

ESTIMADORES DE MÍNIMA DESVIACIÓN ORTOGONAL; CASO LINEAL

José Frez C.

Depto. de Sismología, División de Ciencias de la Tierra, CICESE
Km 107, Carretera Tijuana-Ensenada
Correo Electrónico: jofrez@cicese.mx

RESUMEN

Este artículo intenta popularizar el criterio según el cual el ajuste de un modelo a observaciones se mide por la distancia ortogonal de los puntos-datos a la superficie que representa al modelo matemático. Con ello, se toma en cuenta los errores tanto de las variables independientes como de la dependiente, situación que se presenta en las Ciencias de la Tierra. En contraste, la familia de los métodos de cuadrados mínimos supone que sólo la variable dependiente tiene error.

El álgebra correspondiente al caso lineal se desarrolla con detalle, al mismo tiempo que se describen procedimientos numéricos para la determinación del estimador central y de la variabilidad estadística correspondiente. De todas estas técnicas, aplicamos una basada en el cálculo de un vector característico para el estimador central y otra de remuestreo, para una medida de variabilidad. Los procedimientos se aplican a un caso tomado de la literatura.

Se aclara la relación entre el estimador de mínima desviación ortogonal y el de máxima verosimilitud con errores gaussianos en las observaciones. Ambos son equivalentes si podemos aplicar un factor de ponderación distinto a cada variable; este factor se asocia al error promedio de las observaciones correspondientes. El método de máxima verosimilitud tiene una mayor generalidad cuando los errores presentan mucha dispersión en al menos una de las variables.

1. INTRODUCCIÓN

Existen una gran cantidad de alternativas y mejoramientos del método estándar de cuadrados mínimos (MCE) para resolver sistemas sobredeterminados de ecuaciones; el objetivo es inferir los valores de los parámetros del modelo a partir de un conjunto de observaciones. Este artículo intenta divulgar un criterio que ha llamado la atención a los analistas de datos desde hace más de 100 años (v. gr., Adcock, 1878). En lo que sigue, se trata el caso de un modelo lineal. Como nomenclatura, utilizamos letras latinas minúsculas para representar vectores; letras latinas mayúsculas, para matrices, y letras griegas, para escalares. La excepción son las letras latinas minúsculas con subíndices que pueden ser vectores o escalares (componentes de vectores); el contexto deja claro el significado correspondiente. El superíndice "T" indica la operación de transposición y el acento circunflejo indica el "valor estimado" de la variable sobre la cual está dicho símbolo.

El método estándar de cuadrados mínimos minimiza la norma cuadrática de los residuales, donde el residual i -ésimo se define por

$$r_i = y_i - u_i^T \hat{x} - \hat{x}_0 \quad (1)$$

donde y_i es un elemento del vector de observaciones. Las compo-

nentes del vector $\begin{Bmatrix} x \\ x_0 \end{Bmatrix}$ corresponden a los coeficientes (incógnitas) del hiperplano; el subvector x , a las pendientes; el término

inhomogéneo (escalar) x_0 es el parámetro de posición; u_i^T es el vector-fila que corresponde a la "realización" empírica del modelo para el dato i -ésimo.

En vez de medir el ajuste o desajuste del modelo a los datos a través del vector de residuales, la alternativa que manejaremos minimiza la distancia ortogonal de los puntos-datos al hiperplano que representa al modelo. Suponemos adicionalmente que la matriz formada por todos los vectores filas $\{u_i^T\}$ tiene sus columnas linealmente independientes entre sí y con respecto al vector y ; es decir, suponemos que tenemos suficiente información para estimar independientemente los elementos del vector de incógnitas.

De este modo, minimizamos la distancia del punto $\{y_i, u_i\}$ al hiperplano descrito por

$$y = u^T x + x_0 \quad (2)$$

El vector u representa las coordenadas de las variables independientes mientras que la letra " x " denota "incógnita"; la variable y se toma como dependiente aunque la idea básica del método es introducir una simetría entre las variables, ya que todas participan en la medición de distancia que es necesario minimizar; ésto, en contraste con el método de cuadrados mínimos que considera sólo un ajuste en la variable dependiente.

Designamos al criterio como de “mínima distancia ortogonal” (MDO); la literatura estadística también lo llama de “regresión ortogonal”. Dejando explícito el uso de la norma cuadrática, también recibe el nombre de “mínimos cuadrados total” (TLS, por las siglas en inglés) desde que lo hicieran Golub y Van Loan (1980). Este criterio se relaciona al de “regresión con errores en todas las variables” que considera que todas las variables de la regresión están sujetas a error; esto implica una ponderación implícita en la estimación; este tópico es discutido en la Sección 6. Hasta esa sección, no incluimos, por simplicidad notacional y de exposición, a una ponderación de las mediciones que se base en la bondad relativa de ellas. El lector debe recordar en este respecto que el método de cuadrados mínimos ponderados (v. gr., Lawson y Hanson, 1973) generaliza al estándar (MCE) con una ponderación de las ecuaciones a través de la matriz de covariancias de los errores de la variable dependiente. Finalmente, por simplicidad, utilizamos la norma cuadrática en vez de la menos popular, pero más robusta, l_1 ; una reciente referencia en esta última dirección es Watson (2002). En resumen, el criterio MDO a desarrollar en la sección siguiente es una versión cuadrática, bien condicionada y no ponderada.

Aunque existen procedimientos computacionales probados (Späth, 1987; Van Huffel y Vandewalle, 1991; Kent *et al.*, 1990, entre otras referencias), hay pocas aplicaciones publicadas en Ciencias de la Tierra del criterio MDO o, más en general, del criterio de regresión que considera errores en todas las variables. En su revisión, Van Huffel y Vandewalle (1991) mencionan sólo un artículo (Fisher, 1989) que pertenece a estas ciencias. Una excepción a esta aparente falta de interés la constituye el trabajo para construir isocronas, donde es necesario tomar en cuenta los errores de todas las variables (v.gr., McIntyre *et al.*, 1966; Kent *et al.*, 1990). Pienso que la divulgación del criterio MDO puede ser beneficiosa, particularmente si se toma en cuenta que, por ejemplo, en Geología, Sismología Estadística e Ingeniería Sísmica se hace un gran uso de regresión lineal, a veces en escalas semilog o log-log, en que todas las variables están sujetas a error. No es la intención de este artículo hacer una revisión exhaustiva del tema; más bien se trata de presentar un método poco utilizado, explorar su lógica, compararlo con otros y guiar al lector en su aplicación, entregando referencias adecuadas. Dado que el tratamiento del caso de la línea recta es trivial con los recursos modernos de computación (Sección 8), trataremos el caso multivariante; el precio es un manejo de álgebra lineal (v.gr., Noble y Daniel, 1989) con las expresiones matriciales en notación compacta e indicial. Además de sus propósitos de divulgación, la descripción y discusión de los métodos que se acompañan pueden ser útiles en la enseñanza ya que los textos más generales no incluyen los tópicos aquí cubiertos.

