$ con $n \neq m$, por lo que en este caso no existe \mathbf{G}^{-1} y no puede aplicarse la ecuación (23). En concreto, resolveremos la ecuación por medio del método de ajuste *lineal de mínimos cuadrados*, de modo que los parámetros del modelo que obtengamos serán aquellos que mejor ajusten el modelo a una “media” de los datos.

En este método se define el mejor ajuste como aquel en el que los *residuos* (diferencia entre datos observados y predichos) son menores. En particular intentaremos minimizar la expresión del error cuadrático medio. Esta expresión viene dada por:

$$(24) E^2 = \sum_{i=1}^n \left(d_i - \sum_{j=1}^m G_{ij} m_j \right)^2$$

Para hacer este valor mínimo, lo derivamos con respecto a los parámetros del modelo y lo igualamos a cero:

$$(25) \frac{\partial E^2}{\partial m_k} = 2E \frac{\partial E}{\partial m_k} = -2 \sum_{i=1}^n \left(d_i - \sum_{j=1}^m G_{ij} m_j \right) G_{ik} = 0$$

Con lo cual:

$$(26) \sum_{i=1}^n d_i G_{ik} = \sum_{i=1}^n \left(\sum_{j=1}^m G_{ij} m_j \right) G_{ik}$$

Y escribiendo la ecuación (26) en forma matricial:

$$(27) \mathbf{G}^T \mathbf{d} = \mathbf{G}^T \mathbf{G} \mathbf{m}$$

Obtenemos la expresión denominada como *ecuaciones normales*. En este caso, $\mathbf{G}^T \mathbf{G}$ es una matriz cuadrada de dimensión ($m \times m$) y por lo tanto tiene inversa, siempre que no sea singular⁵. Además es una matriz simétrica y por lo tanto tiene autovalores reales y sus autovectores son ortogonales. Por lo tanto podemos escribir la ecuación (27) en la forma:

$$(28) \mathbf{m} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T \mathbf{d}$$

Donde $\mathbf{G}^{-g} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T$ se conoce como *inversa generalizada* y nos proporciona la solución a la ecuación (22) siempre que $\mathbf{G}^T \mathbf{G}$ no sea singular.

Sin embargo es frecuente que ocurra que todas las filas de \mathbf{G} no sean linealmente independientes, con lo cual el determinante de $\mathbf{G}^T \mathbf{G}$ es igual a cero y por lo tanto esta matriz no tiene inversa. Para estos casos se ha desarrollado la teoría que ahora presentamos:

En el caso general $\mathbf{G}^T \mathbf{G}$ es una matriz simétrica de dimensión $m \times m$ que puede descomponerse usando su matriz de autovalores $\mathbf{\Lambda}_{m(2)}$ y de autovectores \mathbf{V} :

$$(29) \mathbf{G}^T \mathbf{G} = \mathbf{V} \mathbf{\Lambda}_{m(2)} \mathbf{V}^T$$

⁵ Las matrices singulares son aquellas cuyo determinante es igual a cero.

Donde las columnas de \mathbf{V} son cada uno de los m autovectores de $\mathbf{G}^T\mathbf{G}$ y $\mathbf{\Lambda}_{m(2)}$ es una matriz diagonal cuyos términos en la diagonal principal son los m autovalores λ_i^2 de $\mathbf{G}^T\mathbf{G}$ y el resto de los términos son nulos⁶.

Dado que los autovectores son ortonormales tenemos que:

$$(30) \mathbf{V}\mathbf{V}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}$$

Por lo tanto $\mathbf{V}^T = \mathbf{V}^{-1}$ y $[\mathbf{V}^T]^{-1} = \mathbf{V}$

Y si $\mathbf{G}^T\mathbf{G}$ tiene inversa, esta vendrá dada por:

$$(31) [\mathbf{G}^T\mathbf{G}]^{-1} = [\mathbf{V}\mathbf{\Lambda}_{m(2)}\mathbf{V}^T]^{-1} = \mathbf{V}\mathbf{\Lambda}_{m(2)}^{-1}\mathbf{V}^T$$

Donde $\mathbf{\Lambda}_{m(2)}^{-1}$ es una matriz diagonal cuyos términos en la diagonal principal son los inversos de los autovalores de $\mathbf{G}^T\mathbf{G}$, $1/\lambda_i^2$.

En el caso de que $\mathbf{G}^T\mathbf{G}$ sea singular, al menos uno de sus autovalores ha de ser cero. En ese caso utilizamos los p autovalores ($p < m$) distintos de cero y sus autovectores asociados para definir las matrices $\mathbf{\Lambda}_{p(2)}$ y \mathbf{V}_p de dimensiones $p \times p$ y $m \times p$ respectivamente:

$$(32) \mathbf{\Lambda}_{p(2)} = \begin{pmatrix} \lambda_1^2 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \lambda_p^2 \end{pmatrix}$$

$$(33) \mathbf{V}_p = \begin{pmatrix} v_1^{(1)} & \dots & v_1^{(p)} \\ \vdots & \ddots & \vdots \\ v_m^{(1)} & \dots & v_m^{(p)} \end{pmatrix}$$

De manera similar la matriz $n \times n$ $\mathbf{G}\mathbf{G}^T$ puede descomponerse como:

$$(34) \mathbf{G}\mathbf{G}^T = \mathbf{U}\mathbf{\Lambda}_{n(2)}\mathbf{U}^T$$

$\mathbf{G}\mathbf{G}^T$ tiene los mismos p autovalores distintos de cero que $\mathbf{G}^T\mathbf{G}$ por lo que de forma similar definimos la matriz de dimensión $n \times p$, \mathbf{U}_p formada por los p autovectores asociados a los autovalores distintos de cero.

Es posible demostrar que la matriz \mathbf{G} puede descomponerse como:

$$(35) \mathbf{G} = \mathbf{U}\mathbf{\Lambda} \mathbf{V}^T = \mathbf{U}_p\mathbf{\Lambda}_p\mathbf{V}_p^T$$

Donde:

$$(36) \mathbf{\Lambda}_p = \begin{pmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \lambda_p \end{pmatrix}$$

⁶ Por conveniencia expresamos los autovalores de GTG como $\lambda_{12}, \lambda_{22}, \dots, \lambda_{m2}$

De modo que podemos definir la inversa generalizada como:

$$(37) \mathbf{G}_p^{-g} = \mathbf{V}_p \mathbf{\Lambda}_p^{-1} \mathbf{U}_p^T$$

Siendo $\mathbf{\Lambda}_p^{-1}$:

$$(38) \mathbf{\Lambda}_p^{-1} = \begin{pmatrix} 1/\lambda_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & 1/\lambda_p \end{pmatrix}$$

La cual nos da una solución a la ecuación (22) y nos indica algunas de sus limitaciones. En concreto si $p < m$ la solución del problema inverso no es única y además la solución encontrada no es exacta.

La inversa generalizada \mathbf{G}_p^{-g} coincide con \mathbf{G}^{-g} cuando $\mathbf{G}^T \mathbf{G}$ no es singular.

La descomposición llevada a cabo anteriormente por medio de las matrices de autovalores y autovectores (descomposición en valores singulares) es útil además para comprender el significado de la inversa.

En primer lugar podemos definir la relación entre el modelo derivado de las observaciones, $\mathbf{m}_p = \mathbf{G}_p^{-g} \mathbf{d}$, y el modelo real, \mathbf{m} , teniendo en cuenta la ecuación (22):

$$(39) \mathbf{m}_p = \mathbf{G}_p^{-g} \mathbf{G} \mathbf{m} = \mathbf{V}_p \mathbf{\Lambda}_p^{-1} \mathbf{U}_p^T \mathbf{U}_p \mathbf{\Lambda}_p \mathbf{V}_p^T \mathbf{m} = \mathbf{V}_p \mathbf{V}_p^T \mathbf{m}$$

Puesto que $\mathbf{U}_p^T \mathbf{U}_p = \mathbf{I}$ (las columnas de \mathbf{U}_p son vectores ortonormales), pero si $p \neq m$ $\mathbf{V}_p \mathbf{V}_p^T \neq \mathbf{I}$ (las filas de \mathbf{V}_p ya no son ortonormales).

La matriz $\mathbf{R} = \mathbf{V}_p \mathbf{V}_p^T$ se denomina *matriz de resolución* e indica como cada parámetro del modelo real se difumina entre los distintos parámetros del modelo obtenido, expresando con que exactitud puede ser resuelto el problema inverso para un conjunto de observaciones sin error. Idealmente esta matriz es la matriz identidad, lo cual ocurre cuando la matriz $\mathbf{G}^T \mathbf{G}$ no es singular y por lo tanto $p = m$.

Dado que las observaciones, además, contienen errores, se define la *matriz de covarianza C*:

$$(40) \mathbf{C} = \mathbf{V}_p \mathbf{\Lambda}_{p(2)}^{-1} \mathbf{V}_p^T = [\mathbf{G}^T \mathbf{G}]^{-1}$$

Esta matriz interviene en el cálculo de la inversa generalizada y expresa la variabilidad de la solución. Los autovalores de $\mathbf{G}^T \mathbf{G}$ aparecen en el denominador y por lo tanto autovalores pequeños tendrán un gran efecto en la solución produciendo gran variabilidad en ésta.

En los problemas reales lo normal es encontrar autovalores pequeños más que autovalores nulos. Una opción consiste en suponer que estos autovalores son iguales a cero, aumentando así la estabilidad de la solución, sin embargo esto trae consigo una disminución de la resolución.

Las varianzas y covarianzas de la solución, \mathbf{m} , son los elementos de $\sigma^2\mathbf{C}$, donde σ^2 es la varianza de una observación.

Por el momento hemos considerado que nuestro problema puede definirse por la expresión (22), correspondiente a una particularización de (1) para el caso *lineal*. Sin embargo en muchas ocasiones las ecuaciones que definen nuestro modelo no son lineales, como ocurre en el caso de la localización hipocentral, en cuyo caso debemos *linealizar* el problema para así poder utilizar la teoría de ajuste lineal por mínimos cuadrados expuesta anteriormente.

El procedimiento común para llevar esto a cabo consiste en proponer una solución inicial \mathbf{m}^0 del que suponemos que es parecida a la solución real. Una vez hecho esto, resolvemos el problema directo para calcular los datos que se obtendrían, \mathbf{d}^0 , si los parámetros del modelo fueran \mathbf{m}^0 .

$$(41) \mathbf{d}^0 = \mathbf{G}(\mathbf{m}^0)$$

A continuación expandimos la función $G(m)$ por medio de un desarrollo de Taylor en torno a \mathbf{m}^0 , conservando tan solo los términos lineales.

$$(42) d_i = G(m)_i \approx G(m^0)_i + \sum_j^m \left(\frac{\partial G(m)_i}{\partial m_j} \Big|_{m_j^0} \Delta m_j^0 \right)$$

Con $\Delta m_j^0 = m_j - m_j^0$.

Teniendo todo esto en cuenta, la diferencia entre los datos reales, \mathbf{d} , y los datos calculados mediante nuestra suposición inicial, \mathbf{d}^0 , será:

$$(43) \Delta d_i^0 = d_i - d_i^0 \approx G(m^0)_i + \sum_j^m \left(\frac{\partial G(m)_i}{\partial m_j} \Big|_{m_j^0} \Delta m_j^0 \right) - G(m^0)_i = \sum_j^m \left(\frac{\partial G(m)_i}{\partial m_j} \Big|_{m_j^0} \Delta m_j^0 \right)$$

Si definimos: $\Delta \mathbf{d}^0 = \mathbf{d} - \mathbf{d}^0$, $G_{ij}^0 = \frac{\partial G(m)_i}{\partial m_j} \Big|_{m_j^0}$ y $\Delta \mathbf{m}^0 = \mathbf{m}^1 - \mathbf{m}^0$ (donde \mathbf{m}^1 es la matriz de

incógnitas) tenemos la ecuación:

$$(44) \Delta \mathbf{d}^0 = \mathbf{G}^0 \Delta \mathbf{m}^0$$

Ecuación lineal que relaciona la variación en las observaciones debido a una variación en los parámetros del modelo (en el entorno de \mathbf{m}^0). Esta ecuación puede resolverse por medio de la inversa generalizada, como explicamos anteriormente, obteniendo $\Delta \mathbf{m}^0$ a partir de donde podemos calcular \mathbf{m}^1 .

El proceso se repite iterativamente: a partir de \mathbf{m}^1 calculamos \mathbf{d}^1 , $\Delta \mathbf{d}^1$, \mathbf{G}^1 , planteamos la ecuación $\Delta \mathbf{d}^1 = \mathbf{G}^1 \Delta \mathbf{m}^1$, la resolvemos por mínimos cuadrados para obtener \mathbf{m}^2 y así sucesivamente hasta que la solución \mathbf{m}^i del problema inverso converja en un valor dado.

Bibliografía

- [1] Lay, T. y Wallace, T. C. Modern Global Seismology. Academic Press, 1995. pág. 221-232, 238-240, 413-417.
- [2] Stein, S. y Wysession, M. An Introduction to Seismology, Earthquakes, and Earth Structure. Blackwell Publishing, 2003. pág. 96-99, 415-442.
- [3] Udías, A. Principles of Seismology. Cambridge University Press, 1999. pág. 130-132, 149-150, 224-238, 371-373.

Tema 15. Estructura de la corteza terrestre. Perfiles sísmicos. Interpretación. Sismogramas sintéticos. Método gravimétrico y magnético. Anomalías locales y regionales. Estructuras tipo de corteza.

15.1. Estructura de la corteza terrestre

El conocimiento de las propiedades elásticas de los materiales que forman la corteza terrestre es uno de los objetivos prioritarios de la investigación en las ciencias de la Tierra. La razón reside en que no es únicamente una meta científica sino que también es una necesidad ya que el conocimiento de la estructura de los materiales terrestres permite ahondar en otros problemas más prácticos, como pueden ser la de búsqueda de recursos energéticos.

Los datos sismológicos son de vital importancia para el conocimiento de esta estructura terrestre que generalmente será expresada en términos de las velocidades de las ondas P y S. Esto es, a partir del estudio de la propagación de estos tipos de ondas y a veces también de las ondas superficiales se trata de deducir una estructura de velocidades que sea compatible con esta propagación. Los resultados serán más o menos precisos dependiendo de la instrumentación utilizada y del método empleado para deducir estos valores. Estos serán velocidades de propagación y espesores en los que estas velocidades son aplicables.

Existen dos formas de estudiar la propagación de estas ondas internas: refracción y reflexión. La diferencia en la utilización de una u otra dependerá de que tipo de ondas estemos interesados en estudiar, siendo en general utilizadas las técnicas de refracción para la obtención de modelos promedio mientras que las de reflexión se centran en una determinación de gran precisión tanto en velocidad como en profundidad de las discontinuidades.

Existen en la Geofísica otras disciplinas, relacionadas con otros campos como es la Gravimetría en cuya formulación interviene la densidad de los materiales y la distancia a la que estos se encuentran. Es por tanto una alternativa a los métodos sísmicos puesto que la densidad es uno de los parámetros elásticos que pueden compararse con los resultados obtenidos por los métodos sísmicos. Esta metodología usando la Gravimetría no es en general alternativa de la anterior sismológica sino que son complementarias. Esto permite que si con uno de los métodos no existe capacidad de resolución del problema que vamos buscando, el otro puede resolver esa indeterminación ya que estos métodos están basados en diferentes propiedades de la materia; en uno es su capacidad de deformarse al paso de las ondas y en el otro es la atracción que ejerce un

elemento de materia sobre un punto. En ambos casos interviene la densidad del material, pero utilizando distintas propiedades de la Física.

15.2. Perfiles sísmicos. Interpretación. Sismogramas sintéticos

Cuando estamos interesados en estudiar la propagación de ondas es necesario tener una fuente generadora de ellas. En general, los terremotos son una fuente inagotable de energía, pero tienen el inconveniente de que no conocemos ni el instante de su producción ni el emplazamiento exacto en el que se va a producir. Por tanto, en general se considerará como fuente de energía productora de las ondas sísmicas explosivos u otros sistemas artificiales de liberar energía como son: cañones de aire, vibradores sobre el terreno, etc. En cuanto a los instrumentos de recepción serán sismógrafos situados convenientemente según la técnica a utilizar para registrar las ondas que se consideren más apropiadas. Estos sismógrafos han de tener un sistema de tiempo que permita su comparación con la del resto de ellos y con el que permite determinar el instante de producción de la energía liberada por la explosión. También han de poseer una calibración que permita el conocimiento de la respuesta en amplitudes del registro de las ondas. Con esta instrumentación y fuente de energía vamos a intentar resolver el problema directo utilizando ondas refractadas y reflejadas (Método de Refracción) o bien utilizando exclusivamente ondas reflejadas (Método de Reflexión). En ambos casos el resolver el problema directo, quiere decir que suponemos conocida una cierta estructura y queremos determinar que propiedades se pueden obtener de los parámetros como tiempo de recorrido de una onda o la amplitud de ella, en función de las propiedades elásticas del medio. En lo que sigue y para no repetir las propiedades de propagación de las ondas estudiadas en los temas de propagación, nos fijaremos únicamente en los resultados finales que tienen que ver con el objetivo que perseguimos.

Método de Refracción. Modelo de capas planas y velocidad constante. Para distancias entre el foco y las estaciones menores de 500 km, los rayos sísmicos penetran sólo en la corteza y la parte superior del manto. Los principales tipos de ondas que se estudian son: la onda directa, la onda refractada en cada capa y las reflejadas también en cada capa. Para eliminar las posibles heterogeneidades del material terrestre se fuerza al estudio de la propagación sobre un plano vertical, por lo que las estaciones de registro han de estar alineadas con el foco emisor, constituyendo lo que se denomina un *Perfil Sísmico*.

Los intervalos entre estaciones han de seleccionarse en función de donde se espere van a aparecer los distintos tipos de ondas a considerar en el estudio. De la misma forma la longitud total del perfil dependerá de hasta que profundidad estamos interesados en investigar. Si lo que se pretende es llegar a la discontinuidad de Moho, la longitud será tal que se asegure una buena

propagación de la onda refractada en esa discontinuidad así como la reflejada en dicha capa. Esto es deberíamos establecer una longitud entre 150-200 km. El espaciamiento también será decidido en función del tramo del perfil en el que estemos. Así, en los primeros kilómetros donde coexisten la onda que viaja por los sedimentos y la onda correspondiente a la primera capa la distancia será de entre 1-2 km, posteriormente puede aumentarse el espaciamiento hasta los 80-90 km donde se vuelve a necesitar una mayor discriminación par distinguir las ondas reflejada y refractadas que se identifican en dicho intervalo.

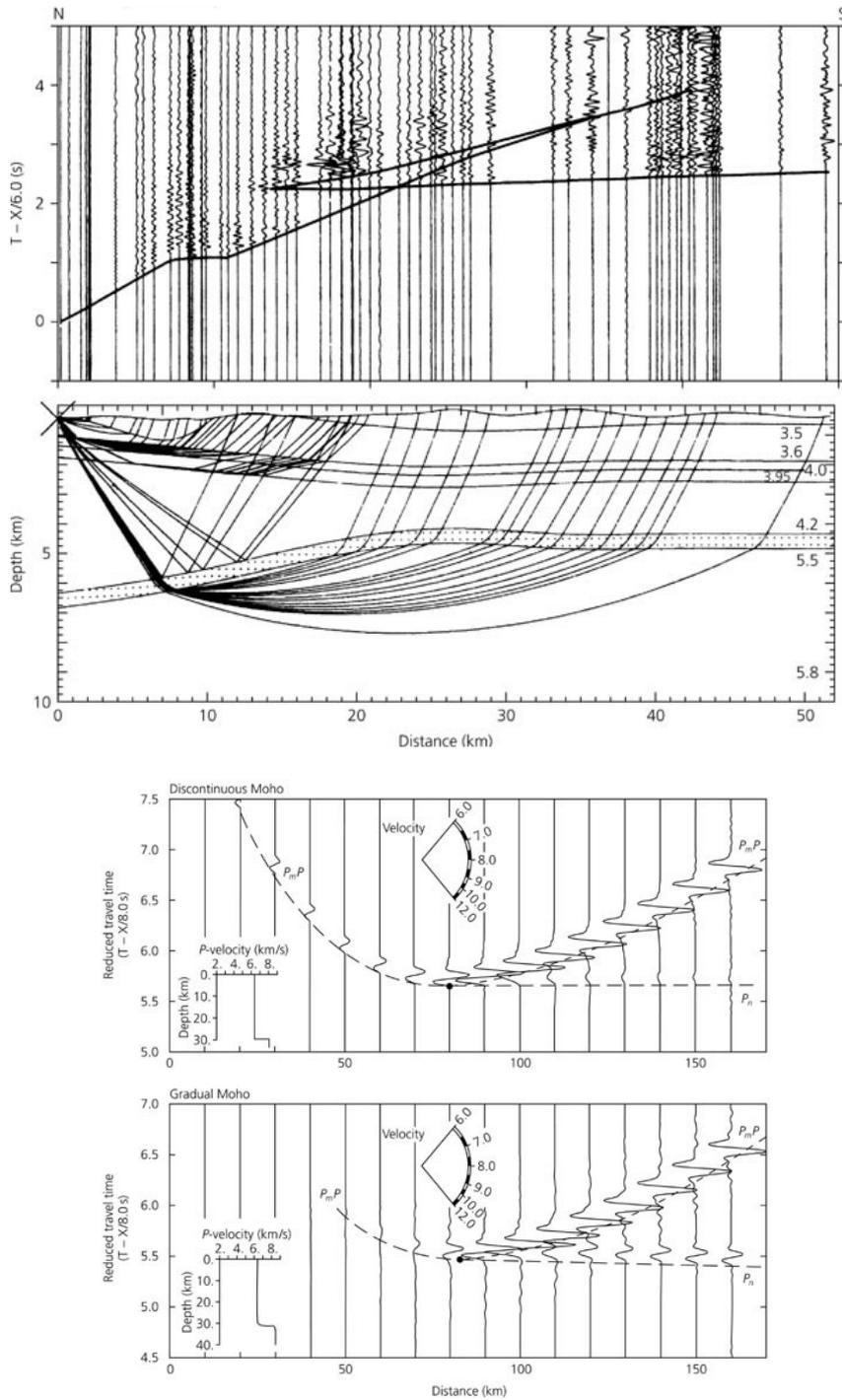


Figura 1.

La representación de los registros del perfil se realizan en el sistema de ejes (x, t) que se denomina dromocrónica o dromocrona. A veces y con el fin de resaltar los cambios de pendiente o lo que es lo mismo los cambios de velocidad se utilizan la representación llamada dromocrónica reducida. Consiste en que para distancia x , el tiempo correspondiente t viene reducido en un tiempo x/v siendo v la velocidad de reducción. Se consigue además separar pendientes positivas de negativas a partir de la correspondiente a la velocidad de reducción v . Así si la velocidad de reducción es $v=6$ km/s, las velocidades menores serán con pendiente negativa y las superiores pendientes positivas.

Por último en los modelos de refracción desarrollados hasta ahora las capas eran paralelas. Sin embargo a veces esto no ocurre con lo que la interpretación sería errónea. Para llevar a cabo la comprobación de si la capa de interés tiene una cierta inclinación es necesario realizar otro perfil en sentido opuesto llamado *perfil inverso*. Es decir donde está el final del perfil directo se realiza la explosión y el final del perfil inverso estará en el inicio del perfil directo. Con estos dos perfiles se obtiene el ángulo de buzamiento de la capa. (Figura 2)

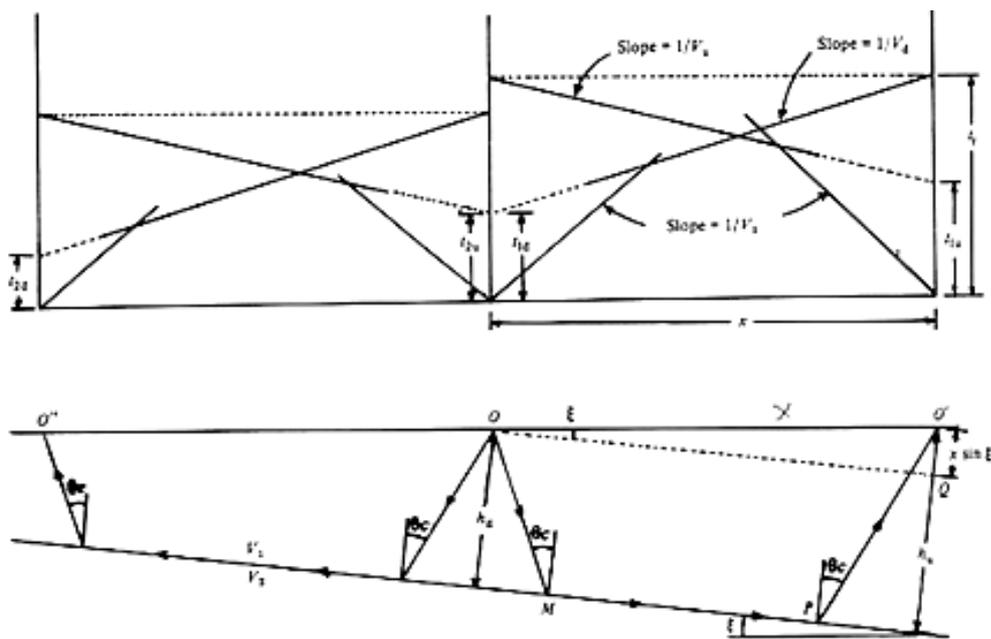


Figura 2. Dromocronas directa e inversa para una capa inclinada.

La expresión del tiempo de recorrido en sentido directo será:

$$t_d = \frac{x \cos \theta \operatorname{sen} i_c}{v_0} + \frac{(2h_d + x \operatorname{sen} \theta)(1 - \operatorname{sen}^2 i_c)}{v_0 \cos i_c}$$

$$= \frac{x \operatorname{sen}(i_c + \theta)}{v_0} + \frac{2h_d \cos i_c}{v_0} = \frac{x}{v_d} + \tau_d$$

Y en sentido inverso:

$$t_i = \frac{x \operatorname{sen}(i_c - \theta)}{v_0} + \frac{2h_i \cos i_c}{v_0} = \frac{x}{v_i} + \tau_i$$

Así, las velocidades observadas tanto directa como inversa son aparentes, siendo la correspondiente a la onda que encuentra la capa en la dirección hacia arriba mayor que si no hubiese buzamiento y si encuentra la capa con dirección hacia abajo la velocidad será menor. Con el conocimiento de ambas velocidades se puede determinar el buzamiento de la capa.

Hasta ahora hemos considerado únicamente los tiempos de recorrido de las distintas ondas. Esto hace que no podamos distinguir entre muy variadas estructuras de velocidad que tienen la misma curva dromocrónica. Para intentar resolver este problema podemos trabajar, de forma adicional con las amplitudes de las ondas. Estas se pueden calcular por algún método como el de reflectividad que evita las limitaciones de los análisis de ondas y rayos planos. El resultado de estos registros teóricos se denomina de sismogramas sintéticos y además de dar información de los tiempos de recorrido, suministran información sobre las amplitudes que se esperan a cada distancia para las distintas ondas consideradas como se vio en la Figura 1.

Método de Reflexión. Este método se utiliza fundamentalmente para la determinación de la estructura con gran detalle y profundidades no muy elevadas, aunque es posible realizar estudios de toda la corteza. En general debido a un mayor coste de operación se restringe este tipo de investigaciones a la búsqueda de recursos naturales tales como el petróleo. El elevado costo reside en que la instrumentación debe ser más numerosa con alta resolución en tiempo y con una precisa respuesta en amplitudes. Además por el sistema de realización es necesario realizar muchas explosiones controladas y además con perforación en el subsuelo lo que precipita el costo a valores muy altos.

Consideremos una geometría muy simple de una capa inclinada, donde la fuente está en S, Figura. Los tiempos de recorrido de las ondas reflejadas se pueden obtener a partir de la

consideración de la teoría de las imágenes, que supone una fuente imaginaria en I, tal que aplicando la ley de los cósenos al triángulo RIS:

$$T^2 = [x^2 + 4h^2 - 4hx \cos(\theta + \pi/2)] / v_0^2$$

$$= [x^2 + 4h^2 - 4hx \text{sen} \theta] / v_0^2$$

El resultado es una hipérbola con un mínimo en $-2h \text{sen} \theta$, es decir no es simétrica respecto al punto $x=0$.

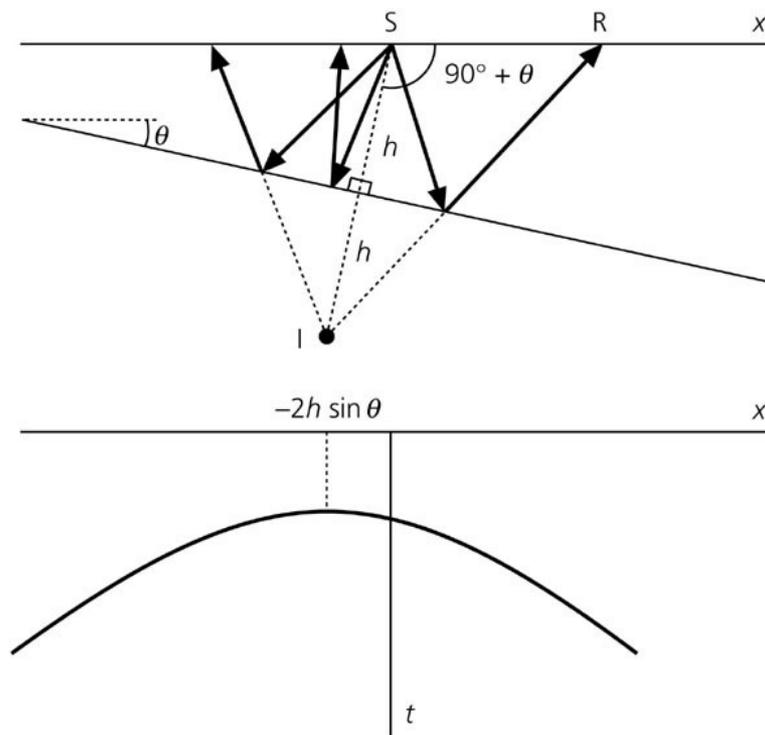


Figura 3. Teoría de imágenes para calcular la hipérbola de la onda reflejada.

Fuente: Stein y Wysession, 2003.

Una característica del método de reflexión es la disposición geométrica multicanal que se puede emplear entre los puntos emisores de ondas y los receptores. Para explicar esta posible disposición geométrica utilizaremos un ejemplo sencillo compuesto de una fuente sencilla y ocho receptores. En este caso en la figura se pueden observar cuatro diferentes sismogramas correspondientes a un punto según se tomen las fuentes $s=9$ con el receptor $r=7$ o las parejas (10,6), (11,5) y (12,4). Todas estas parejas dan información en el mismo punto central pero con distintos ángulos de reflexión en la capa. Si la capa es horizontal el registro así obtenido se denomina *punto medio*. Si por el contrario consideramos las parejas (7,1), (6,1), (5,1) etc. estaremos en que la fuente varía pero el registro es siempre en el mismo punto llamado *receptor*

común, o el inverso que sería *fente común*. Así se pueden formar diferentes situaciones geométricas.

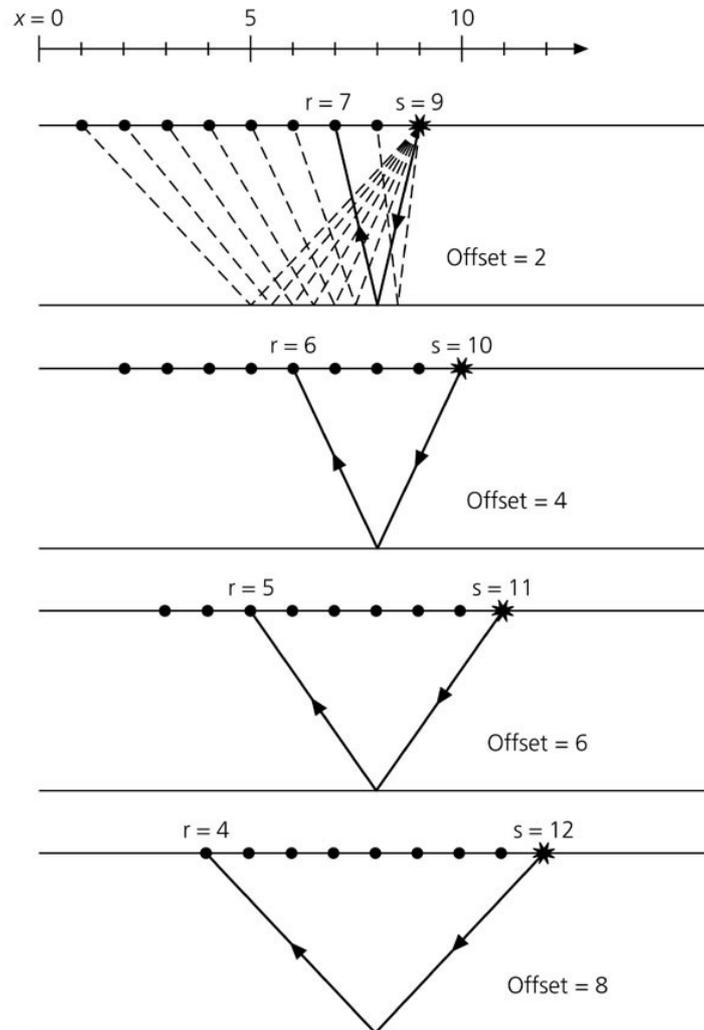


Figura 4. Registro de perfil de reflexión con diferentes alternativas de origen de energía (s) y de registro (r).

Fuente: Stein y Wysession, 2003.

Existe además en la sismica de reflexión lo que se denomina *Apilamiento en punto medio común* (CMP) que consiste en sumar las señales que llegan en una situación de punto común aquellas otras que conociendo que tienen un desplazamiento en el tiempo por corresponder a otra disposición fuente – receptor , puedan llevarse sobre la a la anterior. Esta operación tiene la ventaja de realzar la información que tienen en común y por tanto anular aquella que no sea relevante a la geometría considerada. Puesto que el punto en el que se reflejan es el mismo a la

misma profundidad este método se denomina también de *Apilamiento a profundidad común* (CDP).

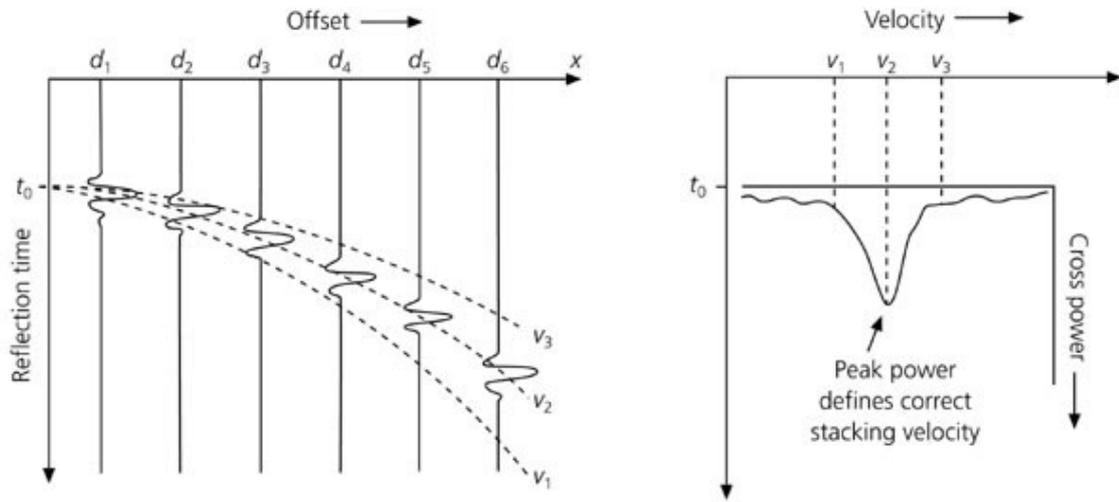


Figura 5. Técnica de apilamiento CMP y análisis de velocidad correspondiente.

Fuente: Stein y Wysession, 2003.

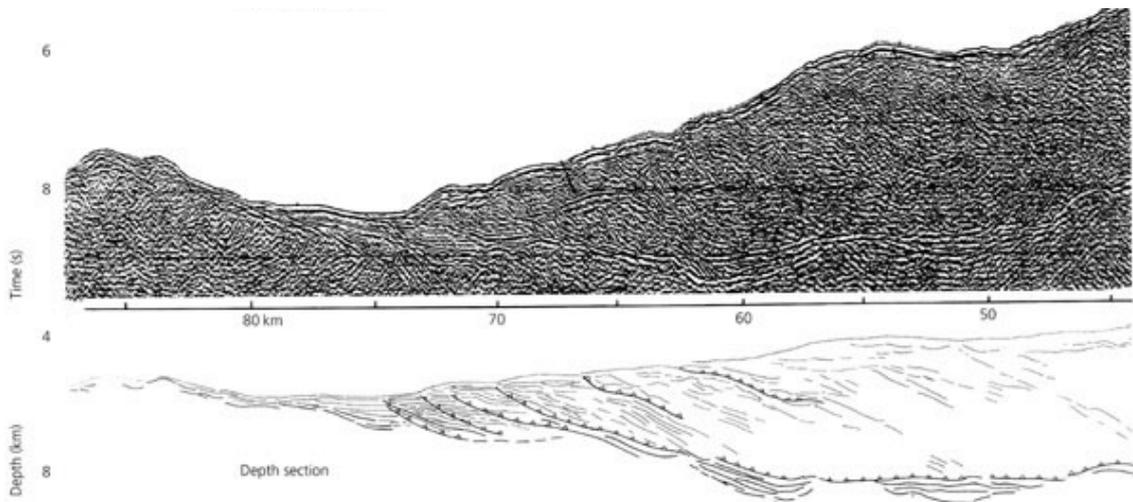


Figura 6. Sección sísmica de reflexión perpendicular a la trinchera en Perú y su interpretación.

Fuente: Stein y Wysession, 2003.

En la Figura se puede ver el registro utilizando esta técnica en un perfil sísmico de reflexión perpendicular a la trinchera frente a la costa de Perú.

15.3 Método gravimétrico y magnético. Anomalías locales y regionales

Como ya se ha comentado se puede utilizar datos de gravimetría para complementar los estudios de las propiedades elásticas de los materiales terrestres. Además, también se pueden utilizar los valores de la observación del campo magnético que reflejará también las propiedades magnéticas de los materiales terrestres. De esta forma se pueden complementar las diferentes propiedades de los materiales terrestres y así poder determinar sus valores y profundidades a las que se encuentran. Es necesario recordar aquí que debido a que vamos a considerar tanto el campo gravimétrico como el magnético, ambos son campos potenciales y por tanto cumplen el Teorema de Stokes. Esto supone que infinitas situaciones de contraste de densidad o propiedades magnéticas, forma y profundidad tienen una respuesta idéntica en ambos campos en la superficie. En otras palabras, existe una indeterminación en la resolución de cualquier problema inverso, dado que existen infinitas situaciones que proporcionan el mismo campo en la superficie. Esta propiedad, que en principio podría pensarse desecha estos métodos no lo es ya que es posible dentro de la indeterminación poder seleccionar de entre todos unos cuantos modelos que sean aplicables. Este inconveniente de los métodos puede ser contrarrestado por la sencillez y economía de estos métodos que los hace imprescindibles en cualquier estudio de la estructura de la tierra.

Tanto en el método gravimétrico como en el magnético se abordará el problema de la obtención de las propiedades de los materiales terrestres a partir de lo que se ha convenido en llamar resolución del problema directo. Es decir, intentaremos plantear modelos de tierra que respondan a las propiedades elásticas o magnéticas de los materiales terrestres y compararemos estos resultados con las observaciones sobre la superficie o a una cierta altitud sobre ella si la observación se hace desde avión o satélite. La no coincidencia entre ambos valores se denomina anomalía gravimétrica o magnética. Estas anomalías resultado de la comparación de los efectos totales en ambos campos se denominan *anomalías regionales*. Estas anomalías reflejan de alguna manera las diferencias del modelo a escala regional con la realidad a esa escala. Estas anomalías tienen superpuestas a ellas las influencias que las partes superficiales de la corteza tienen con respecto al modelo seleccionado. Como estas partes superficiales tienen una distancia menor a la superficie, producen unas variaciones respecto al modelo de muy corta longitud de onda y por tanto son las denominadas *anomalías locales*. Dependiendo de lo que estemos interesados en investigar será conveniente eliminar una u otra. Así si buscamos efectos producidos por las partes más profundas de la corteza deberíamos trabajar con las anomalías regionales, mientras que si lo que nos interesa está en la parte superficial de la misma, deberíamos eliminar las anomalías regionales y trabajar con las locales.

Método Gravimétrico. Cuando se estudia el campo gravimétrico, se considera también la resolución del problema directo. Es decir se trata, a partir de un modelo de tierra con una gravedad teórica compararla con la observada real después de haber sido corregida de las diferencias que conocemos tiene que tener como son: diferencia de altitudes entre el modelo elipsoidal y el relieve, efecto de la masa contenida entre ambas altitudes, efecto de la topografía que sobrepasa o no alcanza la altitud de corrección, efecto del modelo de isostasia aplicable y por último el efecto de masas anómalas que se desvían de la consideración homogénea de la masa de la tierra. Es precisamente este último efecto, el de masas anómalas el que nos permitirá investigar, a partir de las anomalías del campo de gravedad deducir la existencia de ellas y poder valorar su extensión y contraste de densidad con relación al medio que las rodea.

Como siempre intentaremos resolver el problema directo, es decir determinar las anomalías en el valor de gravedad que causan distintos cuerpos situados a diferentes profundidades bajo la superficie terrestre y posteriormente intentaremos relacionarlas con las características geométricas del cuerpo así como del contraste de densidades.

Supongamos un caso sencillo formado por una esfera de radio R y con centro a una profundidad d cuya densidad ρ es distinta a la del medio en la que se encuentra ρ_0 .

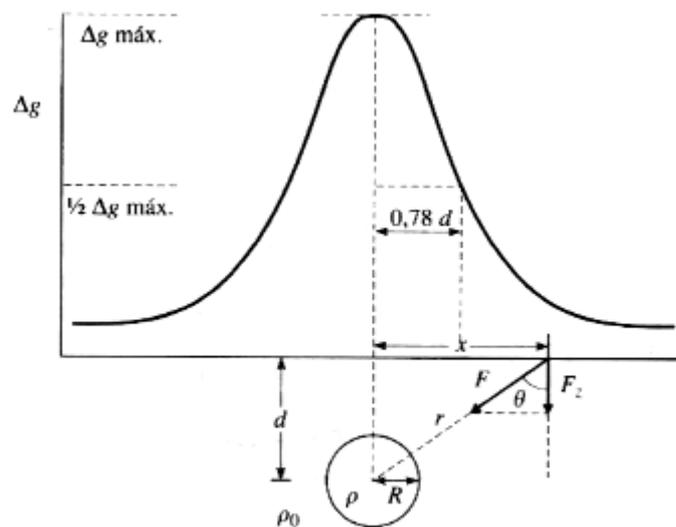


Figura 7. Anomalía gravimétrica de una esfera enterrada. Fuente: Udias y Mezcuca, 1997.

La masa anómala es:

$$\Delta M = \frac{4}{3}(\rho - \rho_0)\pi R^3$$

El potencial de la masa anómala para un punto de coordenadas (x, z) donde el origen de x está sobre el centro de la esfera y z es positivo hacia abajo:

$$U = \frac{G\Delta M}{r} = \frac{G\Delta M}{\sqrt{x^2 + (z-d)^2}}$$

Las anomalías para puntos del eje x, sobre la superficie (z=0):

$$\Delta g(x) = \frac{\partial U}{\partial z} = \frac{G\Delta M}{(x^2 + d^2)^{3/2}}$$

El valor máximo de la anomalía es para x=0:

$$\Delta g_{\max} = \frac{G\Delta M}{d^2}$$

El punto de inflexión de la curva de anomalías corresponde a $x_{\text{inflexión}} = \pm d/2$ y el valor de x para el que el valor de Δg es la mitad del máximo:

$$x_{1/2} = \sqrt{2^{2/3} - 1} = 0.77d$$

Esta relación entre la anchura de la anomalía y la profundidad máxima indica que las anomalías muy estrechas corresponden a masas anómalas muy superficiales. Por tanto si se conoce $\Delta g(x)$, las anomalías en la superficie, para una masa anómala esférica, se puede deducir su profundidad d y su masa ΔM mediante la expresión $d = 2x_{\text{inflexión}} = 1.305x_{1/2}$ y la masa anómala $\Delta M = \Delta g_{\max} d^2 / G$.

De la misma forma se pueden estudiar otros cuerpos sencillos y así tener una serie de expresiones a las que es posible recurrir para su interpretación cuando sospechemos por alguna otra evidencia de la forma del cuerpo que pretendemos investigar.

Método magnético. Cuando observamos en la superficie de la tierra el campo geomagnético, sus valores reflejan las propiedades magnéticas de la parte superior de la corteza además de por supuesto las características del campo interno y externo en ese punto. Por tanto se puede definir el concepto de anomalía $\Delta \mathbf{B} = \mathbf{B} - \mathbf{B}_0$ donde \mathbf{B} es la inducción del campo medido en un punto, corregido por las variaciones del campo externo y \mathbf{B}_0 la correspondiente al campo interno de la tierra. Como las anomalías así definidas son una cantidad vectorial, se suelen usar sus componentes escalares ΔF del campo total, ΔH de su componente horizontal en la dirección del meridiano magnético o bien directamente las componentes sobre un sistema de ejes rectangulares. Como el campo magnético es una variable que está cambiando constantemente es necesario realizar una corrección temporal a todas las observaciones para referirlas a un instante común.

Se procederá de forma similar al caso gravimétrico, es decir supondremos la existencia de cuerpos sencillos y calcularemos la anomalía que producen en el campo magnético en la superficie. El más sencillo de todos es un dipolo magnético de momento magnético m con el polo positivo hacia arriba, enterrando a una profundidad d bajo el origen de la coordenada x , como puede verse en la Figura 8.

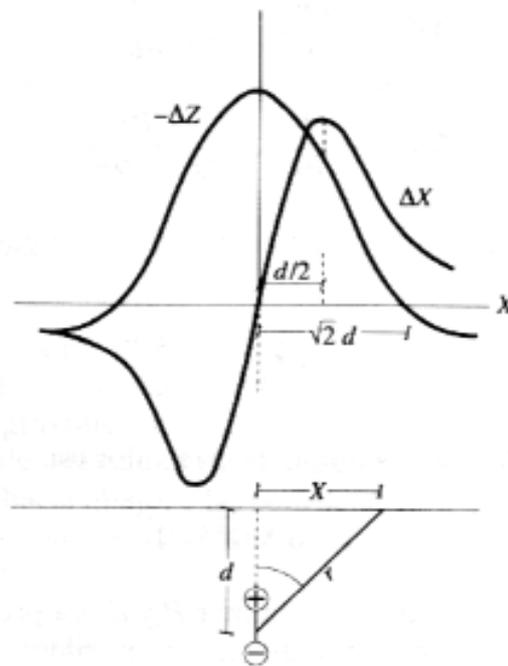


Figura 8. Anomalía producida por un dipolo enterrado. Fuente: Udias y Mezcua, 1997.

El potencial de este dipolo para puntos sobre una línea que pase por encima de él en la dirección N-S, tomando como positivo z hacia abajo y el ángulo θ medido desde el polo positivo, viene dado por :

$$\Phi_A = \frac{Cm \cos \theta}{r} = \frac{Cm(z-d)}{[x^2 + (z-d)^2]^{3/2}}$$

Para $z=0$, los componentes vertical y horizontal de la intensidad del campo son:

$$\Delta Z = -\frac{\partial \Phi_A}{\partial z} = \frac{Cm(2d^2 - x^2)}{(x^2 + d^2)^{5/2}}$$

$$\Delta H = -\frac{\partial \Phi_A}{\partial x} = \frac{3Cmxd}{(x^2 + d^2)^{5/2}}$$

La curva ΔZ tiene el máximo sobre el dipolo, mientras que el horizontal es nulo en dicho punto. La profundidad a la que está enterrado el dipolo puede deducirse del valor de x , para el que ΔZ es la mitad de su valor máximo, en el que cambia de signo, o también de los valores de x donde se encuentran los máximos de ΔX .

El máximo de ΔZ viene dado por $\Delta Z_{\max} = -2Cm/d^2$ y el punto en que la curva ΔZ cambia de signo es $x = \sqrt{2d}$.

El valor de x para el que ΔX es máximo y del que también se puede calcular d es $x=d/2$.

Si la polaridad del dipolo causante de la anomalía es tal que produce un campo en el mismo sentido que el de la Tierra, las anomalías son positivas y en caso contrario negativas. En el hemisferio norte, las anomalías positivas son producidas por dipolos con el polo negativo hacia arriba y al contrario en el hemisferio sur.

15. 4. Estructuras tipo de Corteza

Los estudios mediante perfiles sísmicos tanto de refracción como de reflexión indican un modelo promedio para una corteza continental constituido por capas planas y de velocidad constante en un primera aproximación. Este modelo está formado por una capa superior (granítica) de velocidad 6 km/s y unos 18 km de espesor, y otra inferior (basáltica) de 6.8 km/s y 14 km sobre el Manto de velocidad 8 km/s. En la superficie hay una delgada capa de sedimentos (2 km) en la que la velocidad varía de 5 a 6 km/s. Las fases más destacadas en las dromocrónicas son la P_g , que se transmite por la capa superior, la P_n refractada crítica en la parte superior del manto y la PMP reflejada en la discontinuidad entre corteza y manto y que se denomina Moho. Menos conspicuas son las ondas refractadas críticas y las reflejadas en la interfaz entre la corteza superior e inferior. El tiempo de intersección de la P_g nos da una indicación del espesor de los sedimentos, el de la P_n y la distancia crítica de la PMP del espesor de la corteza.

Los valores para cortezas oceánicas arrojan valores de espesor más delgadas que la deducida para una estructura tipo continental, la cual está formada por una sola capa de material basáltico de espesor entre 5 y 10 km. En mares interiores como el Mediterráneo, en algunas zonas, la corteza puede tener una estructura intermedia con espesor del orden de 15 km. En la Figura 9 se muestra esquemáticamente la estructura de tres cortezas típicas para zonas oceánicas, escudos

continentales y orogénicos. La diferencia más notable es la ausencia de la capa granítica en la corteza oceánica y el mayor espesor de esta capa en las zonas orogénicas recientes, donde a veces la corteza inferior no está diferenciada. También se representan en la figura los modelos obtenidos a partir de perfiles sísmicos profundos para la corteza de la península Ibérica en las zonas del Mar de Alborán, costa Bética y Meseta. Se puede apreciar el cambio de espesor y en estructura desde la correspondiente a una corteza semi-oceánica (mar de Alborán), a una continental estable (meseta). En los tres casos, la corteza está formada por varias capas de distinta velocidad y en la de la Bética y meseta existe una capa intermedia de baja velocidad.

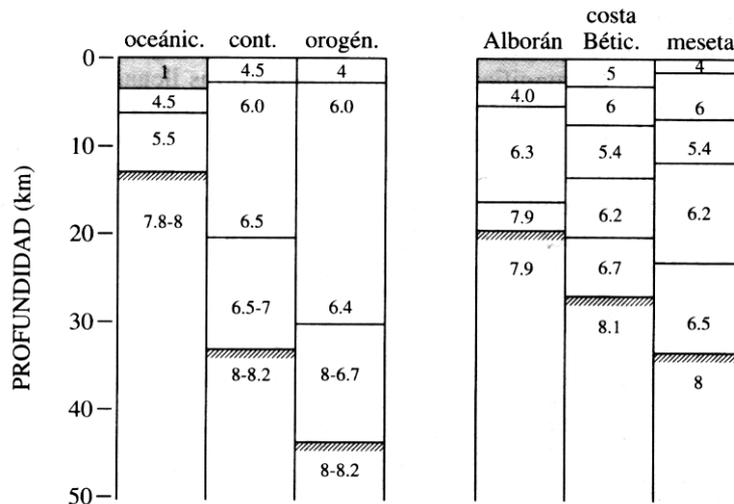


Figura 9. Modelos de estructura de corteza tipo y la determinada para la Península Ibérica.

Fuente: Udias y Mezcuca, 1997.

Bibliografía

- [1] A. Udias y J. Mezcuca. (1997). Fundamentos de Geofísica. Ed. Alianza Universidad. Cap.4., 9., 17.
- [2] Stein, S y M. Wyssession. (2003). An introduction to Seismology, Earthquakes, and Earth Structure. Ed. Blackwell. Cap. 3.
- [3] W. Lowrie. (2000). Fundamentals of Geophysics. Ed. Cambridge Press. CAP.2, .3.

Tema 16. Representación de la fuente sísmica: modelos cinemáticos y dinámicos. Fuerzas equivalentes. Fuente puntual. Fracturas y dislocaciones.

16. 1. Representación de la fuente sísmica: modelos cinemáticos y dinámicos

Los estudios sobre la fuente sísmica están orientados a la determinación de los procesos físicos que se producen en el foco del terremoto y al proceso de generación de las ondas sísmicas que acompañan al sismo. Podemos decir que es a partir del terremoto de San Francisco de 1906 cuando la observación en el terreno de las deformaciones producidas permitió a H. Reid la formulación de lo que ha venido en llamarse Modelo del Rebote Elástico. Según este modelo la generación de un terremoto es consecuencia de la existencia de unas fuerzas que conforman el modelo físico de fuente. Este modelo también representa al mecanismo caracterizado por el desplazamiento producido a ambos lados de la falla. En estos casos, el modelo de fuente sísmica así descrito por el desplazamiento sin relacionarlo con los esfuerzos/ fuerzas actuantes sobre la falla responsables de él es un modelo cinemático. En otras palabras el modelo de fractura causante de la rotura sobre el plano de falla viene dado en función del vector deslizamiento que se genera como una función del tiempo y de la geometría del problema. A partir de estos modelos seremos capaces de deducir el campo de desplazamientos inducido en todo el espacio y en particular, sobre la superficie de la tierra. De esta forma podremos comparar estos desplazamientos con los observados en la realidad. Una coincidencia de ambos campos permitirá inferir un modelo cinemático de la fuente sísmica objeto de estudio.

La aproximación dinámica al problema de la representación de la fuente consiste en considerar el proceso completo, relacionando el deslizamiento observado en la falla con los esfuerzos actuantes en el foco sísmico. Esto además implica el conocimiento de las propiedades elásticas del material que conforma la fuente así como las condiciones iniciales y de frontera aplicables al campo de esfuerzos sobre la región focal. Los modelos dinámicos son más difíciles de realizar y las soluciones no se pueden obtener analíticamente sino mediante métodos numéricos complicados. Al final del tema trataremos estos modelos.

16.2. Fuerzas equivalentes. Foco puntual

Cuando la observación se realiza a distancias del foco suficientemente grandes en comparación a las dimensiones de este, podemos considerar que el foco es puntual y que un sistema de fuerzas aplicadas en el foco puede ser representativo de la fractura que allí se produce. Sea por tanto un medio elástico de volumen V limitado por una superficie S y en el interior tenemos una pequeña región de volumen V_0 limitado por la superficie S_0 .

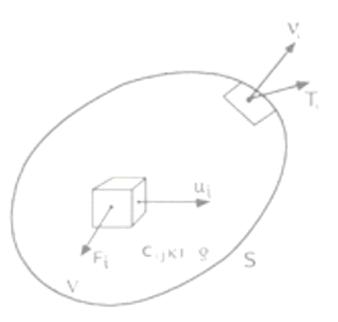


Figura 1. Figura representando el foco sísmico y sus fuerzas actuantes.

La ecuación general de movimiento, si consideramos que no actúan otro tipo de fuerzas en el sistema:

$$\tau_{ij,j} + F_i = \rho \ddot{u}_i$$

Donde τ_{ij} es el tensor de esfuerzos representado por el vector de esfuerzos T_i sobre el elemento de superficie ν_i , F_i las fuerzas actuantes sobre el volumen V_0 y \ddot{u}_i la aceleración de una partícula del terreno en un punto cualquiera de la superficie S .

Si consideramos un medio homogéneo, los esfuerzos pueden expresarse en función de los desplazamientos siguiendo la ley de Hooke:

$$\rho \ddot{u}_i - C_{ijkl} u_{k,lj} = F_i$$

En esta ecuación queda resuelto el problema de conocidas las propiedades elásticas del medio y las fuerzas aplicadas, calcular el valor del desplazamiento del terreno en un punto de la superficie.

Ahora bien, podemos hacer uso del teorema de representación de la elasticidad utilizando el concepto de Función de Green que es la solución de la ecuación de movimiento cuando la

fuerza que actúa es una fuerza impulsiva o delta de Dirac. Esto es, en la ecuación de movimiento anterior sustituimos $F_i = \delta(x - \xi)\delta(t - \tau)\delta_{in}$ resultando:

$$\rho \ddot{G}_{ni} - C_{ijkl} G_{nk,lj} = \delta(x - \xi)\delta(t - \tau)\delta_{in}$$

Aplicando el Teorema de Representación:

$$u_i(x, t) = \int_V d\tau \int G_{in} F_i dV + \int_S d\tau \int (G_{ni} T_i - u_i C_{ijkl} G_{nk,l} v_j) dS$$

Que establece que el campo de desplazamientos en un volumen V limitado por la superficie S consta de dos términos: el primero que es una integral de volumen de todas las fuerzas que intervienen multiplicadas por la función de Green durante todo el tiempo que actúa en la fuente τ y otro término correspondiente a una integral de superficie de todas las tracciones que intervienen y que contiene las funciones de Green y sus derivadas espaciales y también integrada sobre el tiempo de actuación de la fuente.

Si consideramos como caso inicial que no existen tracciones sobre la superficie, el teorema será:

$$u_i(x, t) = \int_{-\infty}^{+\infty} d\tau \int_{V_0} F_k(\xi, \tau) G_{ki}(x, t; \xi, \tau) dV$$

Donde las coordenadas (ξ, τ) se corresponden con puntos en V_0 y (x, t) para puntos sobre la superficie S que engloba a V. Es decir se separan los puntos fuente sísmica de los puntos en la superficie del medio. G_{ki} es la función de Green y se obtiene a partir del teorema de representación cuando la fuerza que actúa es la unidad. El concepto de la función de Green está asociado a la respuesta del medio entre el foco y el punto en la superficie.

Si consideramos el origen de coordenadas coincidente con el punto de aplicación de la fuerza dentro de V_0 , tendremos:

$$u_i(x, t) = \int_{-\infty}^{+\infty} F_k(\xi, \tau) G_{ki}(x, t - \tau) d\tau = F * G$$

Es decir, en este caso el campo de desplazamientos es la convolución del sistema de fuerzas aplicadas con la función de Green del medio.

De lo expuesto hasta ahora se deduce que para obtener el campo de desplazamientos generados en un medio elástico por una fuerza F se puede actuar de dos formas; o bien se resuelve la ecuación primera que es una ecuación diferencial de segundo orden no homogénea en la que se introducen las fuerzas F como una condición de contorno o bien a través del teorema de representación utilizando la última expresión, previo conocimiento de la función de Green. Este segundo método aunque parece en principio más complicado porque es necesario obtener la función de Green, no lo es así, pues una vez determinada para un medio esta no variará para cualquier fuerza.