3. OBTENCIÓN DE LA FUNCIÓN CRITERIO

Sea el modelo lineal aplicado a un conjunto de datos

$$y = A x + e \quad m > n \quad (3)$$

donde e denota el vector de error. El criterio MCE es

$$\min_x \|Ax - y\|_2 \quad (4)$$

Para introducir el criterio MDO, podemos generalizar expresiones para los casos $n = 2$ y $n = 3$. Para $n = 2$ y m datos, tenemos que

$$A = \begin{bmatrix} 1 & u_1 \\ 1 & u_2 \\ \vdots & \vdots \\ 1 & u_m \end{bmatrix}; \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}; \quad x = \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} \quad (5)$$

La distancia de cada punto $P_i(u_i, y_i)$ a la recta

$$y = x_0 + x_1 u \quad (6)$$

es

$$d_i = \frac{y_i - x_1 u_i - x_0}{\sqrt{1 + x_1^2}} \quad (7)$$

Para problemas de regresión, la variable dependiente es y mientras que la independiente es u . Tenemos m pares de mediciones. En el caso $n = 3$, tenemos dos variables independientes (u_1 y u_2) y el modelo está dado por

$$y = x_0 + x_1 u_1 + x_2 u_2 \quad (8)$$

La distancia ortogonal correspondiente a (7) resulta ser

$$d_i = \frac{y_i - x_1 u_{i1} - x_2 u_{i2} - x_0}{\sqrt{1 + x_1^2 + x_2^2}} \quad (9)$$

En el caso multivariante (es decir, cuando el vector de variables independientes u tiene múltiples componentes), las expresiones (7) y (9) se generalizan a

$$d_i = \frac{y_i - u_i^T x - x_0}{(1 + x^T x)^{1/2}} \quad (10)$$

Por lo tanto, la función criterio es

$$\Omega_2(x) = \sum_{i=1}^m \frac{|y_i - u_i^T x - x_0|^2}{1 + x^T x} = \frac{\|y - U^T x - x_0\|_2^2}{1 + \|x\|_2^2} \quad (11)$$

donde la matriz U contiene las coordenadas $\{u_{ij}\}$ de las variables independientes. Observe que el denominador no contiene al término inhomogéneo x_0 .

Una minimización de la función criterio (11) con respecto a x_0 produce

$$m \hat{x}_0 = \sum_{i=1}^m y_i - \left(\sum_{i=1}^m u_i^T \right) x$$

O sea,

$$\hat{x}_0 = y - \bar{u}^T x \quad (12)$$

donde las barras indican valores promedios.

Reemplazando (12) en (11), obtenemos

$$\Omega_2(x) = \frac{1}{1 + x^T x} \sum_{i=1}^m [(y_i - y) - (u_i - \bar{u})^T x]^2 \quad (13)$$

Para el caso de un modelo homogéneo $x_0 = 0$,

$$\Omega_2(x) = \frac{1}{1 + x^T x} \sum_{i=1}^m (y_i - u_i^T x)^2 \quad (14)$$

Las ecuaciones (13) y (14) son matemáticamente equivalentes si aplicamos transformaciones simples.

Si deseamos utilizar la norma p-ésima, la generalización de (14) es

$$\Omega_p(x) = \frac{1}{(1 + x^T x)^{p/2}} \sum_{i=1}^m |y_i - u_i^T x|^p \quad (15)$$

4. UNA SOLUCIÓN NUMÉRICA PARA LA NORMA CUADRÁTICA

Tomemos la solución numérica que describe el texto de Späth (1987) y cuya lógica, con algunas aclaraciones, seguimos en esta sección. La función criterio (14) se puede tomar como la razón entre dos normas cuadráticas con la transformación

$$w = \begin{pmatrix} x \\ 1 \end{pmatrix}; \quad Z = (A; -y) \quad (16)$$

con lo que (14) queda como

$$\Omega_2(x) = \frac{\|Z w\|_2^2}{\|w\|_2^2} = \frac{w^T Z^T Z w}{\|w\|_2^2} \quad (17)$$

El propósito obvio de la transformación (16) es asimilar $\Omega_2(x)$ a un cociente de Rayleigh (v. gr., Noble y Daniel, 1989), con lo cual el problema de minimización equivale al cálculo del vector característico asociado al eigenvalor más pequeño de la matriz positiva definida $Z^T Z$.

Por lo tanto, la minimización de $\Omega_2(x)$ equivale a buscar el valor característico **más grande** en el problema

$$(Z^T Z)^{-1} w = \lambda w \quad (18)$$

para obtener el vector propio correspondiente, que es lo que buscamos. La forma (18) nos permite aplicar el **método de potencias** y obtener el vector característico directamente; ver, por ejemplo, Hildebrand (1965), Acton (1970), Späth (1987). Este método aplica iterativamente la relación

$$w^{(k+1)} = (Z^T Z)^{-1} w^{(k)} \quad (19)$$

donde los superíndices entre paréntesis indican el orden de la iteración. La inicialización puede ser, por ejemplo,

$$w^{(0)} = I \quad \text{o} \quad w^{(0)} = e_i \quad \text{para algún } i, \quad (20)$$

donde el vector I tiene todas sus componentes iguales a uno y $\{e_i\}$ son los usuales vectores bases de un sistema cartesiano de coordenadas. Para solucionar el sistema de ecuaciones (19) para $w^{(k+1)}$, podemos aplicar la descomposición de Choleski a $Z^T Z$, incluyendo la posibilidad de hacerlo como $R^T D R$, donde R es una matriz triangular superior con unos en la diagonal principal y D es una matriz diagonal. Acton (1970) demuestra que el método de potencias converge si los valores característicos no se acumulan cerca del que se busca. En el caso extremo, puede haber más de un vector característico asociado a este eigenvalor. Como criterio de convergencia, parece adecuado utilizar la norma cuadrática de la diferencia entre $w^{(k+1)}$ y $w^{(k)}$ después de normalizar los vectores de tal forma que $\|w^{(k)}\|_2 = 1$. Esta normalización tiene la ganancia extra de evitar aumentos o disminuciones drásticas de los valores numéricos de las componentes del vector al pasar de una iteración a otra; esto sucede en los casos de aplicación que describimos en la sección 7. Obtenida la convergencia, aplicamos la condición (16) para el último elemento del vector obtenido, i.e., $w_{n+1} = 1$.