Como ya se ha dicho el modelo de Rebote elástico plantea como solución más inmediata para el mecanismo físico actuante en un terremoto un par de fuerzas con brazo del par de valor d . Si consideramos el caso particular en que la fuerza del par este dirigida según el eje X_3 y el brazo d en la dirección X_2 , tendremos que el desplazamiento en un punto generado por el par de fuerzas según la expresión en función del teorema de representación a la que multiplicamos y dividimos por d resultará:

$$u_i(x,t) = d \cdot F_3 * \left[\frac{G_{i3}(x,t;\xi + 1/2d,0) - G_{i3}(x,t;\xi - 1/2d,0)}{d} \right] =$$

$$= M_{32} * \frac{\partial G_{k3}(x,t;\xi,0)}{\partial \xi_2}$$

La cantidad $M_{32}(t)$ es *el momento* del par de fuerzas.

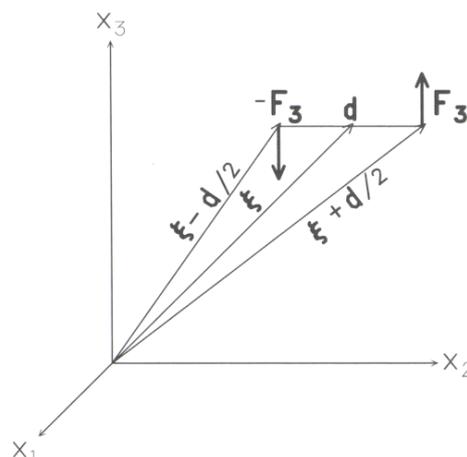


Figura 2. Par de fuerzas actuantes en el foco puntual.

Sin embargo, esta solución al problema no es viable físicamente porque el resultado neto de un par de fuerzas es una rotación. Por esta razón, consideraremos que actúan al menos dos pares de fuerzas, sin momento resultante.

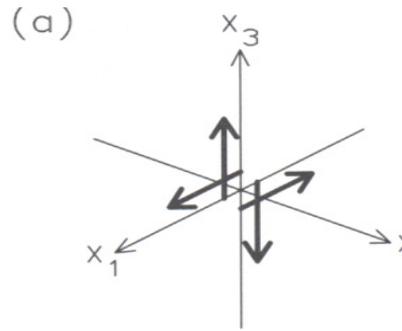


Figura 3. Dos pares de fuerzas sin momento resultante situadas en el plano $X_2=0$

En este caso el campo de desplazamientos generado por este doble par es:

$$u_i(x,t) = \int_{-\infty}^{+\infty} M(\tau) [G_{i1,3}(t-\tau) + G_{i3,1}(t-\tau)] d\tau$$

En general si el doble par tiene las fuerzas en la dirección del vector unitario \vec{n} y el brazo en la dirección \vec{l} ambos perpendiculares, la expresión anterior será:

$$u_i(x,t) = \int_{-\infty}^{+\infty} M(l_k n_l + n_k l_l) G_{ik,l} d\tau$$

Es decir el campo de desplazamientos para una fuente puntual en la que actúan un doble par de fuerzas viene dado por el valor del momento de las fuerzas, de la geometría de estas y de la función de Green.

16.3 Fracturas y dislocaciones

Es un hecho que al producirse un terremoto cercano a la superficie se observa siempre una rotura del terreno. Estas roturas son semejantes a las que en el estudio de sólidos se producen y que fueron estudiadas por Volterra como dislocaciones en un medio elástico.

Una dislocación se define como una superficie interna de un medio tal que a ambos lados de ella existe una discontinuidad de desplazamientos o deformaciones. Es decir que esta discontinuidad en el campo de los desplazamientos va acompañada de una continuidad en los esfuerzos aplicados.

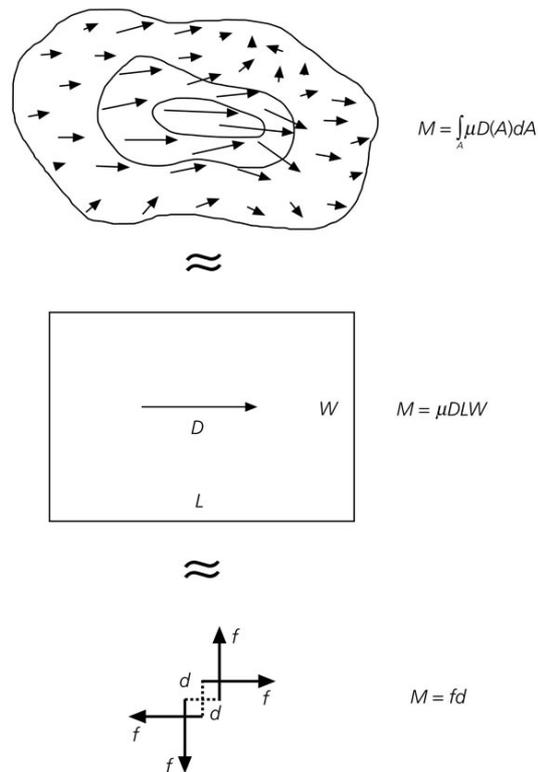


Figura 4. Diferentes aproximaciones al proceso en la fuente sísmica De arriba abajo: Dos pares de fuerzas, desplazamiento sobre una falla rectangular y dislocación. Fuente: Stein y Wyssession, 2003.

Así consideramos que la región focal viene dada por una superficie Σ de dirección n_i y con una discontinuidad en los desplazamientos denominada *deslizamiento* $\Delta u(\xi, \tau)$ y en ausencia de fuerzas, el teorema de representación resulta:

$$u_i(x, t) = \int_{-\infty}^{+\infty} d\tau \int_{\Sigma} \Delta u_i(\xi, \tau) C_{ijkl} n_j G_{nk,l}(\xi, \tau; x, t) dS$$

Si consideramos que el medio es isótropo con coeficientes elásticos λ y μ , y además el vector deslizamiento está dirigido en la dirección l_j resultará:

$$u_i(x, t) = \int_{-\infty}^{+\infty} d\tau \int_{\Sigma} \Delta u(t) [\lambda l_k n_k \delta_{ij} + \mu (l_i n_j + l_j n_i) G_{ki,j}] dS$$

Ahora bien, si consideramos que ambas direcciones n_i y l_j son perpendiculares:

$$u_k(x, t) = \int_{-\infty}^{+\infty} d\tau \int_{\Sigma} \Delta u(t) \mu (l_i n_j + l_j n_i) G_{ki,j} dS$$

A grandes distancias de la fuente:

$$u_k(x, t) = \mu S (l_i n_j + l_j n_i) \int_{-\infty}^{+\infty} \Delta u(\tau) G_{ki,j}(t - \tau) dS$$

Si ahora a partir de las funciones de Green, para un medio infinito homogéneo, elástico e isótropo y considerando sólo el campo lejano, calculamos sus derivadas que intervienen en la expresión anterior:

$$G_{ki,j}^P = \frac{1}{4\pi\rho\alpha^3 r} \gamma_i \gamma_k \gamma_j \dot{\delta}(t - \frac{r}{\alpha})$$

$$G_{ki,j}^S = \frac{-1}{4\pi\rho\beta^3 r} (\gamma_i \gamma_k - \delta_{ik}) \gamma_j \dot{\delta}(t - \frac{r}{\beta})$$

Que corresponden a los dos tipos de propagación que se obtienen de la ecuación de movimiento.

Sustituyendo estas funciones de Green en la expresión que hemos obtenido para el campo de desplazamientos generado por esta dislocación,

$$u_j^P = \frac{\mu S}{4\pi\rho\alpha^3 r} (n_k l_i + n_i l_k) \gamma_i \gamma_k \gamma_j \Delta \dot{u}(t - \frac{r}{\alpha})$$

$$u_j^S = \frac{\mu S}{4\pi\rho\beta^3 r} (n_k l_i + n_i l_k) (\delta_{ik} - \gamma_i \gamma_j) \gamma_k \Delta \dot{u}(t - \frac{r}{\beta})$$

Es decir el campo de desplazamientos generado por la dislocación depende de la velocidad de deslizamiento en vez del deslizamiento en si mismo como parecería lógico esperar. También se puede expresar en función del ya definido momento sísmico si en la expresión anterior agrupamos $M_0 = \mu S \Delta(u)$.

En la expresión deducida anteriormente para el campo de desplazamientos generado por una dislocación es posible suponer a la función temporal del deslizamiento $\Delta u(\xi, \tau)$ distintas formas. Así, en lo que hemos visto hasta ahora la función temporal era una función escalón tal que para un tiempo anterior a la producción de la dislocación la función vale 0 y para el momento inmediatamente después de la ocurrencia de la dislocación se obtiene el valor del deslizamiento; esto es la dependencia temporal es la unidad. Sin embargo existen muchas formas de esta función que se adaptan más a la realidad. Estas son por ejemplo la función rampa

o la función exponencial. Una imagen de estas funciones se puede ver en la Figura y en la parte inferior las derivadas.

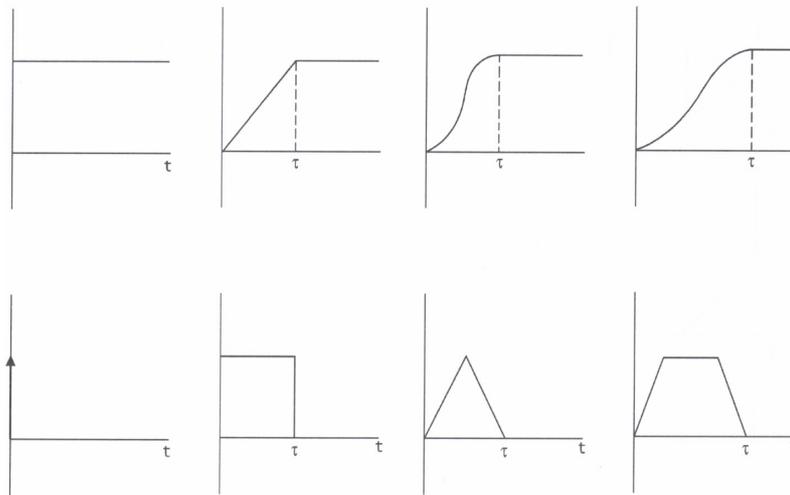


Figura 5. Funciones fuente escalón, rampa y dos exponenciales y sus derivadas temporales: función delta de Dirac, cuadrada, triangular y trapezoidal. Fuente: Udías,1999.

Ejemplos de modelos cinemáticos y dinámicos.

El modelo cinemático más simple se debe a Haskell y considera una falla rectangular de longitud L y anchura W sobre la que el deslizamiento Δu se propaga con velocidad v según la dirección L

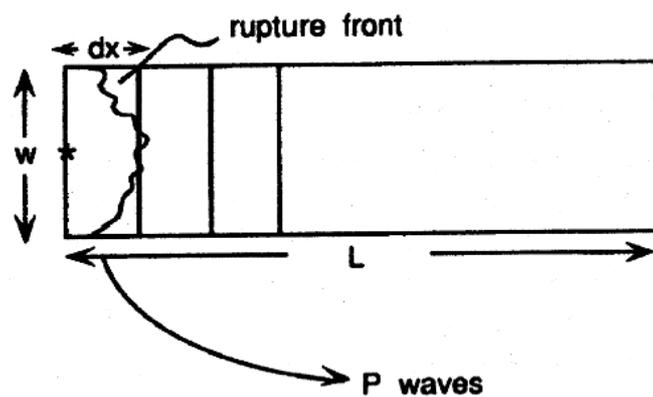


Figura 6. $P(t, \tau)$ Modelo de falla con frente de rotura propagándose a velocidad v . Fuente: Lay y Wallace,1999.

El campo de desplazamientos en un punto en el campo lejano situado a una distancia r de la falla es:

$$u_r(r, t) = \frac{\mu R^P}{4\pi\rho\alpha^3} \frac{w}{r} v_r \dot{u}(t) * P(t, \tau_c)$$

Siendo $H(t)$ la función escalón o de Heaviside y R^P el patrón de radiación para las ondas P y $P(t, \tau_c)$ una función cuadrada de duración el tiempo de ruptura τ_c .

$$P(t, \tau_c) = \left[H(t) - H\left(t - \frac{x}{v_r}\right) \right]$$

Por lo tanto el desplazamiento pueda expresarse como una convolución de dos funciones cuadradas: una de duración τ_c que representa a $\dot{u}(t)$ y otra que es $P(t, \tau_c)$ relacionada con la extensión de la falla. El resultado de esta convolución es una función trapecio.

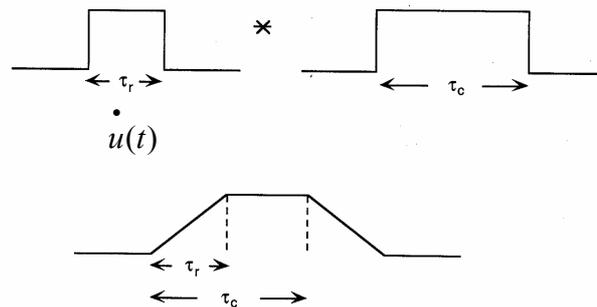


Figura 7.

De este modo la función temporal de la fuente (desplazamiento en la fuente) vendrá dada por:

$$u(t) = M_0 [P(t, \tau_s) * P(t, \tau_r)]$$

Si pasamos del dominio del tiempo al de la frecuencia:

$$A(\omega) = M_0 \left| \frac{\text{sen}(\omega\tau_r / 2)}{\omega\tau_r / 2} \right| \left| \frac{\text{sen}(\omega\tau_s / 2)}{\omega\tau_s / 2} \right|$$

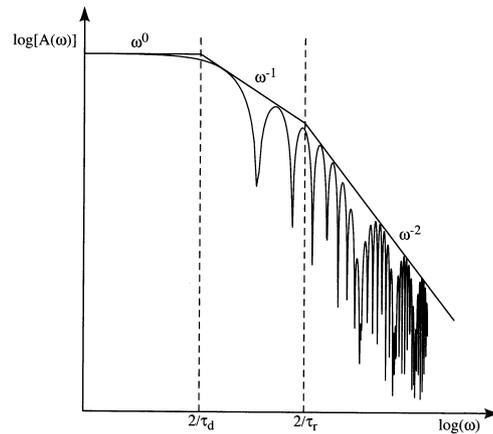


Figura 8. Espectro de Fourier del campo de desplazamientos del modelo de Haskell. Fuente: Lay y Wallace, 1999.

Es decir tenemos que el modelo de Haskell en su representación en el dominio de la frecuencia tiene una forma en representación log-log caracterizado por tres rectas dadas por:

$$A(\omega) = \begin{cases} M_0 & \omega < 2/\tau_s \\ \frac{M_0}{\omega\tau_s/2} & 2/\tau_s < \omega < 2/\tau_r \\ \frac{M_0}{\omega^2(\tau_r\tau_s/4)} & \omega > 2/\tau_r \end{cases}$$

Es decir la parte plana es proporcional a M_0 , después una parte que depende de ω^{-1} y por último una parte que depende de ω^{-2} . Los puntos de cambio de pendientes dependen de $2/\tau_s$ y $2/\tau_r$ respectivamente que son los tiempos de subida y de rotura.

En cuanto al modelo dinámico es necesario resolver la solución de la fractura en el tiempo. Es decir la rotura avanza por el material con una cierta velocidad, por lo que delante del frente de rotura el estado de esfuerzos es nulo.

Si consideramos las direcciones de por una parte el deslizamiento y por otra la dirección de rotura se pueden definir tres modos de fractura en la mecánica del sólido:

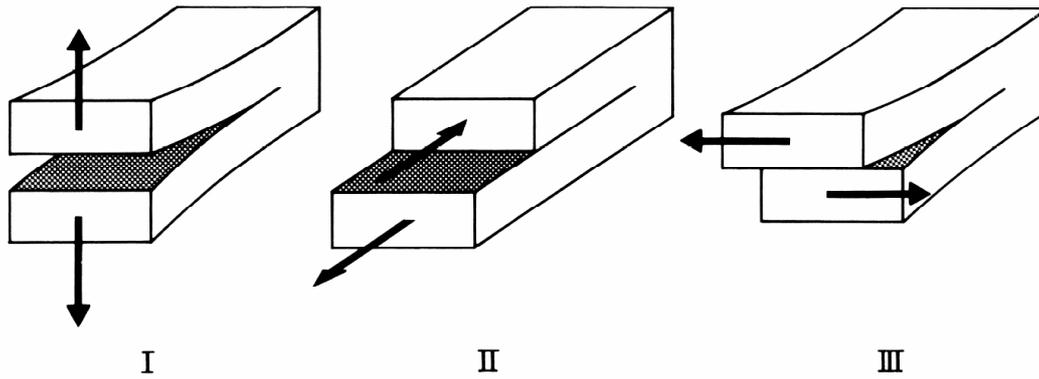


Figura 9. Modos repropagación de fractura.

Modo I es una fractura tensional en la que el deslizamiento es perpendicular al plano de fractura.

Modo II es una fractura de cizalla en el plano.

Modo III es una fractura de cizalla en antiplano.

En general solamente los modos II y III se pueden dar en caso de terremotos.

Para estudiar el problema de propagación se contempla la situación energética del frente de rotura. Para que la fractura avance en un material es necesario que se genere nueva superficie de rotura que consumirá una cierta energía denominada energía de Griffith. Es decir que se necesita un flujo de energía desde la parte no fracturada a la fractura. Esta energía proviene de la caída de esfuerzos $\Delta\sigma$ detrás del frente de rotura.

La relación entre la caída de esfuerzos y el deslizamiento producido es la parte más difícil de la resolución del problema dinámico. La solución para una fractura circular que crece desde el centro fue resuelta por Madariaga siendo:

$$\Delta u(\rho, t) = \frac{\Delta\sigma}{\mu} C(v) v (t^2 - \rho^2 / v^2)^{1/2}$$

El crecimiento de la fractura se realiza mediante un flujo de energía constante desde el material que todavía no se ha fracturado. En esta expresión v es la velocidad de rotura, ρ es el radio de la fractura circular en un instante t y $C(v)$ es un factor con valor próximo a la unidad.

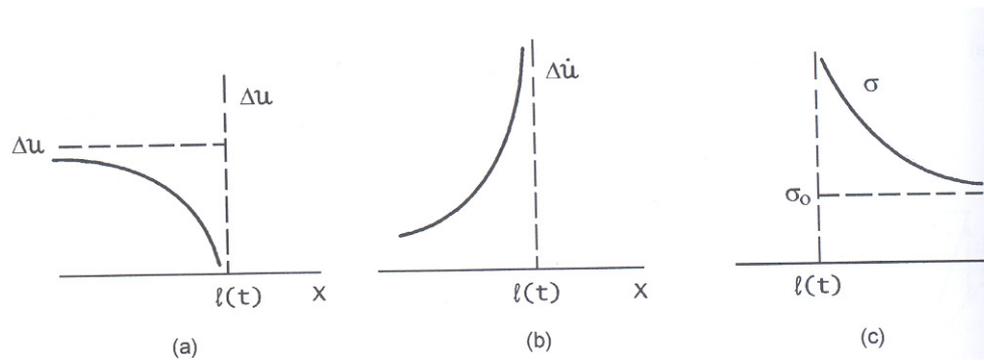


Figura 10. Función del deslizamiento y su tasa en una falla y los esfuerzos fuera de la falla con caída de esfuerzos total. $l(t)$ es la posición del frente de rotura. Udías, 1999.

Bibliografía

- [1] Udías, A (1999). Principles of Seismology. Ed. Cambridge.. CAP. 16., 17., 18.
- [2] R. Madariaga y K. Bak Olsen.(2002). Earthquake dynamics. En Internacional Handbook of Earthquake and Engineering Seismology. Editores. W.H.K. Lee, H. Kanamori, P. C. Jennings y C. Kisslinger. Ed. Academic Press. Cap. 12.

Tema 17. Parámetros focales de los terremotos. Magnitud y energía. Tipos de fallas. Tensor momento sísmico y parámetros de fractura.

17.1. Parámetros focales de los terremotos

Durante el siglo XIX los estudios de campo llevados de forma sistemática después de la ocurrencia de terremotos llevaron a algunos científicos a proponer que los terremotos se producen en un foco puntual a partir del cual se propagan las ondas sísmicas (Mallet, Oldham,...). Asimismo contribuyeron decisivamente a mostrar su relación con las fracturas del terreno. Sin embargo, fue E. Suess, hacia 1875, quien propuso que los terremotos están producidos por el movimiento relativo de los bloques de una falla que resbalan uno con respecto a otro, estableciendo claramente la relación entre fenómenos sísmicos y orogénicos.

El primer modelo mecánico de origen de los terremotos fue propuesto, en 1911, por E. Reid con la teoría del rebote elástico, después de estudiar la fractura de más de 300 Km producida por el terremoto de San Francisco, en 1906. Según esta teoría, las deformaciones elásticas se van acumulando en una región, hasta que se supera la resistencia del material, y se produce una fractura con movimiento relativo de sus dos partes, con la relajación de las deformaciones acumuladas. La energía elástica se almacena de forma lenta, en procesos que pueden durar diez o cien años, relajándose súbitamente en un terremoto que sólo dura algunas decenas de segundos.

Un terremoto de origen tectónico puede considerarse producido por la ruptura de una parte de la corteza terrestre, con un desplazamiento de sus dos lados y la liberación de los esfuerzos elásticos acumulados durante el proceso tectónico. El lugar donde se originan los terremotos se llama región focal o foco y los parámetros que la definen son los que describen el movimiento de la fractura o falla, tal y como veremos más adelante. La localización del foco está dada por sus coordenadas geográficas, su profundidad y la hora origen, si bien debido a que el proceso de ruptura en los terremotos tiene dimensiones y una cierta duración, la asunción de un foco puntual es una simplificación.

17. 2. Magnitud y energía

Aunque las escalas de magnitud vistas en un capítulo anterior dan una medida para comparar diferentes terremotos, existe una forma alternativa de cuantificar el tamaño de los terremotos mediante la energía liberada. Cuando una onda sísmica atraviesa un medio, produce el movimiento de una partícula, inicialmente con energía potencial, lo que le genera una cierta

velocidad y por tanto una energía cinética. La suma de ambas energías, integradas en el tiempo, proporcionarán la energía consumida.

A título de ejemplo consideremos una estación sísmica situada directamente encima de una fuente monocromática de energía sísmica. El desplazamiento del terreno en la estación vendrá dado por:

$$x = A \cos\left(\frac{2\pi t}{T}\right)$$

donde A es la amplitud de la onda de periodo T . La velocidad del terreno será:

$$v = -\left(\frac{2\pi A}{T}\right) \sin\left(\frac{2\pi t}{T}\right)$$

La energía cinética por unidad de masa en la estación de registro es $1/2 \rho v^2$. Si promediamos esta energía para un ciclo completo, obtenemos la densidad de energía cinética:

$$e = \frac{1}{2} \frac{\rho}{T} \int_0^T v^2 dt = \left(\frac{\rho}{2T}\right) \left(\frac{2\pi A}{T}\right)^2 \int_0^T \sin^2\left(\frac{2\pi t}{T}\right) dt = \rho \pi^2 \frac{A^2}{T^2}$$

Tal y como esperábamos, la densidad de energía es proporcional a A^2 . Si integramos sobre el frente de onda esférico para corregir de la extensión geométrica, obtenemos una ecuación de la forma:

$$E = f(r, \rho, c) \left(\frac{A}{T}\right)^2$$

donde r es la distancia de viaje de la onda, ρ es la densidad y c la velocidad del tipo de onda. Esta expresión la podemos modificar hacia una forma similar a la ecuación general que definen las escalas de magnitud:

$$\log E = \log f(r, \rho, c) + 2 \log\left(\frac{A}{T}\right)$$

Por lo tanto es posible relacionar la energía y la magnitud si la función $f(r, \rho, c)$ fuera conocida.

La primera referencia a la energía producida por un terremoto se debe a Bassani, en 1895, en el estudio del terremoto de Florencia del mismo año. Más tarde, Reid, Galitzin y Navarro Neumann, entre 1911 y 1916 hicieron las primeras determinaciones de la energía liberada en terremotos grandes. Una relación empírica entre la magnitud (m_b y M_s) y la energía propagada por las ondas sísmicas E_s (ergios) fue establecida por Gutenberg y Richter en 1956:

$$(1) \log E_s = 5,8 + 2,4 m_b$$

$$\log E_s = 11,8 + 1,5 M_s$$

Por supuesto, la determinación de la energía tiene los mismos problemas que el cálculo de la magnitud. Dado que m_b se satura, la estimación de la energía usando la primera de las ecuaciones anteriores será pequeña para terremotos de magnitud superior a 6,5. La segunda ecuación es más robusta, dado que la magnitud M_S no se satura hasta terremotos de muy alta magnitud. Por otro lado, la energía permite comparar terremotos de distinto tamaño y por ejemplo, la diferencia de energía liberada por un terremoto de magnitud $M_S = 6,0$ y otro de $M_S = 7,0$ es un factor de $10^{1,5} (\cong 32)$. En otras palabras, la energía sísmica liberada por un terremoto de magnitud 7,0 es unas 30 veces superior que la liberada por otro de magnitud 6,0 y también de tres órdenes de magnitud ($\cong 1000$) mayor que la de otro de magnitud 5,0.

En un terremoto, la energía total liberada E es la suma de la energía transmitida en forma de ondas elásticas E_S y la disipada en fenómenos inelásticos y en forma de calor en el foco E_R :

$$E = E_S + E_R$$

El coeficiente de eficiencia sísmica η representa la fracción de la energía total, que se transforma en ondas elásticas:

$$(2) \quad E_S = \eta E$$

El valor de este coeficiente es poco conocido por la dificultad de medir con exactitud la energía total disipada en un terremoto. En el caso de una explosión nuclear, en la que ésta es conocida, el valor de η depende de las condiciones del medio y varía mucho de unos casos a otros.

17. 3. Tipos de fallas

Si consideramos el movimiento de una falla, los parámetros que definen su orientación, dimensiones y la dirección y magnitud de su deslizamiento se describen en la figura 1.

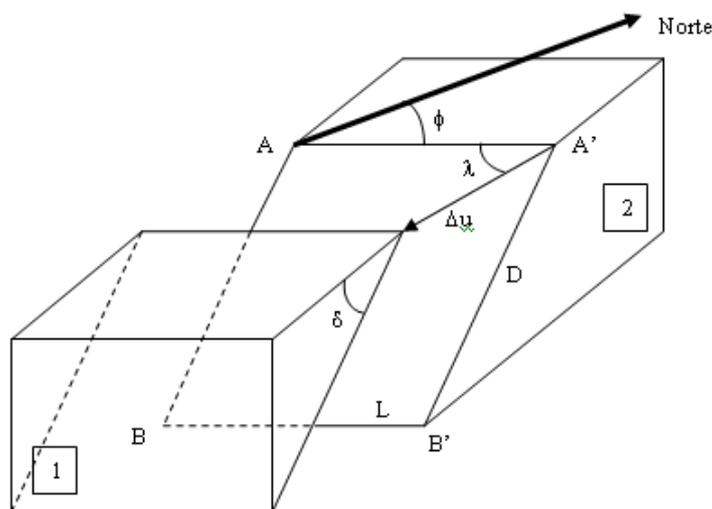


Figura 1. Parámetros que definen el movimiento de una falla.

La intersección del plano de falla (AA'BB') con el plano horizontal, es la traza de la falla (AA'); el ángulo ϕ que forma la traza con el Norte geográfico, es el acimut de la falla, y el ángulo que forma el plano de la falla con el plano horizontal, el buzamiento δ . La superficie del plano de falla, siendo L su longitud (BB') y D su anchura (A'B'), será (suponiendo forma rectangular) $S = L D$. El desplazamiento del bloque (1) con respecto al (2) es Δu , y se le llama también deslizamiento o dislocación. La dirección del deslizamiento viene dada por el ángulo λ , entre la dirección de Δu y la horizontal, medido sobre el plano de falla.

En función de los valores que tengan los ángulos δ y λ , así se conformarían diversos tipos de fallas. Como casos más sencillos estarían (figura 2):

- $\delta = 90^\circ$; $\lambda = 0^\circ$ falla de desgarre – movimiento horizontal;
- $\delta = 90^\circ$; $\lambda = 90^\circ$ falla de gravedad o vertical – movimiento vertical;
- $\delta = 45^\circ$; $\lambda = 90^\circ$ falla normal – movimiento vertical;
- $\delta = 135^\circ$; $\lambda = 90^\circ$ falla inversa – movimiento vertical;

Las fallas normales con buzamiento δ entre 0° y 90° o inversas con buzamiento δ entre 90° y 180° pueden tener además una componente más o menos grande de movimiento horizontal, en cuyo caso el ángulo de desplazamiento λ estaría comprendido entre 0° y 90° .

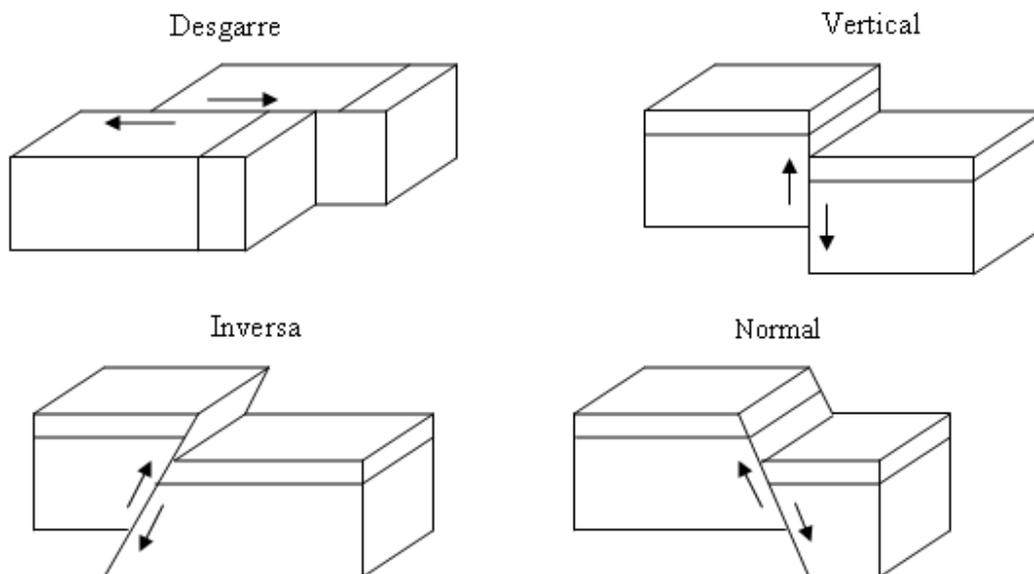


Figura 2. Tipos de fallas.

Hemos visto que la orientación del movimiento de una falla viene dada por los tres ángulos ϕ , δ y λ . Un modelo simplificado de fractura de cizalla lo podemos representar mediante los vectores unitarios n_i y l_i (figura 3).

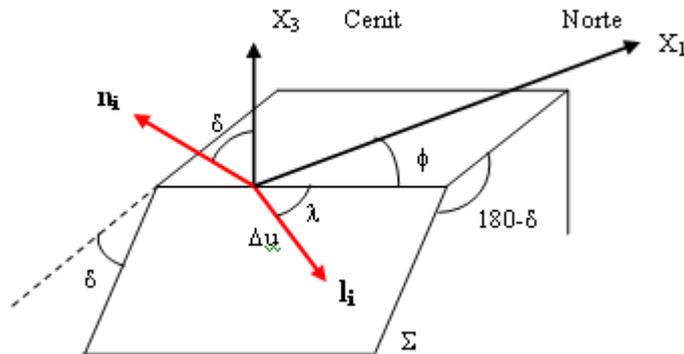


Figura 3. Parámetros de orientación de una falla y vectores unitarios n_i y l_i .

La superficie Σ es plana y su vector unitario normal es n_i . La dirección del desplazamiento Δu viene definida por el vector unitario l_i . Las componentes de los dos vectores unitarios n_i y l_i en función de los ángulos ϕ , δ y λ estarían definidas por:

$$\begin{aligned}
 n_1 &= -\text{sen } \delta \text{ sen } \phi \\
 n_2 &= \text{sen } \delta \text{ cos } \phi \\
 n_3 &= -\text{cos } \delta \\
 l_1 &= \text{cos } \lambda \text{ cos } \phi + \text{cos } \delta \text{ cos } \lambda \text{ sen } \phi \\
 l_2 &= \text{cos } \lambda \text{ sen } \phi - \text{cos } \delta \text{ sen } \lambda \text{ cos } \phi \\
 l_3 &= -\text{sen } \lambda \text{ sen } \delta
 \end{aligned}$$

17.4. Tensor momento sísmico y parámetros de fractura

17.4.1. Momento sísmico escalar, caída de esfuerzos y esfuerzo medio

Hemos visto que la magnitud de un terremoto está relacionada con la energía liberada y es independiente del mecanismo de su generación, sin embargo otra medida del tamaño de un terremoto puede ser el momento sísmico escalar M_0 , que fue introducido por Aki en 1966. Está basado en la idea de que los terremotos son causados por fracturas de cizalla en la corteza terrestre y se define como:

$$M_0 = \mu \bar{\Delta u} S$$

donde μ es el módulo de rigidez o de cizalla, $\bar{\Delta u}$ es el valor medio del desplazamiento del plano de falla y S es el área del plano de falla. Sus unidades en el SI son Nm. El momento sísmico incluye el área de la falla, el desplazamiento y la rigidez del material y por tanto constituye una buena medida física del tamaño de un terremoto.

En un modelo de fractura simplificado, el desplazamiento relativo Δu de los dos bordes de la falla se debe a la actuación de esfuerzos de cizalla, los cuales, en un momento determinado, exceden la rigidez del material o a la fricción que mantiene bloqueada la falla. La diferencia de los esfuerzos de cizalla, antes (σ_0) y después (σ_1) de producirse la fractura (figura 4) permite definir dos nuevos parámetros, el esfuerzo medio $\bar{\sigma}$ y la caída de esfuerzos $\Delta\sigma$:

$$\bar{\sigma} = \frac{1}{2}(\sigma_0 + \sigma_1)$$

$$\Delta\sigma = (\sigma_0 - \sigma_1)$$

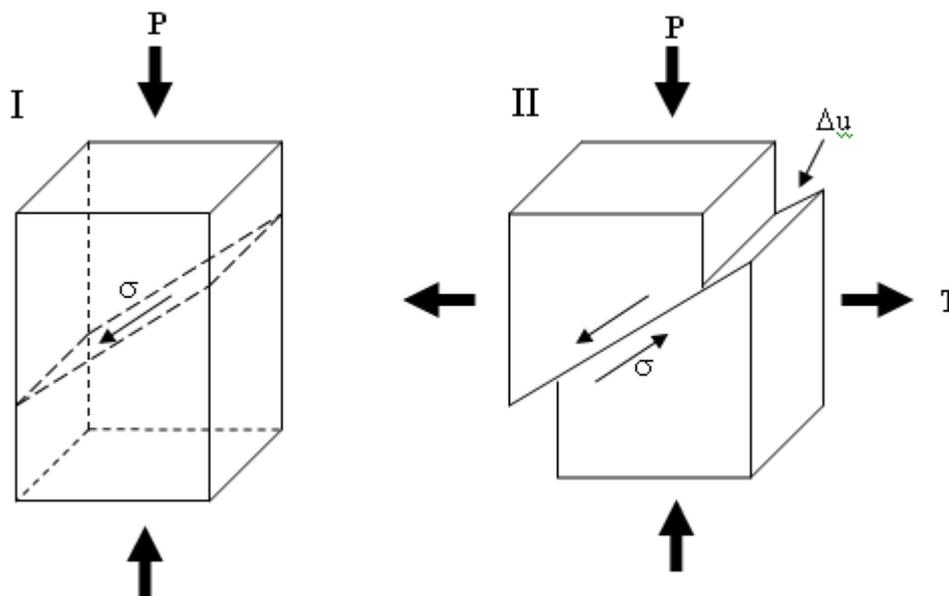


Figura 4. Esfuerzos y desplazamiento en la fractura antes (I) y después (II) del terremoto.

Las unidades usadas para $\bar{\sigma}$ y $\Delta\sigma$ son Pa (SI), bars (1 bar = 10^5 Pa) y dyn/cm², siendo los valores más típicos los comprendidos entre 1 y 10 MPa.

La caída de esfuerzos representa la parte de los esfuerzos actuantes que se emplea en generar el desplazamiento de la falla. Si $\sigma_1 = 0$, la caída de esfuerzos es total y $\Delta\sigma = 2\bar{\sigma}$. Debido a la fricción existente entre ambos lados de la fractura, siempre queda algo de esfuerzo residual después de que la fracturación haya ocurrido.

De una forma simplificada es posible relacionar la energía liberada con el esfuerzo medio o la caída de esfuerzos:

$$E = \bar{\sigma} \bar{\Delta u} S$$

que sustituyendo en la definición del momento sísmico, obtenemos:

$$(3) \quad E = \frac{\bar{\sigma}}{\mu} M_0$$

Si la caída de esfuerzos es total, la expresión quedaría:

$$(4) \quad E = \frac{\Delta\sigma}{2\mu} M_0$$

Esta expresión relaciona la energía total liberada por un terremoto con el momento sísmico y la caída de esfuerzos.

Para una fractura de cizalla, la caída de esfuerzos es proporcional a la deformación de la falla $\Delta\sigma = \Delta u/D$, siendo D la anchura de la falla en el caso de que sea rectangular, o bien $D = a$ para falla circular de radio a . La caída de esfuerzos viene dada, para una falla circular, por:

$$\Delta\sigma = C\mu \frac{\bar{\Delta u}}{a}$$

donde C es un factor adimensional que depende de la forma de la fractura, $C = 7\pi/16$ para una falla circular. La relación entre el momento sísmico y la caída de esfuerzos para este tipo de falla será:

$$M_0 = \frac{16}{7} a^3 \Delta\sigma$$

Por tanto, si se conoce el momento sísmico y las dimensiones de la fractura, se puede determinar la caída de esfuerzos. En la ecuación anterior el radio a está con potencia de tres, lo que implica que pequeños errores en su determinación producirán grandes errores en $\Delta\sigma$.

Usando el valor del área de la falla circular ($S = \pi a^2$), la expresión anterior se puede poner:

$$M_0 = \frac{16 \Delta\sigma}{7\pi^{3/2}} S^{3/2}$$

y tomando logaritmos

$$(5) \log M_0 = \frac{3}{2} \log S + \log \left(\frac{16\Delta\sigma}{7\pi^{3/2}} \right)$$

De esta ecuación se puede extraer la conclusión de que si la caída de esfuerzos es constante para todos los terremotos, entonces $\log S$ es proporcional a $2/3 \log M_0$. Esta hipótesis, válida para un gran rango de magnitudes, ha sido demostrada empíricamente por Kanamori y Anderson (1975). Para terremotos de moderados a grandes ($m > 5$), los valores de $\Delta\sigma$ están comprendidos entre 1-10 MPa y un valor medio de 6 MPa. Estos autores sugieren que los terremotos que suceden en los límites de placa (sismos interplaca) tienen una caída de esfuerzos más baja (3 MPa) que los que tienen lugar en los interiores de las placas (sismos intraplaca) que la tienen alrededor de 10 MPa.

La asunción de que la caída de esfuerzos es constante, va ligada a la definición de la magnitud momento M_w definida por Kanamori. Así, la fórmula que la define ($M_w = 2/3 \log M_0 - 10,7$) está obtenida por la sustitución de la ecuación (4) en la ecuación (1), asumiendo además un valor de $\Delta\sigma/\mu = 10^{-4}$ y despejando el valor de M_s al que renombramos ahora como M_w . Se puede por tanto afirmar que la magnitud momento es la magnitud obtenida a partir del momento sísmico, bajo la hipótesis de una caída de esfuerzos constante que satisface la relación de Gutenberg-Richter entre la magnitud de ondas superficiales y la energía.

La ecuación (5) nos relacionaba el momento sísmico M_0 con la superficie S de la falla, pero también resulta de interés enlazar el momento sísmico con la magnitud M_s . Si sustituimos la energía E en función de la energía sísmica E_s (ecuaciones (2) y (3)), se puede definir el esfuerzo medio aparente:

$$\eta \bar{\sigma} = \mu \frac{E_s}{M_0}$$

Este parámetro está relacionado con la resistencia del material en el que se produce la fractura. En medios o zonas de corteza débil, como pueden ser las zonas de cordilleras oceánicas, se producen valores bajos de $\eta \bar{\sigma}$ y en medios muy resistentes, como en el interior de las placas, valores altos. Usando esta expresión junto con la ecuación (1), se puede relacionar el momento sísmico M_0 y la magnitud M_s :

$$\log M_0 = 1,5M_s + 11,8 - \log \left(\frac{\eta \bar{\sigma}}{\mu} \right)$$

Si $\eta \bar{\sigma}$ es constante, tendremos una relación lineal entre $\log M_0$ y M_s con una pendiente de 1,5, lo que está de acuerdo con las observaciones, aunque existe una cierta dispersión en los datos, que puede ser achacable a los terremotos muy grandes, para los cuales se supone la saturación

de la escala M_S (Kanamori y Anderson, 1975). En este caso, existen también algunas evidencias que señalan que los esfuerzos para los terremotos intraplacas (5 MPa) son mayores que los de terremotos interplacas (1,5 MPa).

Podemos concluir que el momento sísmico M_0 puede ser considerado la mejor medida del tamaño de un terremoto, asumiendo que su mecanismo de generación es de una fractura de cizalla.

17. 4.2. Tensor momento sísmico

El mecanismo de los terremotos corresponde al de una fractura producida por la acumulación de esfuerzos actuando sobre una región de la corteza terrestre. Una formulación muy adecuada a este problema fue la introducida por Gilbert en 1970 con el uso del tensor del momento sísmico M_{ij} , que lo relacionó con la caída total de esfuerzos. Su significado fue posteriormente aclarado por G. Backaus y M. Mulcany, en 1976, precisando que el tensor momento representa solamente una parte que equivale al esfuerzo en exceso del elástico, es decir al disipado en deformaciones inelásticas en la fuente y no a la caída total de esfuerzos.

El tensor momento sísmico M_{ij} , y el tensor de densidad de momento sísmico m_{ij} por unidad de volumen o de superficie están relacionados mediante

$$M_{ij} = \int_V m_{ij} dV = \int_S m_{ij} dS$$

Si consideramos un medio elástico en el que solamente suceden procesos elásticos, la ecuación de movimiento, en ausencia de fuerzas será:

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \frac{\partial \tau_{ij}}{\partial x_j}$$

que con la notación de subíndices queda:

$$\rho \ddot{u}_i = \tau_{ij,j}$$

Sin embargo, la situación real es que además de procesos elásticos ocurren procesos inelásticos y si designamos con σ_{ij} el esfuerzo total, τ_{ij} corresponderá solamente al modelo puramente elástico. La ecuación de movimiento para la situación real será:

$$\rho \ddot{u}_i = \sigma_{ij,j}$$

Si definimos la densidad de momento sísmico m_{ij} como el exceso de esfuerzo del puramente elástico, vendrá dado por tanto por la diferencia:

$$m_{ij} = \tau_{ij} - \sigma_{ij}$$

que sustituida en la ecuación de movimiento resulta:

$$\rho \ddot{u}_i = \tau_{ij,j} - m_{ij,j}$$

Si comparamos esta ecuación con la general del movimiento (ver capítulo 16) en la que la fuente sísmica estaba representada por fuerzas equivalentes:

$$F_i = -m_{ij,j}$$

Por lo tanto, las fuerzas pueden ser obtenidas a partir del tensor momento y ambos permiten representar la fuente sísmica. Asimismo, las fuerzas equivalentes se corresponden únicamente con los esfuerzos responsables de los procesos inelásticos en la región de la fuente. Además, igual que sucede con las fuerzas que actúan sobre el volumen, el momento sísmico está definido solo dentro de la región focal, que es donde se producen los procesos inelásticos, y es cero fuera de ella.

Los desplazamientos elásticos u_i fuera de la región focal los podemos expresar en términos de m_{ij} y la correspondiente función de Green, sustituyendo el valor de F_i en la ecuación del desplazamiento en la que son nulos los esfuerzos y desplazamientos sobre la superficie (ver capítulo 16) quedando:

$$u_i = \int_{-\infty}^{\infty} d\tau \int_{V_0} -m_{kj,j} G_{ik} dV$$

Integrando por partes, con ausencia de fuerzas externas y momentos, con la suma de fuerzas internas y momentos nula y con una apropiada elección del origen de coordenadas, se obtiene:

$$u_i = \int_{-\infty}^{\infty} d\tau \int_{V_0} m_{kj} G_{ik,j} dV$$

si el tensor momento está definido solo sobre la superficie Σ , la ecuación es una integral de superficie:

$$(6) \quad u_i = \int_{-\infty}^{\infty} d\tau \int_{\Sigma} m_{kj} G_{ik,j} dS$$

Las dos últimas ecuaciones muestran que los desplazamientos fuera de la región focal pueden obtenerse a partir del tensor momento sísmico y de las derivadas de la función de Green integradas sobre la región focal (V_0 ó Σ).

Si comparamos la ecuación (6) con la obtenida respecto a los desplazamientos en el teorema de representación (ver tema 16), podemos definir la densidad de momento sísmico correspondiente a una dislocación con desplazamiento Δu sobre una superficie Σ de vector normal n como:

$$m_{ij} = C_{ijkl} \Delta u_k n_l$$

siendo C_{ijkl} el tensor de las constantes elásticas definido por:

$$C_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \mu (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk})$$

Para un medio isótropo con la dirección de la dislocación dada por el vector unitario l quedaría:

$$(7) \quad m_{ij} = \Delta u \left[\lambda l_k n_k \delta_{ij} + \mu (l_i n_j + l_j n_i) \right]$$

A partir de esta expresión es posible encontrar el tensor momento para varios tipos de fuentes, especificando la orientación de los vectores n_i y l_i . Los términos $i = j$, representan dilataciones (o compresiones) en la dirección de los ejes y los términos $i \neq j$, representan el movimiento de cizalla.

En el caso de una fuente explosiva, que puede considerarse como una expansión a lo largo de los tres ejes coordenados, se puede representar por medio de tres dipolos lineales a lo largo de cada eje, con n_i y l_i en la misma dirección (figura 5 (a)). El tensor de densidad de momento será:

$$m_{ij} = K \Delta u \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

donde $K = \lambda + 2/3\mu$ es el módulo de compresibilidad. La suma de los elementos de la diagonal principal proporciona el incremento en volumen por unidad de volumen:

$$m_{11} + m_{22} + m_{33} = 3 K \Delta u$$

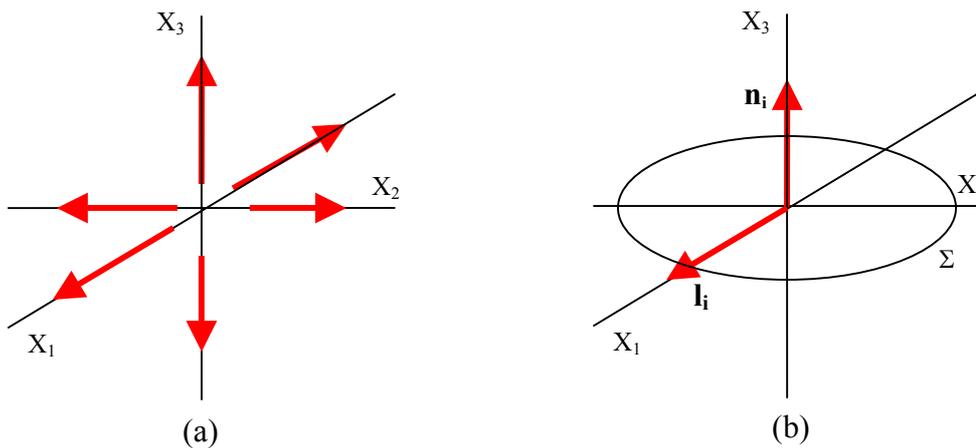


Figura 5. Representación de las fuentes: (a) explosión y (b) fractura de cizalla

En el caso de una fractura de cizalla, Δu está situado a lo largo del plano de falla Σ y los vectores n_i y l_i son perpendiculares. Usando la ecuación (7) y la definición del tensor momento sísmico M_{ij} y el escalar M_0 , se obtiene:

$$(8) \quad M_{ij} = \mu \Delta u S (l_i n_j + l_j n_i) = M_0 (l_i n_j + l_j n_i)$$

En el caso particular en que Σ esté situada en el plano (x_1, x_2) , los vectores serán $n = (0,0,1)$ y $l = (1,0,0)$ (figura 5 (b)):

$$M_{ij} = M_0 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

En este caso, la suma de la diagonal principal es nula, lo que indica que no hay cambio de volumen. Conocidos los valores de n_i y l_i en función de los ángulos ϕ , δ y λ que hemos visto anteriormente, es posible mediante la ecuación (8) determinar las componentes del tensor momento en función de los ángulos de orientación. Por otro lado, en la ecuación (8) se puede observar que, para este caso de fractura de cizalla, si se intercambia el significado de n_i y l_i , el resultado es el mismo, es decir, la expresión es simétrica con respecto a la normal del plano de fractura y la dirección de la dislocación, lo que introduce una ambigüedad en la resolución del problema inverso.

Valores y vectores propios. Tipos de fuentes

El tensor momento sísmico puede ser estudiado en términos de los vectores y valores propios. Debido a que el tensor es simétrico, sus valores propios son reales y sus vectores propios son mutuamente ortogonales y satisfacen la ecuación:

$$(M_{ij} - \delta_{ij}\sigma)v = 0$$

Los tres valores propios σ_1 , σ_2 y σ_3 son las raíces del determinante:

$$\text{Det}[M_{ij} - \delta_{ij}\sigma] = 0$$

y los valores propios v_1 , v_2 y v_3 se obtienen sustituyendo cada valor propio en la ecuación anterior y forman los ejes principales. Respecto a estos ejes, el tensor momento sísmico tiene la forma:

$$M_{ij} = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix}$$

En este sistema, el tensor momento sísmico está formado por tres dipolos lineales en la dirección de los ejes principales y por tanto representa los esfuerzos principales. La suma de los elementos de la diagonal principal es el primer invariante del tensor y tiene el mismo valor para cualquier sistema de referencia:

$$M_{11} + M_{22} + M_{33} = \sigma_1 + \sigma_2 + \sigma_3$$

Esta suma representa el cambio de volumen, tal y como vimos en el caso de una fuente explosiva. En consecuencia, podemos definir la parte isotrópica del tensor momento como:

$$\sigma_0 = \frac{1}{3}(\sigma_1 + \sigma_2 + \sigma_3)$$

Si restamos del tensor M_{ij} la parte isotrónica, obtenemos el tensor desviatorio M'_{ij} , cuya suma de los elementos diagonales es siempre cero y no incluye por tanto cambios de volumen:

$$M'_{ij} = M_{ij} - \delta_{ij}\sigma_0$$

Por lo tanto, el tensor momento sísmico puede ser separado en dos tensores, uno isotrónico $M_{ij}^0 = \delta_{ij}\sigma_0$ y otro desviatorio M'_{ij} :

$$M_{ij} = M_{ij}^0 + M'_{ij}$$

Podemos decir que el tensor momento que representa una fuente explosiva es puramente isotrónico y el que representa a una fractura de cizalla, o de doble par, puramente desviatorio.

En una forma más general, la parte desviatoria del tensor momento puede a su vez estar dividida en otras dos, una correspondiente a la propia de fractura de cizalla y otra debida a cambios rápidos en el módulo de cizalla en una dirección normal al plano de falla y sin cambios de volumen. En resumen, una fuente sísmica puntual de tipo general puede estar representada por un tensor momento, que se divide en tres partes:

$$M = M^0 + M^{DC} + M^{CLVD}$$

Esta partición separa los cambios de volumen, la fractura de cizalla (o doble par) y los cambios en la rigidez de la fuente.

Inversión del tensor momento sísmico

Para una fuente puntual, los desplazamientos elásticos son funciones lineales de las componentes del tensor momento sísmico y de las derivadas de las funciones de Green. Por tanto, es posible determinar las seis componentes del tensor a partir de las observaciones de los desplazamientos elásticos mediante inversión lineal. Para esta fuente, la ecuación (6) puede ser expresada en el dominio del tiempo como una convolución:

$$u_k(t) = M_{ij}(t) * G_{ki,j}(t)$$

que en el dominio de la frecuencia sería:

$$u_k(\omega) = M_{ij}(\omega) * G_{ki,j}(\omega)$$

Si conocemos las derivadas de las funciones de Green, G_{kij} y u_k , que son observaciones, el calculo de M_{ij} no resulta complicado si invertimos una de las dos ecuaciones anteriores, que en forma matricial queda:

$$\mathbf{u} = \mathbf{G} \mathbf{M}$$

donde \mathbf{u} es un vector de dimensión n (número de observaciones) y \mathbf{G} ($n \times 6$). Las seis componentes del tensor se pueden reducir a cinco si asumimos la condición de que no exista cambio de volumen $M_{11} + M_{22} + M_{33} = 0$. La solución puede obtenerse mediante mínimos cuadrados a partir de la ecuación:

$$\mathbf{M} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{u}$$

Los datos utilizados pueden ser ondas superficiales, internas, oscilaciones libres, etc. Las derivadas de las funciones de Green se calculan para modelos teóricos de Tierra y dependen de la posición de la fuente y del observador.

Bibliografía

- [1] Udias, A. Principles of Seismology. Ed. Cambridge University Press. 1999. Cap. 15, 17.
- [2] Lay, T. and Wallace, T.C. Modern Global Seismology. Ed. Academic Press. 1995. Cap. 9.
- [3] - Udias, A. y Mezcua, J. Fundamentos de Geofísica. Ed. Alhambra. 1986. Cap. 11.
- [4] - Kanamori, H. and Anderson, D.L. Theoretical basis of some empirical relations in seismology. Bull. Soc. Seism. Am.65, 1073-1095.

Tema 18. Volcanología: Aspectos geodésicos y geofísicos de la actividad volcánica. Terremotos volcano-tectónicos. Tremor volcánico. Precursores sísmicos de la actividad volcánica. Otros precursores.

18.1. Aspectos geodésicos y geofísicos de la actividad volcánica

Un volcán se puede definir como un punto de la superficie terrestre donde tiene lugar la salida de material rocoso fundido generado en zonas de la corteza terrestre o del manto superior y, en ocasiones, de material no magmático. Un sistema volcánico puede estar durante mucho tiempo sin dar muestras de actividad, mostrando solamente los restos de la actividad pasada, como las fumarolas, las anomalías térmicas debidas al enfriamiento del sistema o los terremotos producidos por el asentamiento del edificio. En el momento en que se produce la inyección de nuevo magma, se empiezan a generar apertura de fracturas, iniciándose la deformación del edificio volcánico, yendo asociado a estos procesos multitud de fenómenos físicos y químicos.

Los aspectos geodésicos y geofísicos de la actividad volcánica representan un papel fundamental en el conocimiento del sistema y en su evolución, pudiendo ser además determinantes en una predicción volcánica (Tabla 1). Dentro de estos procesos podemos citar la deformación del suelo como uno de los principales precursores de una erupción y que se produce por un aumento de presión en la cámara magmática o en otros conductos asociado, aunque también puede originarse por una sobrepresión en un acuífero cerrado debido a un foco calorífico. Estas deformaciones en las tres dimensiones pueden medirse en superficie mediante técnicas geodésicas, como la nivelación, redes geodésicas, GPS, etc.

El campo gravitatorio puede sufrir perturbaciones debido al movimiento del magma bajo la superficie, que puede originar aumentos o disminuciones de masa o de densidad. Asimismo, es posible observar modificaciones en el campo electromagnético natural en las proximidades del volcán, que alteran la intensidad del campo magnético a causa de las variaciones en la composición y propiedades magnéticas del material interior. Otro proceso observable es la variación del potencial espontáneo, que se asocia a la circulación de fluidos sobre el edificio volcánico. En el caso de la resistividad de las rocas, esta depende de la salinidad, viscosidad, permeabilidad, porosidad y temperatura y por tanto pueden variar en el proceso evolutivo del volcán.

Proceso	Técnica para estudio
Inyección de magma Cambios de densidad	Gravimetría
Deformación	Nivelación, redes geodésicas, extensómetros y clinómetros. GPS
Fractura	Sismicidad volcánica
Transmisión de calor	Termometría: suelos, fuentes termales, fumarolas
Termodinámica de gases	Análisis de gases en fumarolas Difusión de CO ₂
Modificación campo electromagnético	Medida campo magnético
	Potencial espontáneo
	Estudio corrientes telúricas
Cambio propiedades electromagnéticas	Resistividad Magnetotelúrica
Radioactividad	Medida radón

Tabla 1. Fenómenos físicos relacionados con la actividad volcánica (Ortiz, R. en Astiz y García Ed. 2000)

El incremento de presión en el sistema volcánico y el movimiento del magma y de los fluidos asociados, son capaces de producir fracturas y generar terremotos de diferentes características en función del mecanismo de generación. En general, la actividad sísmica de un volcán, incluso en las fases más activas, suele ser de moderada magnitud, generalmente no superior a 4, aunque en la zona se produzcan grandes terremotos asociados a otros procesos geodinámicos. Los eventos que suceden son difíciles de clasificar, existiendo diferencias debido a la forma de describir la señal, bien por su apariencia y contenido en frecuencia o bien por el mecanismo de la fuente. Una forma de clasificarlas es por la duración de la señal, existiendo dos categorías: las transitorias y las continuas. Dentro de las señales transitorias estarían los terremotos volcano-tectónicos, los eventos de baja frecuencia, los híbridos y las explosiones y como señales continuas se incluyen los tremor volcánicos.

18.2. Terremotos volcano-tectónicos

Los terremotos volcano-tectónicos (VT) se pueden dividir en dos clases dependiendo de la profundidad a la que se generan. Los más profundos, denominados VT-A, tienen su hipocentro

por debajo de los 2 kilómetros y su forma de onda aparece con llegadas muy claras e impulsivas de las fases P y S y con un alto contenido en frecuencia por encima de los 5 Hz, pudiendo incluso superar los 30 Hz, por lo que también se les denomina eventos de alta frecuencia (HF) (figura 1). Su mecanismo focal se origina por un fractura de cizalla causada por una acumulación de esfuerzos, dando lugar a una dislocación sobre el plano de falla similar a la fracturas de origen tectónico. La incógnita es saber el origen de esa acumulación de esfuerzos, pudiendo ser causado por la tectónica regional o bien por una interacción con la dinámica de los fluidos presentes en el sistema volcánico.

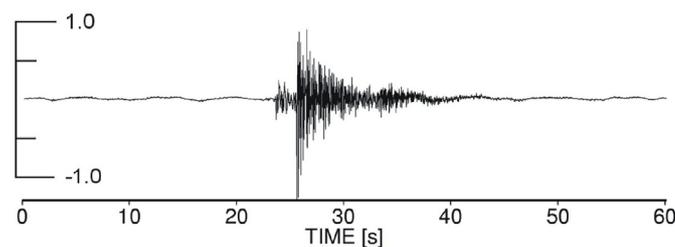


Figura 1. Terremoto volcánico-tectónico VT del tipo A registrado en el volcán Merapi, Indonesia.