5. ALGUNAS PROPIEDADES DEL ESTIMADOR

La ecuación (8) equivale a

$$\lambda \begin{pmatrix} A^T A & -A^T y \\ -y^T A & y^T y \end{pmatrix} \begin{pmatrix} \hat{x} \\ 1 \end{pmatrix} = \begin{pmatrix} \hat{x} \\ 1 \end{pmatrix} \quad (21)$$

o sea, dos ecuaciones de las cuales podemos eliminar a λ . El resultado es

$$(A^T A - y^T r I_n) \hat{x} = A^T y \quad (22)$$

I_n es la matriz unidad cuya dimensión está indicada por el subíndice n ; r es el vector de residuales. El sistema (22) no es lineal en las incógnitas y puede utilizarse para formar un esquema iterativo que

determine al vector de incógnitas; la no-linealidad se evita con el uso del residual de la iteración anterior y todo el esquema es iniciado con la solución MCE. Los resultados de este procedimiento computacional son expuestos y discutidos en la sección 7. El sistema de ecuaciones (22) tiene una estabilidad distinta a la de las ecuaciones normales para MCE con el que se diferencia por el segundo término dentro del paréntesis. Obviamente, el estimador de MDO es sesgado pero, si los elementos de A tienen error, también es sesgada la solución MCE.

Otras dos propiedades se pueden obtener con facilidad utilizando la ecuación (12) para el término libre. Específicamente, la suma de los residuales y la suma de las distancias ortogonales tienen, cada una, valores nulos para los estimadores MCE y MDO, respectivamente.

Para otras propiedades de los estimadores de MDO y MCE, se puede consultar el texto de Van Huffel y Vandewalle (1991). Algunos detalles adicionales son expuestos en las secciones 6 y 7.

La obtención de regiones de variabilidad de los parámetros, debido a la propagación del error observacional a través de un operador o procedimiento de inversión, es un problema complicado pero de importancia crucial. El único caso relativamente simple, suponiendo estabilidad, es el lineal para estimadores mínimo-cuadráticos ya que el operador pseudoinverso resulta lineal; en este caso, es posible calcular fácilmente las matrices de covariancias del estimador y utilizar la propiedad de que el carácter gaussiano del error observacional se transmite al del estimador. Para estimadores que no son lineales y, peor, sin la suposición de que los errores observacionales tienen distribución gaussiana, se cuenta con las siguientes líneas de acción. Una opción es buscar alrededor de la solución central a las soluciones que satisfacen, dentro de su error, a los datos. La búsqueda exhaustiva tiene un alto costo computacional a medida que aumenta el número de incógnitas. Ello lleva a aplicar astucia en el modo de barrer el espacio de soluciones alrededor de la solución "central". La técnica de búsqueda "puercoespín" ("hedgehog", en inglés; v. gr., Biswas y Knopoff, 1974) es recomendable en este caso. Otras tácticas de búsqueda utilizan técnicas de remuestreo; la de "bootstrap" (Efron y Tibishirani, 1986) ha sido aplicada a nuestro problema por Kirby (1991); otras técnicas de remuestreo son la de validación cruzada (Golub *et al.*, 1979) y "jackknife" (Miller, 1974). La idea básica del remuestreo es seleccionar al azar muchos subconjuntos de datos y, luego, calcular la variabilidad de los resultados. Una última opción acude al carácter aproximadamente gaussiano de la variabilidad de la solución y ajusta una (hiper)superficie de segundo grado alrededor de la solución central en la función criterio, con lo que se determina analítica o numéricamente la conducta asintótica gaussiana correspondiente; el capítulo 8 de Van Huffel y Vandewalle (1991) ofrece expresiones útiles para esta última opción.

En nuestros ejercicios de aplicación, utilizamos la técnica "bootstrap" para estimar la variabilidad del vector solución. Esta técnica consiste en producir un número alto de remuestreos, con repetición, del mismo número de mediciones que el original. Luego, podemos calcular promedios y desviaciones estándares de las estimaciones obtenidas, para lo cual los histogramas correspondientes

son de gran ayuda. La robustez del resultado con respecto a las observaciones se mide directamente; por ejemplo, una recta que depende críticamente de un punto aislado obtendrá un grado alto de variabilidad de sus parámetros al utilizar "bootstrap" ya que habrá casos en que ese punto aislado no se toma en cuenta. Ante la existencia de un número pequeño de mediciones que se apartan de la conducta central predicha por el modelo, pueden aparecer máximos secundarios en los histogramas por las características señaladas del remuestreo. Ello puede plantear la necesidad de filtrar los histogramas, eliminando dichos máximos secundarios si se considera que no son representativos. Si el máximo global tiene un valor mucho más alto que el de los locales y está claramente separado de ellos, tomamos dicho máximo global y su dispersión como representativos; con ello, evitamos que datos aislados y sesgados predominen en la estimaciones.

6. EL CRITERIO DE MDO COMPARADO CON EL DE MÁXIMA VEROSIMILITUD EN QUE TODAS LAS VARIABLES ESTÁN SUJETAS A ERROR GAUSSIANO

En esta sección, usamos los errores de medición para formar criterios de estimación. Para ello, necesitamos un modelo de error y un criterio; es común utilizar el modelo gaussiano y el criterio de máxima verosimilitud (MV), respectivamente. Siguiendo un procedimiento estándar (v. gr., Kent *et al.*, 1990), el logaritmo de la función de verosimilitud resulta ser proporcional a

$$-\frac{1}{2} \sum_{i=1}^m (u_i - \mu_i)^T C_{ii}^{-1} (u_i - \mu_i) \quad (23)$$

con la condición lateral, dada por el modelo,

$$u_i^T x = 0 \quad (24)$$

y hemos adoptado $u_{i,n+1} = y_i$ y $x_{n+1} = -1$ para hacer la notación más compacta. La ecuación (24) corresponde al modelo lineal utilizado en la Sección 2. La matriz simétrica C_{ii} es la de covariancias de los errores de observación. De este modo, maximizamos la siguiente expresión con respecto a los valores promedios μ_i

$$-\frac{1}{2} (u_i - \mu_i)^T C_{ii}^{-1} (u_i - \mu_i) + \lambda_i u_i^T x \quad (25)$$

donde λ_i es un multiplicador de Lagrange. La maximización permite obtener la función criterio que debe optimizarse con respecto a los coeficientes del modelo; ella (Kent *et al.*, 1990) es proporcional a

$$L = -\frac{1}{2} \sum_{i=1}^m \frac{u_i^T x}{x^T C_{ii} x} \quad (26)$$

Si los errores son iguales para todas las variables, resulta que L es proporcional a

$$\frac{1}{x^T x + 1} \sum_{i=1}^m (u_i^T x - y_i)^2 \quad (27)$$

donde hemos vuelto a utilizar la simbología de las secciones anteriores. Así, el criterio de MV (con errores gaussianos) se reduce al de MDO.