Fuente: Wasserman, J. Volcano Seismology. New Manual of Seismological Observatory Practice

Una de las diferencias que marcan este tipo de eventos de los tectónicos es que frecuentemente ocurren en forma de enjambres, donde los eventos de mayor magnitud son muy similares y no ocurren al principio de la secuencia, por lo que no siguen la típica distribución de terremotos tectónicos de principal mas replicas.

Los más superficiales, entre 1 y 2 Km de profundidad (llamados VT-B) muestran una onda P más emergente y en algunos casos no es posible detectar la llegada de la fase S. La banda espectral de este tipo de eventos está desplazada hacia las bajas frecuencias (1-5 Hz) y su mecanismo en el foco puede consistir en una fuente simple de doble par, si bien estudios recientes realizados mediante la inversión del tensor de momento sísmico muestran una importante contribución con parte de no doble par. Este tipo de terremotos se pueden registrar de forma muy normal en los volcanes independientemente de su actividad, siendo además el único síntoma que se puede presentar en regiones volcánicas consideradas en reposo.

18. 3. Eventos de baja frecuencia

Los eventos de baja frecuencia (LF, también llamados de largo periodo LP) no muestran llegadas de ondas S, teniendo una señal muy emergente, lo que dificulta definir el principio de la señal. El contenido de frecuencias está la mayoría de las veces restringido a una estrecha

banda entre 1 y 3 Hz y las fuentes que los generan están situadas generalmente en la parte superficial del volcán (< 2 Km), aunque se han detectado señales de este tipo originadas a mas de 20 Km de profundidad (volcanes Kilauea o Pinatubo) (figura 2).

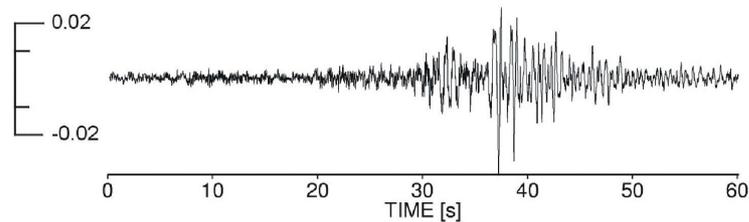


Figura 2. . Evento LF registrado en el volcán Merapi, Indonesia.

Fuente: Wasserman, J. Volcano Seismology. New Manual of Seismological Observatory Practice

En algunas regiones aparecen unos eventos muy característicos que por su forma especial reciben el nombre muy descriptivo de “tornillos” (figura 3).

En general, debido a su forma de onda, estos eventos LF son difíciles de localizar, por lo que se utilizan otras técnicas diferentes de las clásicas, como curvas amplitud-distancia realizadas a partir de terremotos bien localizados, o mediante técnicas de semejanza a partir del movimiento de la partícula en estaciones de tres componentes y banda ancha. Un modelo de fuente que intenta explicar su origen es la apertura de una grieta y su entrada en resonancia cuando el magma asciende hacia la superficie, o bien a la existencia de presiones transitorias dentro de la mezcla gas-fluido que causan fenómenos de resonancia en el mismo magma. La ocurrencia temporal de eventos de tipo LF (o LP) es a menudo en forma de enjambre, con un número elevado de eventos por hora y se ha podido observar una fuerte relación entre la ocurrencia de este tipo de eventos y una erupción volcánica próxima.

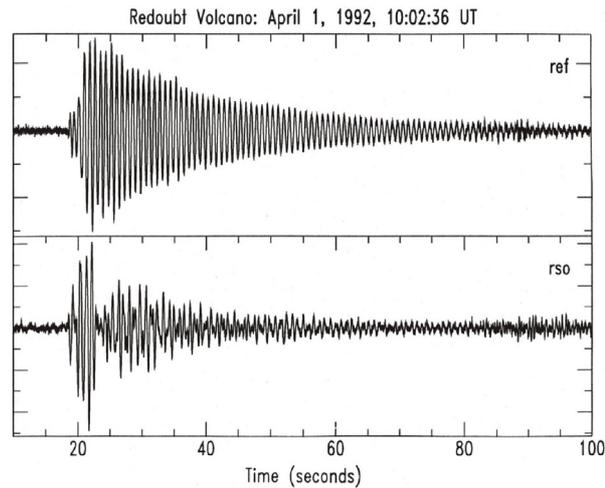


Figura 3. Evento LF conocido como tornillo registrado en dos estaciones en el volcán Redoubt (Alaska)
Fuente: Wasserman, J. Volcano Seismology. New Manual of Seismological Observatory Practice

18.4. Eventos híbridos

Algunas señales sismo volcánicas presentan las características de los eventos de baja frecuencia, LF, y también las de los tectono-volcánicos, VT-(A,B), por lo que se denominan *híbridos* (figura 4). Estos eventos presentan un comienzo de señal de altas frecuencias, en una banda hasta de 10 Hz, donde es posible identificar las fases P y S, seguidas a continuación de una señal parecida en su forma de onda, contenido espectral y duración a la de los terremotos de baja frecuencia. La presencia de estos eventos, tanto en el espacio como en el tiempo, es similar a la de los de tipo LF, por lo que se asocian a episodios pre-eruptivos muy próximos.

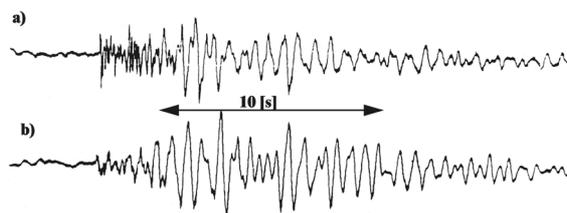


Figura 4. a) Evento híbrido. b) Evento tectono-volcánico VT-B para comparar.
Fuente: Wasserman, J. Volcano Seismology. New Manual of Seismological Observatory Practice

Como los eventos híbridos son la suma de un pequeño terremoto y un evento LF ocurridos simultáneamente, se puede asumir como modelo de la fuente una fractura sellada y sometida a presión de los fluidos volcánicos, cuyo aumento produce la ruptura de la zona generando un terremoto y rellenándose dicha fractura de fluido cuya entrada en resonancia, siguiendo el modelo de LF, produciría la señal monocromática a bajas frecuencias.

18. 5. Tremor volcánico

En la clasificación de las señales volcánicas habíamos considerado como continua al denominado tremor volcánico debido a que su señal puede durar entre algunos minutos y días (figura 5). En estas señales no es posible identificar fases P y S y algunos experimentos han permitido caracterizar el campo de ondas del tremor volcánico como una superposición de ondas superficiales de tipo Lg y Rg. Debido a esta ausencia de fases, la localización es complicada y hay que recurrir a antenas sísmicas, con mayor capacidad para determinar la posición espacial.

Su espectro está situado en una banda muy estrecha (entre 1 y 5 Hz) y las frecuencias centrales de sus máximos son generalmente estables y no vienen condicionadas por las características de la erupción. Asimismo, presentan grandes similitudes con los espectros de los eventos de baja frecuencia (LF), lo que obliga a proponer mecanismos en la fuente también similares. Algunos volcanes que contienen lava de alta viscosidad (Semeru en Indonesia o Lascar en Chile) pueden presentar en su espectro sobretonos o armónicos.

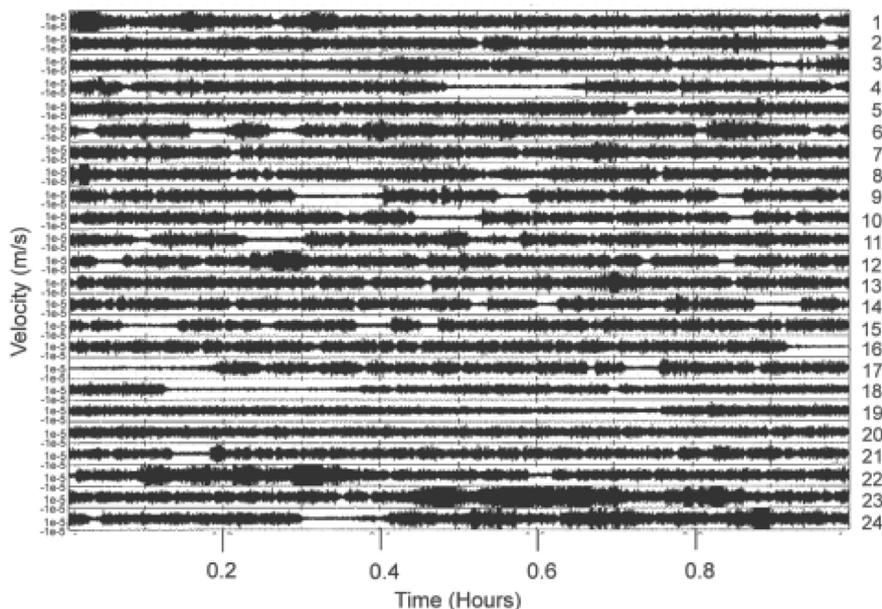


Figura 5. Tremor volcánico en el volcán Bromo (Indonesia) durante una fase de gran actividad.

Fuente: Wasserman, J. Volcano Seismology. New Manual of Seismological Observatory Practice

Atendiendo a su contenido espectral, este tipo de señales se pueden dividir en tres subgrupos:

T_H . Altas frecuencias (> 6 Hz).

T_I . Frecuencias intermedias (1- 6 Hz).

T_L . Bajas frecuencias ($< 0,5$ Hz).

En algunos volcanes el contenido espectral está relacionado con el tipo de actividad y por tanto con la fuente, estando centrado a bajas frecuencias (1 Hz) en fases tranquilas y con valores de alrededor de 3 Hz en fases pre-eruptivas.

Las señales de algunos tremor presentan grandes picos de amplitud, pero muy cortos, mientras otras señales son bastantes estacionarias durante días, o incluso meses. Las fuertes fluctuaciones de amplitud, como sucede en el volcán Etna, se asocian a manantiales de lava o posteriores a la apertura de una fisura lateral. En volcanes como el Stromboli parece existir una correlación entre la amplitud y los cambios de presión atmosférica, lluvia o mareas marinas.

La dificultad de conocer la localización de la fuente de este tipo de eventos implica también la dificultad de establecer modelos que expliquen su generación y al mismo tiempo, estos modelos no deben ser únicos dado la gran variedad en el contenido espectral. En el caso del tremor T_H su origen puede ser explicado asociándolo a avalanchas, caída de rocas desde el frente de lava o grandes desgasificaciones de cráteres. El clasificado como T_L estaría asociado al transporte de magma, a oscilaciones del conjunto de la cámara magmática o a resonancia del tipo tubo de órgano de grandes conductos.

Por último, el tercer tipo de tremor, T_I , es el que más se suele registrar y sobre el que más modelos existen pero también más desconocimiento. Los modelos de resonancia de conductos abiertos son capaces de explicar numerosos episodios de tremor, sin embargo a veces no son capaces de integrar datos sobre la reología de los fluidos y su dinámica. Actualmente, algunos resultados muestran evidencias de que temores de ciertas regiones volcánicas son la suma temporal de eventos LF y en este caso los mecanismos del tremor serian los mismos que los de LF, pero con los transitorios de presión siendo continuos en el tiempo.

Un tipo de señal transitoria, que generalmente acompaña a una señal de tremor, es el de las explosiones en los edificios volcánicos. Su registro aparece superpuesto al tremor de fondo y son identificables por un aumento brusco de la amplitud y por un incremento de la frecuencia (figura 6). La mayoría de las veces, esta señal presenta dos llegadas diferentes y claras, la primera asociada a la propagación de ondas internas o superficiales debidas a la explosión. La segunda, son ondas de aire, de choque y ondas sonoras, con la velocidad de propagación típica del sonido (330 m/s).

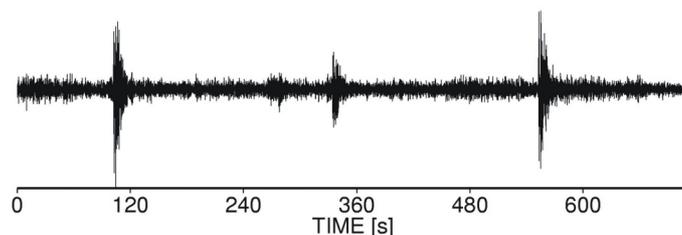


Figura 6. Explosiones superpuestas a tremor volcánico en el volcán Stromboli (Italia).
Fuente: Wasserman, J. Volcano Seismology. New Manual of Seismological Observatory Practice

18. 6. Precursores sísmicos de la actividad volcánica

Dentro de los fenómenos que suceden asociados a un volcán, podemos considerar la actividad sísmica como uno de los precursores más determinantes para diagnosticar una posible erupción posterior. Las características a tener en cuenta son el tipo de señal sísmica que se registra, su contenido espectral, la cuantificación de energía liberada y sus distribuciones espaciales y temporales.

Como tipo de señal más determinante está el tremor volcánico, que es siempre un signo de alta actividad, aunque como su exacto mecanismo es todavía desconocido, la importancia y el tiempo entre su primera aparición y una posible actividad eruptiva sigue siendo materia de discusión.

Los cambios en las propiedades espectrales de las señales sísmicas sirven para caracterizar el estado de la actividad, existiendo varias técnicas para su cálculo como SSAM (*Seismic Spectral Amplitude Measurement*), transformadas de Fourier o la estima de la densidad de potencia espectral (figura 7).

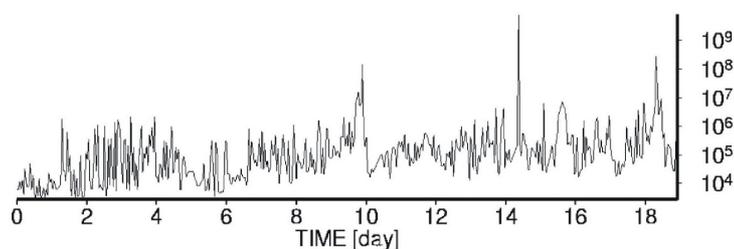


Figura 7. Potencia total (cada 60 minutos) calculada en la banda de frecuencia 0,6-3,0 Hz desde el 1 al 19 de julio de 1998 en el volcán Merapi (Indonesia). Dos de los picos más visibles (días 9 y 18) están asociados a flujos piroclásticos, mientras que el mas fuerte (día 14) corresponde a un terremoto regional. Fuente: Wasserman, J. Volcano Seismology. New Manual of Seismological Observatory Practice

Para determinar la energía que se va liberando en un proceso volcánico y observar los cambios de actividad, existe una técnica en el dominio del tiempo llamada RSAM (*Real Time Seismic Amplitude Measurement*), que viene definida por:

$$RSAM(T) = \frac{1}{T} \sum_{t=T-\frac{T}{2}}^{T+\frac{T}{2}} |s(t)|$$

donde T es el intervalo para promediar y $s(t)$ los valores muestreados de la traza de la señal (figura 8).

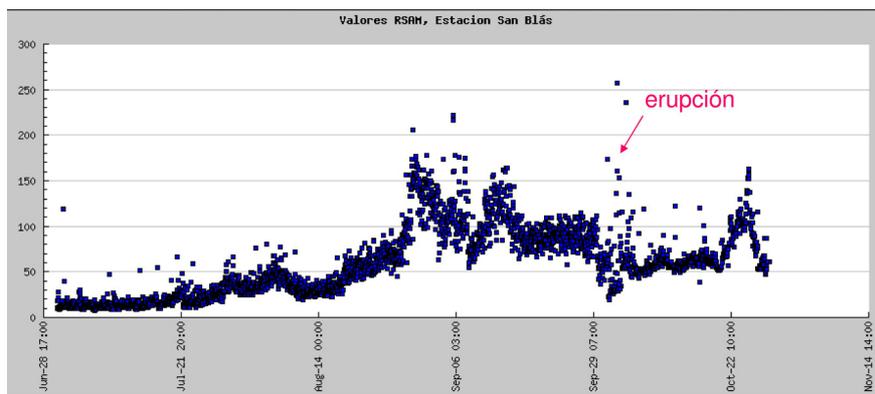


Figura 8. Valores de RSAM durante la fase de actividad en el volcán Santa Ana (El Salvador). (Cortesía de B. Benito)

La determinación del RSAM puede realizarse solamente con estaciones de una sola componente, sin embargo en el caso de disponer de estaciones con tres componentes es posible analizar los cambios de actividad volcánica utilizando técnicas de polarización. Esta metodología permite calcular los parámetros del movimiento de la partícula de las señales sísmicas y en particular el ángulo de incidencia y el acimut y su variación a lo largo del tiempo.

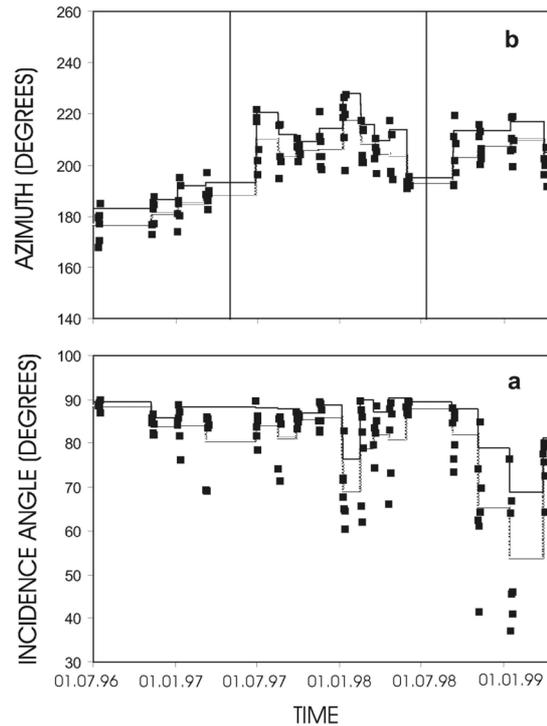


Figura 9. . Variaciones para largos periodos de tiempo del ángulo de incidencia (a) y acimut (b) en el volcán Stromboli (Italia).

Fuente: Wasserman, J. Volcano Seismology. New Manual of Seismological Observatory Practice

Uno de los aspectos más significativos de la actividad volcánica es el conocimiento de donde se producen los eventos sísmicos, es decir sus coordenadas epicentrales y la profundidad. Su distribución aporta una excelente información sobre el comportamiento, las características y la geometría del volcán. Asimismo, la migración o movimiento de los hipocentros durante la actividad puede dar información de una inminente erupción (figura 10). Mientras las señales continuas, tales como los tremor o las transitorias de eventos LF, que no tienen una llegada clara de fases, tienen problemas para su localización, otro tipo de eventos como los tectono-volcánicos o explosiones, pueden ser localizados mediante técnicas clásicas a partir de una red de estaciones sísmicas distribuidas alrededor del volcán.

Uno de los problemas en la localización es el uso de buenos modelos de velocidad, los cuales en la mayor parte de los volcanes, no satisfacen los criterios necesarios para obtener buenos hipocentros. Actualmente, modelos de velocidad 1D o 2D están produciendo mejores resultados, así como localizaciones relativas mediante similitudes en la forma de onda.

Para la localización de eventos como el tremor volcánico, o los de baja frecuencia, donde los procedimientos clásicos no funcionan, es posible utilizar otras técnicas como la aportada por las antenas sísmica (o arrays). Su configuración permite determinar la posición de las fuentes que

originan esas señales, teniendo la ventaja adicional de poder ubicarse en un lateral del volcán, sin necesidad de tenerlo rodeado. Asimismo, esta técnica permite analizar señales muy debiles.

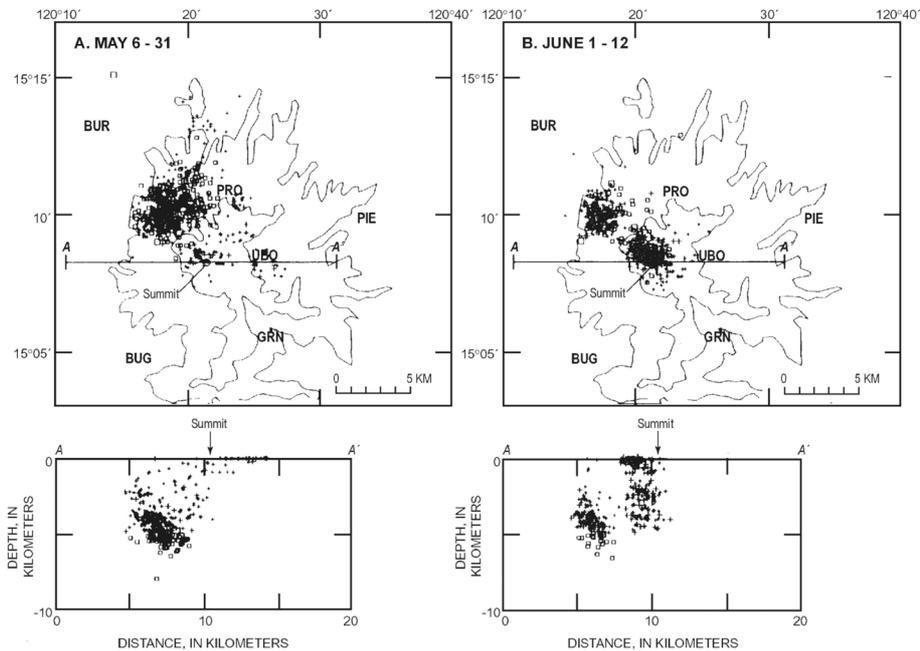


Figura 10. A) Sismicidad en el volcán Pinatubo entre los días 6 y 31 de mayo, con una concentración epicentral hacia el noroeste del centro del volcán. B) La sismicidad entre el 1 y 12 de junio, donde se aprecia un desplazamiento de los hipocentros hacia la superficie y la actividad más concentrada en la cima del volcán. Wasserman, J. Volcano Seismology. New Manual of Seismological Observatory Practice

Los parámetros que aportan las antenas son el ángulo de incidencia, el azimut y la lentitud (o velocidad aparente) y si por el tipo de señal no se pudiese determinar la distancia mediante diferencias de fases, se necesitaría otra antena para determinar la posición de la fuente.

Una de las formas más usuales de disponer de información del volcán es realizar estadísticas del número de eventos, bien por su tipo o por su distribución en el tiempo (figura 11), lo que nos puede permitir observar su evolución y caracterizar el estado del volcán.

Asimismo, para determinar la energía que se va liberando en un proceso eruptivo o representar una distribución en el tiempo del tamaño de los eventos, es necesario calcular su magnitud. En el caso de terremotos tectono-volcánicos, como señales transitorias, se usan expresiones en función de la duración de la señal:

$$M_{\tau} = 2 \log \tau + 0,0035 \Delta - 0,87$$

donde τ es la duración de la señal y Δ la distancia epicentral en kilómetros. En el caso de tremor volcánico, donde resulta una duración excesiva hay que recurrir a expresiones de la magnitud local basadas en la amplitud:

$$M_L = \log A - \log A_0$$

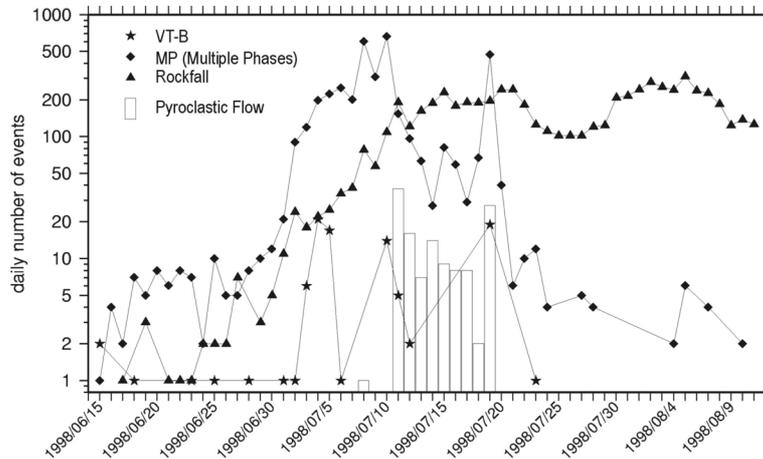


Figura 11. Distribución del número de eventos por día durante un periodo de alta actividad (julio 1998) en el volcán Merapi. Fuente: Wasserman, J. Volcano Seismology. New Manual of Seismological Observatory Practice

18. 7. Otros precursores

Aunque el uso de la actividad sísmica es generalmente el más importante elemento predictivo de una erupción volcánica, su análisis exclusivo resulta inadecuado para determinar con la fiabilidad exigida el estado y evolución de un volcán. Existen otros fenómenos geofísicos, geodésicos, geoquímicos o geológicos que permiten predecir la situación y posible erupción de un sistema volcánico. En la tabla 1 se señalaban los aspectos físicos que un volcán lleva asociado a su proceso y algunos de ellos, para determinados volcanes y tipos de actividad, pueden ser considerados como precursores.

Muy asociado a la actividad sísmica de un proceso eruptivo estaría la deformación causada por inyección de magma o por presurización hidrotermal en el edificio volcánico. En algunos volcanes se ha observado una buena correlación entre la deformación y la aparición de señales sísmicas. Por ello, una deformación medida en superficie y su evolución temporal, puede ser un elemento predictivo de primer orden. Para la correcta medida de las deformaciones deben eliminarse las originadas por efectos indirectos como la temperatura, el viento, las mareas, los cambios de presión, etc., lo que requiere que se utilicen técnicas capaces de medir con gran resolución, como inclinómetros, extensómetros, medida electrónica de distancias o más

modernas como las redes de estaciones GPS. Actualmente se están desarrollando metodologías de interferometría de radar, lo que permite realizar un control remoto de la deformación.

También asociados a la actividad volcánica, se pueden generar cambios en la densidad, volumen y localización del material magnético, lo que da lugar a modificaciones del campo gravitatorio observables previamente a la erupción. Sin embargo, estos cambios también pueden deberse a las deformaciones de la superficie del volcán, por lo que se considera necesario acompañar los estudios de gravedad con la nivelación de alta precisión.

En ciertas regiones volcánicas, las erupciones están precedidas por señales magnéticas de varios días de duración, apareciendo progresivamente en los últimos días previos a las erupciones. Para su observación y caracterización se diseñan redes de estaciones magnéticas próximas al volcán, a las que se añade una estación de referencia lo suficientemente alejada para no detectar el campo anómalo creado por la actividad directa del volcán. Por ello, las señales identificadas en las estaciones próximas al cráter y no en la estación de referencia, se consideran como propias del volcán y permiten analizar su evolución.

Otro tipo de fenómenos geofísicos que habíamos visto anteriormente asociados a un sistema volcánico, son también susceptibles de ser considerados como precursores, al menos en ciertos tipos de volcanes. Estos serían los asociados a cambios en el potencial espontáneo, en la resistividad de las rocas o en las emisiones de radón.

La composición química de las aguas y de los gases presentes en las emisiones de un volcán, son reflejo de su actividad. Se considera que las emisiones gaseosas están en equilibrio cuando el volcán se encuentra en reposo y cuando evoluciona, se produce un desequilibrio en la composición de los gases. El gas volcánico tiene como componentes, además del vapor de agua, SO₂, H₂S, CO₂, CO, HCl, HF y otros en concentración mucho menor. De todos estos el SO₂ y el CO₂ se consideran los componentes más significativos de la presencia de magma. Además de la composición, existen otros parámetros de los gases que su variación precede a una erupción, como son el volumen, la velocidad de emisión o la temperatura.

Bibliografía

- [1] Astiz, M. y Garcia, A. (Editores científicos). Curso Internacional de Volcanología y Geofísica Volcánica. Serie Casa de los Volcanes nº 7. Ed. Cabildo Insular de Lanzarote. 2000.
- [2] Wasserman, J. Volcano Seismology. New Manual of Seismological Observatory Practice. P. Bormann Ed., GeoForschungsZentrum Potsdam (GFZ). Cap. 13.

Tema 19. Análisis armónico del campo geomagnético. El campo geomagnético internacional de referencia. Cartografía magnética. Anomalías magnéticas. Levantamientos magnéticos: terrestres, aéreos y marítimos.

19.1. Análisis armónico del campo geomagnético

El campo que se mide en la superficie terrestre, es la suma vectorial de un número de diferentes campos constituyentes, cada uno de ellos originados en un lugar diferente y varía en el tiempo y en el espacio de maneras distintas. De las observaciones se deduce que el campo está formado fundamentalmente por dos constituyentes: uno global de gran amplitud, cuya fuente debe estar situada en el interior de la tierra y otro, de amplitud mucho menor y que debe ser creado por fuentes mucho más superficiales. Conviene tratar a estos como dos campos diferentes, y se les suele llamar "campo fundamental" al de origen interno y "campo local" al de origen superficial, superpuesto a estos, está el "campo externo", producido por corrientes eléctricas que circulan por la Ionosfera y Magnetosfera, que al variar en el tiempo generan campos inducidos, que se suman al "campo anómalo o cortical", originado por rocas magnetizadas, bien inducida, bien remanente, para crear lo que hemos llamado campo local.

El campo fundamental puede ser definido como el campo medio medido durante un año y sobre un área de 10^6 km^2 (Parkinson, 1983). Este campo se representó por primera vez en el siglo XVII, como el creado por un dipolo en el centro de la Tierra y orientado según su eje de rotación, siendo una mejor aproximación la de inclinar 11° dicho dipolo, "dipolo inclinado" o "dipolo centrado"; por las cartas magnéticas se ve que esta aproximación tampoco se ajusta bien, obteniendo una aproximación aun mejor si desplazamos el dipolo del centro de la Tierra hacia el Pacífico occidental, "dipolo excéntrico". Esta última aproximación puede llegar a representar el 97% del campo real, (Torta, 1992), pero aun no explica algunas características que aparecen en los mapas, necesitando añadir nuevos términos. Una manera de hacerlo es añadiendo dipolos en el interior de la Tierra en diferentes posiciones, orientaciones y momentos dipolares, métodos propuestos por Bochev(1969), Alldredge y Hurwitz (1964), pero resulta más sencillo de manejar analíticamente la representación multipolar, que consiste en añadir multipolos magnéticos en el centro de la Tierra.