Otro caso, de mayor interés, es cuando todas las mediciones de la misma variable tienen el mismo error y no tienen correlación, aunque cada variable está sujeta a un error distinto. Llamamos e_k y e_y a los errores de las variables independientes (u_k) y dependiente (y), respectivamente. Así, (26) se puede escribir como

$$\frac{1}{e_y^2 \sum_{k=1}^n x_k^2 e_k^2 + 1} \sum_{i=1}^m \left(\frac{1}{e_y} \sum_{k=1}^n \frac{u_{ik}}{e_k} x_k e_k - \frac{y_i}{e_y} \right)^2 \quad (28)$$

que, con las transformaciones

$$x_k \leftarrow \frac{x_k e_k}{e_y}; \quad u_{ik} \leftarrow \frac{u_{ik}}{e_k}; \quad y_i \leftarrow \frac{y_i}{e_y} \quad (29)$$

toma también la forma (27). Este es el caso más general que hace equivalente los dos estimadores considerados en esta sección. La condición general para que exista equivalencia es que el índice asociado a cada medición (i , en las ecuaciones 26 y 28) no puede participar en la ponderación de las variables independientes.

Concluimos que el criterio de MV, unido a la hipótesis gaussiana para los errores de observación, toma mejor en cuenta a estos errores en la estimación de los parámetros. Si los errores de medición de una variable son parecidos, se puede tomar el error promedio para ponderar; si ello se justifica para cada una de las variables, el método de MDO es recomendable. Si los errores individuales tienen mucha variabilidad, es preferible maximizar (26); por ejemplo, con un esquema numérico iterativo del tipo Newton-Raphson como lo hacen Kent *et al.* (1990). Este esquema puede adaptarse opcionalmente para calcular el estimador MDO; Watson (2002) usa uno del mismo tipo al aplicar la norma l_1 para la regresión ortogonal.

7. APLICACIÓN

Los resultados de la aplicación del método de potencias para resolver (18) son expuestos a continuación; luego, describimos los resultados de la aplicación del algoritmo que se basa en (22). Los resultados de un método de búsqueda directa en el espacio de soluciones son expuestos al final de esta sección.

La Tabla 1 (Middleton, 2000), que simplifica la de Forsyth y Uyeda (1975), contiene valores de una variable dependiente (velocidades de placas) e independientes (u_1 : área total de la placa; u_2 : área de la parte continental; u_3 : longitud efectiva de la cordillera oceánica, y u_4 : longitud efectiva de los límites con las trincheras correspondientes). Las tablas 2 a 5 muestran los resultados del ajuste de los criterios de mínima distancia ortogonal y de cuadrados mínimos estándar en los casos que se describen a continuación. No es la intención de esta sección reemplazar el extenso análisis hecho en la publicación original. El propósito de utilizar los datos de la Tabla 1 es discutir aspectos de implementación del criterio MDO y compararlos con los resultados del criterio MCE.

La Tabla 2 resume los resultados de la regresión para los métodos MDO y MCE cuando se utilizan las cuatro variables independientes. El método MDO necesita cuatro iteraciones para obtener 5 cifras significativas en los coeficientes. No se ven diferencias en los resultados de ambos métodos; esto se explica si se observan los valores de las pendientes obtenidas. El hiperplano de regresión es prácticamente perpendicular al eje de la variable dependiente; por lo tanto, ambos métodos (MDO y MCE) coinciden aproximadamente. Esto plantea el problema de escalamiento (sección 6) y el hecho de que la ortogonalidad que aplicamos no es invariante a cambios de escala. Es decir, un simple cambio de unidades en uno de los ejes puede producir resultados diferentes.

La escala más adecuada para las variables de un problema de regresión resulta al normalizarlas en unidades de los errores de observación correspondientes. En la sección 6, hemos justificado este procedimiento si, para cada variable, los errores son estocásticamente independientes y con un mismo o muy parecido valor. Cuando no tenemos información acerca del error de una variable, supondremos que es del orden de unas pocas unidades en su última cifra significa-

TABLA 1. Información Procesada

Placas	Área Total x 10 ⁶ km ²	Área Continental x 10 ⁶ km ²	Longitud Efectiva Cordillera Océánica x 10 ² km	Longitud Efectiva Zona Subducción x 10 ² km	Velocidad Placa cm/año
	u_1	u_2	u_3	u_4	y
NA	60	36	86	10	1.1
SA	41	20	71	3	1.3
PAC	108	0	119	113	8.0
ANT	59	15	17	0	1.7
IND	60	15	108	83	6.1
AF	79	31	58	9	2.1
EUR	69	51	35	0	0.7
NAZ	15	0	54	52	7.6
COC	2.9	0	29	25	8.6
CAR	3.8	0	0	0	2.4
PHIL	5.4	0	0	30	6.4
ARAB	4.9	4.4	27	0	4.2

Variables independientes (u_1 a u_4) y dependiente (y) utilizadas en la aplicación del procedimiento computacional. Tomado de Middleton (2000) que a su vez, lo adapta de Forsyth y Uyeda (1975).

Tabla 2. Comparación: Métodos MCE y MDO

y	MCE			MDO		
	yhat	res	d	yhat	res	d
1.1	1.30	-0.200	-0.1988	1.30	-0.195	-0.1944
1.3	1.91	-0.614	-0.6113	1.90	-0.605	-0.6023
8.0	7.84	0.156	0.1556	7.84	0.156	0.1557
1.7	1.85	-0.150	-0.1495	1.84	-0.143	-0.1428
6.1	7.17	-1.070	-1.0658	7.18	-1.077	-1.0729
2.1	0.98	1.116	1.1115	0.98	1.122	1.1174
0.7	0.66	0.037	0.0366	0.67	0.030	0.0301
7.6	7.38	0.218	0.2169	7.39	0.212	0.2110
8.6	6.03	2.571	2.5607	6.03	2.568	2.5578
2.4	4.40	-2.004	-1.9957	4.40	-2.003	-1.9943
6.4	6.75	-0.355	-0.3532	6.76	-0.364	-0.3622
4.2	3.91	0.294	0.2930	3.90	0.298	0.2969
Normas Medias:		1.07	1.07	1.07	1.07	
Soluciones:						
$X_{MCE} =$		4.54710	-0.03767	-0.01529	-0.01442	0.08037
$X_{MDO} =$		4.54645	-0.03785	-0.01479	-0.01459	0.08072
$DE_{MDO} =$		1.1	0.021	0.036	0.013	0.021