A continuación repasaremos unos pocos conceptos generales de magnetismo y del análisis de armónicos esféricos, estos conocimientos vienen en cualquier libro de magnetismo (p.e. Reitz y Milford, 1972; Lorrain y Corson, 1972) y de geomagnetismo (Chapman y Bartels, 1940; Parkinson, 1983; Jacobs, 1987).

Considerando que las medidas magnéticas, en las que se basa el análisis del campo fundamental, se realizan en la atmósfera inferior, y aceptando que la atmósfera es un aislante, entonces las medidas se realizan en una zona libre de corrientes, por lo cual las ecuaciones de Maxwell para el campo magnético, se pueden escribir en la forma:

$$\begin{aligned} \vec{\Delta} \times \vec{B} &= 0 \\ (1) \quad \vec{\Delta} \cdot \vec{B} &= 0 \end{aligned}$$

por lo que:

$$(2) \quad \vec{B} = -\Delta V$$

de modo que V debe satisfacer la ecuación de Laplace:

$$(3) \quad \Delta^2 V = 0$$

En coordenadas esféricas la solución de esta ecuación es de la forma:

(4)

$$\begin{aligned} V(r, \theta, \phi) &= a \left[\sum_{n=1}^{\infty} \sum_{m=0}^n \left(\frac{r}{a}\right)^n (b_n^m \cos m \phi + c_n^m \sin m \phi) P_n^m(\cos \theta) \right. \\ &\quad \left. + \sum_{n=1}^{\infty} \sum_{m=0}^n \left(\frac{a}{r}\right)^{n+1} (g_n^m \cos m \phi + h_n^m \sin m \phi) P_n^m(\cos \theta) \right] \end{aligned}$$

siendo:

a, el radio de la misma, igual aproximadamente al radio medio (6371.2km); r, la distancia radial desde el centro de la Tierra; θ, ϕ la colatitud y la longitud geocéntricas; $P_n^m \cos(\theta)$ las funciones asociadas de Legendre de grado n y orden m del tipo Schmidt quasi-normalizadas (Chapman y Bartels, 1940); N el máximo grado y orden; y $g_n^m, h_n^m, b_n^m, c_n^m$ los coeficientes armónicos esféricos en nanoteslas, que se determinaran por ajuste con el campo observado.

Debido a las condiciones de contorno, el sumatorio comienza en n=1 puesto que n=0 en el primer término correspondería a V= constante, campo nulo, y n=0 en el segundo correspondería a un monopolio en el centro de la tierra, que violaría las ecuaciones de Maxwell. También están excluidas las funciones de Legendre de segunda clase Q_n^m que implicaría un punto singular del campo en el polo .

Dividiendo el espacio en tres regiones: R_1 correspondiente al interior de la tierra, R_2 una fina capa justo por encima de la superficie y en la que se realizan las medidas, y R_3 la región externa a la anterior. Ninguna de las fuentes están en R_2 , por lo que podremos expresar el potencial en esta región como:

$$(5) \quad V = V_i + V_e$$

donde V_i es el potencial del campo originado en R_1 y V_e es el creado en R_3 . Dado que V_i no tiene fuentes en R_3 debe obedecer la ecuación de Laplace en R_2 y en R_3 , pero si lo expresamos mediante la ecuación (4), todos los términos del primer sumatorio crecen indefinidamente al crecer r salvo si todos los coeficientes b_n^m y c_n^m son cero. Es decir la expresión armónica para V_i contendrá el segundo sumatorio pero no el primero. De manera similar el campo descrito por V_e no tiene fuentes en R_2 y en R_1 y por tanto V_e debe obedecer la ecuación de Laplace en todo punto de R_1 , incluido $r=0$, por tanto en la expresión (4) para V_e todos los coeficientes g_n^m y h_n^m han de ser nulos, quedando solo el primer sumatorio. De aquí se deduce que si el potencial de un campo observado se expresa mediante (4) se determinan los coeficientes, el primer sumatorio debe describir un campo cuyas fuentes están todas en R_3 y el segundo debe describir uno cuyas fuentes estén en R_1 . Es decir el primer sumatorio expresa un campo externo y el segundo uno interno, por lo que para determinar los coeficientes podemos dividir el campo en dos partes una de origen interno y otra de origen externo.

El potencial sin embargo no es un observable, pudiendo solo medir las componentes del campo, cuyas expresiones son:

$$\begin{aligned} X &= -B_\theta = \frac{1}{r} \frac{\partial V}{\partial \theta} \\ (6) \quad Y &= B_\phi = -\frac{1}{r \sin \theta} \frac{\partial V}{\partial \phi} \\ Z &= -B_r = \frac{\partial V}{\partial r} \end{aligned}$$

obteniendo fácilmente de ellas los otros elementos de campo

$$\begin{aligned} F &= (X^2 + Y^2 + Z^2)^{1/2} \\ (7) \quad H &= (X^2 + Y^2)^{1/2} \\ D &= \arctg Y/X \\ I &= \arctg Z/H \end{aligned}$$

siendo F , la intensidad total, H la componente horizontal, D , la declinación e I , la inclinación.

Diferenciando 6 y sustituyendo $r=a$, obtenemos en la superficie:

$$\begin{aligned} (8) \quad X &= \sum_{n=1}^{\infty} \sum_{m=0}^n [(b_n^m + g_n^m) \cos m \phi + (c_n^m + h_n^m) \sin m \phi] (d/d\theta) P_n^m(\cos \theta) \\ Y &= \sum_{n=1}^{\infty} \sum_{m=0}^n [m(b_n^m + g_n^m) \sin m \phi - m(c_n^m + h_n^m) \cos m \phi] P_n^m(\cos \theta) / \sin \theta \\ Z &= \sum_{n=1}^{\infty} \sum_{m=0}^n [(n b_n^m - (n+1) g_n^m) \cos m \phi + (n c_n^m - (n+1) h_n^m) \sin m \phi] P_n^m(\cos \theta) \end{aligned}$$

Los resultados de los ajustes (Hurwitz et al., 1966) indican que los coeficientes b y c son mucho menores que los g y h , que corroboran la hipótesis de un campo fundamental de origen interno.

Así mismo, los términos de la expansión en armónicos esféricos representan la aproximación multipolar antes mencionada, el primero ($n=1, m=0$) corresponde a un dipolo situado en el centro de la Tierra y siguiendo el eje de rotación de la misma, los términos $n=1$ representan al dipolo inclinado, los $n=2$ el cuadrupolo, $n=3$ al octopolo, etc.

Analicemos el segundo sumatorio de la ecuación (4), es decir consideremos solo el campo de origen interno, expresando explícitamente los polinomios asociados de Legendre, tenemos:

$$(9)$$

$$V = g_1 (a^3/r^2) \cos\theta + (a^3/r^2)(g_1^1 \cos\phi + h_1^1 \sin\phi + (a^4/r^3) g_2^{1/2} (3\cos^2\theta - 1) + \\ + (a^4/r^3) g_2^1 (3\cos\phi + h_2^1 \sin\phi) 3^{-1/2} \sin\theta + \\ + (a^4/r^3) g_2^2 (\cos 2\phi + h_2^2 \sin 2\phi) (3/4)^{1/2} \sin^2\theta + \dots$$

los coeficientes se denominan “coeficientes de Gauss”.

El primer término $n=1, m=0$ representa un dipolo con un momento magnético $m = (4\pi/\mu_0) a^3 g_1$ cuyo eje coincide con el eje del sistema de coordenadas, es decir el eje de rotación de la tierra.

El segundo término, $n=1, m=1$, $n=$ representa un dipolo con el eje en el plano ecuatorial en una dirección con longitud $\tan^{-1}(h_1^1 / g_1^1)$, es decir los dos primeros términos se combinan para formar el campo de un dipolo inclinado, con una inclinación respecto al eje de rotación de la tierra $\tan^{-1}[\{(g_1^1)^2 + (h_1^1)^2\}^{1/2}/g_1]$.

Los tres siguientes términos, dependientes de r^{-3} , se combinan para crear un campo cuadripolar. El primero de ellos representa un cuadrupolo con simetría axial, es decir dos dipolos axiales opuestos desplazados a lo largo del eje. El segundo representa dipolos axiales desplazados en el plano ecuatorial y el tercero, representa dipolos ecuatoriales desplazados en el plano ecuatorial. Los siguientes términos en r^{-4} representan el campo de un octopolo. Así más y más detalle se va añadiendo al ir agregando multipolos de mayor grado todos en el centro de la tierra.

Los términos dipolares y cuadripolares especifican un dipolo excéntrico. Si el origen de coordenadas se traslada a la posición del dipolo excéntrico (llamado centro magnético) y se elige un eje que coincida con el del dipolo excéntrico, los coeficientes g_1^1, h_1^1, g_2^1 y h_2^1 pueden anularse, pero no así g_2^2 y h_2^2 . Volviendo al dipolo centrado o inclinado, los puntos donde estos cortan la superficie de la tierra se llaman “polos geomagnéticos”. El polo norte geomagnético tiene por coordenadas geográficas (78.8°N, 70.9°W) de acuerdo con el IGRF de 1980. No hay que confundir los polos geomagnéticos con los polos magnéticos, que son los puntos donde la media temporal del campo magnético es vertical. Los polos geomagnéticos se calculan a partir del análisis en todo el mundo que determinan los dos primeros términos de (9). Los polos magnéticos solo se pueden calcular a partir de medidas reales en el campo, puesto que su posición no solo está influenciada por el cuadrupolo y armónicos mayores del campo fundamental, sino también por el campo local.

El diámetro que une los polos geomagnéticos, se llama “eje geomagnético” y puede usarse como eje de referencia para coordenadas geomagnéticas.

El “campo no dipolar” es simplemente el campo fundamental menos el dipolo centrado. Se puede aproximar mediante todos los términos de la expresión (9) menos los tres primeros. Estos términos al depender de potencias negativas de r , tienen cada vez menos importancia y a suficiente distancia de la superficie de la tierra el campo se reduce al del dipolo centrado.

19.2. El campo geomagnético internacional de referencia

El Campo Internacional de Referencia, IGRF, es una aproximación del campo magnético fundamental de la Tierra, tal como hemos visto en el apartado anterior al campo de origen interno, excluyendo tanto las contribuciones causadas por los materiales magnéticos de la corteza, como las causadas por las corrientes eléctricas inducidas por campos externos.

El IGRF fue concebido por la necesidad de tener un campo de referencia que permitiera crear una base acordada para cálculos del campo fundamental y para unificar resultados en diversos estudios magnéticos, como por ejemplo, obtención de anomalías superficiales, cálculos de corrientes ionosféricas y magnetosféricas, localización de puntos conjugados, etc...(Zmuda, 1971).

Siendo el IGRF una aproximación al campo fundamental y según lo visto en el apartado anterior, su expresión analítica será:

$$(10) \quad V(r, \theta, \phi) = a \sum_{n=1}^N \sum_{m=0}^n \left(\frac{a}{r}\right)^{n+1} (g_n^m \cos m\phi + h_n^m \sin m\phi) P_n^m(\cos \theta)$$

donde los coeficientes se determinan a partir de modelos propuestos por diferentes grupos de trabajo.

EVOLUCIÓN EN EL DESARROLLO DEL IGRF.

Para entender mejor este campo de referencia haremos un breve resumen de la historia de los distintos campos internacionales de referencia (Peddie, 1982).

1ª y 2ª Generaciones del IGRF.

El IGRF original o primera generación fue adoptada en 1968 (IAGA Commission 2 Working Group 4, 1969). Consistía en un modelo del campo fundamental en 1965.0 y en un modelo de su variación secular, que juntos constituían el IGRF 1965. Cada modelo estaba definido por armónicos esféricos de orden y grado octavo, el modelo para la variación secular venía dado de forma que el modelo del campo fundamental pudiera ser extrapolado en el tiempo en los dos sentidos, el límite hacia atrás era 1955.0 y para adelante 1975.0. Zmuda (1971) Y Peddie(1982) describen con todo detalle el desarrollo del IGRF

original, incluyendo evaluaciones de los modelos propuestos, mapas de los elementos magnéticos y una larga lista de referencias.

Debido a la complejidad de la variación secular del campo fundamental, el modelo de esta está incluido en el IGRF 1965, y de hecho en todos los modelos asociados con el IGRF, es simplemente un modelo lineal, que especifican cambios constantes en los coeficientes del campo fundamental. Estos tipos de modelos pueden aproximarse a la variación secular, pero pueden solo para períodos de unos pocos años. Después de una extrapolación de diez años, llegando a 1975, el IGRF 1965 es muy inexacto para algunas regiones (Dawson y Newitt, 1978; Regan y Cain, 1975). La primera revisión del IGRF corrigió el modelo de variación secular pero no corrigió el error que se había acumulado en los coeficientes del campo fundamental después de ser extrapolado a 1975.

La segunda generación de IGRF después de esta primera revisión fue adoptada en 1975 (IAGA Division I Study Group, 1975, 1976, 1977). Consistía en un modelo del campo fundamental para 1975.0 y un modelo de predicción de la variación secular durante 1975-1980, juntos forman el IGRF 1975. El modelo del campo fundamental era simplemente el modelo del IGRF 1965 extrapolado a esta época, mientras que el de la variación secular era nuevo y, como en el modelo anterior estaba definido por armónicos esféricos hasta el octavo orden y grado.

Al principio de los años 70, se sabía con certeza que simplemente revisando periódicamente el modelo de variación secular no mantendría al IGRF próximo al campo fundamental real. Para la segunda revisión se decidió que no se requeriría la continuidad del nuevo IGRF con modelos anteriores, también se decidió que la segunda revisión introduciría una serie de modelos definitivos que serían continuos y que representarían el campo magnético fundamental.

La última generación del IGRF (número 10), fue establecida por la IAGA en Diciembre de 2004, que lo define como “la última versión de una descripción matemática estandar del campo fundamental de la tierra y utilizado ampliamente en estudios del interior, de la corteza, de la ionosfera y de la magnetosfera de la tierra. Está formado por armónicos esféricos hasta el grado y orden 13 y los coeficientes se determinaron mediante ajustes a los datos de observatorios, de levantamientos magnéticos y de satélites.

Este campo teórico debe representar, en el conjunto del Globo, las anomalías regionales, mientras que las anomalías locales, que no vienen representadas en él, se manifiestan como un campo perturbador dentro de este campo regular. De esta forma se adopta, en la mayoría de los levantamientos aeromagnéticos este campo teórico como campo normal, representante del campo de origen profundo, y la diferencia entre el campo observado y el campo de referencia constituye el campo de anomalías superficiales. Sin embargo, este no es el único campo normal que se puede adoptar para una región,

pudiendo, como veremos en los apartados siguientes determinar otros, por aproximaciones polinómicas calculadas a partir de los datos observados en la región.

La IAGA, al implantar este nuevo modelo, recomienda que se utilice el datum y el esferoide del Sistema Geodético 1984 para la transformación de coordenadas en lugar del recomendado anteriormente (International Astronomical Union 1966). Debido a la alta precisión de los datos de satélite y para que el IGRF refleje ésta alta calidad, se extendieron el grado y orden de los coeficientes a 13 y se llega a la precisión de una décima de nanotesla.

Nombre completo		Válido para	Definitivo para
IGRF 10 generación (revisada 2005)	IGRF-10	1900.0-2010.0	1945.0-2000.0
IGRF 9 generación (revisada 2003)	IGRF-9	1900.0-2005.0	1945.0-2000.0
IGRF 8 generación (revisada 1999)	IGRF-8	1900.0-2005.0	1945.0-1990.0
IGRF 7 generación (revisada 1995)	IGRF-7	1900.0-2000.0	1945.0-1990.0
IGRF 6 generación (revisada 1991)	IGRF-6	1945.0-1995.0	1945.0-1985.0
IGRF 5 generación (revisada 1987)	IGRF-5	1945.0-1990.0	1945.0-1980.0
IGRF 4 generación (revisada 1985)	IGRF-4	1945.0-1990.0	1965.0-1980.0
IGRF 3 generación (revisada 1981)	IGRF-3	1965.0-1985.0	1965.0-1975.0
IGRF 2 generación (revisada 1975)	IGRF-2	1955.0-1980.0	-
IGRF 1 generación (revisada 1969)	IGRF-1	1955.0-1975.0	-

Tabla 1. Resumen de la historia del IGRF

La IAGA recomienda también que en cualquier estudio en el que se utilice el IGRF se especifique el nombre completo del modelo utilizado.

19.3. Cartografía magnética

Los valores de las componentes del campo geomagnético, D, H, I y F constituyen la cartografía geomagnética. En ellas se representan las líneas de igual valor de cada una, y reciben los nombres “isodinámicas totales” para F, “isodinámicas horizontales” para H, “isodinámicas verticales ” para Z, “isogonas” para D e “isoclinas” para I (esta última ya no se realiza). La línea de D=0 se llama línea áгона, I=0 constituye el ecuador magnético e I= 90 los polos magnéticos.

El Mapa Geomagnético de España (Península e Islas Baleares) para la época 2005.0, está compuesto por cuatro hojas que representan cada una, las distintas componentes del campo magnético terrestre. Estas componentes son: Declinación magnética, Componente horizontal, Componente vertical y Campo

total. Está realizado, siguiendo las recomendaciones de la I.A.G.A. referentes al intervalo de publicación de cada componente magnética. En el caso de la declinación magnética este intervalo se cifra en cinco años y para el resto de las componentes cada diez años. Las isolíneas de cada componente, han sido trazadas a partir de los valores, referidos a una fecha común del uno de enero de 2005 (2005.0), de las estaciones de mapa disponibles.

Para poder utilizar el mapa, fuera de la fecha indicada anteriormente, se han calculado los modelos isopóricos correspondientes a cada componente, trazándose con estos las correspondientes isopóras, que nos permitan una predicción en épocas posteriores a la de la realización del mapa. También se han utilizado estos modelos isopóricos para efectuar el traslado de los valores obtenidos para el mapa del año 2000 en el caso de la declinación y de 1995 en el caso de las otras componentes.

Instrumentación

Para el registro continuo de las variaciones de las componentes del campo magnético H, D y Z, y campo total F, a lo largo del tiempo, los observatorios disponen de variómetros y equipos de medidas absolutas. El Observatorio de San Pablo, dispone del siguiente instrumental:

Medidas absolutas:

Teodolito Zeiss 010 B y fluxgate Bartington.

Magnetómetros de efecto Overhauser GSM-19 y GSM-90, y magnetómetro de protones G-856AX

Variómetros:

Equipo Geomag M390

Magnetómetro Vector

Fluxgate FGE

Para la observación de las estaciones seculares y de mapa, se han utilizado los siguientes equipos:

Declinómetro / inclinómetro: compuesto de teodolito Zeiss 020B con electrónica y sonda Eda o Bartington.

Magnetómetro de protones G-856 AX

Con esta instrumentación, se determinan de forma directa los valores de la declinación magnética, inclinación y campo total (D, I y F respectivamente) y mediante su descomposición vectorial se obtienen los valores de la componente horizontal y vertical del campo magnético (H y Z).

Para la confección de los mapas de los elementos magnéticos y campo total, para la época 2005.0, se ha utilizado los datos de la red de estaciones geomagnéticas compuesta por: Observatorio de San Pablo de los Montes, estaciones de repetición y estaciones de mapa.

Reducción de los valores de las estaciones de repetición.

Debido a la variación temporal del campo magnético, es necesario reducir los valores a una época común. Las variaciones temporales más importantes son:

- La variación secular, que suele ser de unos nT para la intensidad y de 6 a 10nT para la declinación e inclinación. Esta variación esta relacionada con el campo interno.
- Las variaciones periódicas originadas por el campo externo.
- Las variaciones no periódicas, como las tormentas magnéticas, también de origen externo.
- Las pulsaciones de periodo muy corto y pequeña amplitud.

Debido a estas variaciones se deben dar los datos instantáneos, así como las medias diarias y anuales. Estas medias se obtienen en los observatorios a partir de los registros continuos.

Cuando se hacen medidas magnéticas en estaciones de campo (estaciones de repetición o de mapa) $E(x, t)$ hay que eliminar lo más posible la influencia de estas variaciones temporales. Las dos últimas (tormentas y pulsaciones magnéticas) se notan directamente en los valores medidos y se prescinde de los valores obtenidos durante estos fenómenos. Las variaciones periódicas se eliminan reduciendo a los valores medios anuales para dicho punto.

Hay diferentes métodos de reducción, antiguamente dada la precisión de las medidas de campo, se consideraba que el campo magnético es homogéneo para distancias menores a 200 km, pudiéndose determinar este valor medio a partir de el valor medio del observatorio. Es decir se *consideraba las variaciones diarias y seculares de la estación y el observatorio iguales*. Si llamamos ΔE a la variación temporal de la componente E del campo magnético, desde el instante t_0 al valor medio anual, tendremos:

$$E(x, t) = E(x, t_0) + \Delta E$$

como hemos supuesto que este ΔE es igual que en el observatorio

$$E(x_0, t) = + \Delta E$$

Sustituyendo

$$E(x, t) = E(x_0, t) + E(x, t_0) - E(x_0, t_0)$$

obteniendo así el valor medio en la estación o en cualquier instante (es decir reducimos a época común todos los valores observados en las estaciones en diferentes días). Esta reducción tiene el inconveniente que no siempre el observatorio más cercano esté a menos de 200 km, por lo que hipótesis de una variación igual a más de 600 km (como es el caso de algunas estaciones en España) es aun menos cierta.

Una reducción un poco más precisa es considerar “iguales las variaciones diarias de ambos, pero distintas la variación secular”, la ecuación de reducción será:

$$E(x, t) = E(x_0, t) + (E(x, t_0) - E(x_0, t_0)) + (VSE(x, t) - VS E(x_0, t)) * \Delta t$$

La reducción que la IAGA ha recomendado últimamente, considera distintas ambas variaciones entre la estación y el observatorio de referencia. Para ello se utilizan de estaciones temporales en las que se ubicará un variómetro durante todo el tiempo de las observaciones, que registrará las variaciones del campo. Estas estaciones cubrirán un radio de unos 150 Km., por lo que la hipótesis de un campo homogéneo en ese círculo será más aceptable. También se referirán lo valores de la observación a las 2 h. de T.U., o “calma nocturna” cuando las variaciones magnéticas son mínimas y finalmente se reducirán los valores con el observatorio de referencia. La ecuación de reducción será:

$$E(x, t) = E(x_0, t) + (E(x, t_0) - E(x_0, t_0)) + (VSE(x, t) - VS E(x_0, t)) * + (E(x, t_{2TU}) - E(x_0, t_{2TU}))$$

Modelización Isopórica del campo geomagnético

Hemos visto que la cartografía magnética para que sea útil tiene que representar la variación secular, que también depende de la posición por lo que, si se pretende estudiar este fenómeno, es necesario determinar sus características en la zona a través de los datos obtenidos en estaciones seculares.

La variación secular puede ser separada en primera aproximación en una parte temporal y otra espacial, que pueden ser analizadas independientemente.

Donde E_0 es el valor de la componente en el tiempo origen y los coeficientes A y B se determinarían en su mejor ajuste, estando éstos relacionados con la variación secular.

Ajuste polinomial

Para la parte temporal, se puede asumir que en una estación y durante un corto período de tiempo, la variación de las componentes magnéticas o angulares se ajusta a un polinomio de la forma (De Miguel, 1980):

$$\Delta E_i = A_0 + A_1(\varphi_i - \varphi_0) + A_2(\lambda_i - \lambda_0) + A_3(\varphi_i - \varphi_0)(\lambda_i - \lambda_0) + A_4(\varphi_i - \varphi_0)^2 + A_5(\lambda_i - \lambda_0)^2$$

que es la llamada ecuación isopórica, siendo φ_i y λ_i , la longitud y latitud del punto y ΔE_i su variación secular, obtenida por la expresión anterior, y φ_0 , λ_0 las coordenadas del observatorio de referencia. Los coeficientes A_0 al A_5 se calculan mediante el mejor ajuste a los datos de cada estación.

Posteriormente tras la obtención de la VS en cada estación, se obtiene un modelo de VS mediante ajuste por mínimos cuadrados de la expresión anterior.

Este método presenta dos inconvenientes: en primer lugar cada componente se calcula independientemente por lo que la ecuación de Laplace no se impone y en segundo lugar solo depende de las coordenadas planas, sin considerar la altura.

Análisis armónico: método scha.

SCHA (Spherical Cap Harmonic Analysis) es un nuevo método de cálculo de la ecuación isopórica que pretende subsanar estos defectos. Parte de la ecuación de Laplace, pero en lugar de resolverla en la esfera como es el caso del IGRF la resuelve en un casquete esférico que cubre la región a estudiar. (SCHA, Haines, 1985a).

19.4. Anomalías magnéticas

A partir de registros del campo magnético en diferentes puntos de la Tierra y en diferentes épocas, se observa por la forma de su variación, que el campo que medimos en la superficie de la tierra es la suma vectorial de diferentes campos constituyentes, cada uno de ellos generado en diferentes regiones y varían de manera distinta tanto en el espacio como en el tiempo. Un análisis profundo es capaz de separar cada uno de estos campos en dos partes: uno creado por fuentes en el interior de la Tierra y el otro debido a fuentes externas a ella. El interno a su vez está formado fundamentalmente por dos constituyentes: uno global de gran amplitud, cuya fuente debe estar situada en el interior de la tierra y otro, de amplitud mucho menor y que debe ser creado por fuentes mucho más superficiales. Conviene tratar a estos como dos campos diferentes, y se les suele llamar "campo fundamental" al de origen interno y "campo local" al de origen superficial, superpuesto a estos, está el "campo externo", producido por corrientes eléctricas que circulan por la Ionosfera y Magnetosfera, que al variar en el tiempo generan campos inducidos, que se suman al "campo anómalo o cortical", originado por rocas magnetizadas, bien inducida, bien remanente, para crear lo que hemos llamado campo local, también se le llama campo residual o campo anómalo.

La anomalía en un punto de uno cualquiera de los elementos del campo magnético terrestre se define como la diferencia entre el valor observado en dicho punto y el que corresponde a una cierta distribución teórica de referencia llamada distribución normal, que represente el campo fundamental, de ahí el nombre de "campo residual". No existe una definición precisa y única

para esta distribución, pero en suma se trata de encontrar un compromiso entre la simplicidad de la fórmula que la define y su carácter representativo, de forma que podamos calcular a partir de estas fórmulas unas distribuciones teóricas tan próximas a la observada como queramos.

De esta imprecisión en la definición de la distribución normal resulta que, para regiones vecinas, se puedan proponer campos normales cuyo dominio de validez, el modo de cálculo y el significado físico son muy variados, y que, en muchos casos, se ajusten mal entre sí. El deseo de un campo normal universal, a partir del cual serían uniformemente medidas las anomalías magnéticas, fue cada vez más buscado por los geomagnéticos. Diferentes métodos fueron propuestos, la mayoría basados en el desarrollo del campo normal en armónicos esféricos; finalmente se eligió el IGRF (Campo Internacional de Referencia), como campo normal universal.

Este campo teórico debe representar, en el conjunto del Globo, las anomalías regionales, es decir las variaciones espaciales del campo fundamental, mientras que las anomalías locales, que no vienen representadas en él, se manifiestan como un campo perturbador dentro de este campo regular. De esta forma se adopta, en la mayoría de los levantamientos geomagnéticos este campo teórico como campo normal, representante del campo de origen profundo, y la diferencia entre el campo observado y el campo de referencia constituye el campo de anomalías superficiales. Sin embargo, este no es el único campo normal que se puede adoptar para una región, pudiendo, por aproximaciones polinómicas calculadas a partir de los datos observados en la región.

Propiedades estadísticas de la distribución de anomalías. La opción entre las dos alternativas para la determinación analítica del campo profundo, la de un campo normal definido por una aproximación polinómica o la adopción del campo internacional de referencia, depende fundamentalmente de cual de los dos represente mejor este campo, es decir las anomalías regionales, de forma que al restarlo del campo total observado deje como campo perturbador únicamente las anomalías corticales, por ello antes de calcular un campo normal para nuestro levantamiento, estudiaremos brevemente las propiedades estadísticas de un campo perturbador, engendrado por fuentes corticales (situadas entre la superficie y la isoterma de Curie).

Sea **FN7** el campo de origen profundo, cuya intensidad es del orden de miles de nanoteslas, mientras que **TA8** el campo cortical tiene una intensidad que no suele superar el millar. Las dimensiones de las variaciones, anomalías regionales, de **FN 9** son de decenas de miles de kilómetros, las del campo **TA 10** raramente sobrepasan la centena.

El campo observado es el campo total $\mathbf{FO} = \mathbf{FN} + \mathbf{TA}$ 11. Llamaremos anomalía de intensidad, AT, a la cantidad (FO-FN). Dentro de una región de dimensiones limitadas comparadas con las de la Tierra, podemos suponer que el campo de origen profundo tiene una dirección constante, como TA12 es pequeño frente a FN13, la cantidad AT es en buena aproximación, la proyección de TA 14 sobre FN 15, es decir:

$$AT = \mathbf{TA} \cdot \mathbf{FN} / FN$$

Las conclusiones de este párrafo no serían modificadas de forma sensible si consideráramos la variación lenta de FN 17, pero para simplificar consideraremos realizada esta condición de uniformidad; igualmente consideraremos despreciable la curvatura de la Tierra, Si AT es la anomalía de intensidad engendrada por una estructura cortical de dimensiones finitas, se verifica fácilmente que la integral de AT extendida a todo el plano D de medidas, es nula (Le Mouel, 1969):

$$\int_D AT(x, y) dx dy = 0$$

En realidad no conocemos la función continua AT, sino únicamente sus valores en los puntos de medida (los nudos de una red regular, en nuestro caso la red del levantamiento), por lo que sustituiremos la fórmula de integración por la de sumatorio:

$$\sum at^{ij} = 0$$

la suma extendida a todos los nudos del plano.