Resultados del ajuste mínimo-cuadrático estándar (MCE) y de mínima distancia ortogonal (MDO). El vector de soluciones contiene x_0 y de x_1 a x_4 (asociados a las columnas u_1 a u_4 de la Tabla 1). yhat: observación predicha por el modelo y el ajuste; res: residuales; d: distancia ortogonal. DE: desviación estándar estimada por "bootstrap" (Figura 4).

tiva. Al aplicar este criterio, resulta que las velocidades de placas tienen un factor aproximado de 10 en exactitud menor que las variables independientes. De este modo, podemos simular la normalización aplicando este factor de 10. Los resultados, que deben presentar una situación más realista, aparecen en la Tabla 3, donde se aprecian claras diferencias, incluyendo de signos, entre ambas estimaciones. Como es de esperar, ambos métodos producen valores mínimos de las normas medias en los valores que le corresponden, de residuales para MCE y de distancia ortogonal para MDO. Con 14 iteraciones, se obtienen 5 cifras significativas en la estimación MDO. El costo en preferir los resultados del criterio MDO son, en norma media de residuales, de 10.70 (MCE) a 17.05 (MDO) mientras que la ganancia para las distancias ortogonales es de 7.91 (MCE) a 6.87 (MDO).

Middleton (2000) presenta, además, el resultado de la regresión entre la velocidad de placa y u_4 , cosa que también hemos hecho para agregar un grado de visualización a las tablas numéricas. La Tabla 4 y la Figura 1 entregan los resultados (3 iteraciones para el método MDO) para este caso. Se aprecia, otra vez, que los resultados de ambos métodos son prácticamente indistinguibles y puede observarse que las rectas que resultan de los ajustes serían casi horizontales si, como ya hemos propuesto, las dimensiones de los ejes se calculara en unidades del error de las mediciones y la unidad error tuviera la misma longitud en ambos ejes. Esto implica aplicar al eje vertical un factor aproximado de 10. La Figura 1 incluye curvas simétricas alrededor de la línea recta MCE que corresponden a intervalos de confianza del 95 % y 99 % para el valor central estimado. La Tabla 5 y la Figura 2 presenta los resultados (5 iteraciones para el método MDO) cuando las velocidades de placas se multiplican por un factor de 10; ello se hace, nuevamente, para simular una normalización relativa de los ejes en términos del error. En este caso, hay clara diferencia entre ambas rectas estimadas aunque, para efectos predictivos, la diferencia está dentro de la variabilidad permitida de la solución MCE al nivel del 95%. Las curvas que entregan los intervalos de confianza no cambian relativamente a las de la Figura 1

Tabla 3. Comparación: Métodos MCE y MDO

y	MCE			MDO		
	yhat	res	d	yhat	res	d
11.0	13.00	-1.996	-1.4750	12.86	-1.858	-0.7486
13.0	19.14	-6.138	-4.5349	3.58	9.421	3.7961
80.0	78.44	1.562	1.1540	67.55	12.449	5.0161
17.0	18.50	-1.501	-1.1088	2.25	14.748	5.9424
61.0	71.70	-10.702	-7.9067	87.23	-26.227	-10.5674
21.0	9.84	11.161	8.2458	1.58	19.421	7.8249
7.0	6.63	0.367	0.2715	33.92	-26.916	-10.8450
76.0	73.82	2.178	1.6092	82.49	-6.492	-2.6157
86.0	60.29	25.713	18.9974	62.68	23.321	9.3964
24.0	44.04	-20.040	-14.8058	36.81	-12.811	-5.1619
64.0	67.55	-3.547	-2.6206	81.91	-17.908	-7.2156
42.0	39.06	2.942	2.1739	29.15	12.852	5.1783
Normas Medias:		10.70	7.91	17.05	6.87	
Soluciones:						
$X_{MCE} =$		45.47098	-0.37668	-0.15291	-0.14425	0.80367
$X_{MDO} =$		40.03500	-0.84837	1.34984	-0.46923	1.54848
$DE_{MDO} =$		10.5	0.16	0.28	0.090	0.19

Similar a la Tabla 2, salvo que los valores de la variable y se multiplican por un factor de 10.

Tabla 4. Comparación: Métodos MCE y MDO

y	MCE			MDO		
	yhat	res	d	yhat	res	d
1.1	3.19	-2.092	-2.0881	3.19	-2.089	-2.0851
1.3	2.79	-1.485	-1.4828	2.78	-1.481	-1.4785
8.0	9.17	-1.171	-1.1689	9.19	-1.186	-1.1840
1.7	2.61	-0.911	-0.9096	2.61	-0.906	-0.9048
6.1	7.43	-1.329	-1.3271	7.44	-1.339	-1.3369
2.1	3.13	-1.034	-1.0318	3.13	-1.030	-1.0286
0.7	2.61	-1.911	-1.9079	2.61	-1.906	-1.9031
7.6	5.63	1.970	1.9669	5.63	1.966	1.9625
8.6	4.06	4.538	4.5300	4.06	4.538	4.5303
2.4	2.61	-0.211	-0.2108	2.61	-1.206	-0.2060
6.4	4.35	2.047	2.0439	4.35	2.047	2.0434
4.2	2.61	1.589	1.5862	2.61	1.594	1.5910
Normas Medias:		1.97	1.97	1.97	1.97	
Soluciones:						
$x =$		2.6111	0.05805	$x =$	2.60643	0.05823
				$DE_{MDO} =$	0.63	0.011

Similar a la Tabla 2, salvo que la regresión es sólo sobre los valores de u_4 (longitud efectiva del límite con zona de subducción).

ya que se basan en la dispersión media de los residuales que dependen de la escala de graficación del eje vertical; esta dispersión indica que el error medio (una desviación estándar) de las velocidades de placa es de 0.8 [cm/año] si es que aceptamos que la relación lineal es estructural (basada en la física del problema). Sin embargo, esto no es probablemente cierto y el propósito geofísico de todo el ejercicio es sólo semicuantitativo, es decir, en busca de correlaciones. Obviamente, los errores de las velocidades no pueden ser tan grandes y es el modelo el que sobrestima este error. Para una discusión acerca de las limitaciones y justificación del ajuste por regresión, véase Webster (1998).

Llamamos al algoritmo anterior A1, para designar como A2 al que se deriva de aplicar iterativamente la ecuación (22). La primera aplicación de A2 es indistinguible con los resultados mostrados en la Tabla 2 para A1; la convergencia es rápida: en tres iteraciones se obtiene lo mismo. Cuando lo aplicamos al segundo experimento (la variable dependiente multiplicada por 10); sin embargo, los resultados fluctúan sin alcanzar convergencia en 500 iteraciones. Con el objeto de entender mejor el comportamiento de este algoritmo, los valores de las velocidades de placa se multiplicaron por diversos

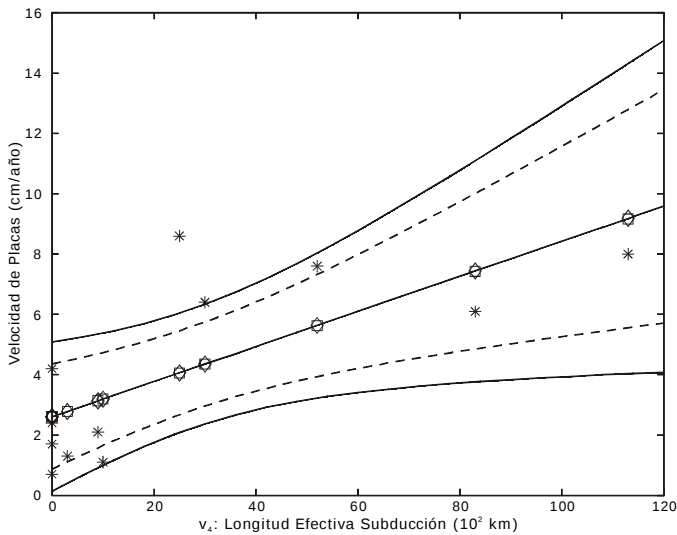


Figura 1. Resultado del ajuste MCE (cuadrados) y MDO (diamantes) para y en función de u_4 (ver tablas 1 y 4). En este caso, ambos ajustes prácticamente coinciden. Los asteriscos representan los datos; los puntos señalados sobre la recta de ajuste, los valores correspondientes predichos, según los parámetros estimados con el modelo lineal. Las curvas señalan intervalos de confianza del 95 % (líneas segmentadas) y 99 % (líneas continuas) alrededor del valor central predicho por el criterio MCE.

factores entre 1 y 10. Para factores mayores a 7.5, no se alcanza convergencia mientras que para factores menores a 7.0 sucede lo contrario. Para un factor de 7.0, los resultados oscilan fuertemente en las primeras 10 iteraciones para luego, obtenerse 4 o 5 cifras significativas en las componentes del vector de incógnitas a 60 iteraciones. No me pareció necesario aplicar alguna técnica de relación para tratar de estabilizar el esquema: el algoritmo A1 parece claramente preferible.

En vez de utilizar métodos numéricos para minimizar una función criterio es posible obtener la solución por una búsqueda directa en el espacio de soluciones. En vez de una búsqueda exhaustiva (muy costosa o llanamente imposible de realizar), se prefiere aplicar una selección aleatoria en la búsqueda, tal como en el método genético. Esta opción me parece más eficiente que utilizar técnicas de programación lineal (simplex). La Figura 3 ilustra el muestreo

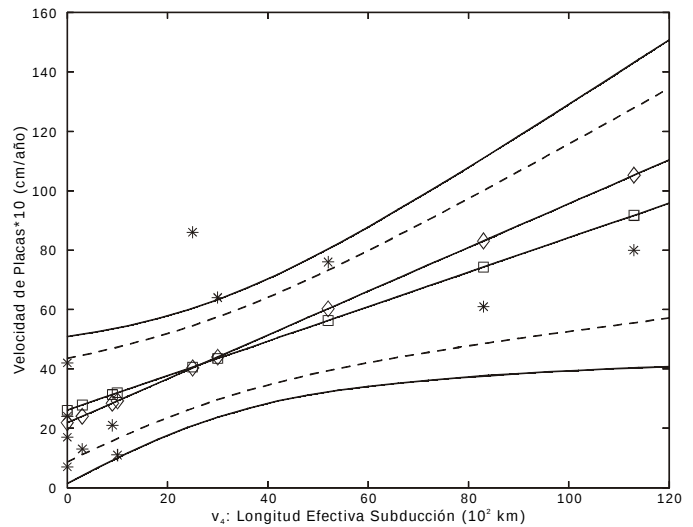


Figura 2. Resultado del ajuste MCE (cuadrados) y MDO (diamantes) para 10 veces y en función de u_4 (ver Tabla 5). Los demás símbolos son iguales a los de la Figura 1.

para el caso presentado en la Tabla 5 y Figura 2; para el cual basta con ampliar un par de veces la zona donde se encuentra la solución para determinar gráficamente los coeficientes con las cifras significativas necesarias; esto es simple si se usa MatLab. El caso de tres o más incógnitas se puede visualizar en proyecciones de combinaciones de pares de incógnitas si el número total de ellas no es demasiado alto. La ganancia adicional es que las líneas de nivel nos indican la variabilidad de estimaciones, incluyendo la correlación entre ellas.

Entregamos, enseguida, detalles de la determinación de la variabilidad de las estimaciones. En el procedimiento numérico-computacional, tomamos 500 muestras de m mediciones; "bootstrap" utiliza cada vez el mismo número de observaciones que en el conjunto original, pero permite repetición de ellas. Un generador de números (seudo)aleatorios elige los elementos de cada muestra. Luego, se estiman los valores de los parámetros y se obtiene la estadística de los promedios y desviaciones estándares. La Figura 3 contiene la mayor parte de las 500 estimaciones de los parámetros (parámetro de posición y pendiente) para el caso de la regresión de la velocidad de placa con u_4 . La Figura 4 presenta cinco histogramas con la distribución de los parámetros obtenidos por "bootstrap". Hemos filtrado las estimaciones que se apartan claramente de la distribución central, es decir, la que entrega máximos globales; en este caso, el número de muestreos considerados baja de 500 a 486. El filtro elimina un vector-solución completo si por lo menos uno de sus parámetros se aparta de la conducta central correspondiente. Las distribuciones estándares resultantes están incluidas en las tablas 2, 3, 4 y 5.