Las relaciones anteriores, válidas en el caso del campo anómalo creado por una fuente aislada finita, también lo son para el caso del campo resultante de una serie de fuentes, pero el plano hay que extenderlo a todo el plano horizontal mientras que el dominio cubierto por las medidas, D, es necesariamente limitado. Si AT es la anomalía engendrada por una estructura no se cumple exactamente la integral nula, pero si consideramos el conjunto de estructuras responsables de la anomalía, las integrales correspondientes extendidas a D serán tanto negativas como positivas y, en ausencia de un conocimiento más preciso de la distribución de fuentes en la periferia de D, estamos obligados a suponer nula la integral resultante.

Las anomalías magnéticas locales son debidas a las rocas magnetizadas de la corteza terrestre. Para poder estudiar estos campos magnéticos podemos hacer uso de potenciales magnéticos anómalos, Φ_A , tal que la anomalía no es más que el menos gradiente de este potencial.

$$\Delta Z = -\frac{\partial \Phi_A}{\partial z} \quad \Delta X = -\frac{\partial \Phi_A}{\partial x}$$

Denotamos a la componente vertical y horizontal de una anomalía magnética como ΔZ y ΔX .

De esta forma se establecen las siguientes relaciones:

Si la anomalía magnética es producida por un dipolo magnético, las expresiones del potencial y las componentes serían:

$$\Phi_A = \frac{Cm \cos \theta}{r^2} = \frac{Cm(d+z)}{[x^2 + (d+z)^2]^{3/2}}$$

$$\Delta X = -\frac{\partial \Phi_A}{\partial x} = \frac{3Cm x (d+z)}{[x^2 + (d+z)^2]^{5/2}}$$

$$\Delta Z = -\frac{\partial \Phi_A}{\partial z} = \frac{Cm[2(d+z)^2 - x^2]}{[x^2 + (d+z)^2]^{5/2}}$$

Donde d es la profundidad del dipolo.

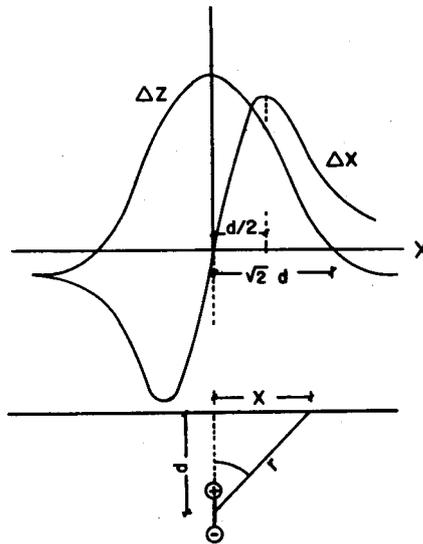


Figura 1.

Para la superficie, $z = 0$ m, ΔZ cambiará de signo a $x = \sqrt{2} d$. El valor máximo de ΔZ es en $x = 0$ m mientras que el valor máximo de ΔX se alcanza a $x = d/2$.

Las curvas de ΔZ y ΔX están representadas en la figura. La componente vertical ΔZ tiene el máximo sobre el dipolo, mientras que el horizontal es nulo en dicho punto. La profundidad a la que está enterrado un polo o dipolo puede deducirse del valor de x , para el que ΔZ es la mitad de su valor máximo, en el que cambia de signo, o también de los valores de x donde se encuentran los máximos de ΔX .

En el caso del polo magnético, el valor del máximo de ΔZ es:

$$\Delta Z_{\max} = CP/d^2$$

y el valor de x en el que ΔZ vale la mitad de este valor es:

$$x = d^2(2^{2/3}-1)^{1/2}$$

De estos dos valores se pueden determinar P y d . La profundidad d se puede obtener también a partir del valor de x en que ΔX es máximo:

$$x = d/2^{1/2}$$

En el caso del dipolo vertical el máximo de ΔZ viene dado por

$$\Delta Z = 2Cm /d^3$$

y el punto en que la curva de ΔZ cambia de signo:

$$x = (2d)^{1/2}$$

El valor de x para el que ΔX es máximo y del que también se puede calcular d es:

$$x = d /2$$

Si la polaridad del polo o dipolo magnético causante de la anomalía es tal que produce un campo en el mismo sentido que el de la Tierra, las anomalías son positivas y en el caso contrario negativas. En el hemisferio norte, las anomalías positivas son producidas por polos negativos, o dipolos con el polo negativo hacia arriba y al contrario en el hemisferio sur. La distribución de anomalías positivas y negativas en los océanos ha proporcionado, como se verá más adelante, importante información sobre su apertura y el movimiento de los continentes.

Una formulación más general para el potencial de un dipolo enterrado con una orientación arbitraria, definida por los cosenos directores (v_1, v_2, v_3), utilizando la notación de subíndices es:

$$\Phi_A = Cm v_i \frac{\partial}{\partial x_i} \left(\frac{1}{r} \right)$$

donde r es la distancia del centro del dipolo al punto en que se mide el potencial. Los componentes de intensidad del campo o anomalías en la dirección de los ejes geográficos x_i vienen dadas por:

$$(\Delta X, \Delta Y, \Delta Z) = \Delta B_i = Cm \frac{\partial}{\partial x_i} \left[v_j \frac{\partial}{\partial x_j} \left(\frac{1}{r} \right) \right]$$

Debido a que las anomalías son pequeñas, en relación con el valor del campo magnético de la Tierra se suele utilizar, como ya se ha mencionado, la componente de éstas en la dirección del campo geomagnético o anomalía de campo total ΔF , y la horizontal en la dirección del meridiano magnético ΔH . Estas se pueden obtener a partir de los componentes en la dirección de las coordenadas geográficas $\Delta X, \Delta Y, \Delta Z$ y los ángulos de declinación D , e inclinación I en la forma:

$$\Delta H = \Delta X \cos D + \Delta Y \sin D$$

$$\Delta F = \Delta H \cos I + \Delta Z \sin I$$

INTERPRETACION DE ANOMALÍAS MAGNETICAS

De forma análoga a las anomalías gravimétricas, las magnéticas son también un instrumento eficaz para estudiar el interior de la Tierra. Sin embargo, las fuentes magnéticas responsables de las anomalías sólo pueden estar cerca de la superficie. Esto es debido a que la temperatura de Curie, por encima de la cual los materiales pierden sus propiedades magnéticas, se alcanza profundidades menores de 100 km. Las anomalías magnéticas, por tanto, sólo dan información sobre estructuras y materiales de la corteza.

Por su extensión, las anomalías magnéticas, como las gravimétricas, se pueden dividir en locales y regionales. Las primeras sirven de base a los métodos de prospección magnética de minerales. Estas anomalías están muy localizadas y pueden tener valores muy altos entre 100 y

1 000 nT. Las anomalías regionales están producidas por la magnetización del zócalo cristalino y nos descubren su estructura, presencia de falla, etc. La gran diferencia en los valores de la susceptibilidad magnética entre los minerales ferromagnéticos, entre 0,07 a 6 y los paramagnéticos, del orden de 0-6, hace que la proporción de los primeros sea la que determina las características magnéticas de una roca. En especial, juega un papel muy importante en este aspecto, debido a su abundancia, la proporción de magnetita presente en las rocas cristalinas.

19.5. Levantamientos magnéticos

Tanto en las observaciones de anomalías locales como regionales, los levantamientos de mapas magnéticos se realizan bien por observaciones en la superficie terrestre, o desde el aire mediante el empleo de aviones y helicópteros. El uso de uno u otro método depende fundamentalmente del coste, la extensión y topografía del área a estudiar. En las observaciones de superficie se utiliza además de la componente de campo total ΔF la vertical ΔZ y, aunque muy raras veces, también la horizontal ΔH . En las observaciones desde el aire, generalmente se usa el campo total ΔF .

Los parámetros de un levantamiento aeromagnético: altura de vuelo, tiempo de muestreo, espaciamento entre líneas de vuelo, ... varían según el objetivo del mismo. Para estudios de prospección, donde se desea un mayor detalle y mayor superficialidad de las anomalías, la altura suele ser de unos pocos cientos de metros, el espaciamento entre líneas pequeño y un tiempo de muestreo pequeño, para no perder pequeñas longitudes de onda. Por el contrario, para fines geofísicos (estudio de estructuras corticales) todos los parámetros aumentan, para eliminar anomalías de pequeña intensidad y corta longitud de onda, pero cercanas a la superficie, que par estos estudios se consideran ruido.

El IGN ha realizado varios levantamientos del territorio español, que a continuación se resumen.

Levantamiento aeromagnético de España peninsular

El levantamiento aeromagnético de la España Peninsular fue el primero de los realizados por el Instituto Geográfico Nacional. La fase de adquisición de datos comenzó en septiembre de 1986, siendo interrumpida durante el invierno (diciembre 1986-abril 1987), finalizando la misma en junio de 1987. (Ardizzone et al., 1989)

El levantamiento consta de 102 líneas de vuelo en dirección Norte-Sur, espaciadas 10 km entre sí; 20 líneas de control en dirección Este-Oeste, con un espaciamento de 40 km y 40 líneas

perimetrales para cerrar casi la totalidad de la red. La altitud barométrica del vuelo es de 3000 m. La velocidad media del avión se mantuvo entre los 240 y 250 km/h, que junto con una velocidad de muestreo de 0,5 s, se obtiene una observación cada 34 m. El magnetómetro va colgado mediante un cable a unos 30 m del avión, con el fin de minimizar la influencia de las masas metálicas del avión sobre el campo magnético. El mapa del campo anómalo o residual se obtiene restando de esta malla el valor del IGRF-85, calculados en los mismos puntos y a la misma altitud.

Posteriormente este levantamiento se unió al levantamiento portugués, obteniendo el mapa de anomalías magnéticas de la Península Ibérica.

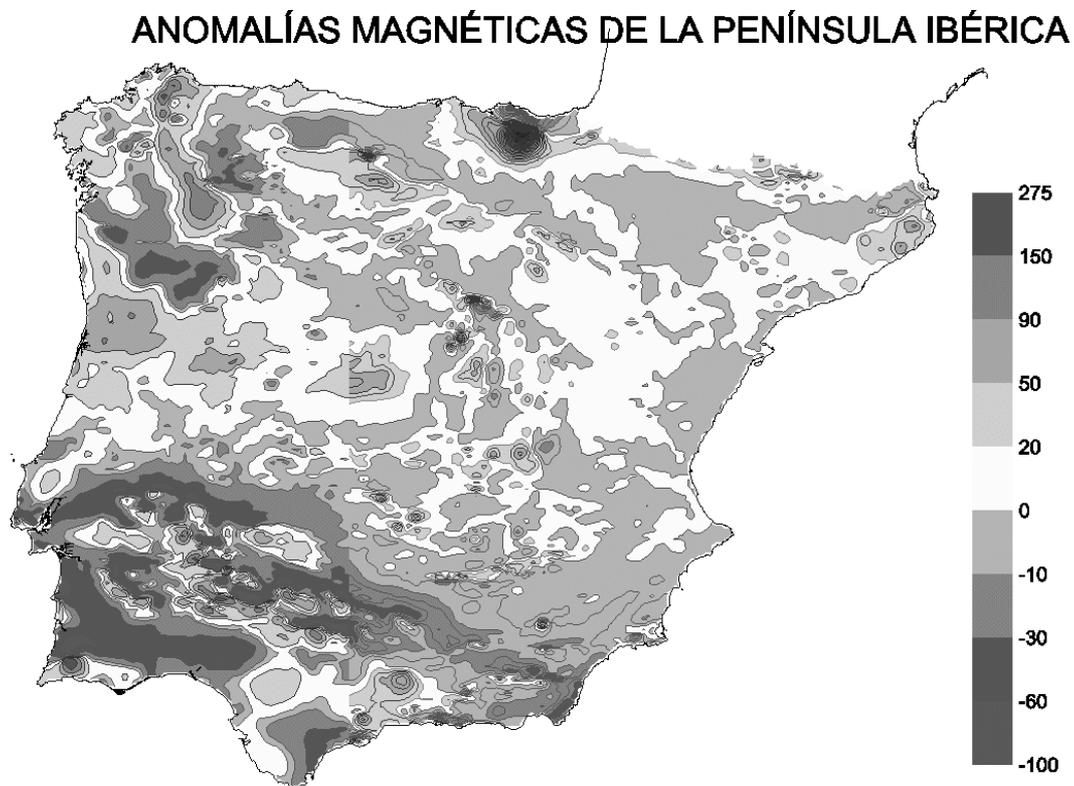


Figura 2.

Levantamiento aeromagnético del Archipiélago Canario.

La red del levantamiento aeromagnético se compone de siete bloques, volados a distintas altitudes. Estas diferencias en las alturas de vuelo, eran debidas a la topografía del terreno, dado que el levantamiento cubría zonas marítimas situadas a más de 3000m por debajo del nivel del mar, hasta zonas montañosas a más de 3000m sobre el nivel del mar. El total de kilómetros volados es de 31894.25 y las altitudes están comprendidas entre los 500m para el bloque marino a 3810m para el

Teide. Las líneas de vuelo van en dirección Norte-Sur, con un espaciamento de 5000m en el bloque marino, 2500m en los bloques sobre islas y de 500m en la zona del Teide, donde se pretendía tener una red de gran detalle, para realizar un estudio geofísico completo del volcán, dentro de un programa financiado por la Comunidad Europea. Las líneas de control, en dirección Este-Oeste, tienen un espaciamento de 20000m salvo en la zona del Teide, donde es de 5000m. En este caso el magnetómetro va colocado en aguijón, es decir al final de una barra fijamente unida al avión, un compensador automático elimina lo influencia de las masas metálicas del avión. La velocidad del avión se mantuvo entre los 280 y 290 km/h, con lo que dado el intervalo de muestreo de 0.1 s, se obtiene una observación cada 10 m aproximadamente. (Socías y Mezcuca , 2002).

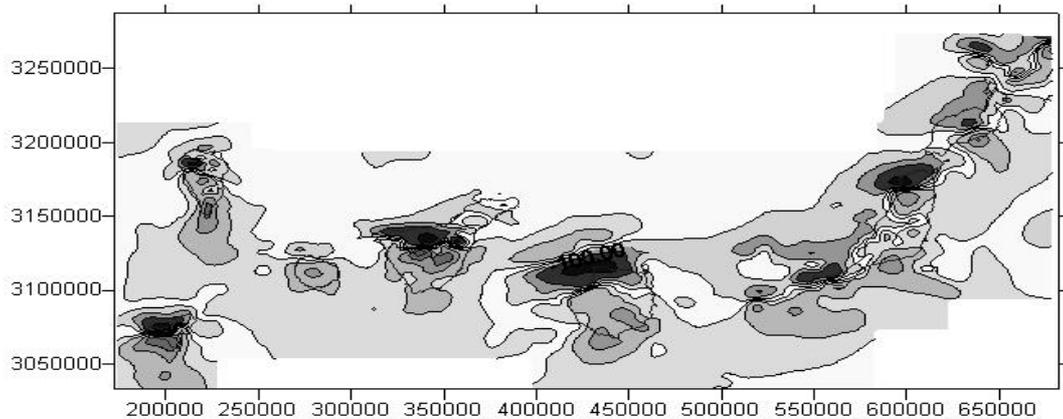


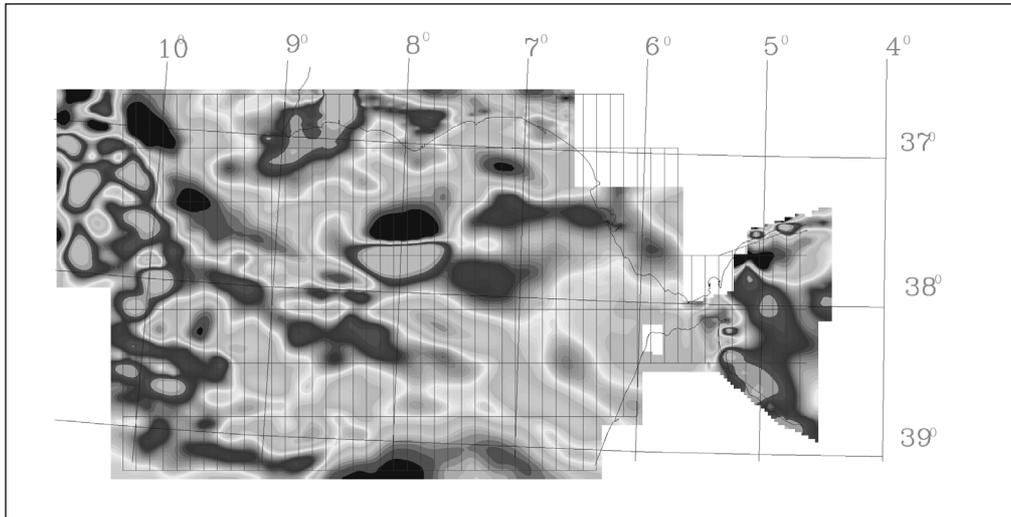
Figura 3.

Levantamiento aeromagnético del Golfo de Cádiz

El levantamiento fue realizado utilizando la aeronave Cessna Titan 404. El levantamiento se realizó en dos etapas, debido a problemas con los permisos de vuelo de los diferentes países. La primera etapa transcurrió del 14 de febrero al 3 de abril del 2005 y la segunda se realizó del 1 de mayo al 26 de junio del 2005. En total se realizaron 23 vuelos para completar 16,066 kilómetros lineales.

El área está situada sobre parte del Atlántico y extendiéndose más allá del estrecho de Gibraltar dentro del Mar de Alborán, entre los países de España, Portugal y Marruecos, extendiéndose hacia el oeste cerca de la longitud 11°Oeste y hacia el este a la altura de la ciudad de Málaga. La parte Norte del área está limitada por las costas de España y Portugal y al Sur por la costa de Marruecos. El levantamiento fue realizado a una altura GPS constante de 700 metros sobre el nivel del mar. En algunos lugares sobre la costa, la altura del levantamiento fue un poco

superior para salvar la topografía. El área del levantamiento se encuentra principalmente sobre el mar, cubriendo también las costas de los países ya mencionados. (Socias, 2007)



Bibliografía

- [1] IAGA Division V, Working Group VMOD: Geomagnetic Field Modeling, (2005): "The 10th-Generation International Geomagnetic Reference Field". *Geophysical Journal International*, 161, 561
- [2] IGN, Servicio de Geomagnetismo: "Mapa Geomagnético de España, 2005.0". IGN 2006.
- [3] Miguel, L. de (1980): *Geomagnetismo*. I.G.N. Madrid
- [4] Parkinson, W.D. (1983): *Introduction to Geomagnetism*. Elsevier, Amsterdam.
- [5] Torta, J.M. (1990): "Modelización regional del campo geomagnético sobre España: campo anómalo, variación secular y campo de referencia". *Publicaciones del Observatorio del Ebro, Memoria 15, Roquetes – Tarragona*.
- [6] Udías, A. y J.Mézcua (1986): *Fundamentos de Geofísica*. Alhambra Universidad. Madrid.
Los levantamientos aeromagnéticos están publicados por el IGN.

Tema 20. Paleomagnetismo. Propiedades magnéticas de la materia. Minerales magnéticos en rocas. Mecanismos de magnetismo remanente. Polos virtuales geomagnéticos. Migración de los polos y de los continentes. Inversiones del campo.

20.1. Paleomagnetismo. Propiedades magnéticas de la materia

Uno de los fenómenos que más ha contribuido al conocimiento de las variaciones del campo magnético terrestre con el tiempo, y a la vez ha servido para revolucionar nuestras ideas sobre la tectónica global, es el del magnetismo remanente de las rocas. Este fenómeno se debe a la propiedad de ciertas rocas de adquirir una magnetización producida por la acción de un campo externo, que permanece estable aunque desaparezca o cambie dicho campo. En general, la magnetización remanente es adquirida durante la formación de la roca y proporciona información sobre el campo magnético terrestre en aquella época. En concreto, las lavas adquieren una magnetización remanente al enfriarse, que permanece estable a no ser que vuelvan a ser calentadas. Esta imanación refleja las características del campo terrestre de la época en la que se formaron las coladas de lava. Se puede considerar éste, por tanto, como un magnetismo fósil. La primera indicación de que la imanación remanente de las rocas está en relación con el campo magnético terrestre del pasado es debida a los estudios de Melloni y Folgerheiter a fines del siglo pasado sobre lavas de volcanes italianos. Los estudios modernos de paleomagnetismo se inician con los trabajos sobre las lavas históricas del Etna por R. Chevallier, en 1925. A partir de 1940, Konisberg, E. Thelie y T. Nagata dieron un gran impulso al paleomagnetismo. Estos trabajos fueron posibles gracias a las teorías sobre el magnetismo remanente de L. Néel y los métodos estadísticos desarrollados por R. A. Fisher hacia los años cincuenta.

Las propiedades magnéticas de la materia dependen del comportamiento de las partículas con carga eléctrica en movimiento en su estructura atómica, es decir, de la órbita y spin de los electrones. Éstos llevan asociado un momento magnético cuyo resultante es el momento magnético total del átomo. A nivel macroscópico se puede hablar de la imanación o magnetización específica M , o momento magnético por unidad de volumen. Las propiedades magnéticas de un material vienen caracterizadas no sólo por el valor de M , sino por la forma en que éste depende del campo magnético externo:

$$M = \chi H$$

donde χ es la susceptibilidad magnética. En general χ es un tensor de segundo rango y sólo en el caso de isotropía es un escalar. La inducción magnética en el interior del cuerpo vendrá dada por:

$$B = \mu_0(H + M) = \mu_0(1 + \chi)H$$

Si χ es pequeña y negativa, el material es diamagnético. El material, cuyo momento magnético resultante es nulo, adquiere una imanación inducida pequeña y en sentido opuesto y el campo en su interior es menor que en el exterior. Si χ es positiva y pequeña, el material es paramagnético y en presencia de un campo externo adquiere una pequeña imanación en el mismo sentido que el campo inductor. Este comportamiento, que corresponde a la mayoría de los materiales, es consecuencia de la precesión de Larmor del movimiento orbital de los electrones de las capas más externas de los átomos. Cada electrón puede considerarse como un dipolo magnético que se orienta en la dirección de las líneas de fuerza del campo inductor.

En ciertos materiales denominados ferromagnéticos, el campo inducido en el material por un campo externo puede ser muy grande, debido a la interacción entre los momentos magnéticos asociados al spin de los electrones no apareados de sus átomos, cuya interacción práctica es muy intensa. En estos materiales, la imanación aumenta al aumentar el campo externo hasta un valor M que corresponde a la imanación de saturación. Al disminuir el campo externo H , M disminuye, pero no se anula totalmente al anularse el campo inductor. A la magnetización M_R que permanece al anularse el campo externo se la denomina imanación remanente

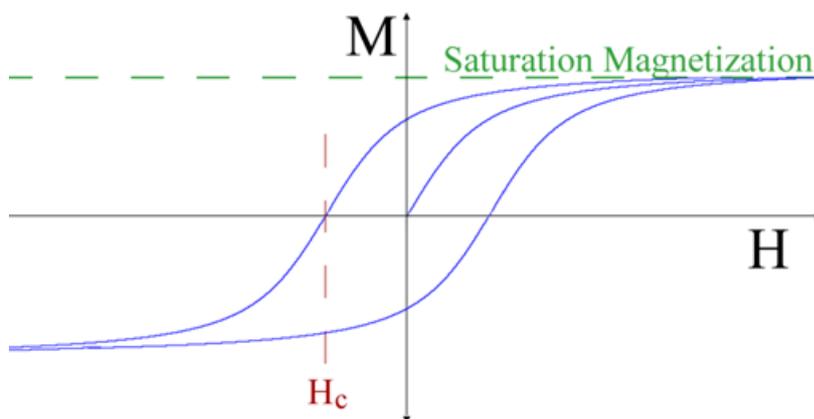


Figura 1. Ciclo de histéresis de un material ferromagnético.

Las interacciones que existen entre los momentos magnéticos en estas sustancias son muy intensas, y ello hace posible que a temperaturas relativamente elevadas, pero por debajo de una temperatura crítica, llamada temperatura de Curie (T_C), se dé un paralelismo entre los momentos magnéticos vecinos, necesario para explicar el elevado valor de la imanación que presentan incluso en ausencia de un campo externo (imanación espontánea).

Estas interacciones fueron explicadas por Weiss, en 1907, que supuso la existencia de un campo molecular, de canje o campo de Weiss H_W . Este campo H_W representa la interacción de un portador de momento magnético con sus vecinos. Weiss consideró que este campo es proporcional a la imanación:

$$H_W = \lambda M$$

siendo λ la corriente de campo molecular. A la acción orientadora se opone la agitación térmica y, así, toda sustancia ferromagnética tiene una temperatura crítica, denominada punto de Curie T_C , por encima de la cual desaparece el ordenamiento de los momentos. Esta temperatura marca la transición entre la fase ferromagnética y la fase paramagnética, y viene dada por:

$$T_C = \frac{\mu_B H_W}{k}$$

donde μ_B es el momento magnético del electrón y k la constante de Boltzmann.

Los materiales ferromagnéticos se dividen de acuerdo con la dirección de los momentos magnéticos intrínsecos de los electrones, en ferromagnéticos, todos los momentos en una misma dirección y sentido; ferrimagnéticos, una dirección predominante y algunos opuestos, y antiferromagnéticos, en los que los momentos están divididos en dos grupos con sentidos opuestos; si su dirección es totalmente paralela, el momento magnético resultante es nulo.

20.2. Minerales magnéticos en las rocas

Las propiedades ferromagnéticas de las rocas se deben a los minerales ferromagnéticos dispersos en ellas, siendo el resto de la roca de propiedades diamagnéticas. Citaremos a continuación los más importantes con sus características más significativas, así como aquellos otros que se obtienen a partir de ellos como combinación.

Óxidos de hierro

Magnetita es el óxido ferroso férrico (Fe_3O_4) y cristaliza en el sistema cúbico. A la temperatura de 118 °K se produce un cambio de fase, cristalizando en el ortorómbico, proceso que está

acompañado por un cambio en sus propiedades magnéticas. Este mineral es el principal constituyente magnético de las rocas ígneas. Su temperatura de Curie es de 578 °C.

Hematites, óxido férrico (Fe_2O_3) es un material fundamentalmente antiferromagnético, aunque debido a impurezas en la red cristalina los momentos magnéticos no son completamente antiparalelos presentando lo que se denomina un ferromagnetismo no compensado o parásito. Este mineral cristaliza en el sistema romboédrico y su temperatura de Curie es de 675 °C. Las propiedades magnéticas de algunas rocas sedimentarias, como son los sedimentos rojos, se deben fundamentalmente a la presencia de hematites.

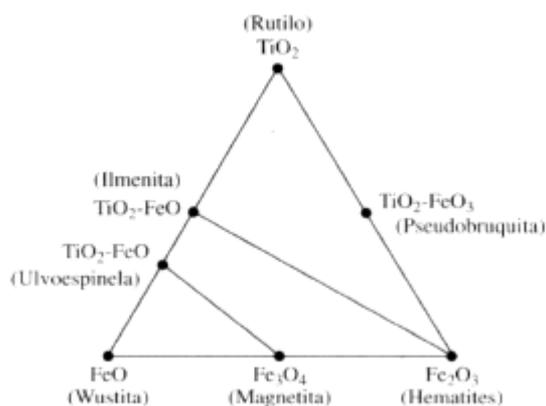


Figura 2. Diagrama de composición de los minerales ferromagnéticos, Udías y Mezcuca, 1997.

Óxidos de titanio-ferroso-férrico

A este grupo pertenecen minerales compuestos por TiO_2 (rutilo), FeO (wustita) y Fe_2O_3 (hematites) en diferentes proporciones.

Su composición puede escribirse en la forma $X\text{TiO}_2$, $Y\text{FeO}$, $Z\text{Fe}_2\text{O}_3$, cumpliéndose que $X + Y + Z = 1$. Si consideramos un triángulo equilátero, en el que la altura es la unidad, la composición de un mineral determinado viene representada por un punto del triángulo, en el que las distancias a cada lado son X , Y , Z . Cada punto del triángulo corresponde a una composición distinta, aunque no quiere decir que ésta exista realmente.

Por ejemplo, el mineral cuya composición corresponde al punto central del lado $\text{FeO}-\text{Fe}_2\text{O}_3$ es la magnetita. De idéntica forma, el punto medio del segmento TiO_2-FeO representa la ilmenita, que es un mineral que cristaliza en el sistema romboédrico y que se comporta como paramagnético a la temperatura ordinaria. El punto situado a los $2/3$ del segmento TiO_2-FeO representa un

mineral denominado ulvoespinela, que cristaliza en el sistema cúbico y que posee propiedades paramagnéticas.

Los minerales denominados titanomagnetitas y titanohematites forman dos series. La primera está representada por puntos sobre la recta que une los minerales ulvoespinela y magnetita, y los de la segunda sobre la que une la ilmenita y la hematites.

Las capas basálticas de la corteza oceánica, responsables de las anomalías magnéticas en las dorsales, deben sus propiedades magnéticas a la presencia de titanomagnetita. Existen finalmente otras variedades, pero, generalmente, no juegan un papel importante en el magnetismo remanente de las rocas.

Otra serie de minerales es la de la titanomaghemita. Con este nombre se denomina a los óxidos de hierro titanio que cristalizan en el sistema cúbico con estructura espinelo inversa y que por su composición química están comprendidas entre las series magnetita-ulvoespinelo y la hematita-ilmenita.