8. DISCUSIÓN Y CONCLUSIONES

El criterio MDO debe aplicarse cuando los errores de las mediciones de las variables independientes son del mismo orden o mayor que los de la variable dependiente. El procedimiento computacional A1, basado en la interpretación como cociente de Rayleigh de la

Tabla 5. Comparación: Métodos MCE y MDO

y	MCE			MDO		
	yhat	res	d	yhat	res	d
11.0	31.92	-20.916	-18.0893	29.23	-18.230	-14.6694
13.0	27.85	-14.853	-12.8453	24.07	-11.065	-8.9042
80.0	91.71	-11.708	-10.1259	105.22	-25.221	-20.2949
17.0	26.11	-9.111	-7.8798	21.85	-4.852	-3.9043
61.0	74.29	-13.293	-11.4965	83.09	-22.087	-17.7736
21.0	31.34	-10.336	-8.9389	28.49	-7.492	-6.0287
7.0	26.11	-19.111	-16.5283	21.85	-14.852	-11.9513
76.0	56.30	19.702	17.0395	60.22	15.784	12.7012
86.0	40.62	45.376	39.2431	40.30	45.704	36.7776
24.0	26.11	-2.111	-1.8260	21.85	2.148	1.7286
64.0	43.53	20.474	17.7064	43.99	20.015	16.1059
42.0	26.11	15.889	13.7412	21.85	20.148	16.2131
Normas Medias:		19.69	17.02		20.47	16.47
Soluciones:						
	x=	26.1113	0.5805	x=	21.8519	0.7378
				DE _{MDO} =	1.7	0.33

Similar a la Tabla 4 salvo que se aplica un factor de 10 a las velocidades de placas.

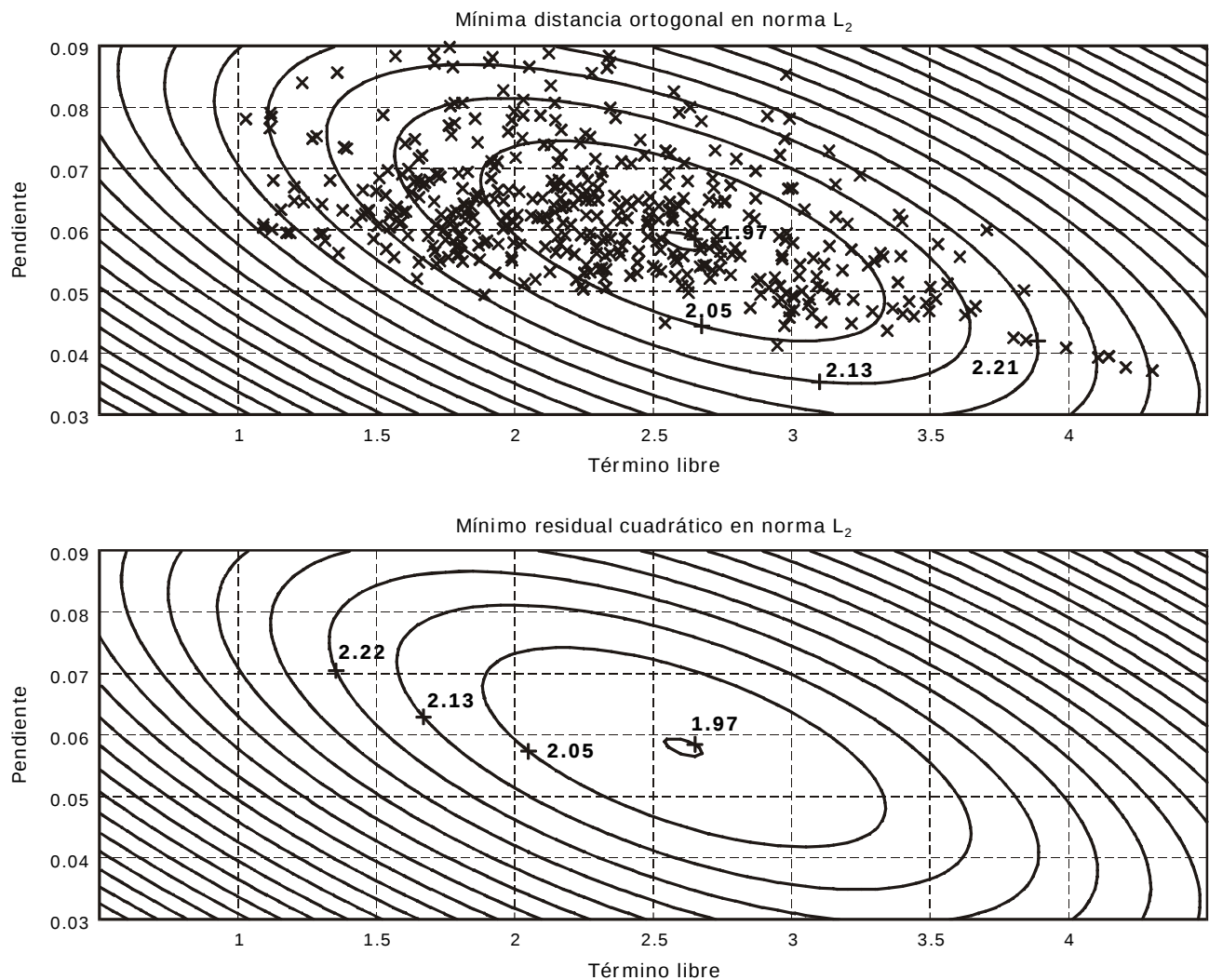


Figura 3. Líneas de nivel de las funciones criterios MDO (figura superior) y MCE (inferior) para el caso tratado de la Figura 2 y Tabla 5. Se han rotulado las cuatro líneas de nivel más cercanas a la solución. Nótese la equivalencia entre los resultados de esta figura y los que se leen en la Tabla 5. La figura superior incluye las soluciones obtenidas por “bootstrap” (cruces).

ecuación (17) y el uso del método de potencias para resolver el problema asociado de valores característico, es fácil de implementar y entrega resultados estables si es que no existe una acumulación de valores característicos cerca del mínimo correspondiente. Con la hipótesis de que el problema está bien propuesto o determinado, para lo cual basta que las columnas de la matriz A en (3) sean linealmente independientes, esta condición está asegurada. Cuando ello no sucede, la literatura (Golub y Van Loan, 1980; Van Huffel y Vandewalle, 1991; entre otras referencias) aconseja el uso de la descomposición de valores singulares.

Dado el enorme aumento en recursos computacionales (CPU y memoria), el uso de métodos de **búsqueda directa** en el espacio de soluciones ha ampliado su aplicación en los últimos años, aunque todavía tiene un límite de alrededor de 8 incógnitas. El caso de ajuste a una línea recta es, por su uso frecuente, el más estudiado (Ward, 1940; Acton, 1966; York, 1966, entre otros) para calcular sus coeficientes y la variabilidad empíricamente permitida de ellos. Dada la facilidad de graficación y de cálculo en computadoras mo-

dernas, es posible dibujar con facilidad las líneas de nivel de la función criterio, señalar su valor óptimo y determinar la línea de nivel que limita la región de variabilidad o confianza. El uso de ampliación de los dibujos directamente en pantalla nos entrega las cifras significativas que necesitamos. Así, el análisis del ajuste a una línea recta se hace inmediato sin necesidad de acudir a recursos matemático-numéricos propios de 20 o más años atrás.

Al final, toda evaluación parte del buen sentido común del analista. Si los datos son extremadamente buenos (con respeto al modelo), cualquier estimador también lo será. Al revés, para datos con mucha dispersión (con respecto al modelo) y donde se plantea un mero problema de correlación como el ejemplo tratado en este artículo, se producirán resultados diferentes con técnicas diferentes y con una significación difícil de estimar. Sin embargo, la presencia de ruido en las variables independientes, a nivel semejante o mayor al de la dependiente, hace preferible la solución MDO o la más general de MV que está descrita en la sección 6.

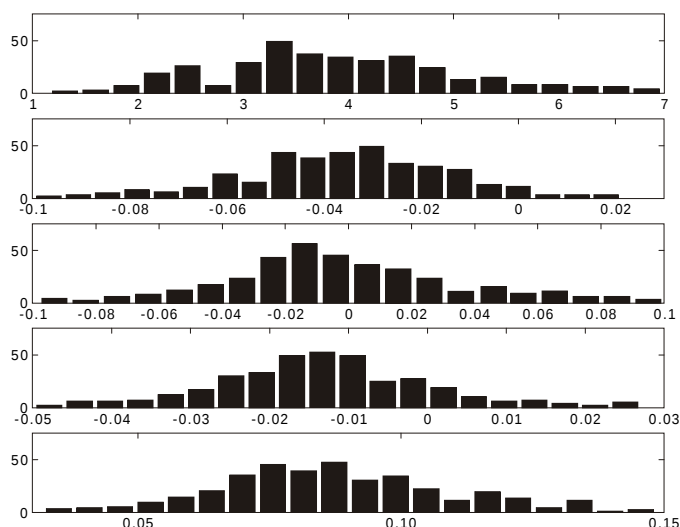


Figura 4. Histogramas de las estimaciones obtenidas al aplicar "bootstrap" (500 soluciones) el criterio de MDO para el caso descrito en la Tabla 2. De arriba hacia abajo, se encuentran las distribuciones de x_0 a x_4 . Se aplicó un filtro que eliminó 14 soluciones que se apartan claramente de la tendencia central en por lo menos una de las componentes del vector de incógnitas.

AGRADECIMIENTOS

Este trabajo resultó de su tratamiento en clases en el CICESE; además, fue impulsado por conversaciones con E. Gómez. R. Eaton colaboró con la edición y D. Sarracino obtuvo con rapidez varias de las referencias que necesitó. Aprecio los comentarios de los árbitros que contribuyeron a mejorar la presentación del artículo y me impulsaron a aclarar varios de los aspectos tratados.

REFERENCIAS

- Acton, F.S., 1966. "Analysis of Straight-Line Data", Dover, 207 pp.
- Acton, F.S., 1970. "Numerical Methods That Work", Harper and Row, 541 pp.
- Adcock, R.J., 1878. "A problem in least squares", *The Analyst*, vol. 5, pp. 53-54.
- Biswas, M.N. and L. Knopoff, 1974. "Structure of the upper mantle under the United States from the dispersion of Rayleigh waves", *Geophysical Journal Royal Astronomical Society*, Vol. 36, pp. 515-539.
- Efron, B. and R. Tibshirani, 1986. "Bootstrap methods for standar errors, confidence intervals, and other measures of statical accuracy", *Statistical Science*, Vol. 1, pp. 54-77. (Incluye comentarios).
- Fisher, G.W., 1989. "Matrix analysis of metamorphic mineral assemblages and reactions", *Contrib. Mineralogy and Petrology*, Vol. 102, pp. 69-77.
- Forsyth, D. and S. Uyeda, 1975. "On the relative importance of the driving forces of plate tectonics", *Geophysical Journal of the Royal Astronomical Society*, Vol. 43, pp. 163-200.

- Golub, G.H., M.T. Heath and G. Wahba, 1979. "Generalized cross-validation as a method for choosing a good ridge parameter", *Technometrics*, Vol. 21, pp. 215-223.
- Golub, G.H. and Ch. F. VanLoan, 1980. "An analysis of the total least squares problem", *SIAM Journal Numerical Analysis*, Vol. 17, pp 883-893.
- Hildebrand, F.B., 1965. "Methods of Applied Mathematics", second ed., Prentice Hall, 362 pp.
- Lawson, C.L. and R.J. Hanson, 1974. "Solving Least Squares Problems", Prentice Hall, 340 pp.
- Kent, J.T., G.S. Watson and T.C. Onstott, 1990. "Fitting straight lines and planes with an application to radiometric dating", *Earth and Planetary Science Letters*, vol. 97, pp. 1-17.
- Kirby, J.M., 1991. "Multiply functional regression. I. Function minimization technique", *Computers and Geosciences*, Vol. 17, pp. 537-547.
- McIntyre, G.M., C. Brooks, W. Compston and A. Turek, 1966. "The statistical assessment of Rb-Sr isochrones", *Journal of Geophysical Research*, Vol. 71, pp. 5459-5468.
- Middleton, G.V., 2000. "Data Analysis in the Earth Sciences Using MatLab", Prentice Hall, N. Jersey, 260 pp.
- Miller, R.G., 1974. "The jackknife: a review", *Biometrika*, Vol. 61, pp. 1-15.
- Noble, B. y J.W. Daniel, 1989. "Algebra Lineal Aplicada", Prentice Hall Hispanoamericana, México, 572 pp.
- Späth, H., 1987. "Mathematical Algorithms for Linear Regression", Academic Press, 327 pp.
- Van Huffel, S. and J. Vandewale, 1991. "The Total Least Squares Problem. Computational Aspects and Analysis", SIAM, Philadelphia, USA, 300 pp.
- York, D., 1966. "Least-squares fitting of a straight line", *Canadian Journal of Physic*, Vol. 44, pp. 1079-1086.
- Watson, G.A., 2002. "On the Gauss-Newton method for l_1 orthogonal distance regression", *IMA Journal of Numerical Analysis*, vol. 22, pp. 345-357.
- Ward, A., 1940. "The fitting of straight lines if both variables are subjects to error", *Ann. Math.*, Vol. 11, pp 208-215.
- Webster, R., 1989. "Is regression what you really want?", *Soil Use and Management*, Vol. 5, pp. 47-53.