20.3. Mecanismos de magnetismo remanente

Las propiedades magnéticas de las rocas dependen de los minerales ferromagnéticos que se encuentran dispersos en una mezcla muy fina, con el resto de los materiales que forman la roca que tienen propiedades diamagnéticas o paramagnéticas. La roca así constituida por una matriz de material diamagnético o paramagnético y una dispersión de material ferromagnético presenta en su conjunto propiedades ferromagnéticas. Desde el punto de vista del paleomagnetismo, interesa el magnetismo remanente, adquirido por la roca en la época de su formación que mantiene sus características, aunque cambie el campo inductor. Para que este magnetismo pueda dar información importante sobre el campo magnético de la Tierra en el pasado, debe ser estable y conservar la imanación de cuando se formó la roca. Su dirección debe ser la del campo inductor y debe conocerse el tiempo en que adquirió esta imanación.

De los distintos mecanismos de adquisición de magnetismo, dos son de particular importancia por su aplicación al paleomagnetismo: el magnetismo remanente térmico, aplicable a las rocas ígneas, y el magnetismo remanente deposicional, de aplicación en las rocas sedimentarias.

Relajación magnética y superparamagnética

En ausencia de un campo inductor, un material magnético formado por partículas de un solo dominio magnético, cuya imanación resultante sea M_0 , experimenta con el tiempo una pérdida debida a la activación térmica. Esta variación es de la forma:

$$M = M_0 \exp(-t / \tau)$$

donde τ es una constante o tiempo de relajación, dado por:

$$\tau = \frac{1}{C} \exp(KV / kT)$$

En esta ecuación C es un factor dependiente de la frecuencia que corresponde a la probabilidad de obtener una inversión de polaridad, debida a la agitación térmica y es del orden 10^8 s^{-1} ; k es la constante de Boltzmann; K , la energía de anisotropía magnética por unidad de volumen (= 10^3 J/m^3 para la magnetita), y V , el volumen de la partícula. El producto KV representa la barrera de potencial que es necesario superar por agitación térmica.

Un aspecto importante de la ecuación es que T varía fuertemente con V/T , de forma que pequeñas variaciones de V y T producen variaciones grandes de τ . Así para $C = 10^8 \text{ s}^{-1}$ y $KV = 18 \text{ kT}$, se deduce un $\tau = 1 \text{ s}$, mientras que para $KV = 57,7 \text{ kT}$, el valor de τ es de 4×10^9 años. Esta última ecuación expresa una relación entre el tiempo de relajación τ y el volumen y temperatura, que es crucial para el entendimiento de los procesos de magnetización termorremanente y química.

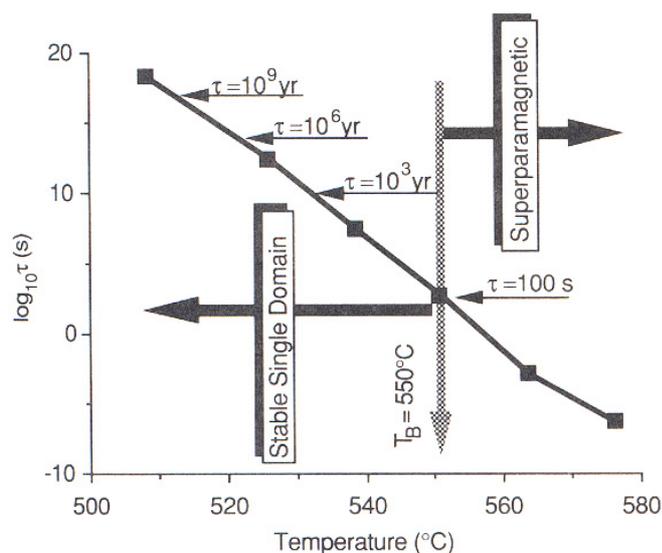


Figura 3. Variación del tiempo de relajación en función de la temperatura. T_b es la temperatura de bloqueo. Butler, 1992.

A temperatura ambiente, los granos pequeños se comportan de forma inestable debido a la agitación térmica y alcanzan rápidamente la posición de equilibrio si actúa un pequeño campo magnético. Al momento magnético adquirido se le denomina magnetización de equilibrio. En esta situación no todos los dominios se encuentran alineados con el campo. Si el campo cesa, la remanencia adquirida por los granos desaparece rápidamente. A este comportamiento magnético se le llama «superparamagnetismo», por la analogía que presenta con el comportamiento paramagnético. Para granos de mayor tamaño, la magnetización resulta ser más estable y éstos presentan un momento magnético bloqueado.

Un hecho importante es la fuerte dependencia que presenta el tiempo de relajación τ con el cociente V/T . Un aumento de un factor de dos en este cociente origina un aumento en el tiempo de relajación de 10^8 , lo que hace que la transición entre el estado superparamagnético y el de «momento bloqueado» sea muy brusca. En la práctica, la separación entre estos dos estados se define mediante los conceptos de «temperatura de bloqueo» y de «volumen crítico». Para un material determinado existe un «volumen crítico» (que depende de la temperatura) que separa los granos en dos clases. Los de volumen menor que el crítico tienen tiempos de relajación pequeños (10 a 60 minutos) y alcanzan pronto la situación de equilibrio orientándose en la dirección del campo externo. Los que tienen volúmenes por encima del crítico permanecen bloqueados. Análogamente, se define la «temperatura de bloqueo» T_b , para un grano determinado, como la temperatura a la cual el tiempo de relajación es pequeño. Para temperaturas superiores a T_b , el grano se encuentra en estado superparamagnético y a temperaturas inferiores presenta un momento bloqueado. Estos conceptos de «temperatura de bloqueo» y «volumen crítico» son básicos para la comprensión de los procesos de imanación remanente térmica y química.

Magnetización termorremanente

Se llama magnetización termorremanente (TRM) a la adquirida por una roca durante su enfriamiento desde la temperatura de Curie hasta la temperatura ambiente en presencia de un campo magnético. Este tipo de remanencia magnética es la que presentan las rocas ígneas alrededor de los 1.000 °C. Durante el enfriamiento, al sobrepasar la temperatura de Curie, aparece la imanación espontánea. Los granos de materiales ferromagnéticos se encuentran en una situación superparamagnética y se alcanza la imanación de equilibrio aunque el campo externo sea débil como el de la Tierra. Si la temperatura disminuye, sobrepasando las temperaturas de bloqueo de los diferentes granos, el tiempo de relajación de los granos aumenta muy rápidamente y la imanación adquirida se bloquea. Los granos pasan a un estado de

«momento bloqueado» y posteriores cambios en la dirección del campo externo, que ocurren a temperaturas por debajo de T_b , no producen cambios en la magnetización de los granos.

Si el tiempo de relajación de un grano es de 10^3 segundos (tiempo de una experiencia de laboratorio) a una temperatura de $530\text{ }^\circ\text{C}$ (temperatura de bloqueo) después de su enfriamiento hasta la temperatura ambiente, el tiempo de relajación aumenta hasta ser de 10^{17} años. Incluso granos que presentan temperaturas de bloqueo de $330\text{ }^\circ\text{C}$ tienen tiempos de relajación de 10^9 años después de enfriarse a temperatura ambiente. La magnetización termorremanente es, por lo tanto, estable en una escala geológica de tiempos. Este hecho es una de las bases de los estudios paleomagnéticos.

Consideremos una muestra de mineral conteniendo partículas de hematites de distinto tamaño de grano. Si a dicha muestra se la somete a un calentamiento por encima de la temperatura de Curie T_C ($675\text{ }^\circ\text{C}$), y posteriormente, a un enfriamiento en presencia del campo magnético terrestre (50.000 nT), ésta adquiere una magnetización remanente de 10^{-2} - 10^{-6} Am^{-1} , dependiendo de la concentración en la muestra.

Para cristales multidominios, el mecanismo es distinto. Los cristales están divididos en dominios magnéticos, cada uno con un momento magnético en distinta dirección y separados por barreras también llamadas paredes de Bloch. El momento magnético resultante es la suma del de todos los dominios. Un campo magnético externo tiende a correr las barreras entre los dominios, de forma que el dominio mayor esté en su misma dirección. A temperatura ambiente, el campo externo necesario para esto ha de ser muy alto (para la magnetita por encima de $0,1\text{ T}$). Sin embargo, si se calienta el material a una temperatura por encima de la de bloqueo, pero por debajo de la de Curie, campos externos no muy grandes pueden lograr este efecto. De esta forma se logra en cristales multidominios una imanación remanente estable.

Magnetización remanente química

La magnetización remanente química se produce cuando una roca, en presencia de un campo magnético, experimenta alteraciones químicas que afectan a sus minerales magnéticos. Estas alteraciones pueden ser debidas a modificaciones diagenéticas, oxidación, formación autogenética de un nuevo mineral, etc.

La teoría anteriormente expuesta para granos monodominios puede explicar este tipo de magnetización remanente que se produce durante el proceso de nucleación de un nuevo mineral. En los primeros estadios de formación los granos del nuevo mineral magnético son pequeños,

por lo que su tiempo de relajación es muy pequeño y el material se encuentra en una situación paramagnética. En presencia de un campo magnético externo, la imanación espontánea alcanza pronto la imanación de equilibrio. Los granos siguen desarrollándose y superan el volumen crítico, con lo que aumenta rápidamente su tiempo de relajación. La magnetización se bloquea y los cambios del campo externo ya no afectan al material. De forma análoga a la termorremanencia, se puede demostrar que la remanencia es estable en una escala geológica de tiempos. La magnetización remanente química es probablemente el origen de la remanencia que presentan las rocas que han sufrido importantes transformaciones químicas.

Magnetización remanente deposicional y post-deposicional

El otro tipo importante de rocas que pueden adquirir una magnetización remanente son las rocas sedimentarias. En algunas de estas rocas se encuentran minerales ferromagnéticos finamente pulverizados. Los granos de estos minerales, al depositarse por sedimentación en el mar en presencia de un campo magnético, se orientan estadísticamente en la dirección de éste, adquiriendo así la roca sedimentaria que se forma una magnetización en la dirección del campo inductor.

La remanencia magnética adquirida por alineamiento de los granos durante la sedimentación se denomina magnetización deposicional (MRD), mientras que la que se adquiere después de la deposición y antes de la consolidación, debida a la rotación de los granos magnéticos se le llama magnetización remanente post-deposicional (MRpD).

El mecanismo puede explicarse de forma sencilla, considerando granos esféricos de material ferromagnético que corresponden a un dipolo magnético de momento m que se depositan en un medio líquido (agua) de profundidad h . Sea ρ la masa de cada grano; v , el coeficiente de amortiguamiento del agua con respecto a su caída, y λ , con respecto a la rotación.

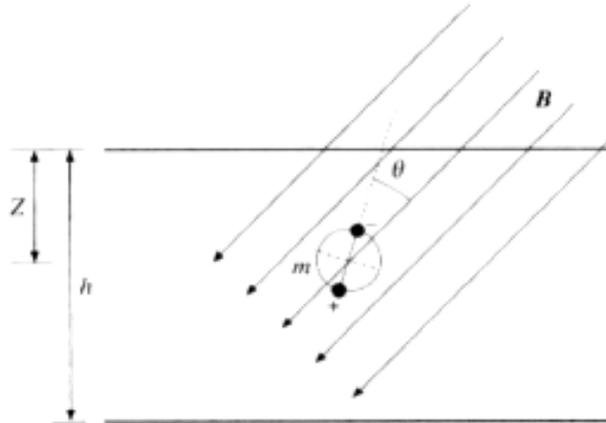


Figura 4. Orientación de un grano esférico magnetizado en su caída en una capa líquida, Udias y Mezcuca, 1997.

La ecuación diferencial de la caída del grano es:

$$\rho \ddot{z} = \rho g - v \dot{z}$$

La de rotación del grano sobre sí mismo es:

$$I \ddot{\theta} = m B \text{sen} \theta + \lambda \dot{\theta}$$

donde I es el momento de inercia del grano, y θ el ángulo entre el eje del dipolo y la dirección del campo inductor B. Si suponemos que el grano cae con velocidad constante, la ecuación correspondiente puede escribirse:

$$dt = \frac{v}{\rho g} dz$$

y su solución es:

$$t = \frac{v}{\rho g} z$$

Suponiendo que para $t = 0$, $z = 0$, el tiempo que tarda en llegar al fondo a profundidad h es:

$$t = \frac{v}{\rho g} h$$

De forma análoga, si en la ecuación de rotación consideramos la velocidad constante, la ecuación quedará:

$$-\frac{Bm}{\lambda} dt = \frac{d\theta}{\text{sen} \theta}$$

cuya solución es:

$$\text{tg} \frac{\theta}{2} = \text{tg} \frac{\theta_0}{2} \exp\left(-\frac{mB}{\lambda} t\right)$$

donde θ_0 es el valor del ángulo θ para $t = 0$ y $z = 0$. El valor del ángulo θ , con que se deposita el grano en el fondo (profundidad h) se obtiene sustituyendo en la ecuación el valor de t obtenido resultando:

$$\operatorname{tg} \frac{\theta}{2} = \operatorname{tg} \frac{\theta_0}{2} \exp\left(-\frac{vmBh}{\lambda\rho g}\right)$$

Cuando $\theta = 0$, el grano está orientado en la dirección del campo magnético inductor. Esta situación se da, cualquiera que sea el ángulo inicial θ_0 , si el exponente es grande, lo que se logra si el campo B es intenso, o la profundidad h muy grande. También favorece el que la viscosidad ν , o el momento de cada grano, sean grandes, o por el contrario que la masa de los granos y el amortiguamiento al giro sean pequeñas. Cualquier combinación de estos factores, que favorezca un exponente grande, permitirá el que la caída del grano sea más lenta y dé tiempo a que el grano se oriente en la dirección del campo inductor.

Una vez que los granos de mineral ferromagnético se depositan y se consolida la roca, sus momentos magnéticos quedan orientados en la dirección del campo inductor. Al perder su libertad de reorientarse, la roca queda así con una magnetización estable remanente, que no cambiará aunque cambie la dirección del campo inductor. Este mecanismo explica la imanación remanente de las rocas sedimentarias. Como la forma de los granos no es esférica, pueden introducirse errores a veces notables en el ángulo de inclinación del magnetismo total de la roca. Después de la deposición hay un periodo durante el cual debido a la poca compactación de la roca, los granos tienen todavía cierta libertad de movimiento y pueden orientarse en la dirección del campo terrestre. La imanación adquirida de esta forma se denomina imanación postdeposicional.

20.4. Polos virtuales paleomagnéticos

Ya se ha visto cómo algunos tipos de rocas pueden adquirir una magnetización estable en la dirección del campo inductor en la época de su formación, ya sea de rocas ígneas por enfriamiento a partir de lavas, o por deposición de rocas sedimentarias u otros mecanismos. Para rocas en la superficie de la Tierra, el campo inductor es el geomagnético. Si suponemos que este campo es esencialmente el de un dipolo centrado, es fácil determinar su orientación. Uno de los resultados más importantes del estudio de la imanación de rocas antiguas es que el campo inductor no corresponde siempre a la orientación actual del dipolo terrestre.

Se llaman *polos virtuales paleomagnéticos* a la orientación de los polos del dipolo responsable de la magnetización de una roca determinada. Supongamos una muestra de una roca tomada en un punto A, cuyas coordenadas son ϕ_A, λ_A , formada en la época t_A que ha adquirido una magnetización en la dirección del vector M, con ángulos de inclinación I^* y declinación D^* . Por la relación de la ecuación entre el ángulo θ y el de inclinación I:

$$\cot g\theta^* = \frac{1}{2} tgl^*$$

donde el ángulo θ^* es el ángulo entre el radio vector al punto A y el polo del dipolo responsable de la magnetización de la muestra. Si se toma en el punto A un ángulo D^* con la dirección norte y sobre el círculo máximo en esa dirección un arco θ^* , el punto P^* definido de esta forma es el polo virtual paleomagnético, que corresponde a la magnetización de la roca A, en el tiempo t_A

Las medidas de estos polos virtuales han de tratarse estadísticamente, ya que tanto en los valores de θ^* y D^* , como en la determinación de la edad t_A de la roca, se pueden introducir errores. El problema de establecer con exactitud la edad de las rocas sólo se ha podido resolver con el desarrollo de los métodos radiométricos de datación.

Las coordenadas ϕ^* y A^* del polo virtual P^* se pueden obtener resolviendo el triángulo esférico P^*NA :

$$\begin{aligned} \text{sen}\phi^* &= \cos\theta^* \text{sen}\phi_A + \text{sen}\theta^* \cos\phi_A \cos D^* \\ \text{sen}(\lambda^* - \lambda_A) &= \text{sen}D^* \text{sen}\theta^* / \cos\phi^* \end{aligned}$$

En teoría, si el campo magnético terrestre ha mantenido a lo largo de las épocas geológicas su carácter dipolar y los continentes no han variado su posición, para una época determinada todas las rocas deben dar polos virtuales en la misma posición. La variación de estos polos con el tiempo representaría la *migración de los polos magnéticos*. Sin embargo, el mismo efecto se lograría manteniendo los polos fijos y haciendo migrar los continentes. Ya veremos más adelante cómo los datos deducidos de los polos virtuales han sido los más definitivos para establecer la deriva de los continentes.

Los datos paleomagnéticos no sólo permiten determinar la posición de los polos, sino también la magnitud del momento magnético del dipolo correspondiente. Como ya se dijo, la magnetización es proporcional a la intensidad del campo inductor; por tanto, sus módulos están relacionados por:

$$M = \chi H$$

Para materiales con magnetización remanente térmica, si se calienta la muestra por encima del punto de Curie y se enfría en presencia del campo geomagnético actual H_0 , adquiere una nueva magnetización.

$$M' = \chi H_0$$

En principio, la susceptibilidad aparente χ debe ser la misma en las dos ecuaciones, por lo que se puede despejar H:

$$H = \frac{M}{M'} H_0$$

en función del campo actual H_0 y de las imanaciones actual y fósil observadas. Esta relación sólo es válida si el valor de χ no varía en el periodo de calentamiento y enfriamiento. Esto no es muy exacto y, en general, las medidas de la intensidad del campo no son tan fiables como las de orientación del dipolo.

20.5 Migración de los polos y de los continentes.

Los resultados de las determinaciones de polos virtuales geomagnéticos para distintas épocas, deducidos de rocas de distintos continentes, han llevado a descubrimientos de gran importancia para la historia de la evolución de la Tierra. Prescindiendo, de momento, de la polaridad de estos polos, los resultados pueden resumirse de la siguiente manera:

Si se consideran los polos virtuales, deducidos de rocas relativamente modernas (del Cuaternario hasta el presente), el resultado es que como valor medio para todos los datos, el campo geomagnético es relativamente estable y coincide con un dipolo centrado en la dirección del eje de rotación de la Tierra. Esto puede verse en la figura, donde se han representado polos virtuales geomagnéticos de rocas hasta veinte millones de años. En la figura hay una cierta dispersión, pero el valor medio coincide con el eje actual de rotación de la Tierra. Algunos autores establecen para los polos del Cuaternario un cierto movimiento de giro del polo magnético con respecto al de rotación, con una amplitud de unos 20° y un periodo de 10^4 a 10^6 años. Esto representaría un movimiento de precesión del eje del dipolo con respecto al eje de rotación.

Los análisis de muestras de rocas antiguas de una misma época tomadas de diversas regiones de un mismo continente proporcionan, como resultado, una agrupación de todos los polos virtuales, lo que indica el carácter dipolar del campo. Sin embargo, esta agrupación no coincide con la del eje actual de la Tierra. Por ejemplo, para muestras del periodo Pérmico (280 ma) tomadas en todo el continente Euroasiático desde Francia a Siberia, los polos virtuales están agrupados en torno a un punto al sur de la península de Kanchatka. Para rocas de periodos cada vez más modernos de un mismo continente, los polos virtuales siguen una trayectoria desde una posición más alejada hasta el polo actual de la Tierra

Si se toman muestras de otros continentes, por ejemplo, América del Norte, para el periodo Pérmico, los polos están agrupados en torno a un punto situado al NW de China, con la misma latitud, pero separado en longitud del polo determinado por las rocas de Eurasia del mismo periodo en 65° . Si tomamos rocas de otros periodos, vemos cómo forman una trayectoria distinta que la seguida por los polos deducidos de las rocas de Eurasia. Si el campo es dipolar, para una época ha de existir un único polo geomagnético y la divergencia es debida a que los continentes no ocupan ahora la misma posición relativa que en periodos geológicos pasados. En el caso de Europa y América del Norte, las divergencias en los polos anteriores al periodo Jurásico se pueden eliminar si se mueven estos dos continentes cerrando el océano Atlántico. De esta forma, los datos del paleomagnetismo permiten reconstruir la posición de los continentes en el pasado. Estos resultados se aplican tanto a los continentes del hemisferio norte, Norteamérica y Eurasia, como los del sur, África, América del Sur, Australia, India y Antártica.

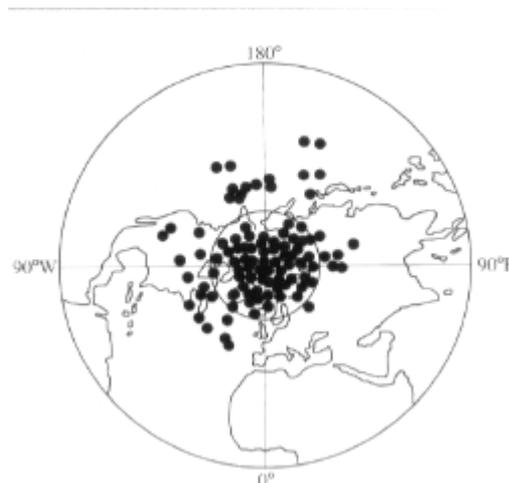


Figura 5. Distribución de polos virtuales paleomagnéticos desde el cuaternario al presente, Udías y Mezcua, 1997.

La diferencia en los polos virtuales de distintas épocas no sólo ha servido para establecer la migración y cambio de situación relativa entre continentes, sino también la de movimientos de pequeños bloques aislados, como pueden ser el movimiento de Japón con respecto a la costa asiática, el de la península Ibérica con respecto a Europa y la rotación de Córcega y Cerdeña.

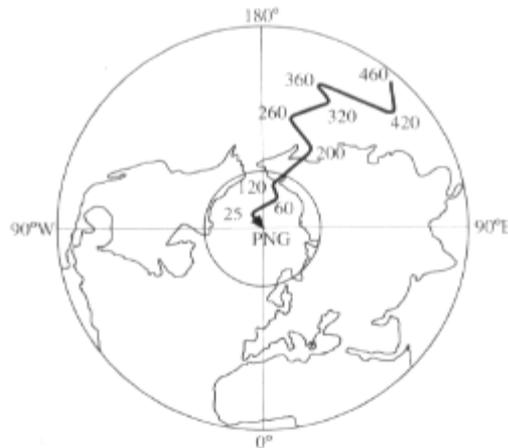


Figura 6. Posición de los polos virtuales de rocas de Eurasia desde el Cámbrico hasta nuestros días. Udías y Mezcua, 1997.

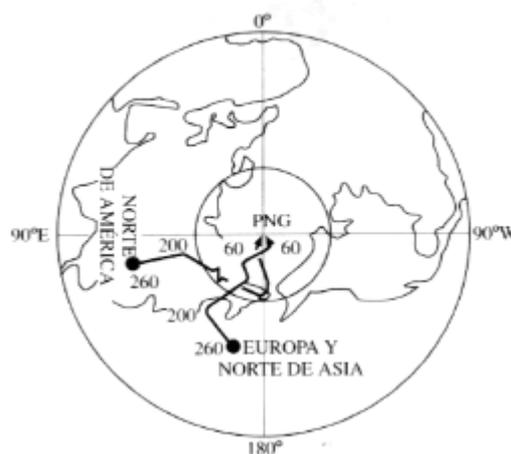


Figura 7. Comparación de la trayectoria de los polos virtuales determinados con rocas de América del Norte y Eurasia. Los valores representan la edad en millones de años. Udías y Mezcua, 1997.

20.6. Inversiones del campo magnético.

En 1905, B. Brunhes, estudiando el magnetismo remanente de lavas en la región cercana a Clermont-Ferrand, detectó que algunas coladas tenían la polaridad invertida con respecto al campo actual terrestre y a las lavas de otras coladas. En ambos casos, la dirección de imanación coincidía, aproximadamente, con la del campo actual. Estas lavas pertenecen al cuaternario superior, con edades entre setecientos mil y dos millones quinientos mil años. Otros autores,

entre los que conviene destacar a Matuyama, que en 1929 estudió lavas de los volcanes de Japón, constataron que no se trataba de un fenómeno local, sino a escala mundial. Se estableció, por tanto, que dependiendo de las épocas, las rocas presentan polaridades normales (en la dirección del campo actual) o inversas (en la dirección opuesta). Ese cambio de polaridad de unas épocas a otras no puede explicarse por un fenómeno de autoinversión en la roca misma. Este fenómeno no es imposible, aunque sí muy excepcional y además improbable que se dé para todas las rocas de una misma época a nivel global. La conclusión, por tanto, sólo puede ser que el campo magnético terrestre ha sufrido inversiones de polaridad en el transcurso del tiempo.

Estas inversiones de polaridad se han detectado también en las muestras de sedimentos marinos obtenidas en sondeos y en secciones sedimentarias en tierra. La polaridad magnética de estos sondeos es altamente coincidente con las épocas de polaridad normal e inversa, obtenida del estudio de lavas, lo que ha confirmado el carácter global de las inversiones. La correlación de los datos deducidos tanto de lavas como de sedimentos marinos ha dado como resultado el establecer una cronología de las épocas en que el campo magnético terrestre es normal o inverso. Para los últimos cuatro millones de años se han establecido cuatro épocas que se denominan de Brunhes (normal), Matuyama (inversa), Gauss (normal), Gilbert (inversa). Sin embargo, en cada una de estas épocas hay periodos más o menos cortos denominados sucesos, en los que la polaridad tiene el signo opuesto a la correspondiente de la época. Esto complica la situación multiplicando el número de inversiones. Modernamente se ha introducido la terminología de supercrono, crono y subcrono para periodo, época y evento. La figura muestra las inversiones de polaridad del campo magnético terrestre en los últimos cinco millones de años con las épocas (cronos) de Brunhes, Matuyama, Gauss y Gilbert junto con algunos eventos (subcronos) como Jaramillo, Gilsa y Olduvai.

Otro tipo de observaciones que ha permitido extender la historia de las inversiones del campo magnético hasta épocas geológicas muy antiguas es el de las anomalías magnéticas observadas a ambos lados de las dorsales oceánicas. Estas anomalías se presentan en forma de franjas paralelas a ambos lados de dichas dorsales, con valores alternantes positivos y negativos, y se han interpretado como debidas a franjas de material con magnetización alternativamente normal (positivas) e inversa (negativas). El mecanismo de la formación de estas franjas de material está basado en el mecanismo de extensión del suelo oceánico. Brevemente podemos citar aquí que el material del manto surge en el eje de las dorsales y se va separando de éste al surgir nuevo material. En el momento de depositarse y enfriarse, el material adquiere la imanación con la polaridad correspondiente a esa época. Al continuar el proceso de extensión del suelo oceánico, se van sucediendo franjas paralelas de material de polaridad alternante. Estas observaciones han permitido extender la escala de inversiones hasta épocas geológicas muy antiguas. Debido a que

las dataciones, al aumentar la edad de las rocas, son cada vez menos exactas, la escala de inversiones lo es también.

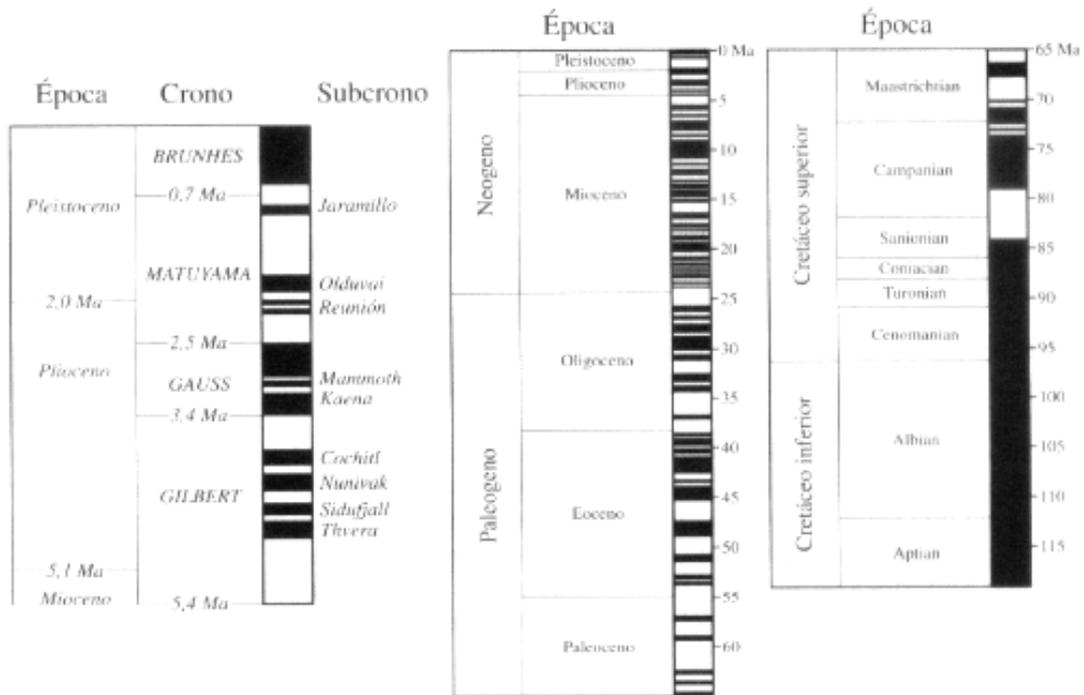


Figura 8. Épocas de polaridad normal (negro) e invertida (blanco), Udías y Mezcuca, 1997.

Bibliografía

- [1] A. Udías y J. Mezcuca. (1997). Fundamentos de Geofísica. Ed. Alianza Universidad. CAP.18.
- [2] W. Lowrie. (2000). Fundamentals of Geophysics. Ed. Cambridge Press. CAP. 5