Vol. **17**, **1**, 3–17 (2001)

Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería

Estimadores robustos de autocorrelación espacial basados en la varianza muestral

Alejandro F. Saccomano, Gabriela B. Savioli y M. Susana Bidner

Laboratorio de Ingeniería de Reservorios - Facultad de Ingeniería
Universidad de Buenos Aires, Pabellón Industrias, Ciudad Universitaria
1428, Buenos Aires, Argentina
Tel.: 54-11-4784 8085/4576 3240, Fax: 54-11-4780 0145/4576 3241
e-mail: asaccom@di.fcen.uba.ar, gsavioli@di.fcen.uba.ar, sbidner@di.fcen.uba.ar

Resumen

La Geoestadística se ha convertido en una herramienta muy útil en las ciencias de la tierra. La aplicaremos para estudiar las características espaciales de las rocas reservorios de petróleo, desde la perspectiva de la ingeniería de reservorios. El comportamiento espacial de la permeabilidad y de la porosidad se describe en este trabajo mediante la autocorrelación. Una medida tradicional de la autocorrelación es el semivariograma. En este trabajo se proponen dos nuevos estimadores de la autocorrelación espacial. El primero es LV (Varianza Local), basado en la varianza de la distribución. Se deduce que LV es también un estimador del valor medio del semivariograma dentro de una región. LV es muy robusto y resistente a los valores extremos. Es también muy suave y por eso fácil de ser representado por un modelo teórico. El segundo, SLV (Semivariograma desde la Varianza Local), está relacionado con las derivadas del primero. Es un estimador del semivariograma propiamente dicho. LV y SLV se aplican a conjuntos de datos reales y sintéticos. Sus resultados se comparan con los de otros estimadores clásicos e integrales. LV y SLV se comportan mejor que el estimador clásico punto a punto de Matheron, que es impreciso a distancias de separación moderadas a grandes. Además se comportan de manera similar a los estimadores integrales de Li y Lake, siendo su principal ventaja el tiempo de cómputo, que es tres veces menor.

ROBUST SPATIAL AUTOCORRELATION ESTIMATORS BASED ON THE SAMPLE VARIANCE

Summary

Geostatistics has become a useful tool in earth sciences. We will apply it to study spatial oil reservoir rock properties, from a reservoir engineering outlook. The spatial behavior of permeability and porosity is described herewith by the autocorrelation. A traditional autocorrelation measure is the semivariogram. In this paper, two new estimators of space autocorrelation are put forward. The first is LV (Local Variance), based on the variance of the distribution. It is also an estimator of the average value of the semivariogram within a region. LV is very robust and resistant to extreme values. It is also very smooth and easy to be represented by a theoretical model. The second, SLV (Semivariogram from the Local Variance), is related to the derivatives of the former. It is an estimator of the semivariogram itself. LV and SLV are applied to real and synthetic data sets. Their results are compared to those of other classical and integral estimators. They perform better than the classical point to point estimator of Matheron, which is imprecise at moderate to large lag distances. LV and SLV perform similarly to Li and Lake's integral estimators, and their main advantage is that they are three times faster to compute.

 ISSN: 0213–1315 Recibido: Enero 2000

INTRODUCCIÓN

Actualmente, los simuladores numéricos de reservorios son la herramienta más potente para predecir la producción de petróleo y gas de los yacimientos y para analizar distintas alternativas de explotación[1]. Los simuladores se alimentan con datos que caracterizan al reservorio: la descripción geológica, las propiedades de la roca y las de los fluidos. Específicamente, nos interesa estudiar la heterogeneidad de las dos propiedades más importantes de la roca: la permeabilidad y la porosidad. Estas dos propiedades sólo pueden medirse en los pozos[2] y se ignoran para las grandes zonas intermedias. A fin de alimentar el simulador, se debe contar con valores plausibles de estas propiedades entre los pozos. Para eso se aplican técnicas geoestadísticas. Uno de los aspectos más críticos de la geoestadística es la determinación de correlaciones espaciales de las propiedades que caracterizan a los reservorios.

La autocorrelación, o grado de similitud entre datos separados en el espacio, se mide usualmente con el semivariograma. El estimador clásico del semivariograma (CSV) fue definido por Matheron[3−6] en 1962. Su mayor inconveniente es la falta de precisión para largas distancias de separación. Eso se debe a que el número de pares de mediciones disminuye al aumentar la distancia de separación.

A fin de superar este inconveniente, Li y Lake[4] introdujeron en 1994 dos estimadores integrales que denominaron Moving Window Semivariance Estimators (MW1 y MW2). Al ser estimadores integrales, el número de pares de mediciones no disminuye al aumentar la distancia de separación. Por eso, MW1 y MW2 son más robustos que CSV.

En un trabajo anterior, Savioli *et al.*[7] estudiaron varios estimadores punto a punto derivados del clásico de Matheron y los estimadores integrales MW1 y MW2. Se comparó el comportamiento de cinco estimadores para deducir cuál era más conveniente para una aplicación determinada. La comparación fue realizada con datos de permeabilidad de rocas de pozos petrolíferos medidos en función de la profundidad. Se concluyó que los estimadores integrales son robustos, resistentes a la contaminación y confiables para todas las distancias de separación. Sin embargo, su cálculo es mucho más lento que el de los estimadores punto a punto.

En este trabajo presentamos dos nuevos estimadores integrales de autocorrelación espacial. El primero de ellos, al que llamamos LV (Varianza Local), es un estimador de la varianza muestral dentro de una región de tamaño variable. Se deduce que LV es también un estimador del valor medio del semivariograma en dicha región. El segundo estimador está relacionado con las derivadas del primero, se denomina SLV (Semivariograma desde la Varianza Local) y es un estimador del semivariograma propiamente dicho.

Para evaluar el comportamiento de estos dos nuevos estimadores, los aplicamos a diferentes conjuntos de datos, tanto reales como sintéticos. Además, los comparamos con CSV, MW1 y MW2. El comportamiento de LV es análogo al de MW1, siendo ambos suaves y fáciles de parametrizar con modelos teóricos. La ventaja de LV es que consume menos tiempo de cálculo. Su robustez se traslada parcialmente a SLV.

TEORÍA

Si tenemos una variable aleatoria Z que es función de la posición espacial x, es decir $Z = Z(x)$, se pueden definir diferentes funciones de autocorrelación espacial. Esto es de suma utilidad para poder detectar la presencia de una estructura en la distribución espacial de la variable Z. Queremos obtener un estimador de esta autocorrelación espacial.

Una medida de autocorrelación espacial es el semivariograma γ, que se define como

$$\gamma(x,y) = \frac{1}{2}E\left\{(Z(x) - Z(y))^2\right\} \tag{1}$$

donde $E\{\}$ es la esperanza matemática y $Z(x)$ el valor de Z en la posición x. Cuando la variable Z es estacionaria de segundo orden, el semivariograma es función sólo de la separación entre x e y resultando

$$\gamma(x+h,x) = \gamma(h) = \frac{1}{2}E\left\{(Z(x+h) - Z(x))^2\right\} \tag{2}$$

donde $h = |x-y|$. Un gráfico de γ versus h es la manera tradicional de representar al semivariograma. En la ecuación (2) se supone que la distribución espacial de Z es isótropa. Si la distribución fuera anisótropa, se podría construir un semivariograma para cada una de las direcciones de interés.

Nuestra nueva medida de autocorrelación

Se define la varianza de dispersión de Z en un volumen V como[3]

$$\sigma_V^2 = E\left\{\frac{1}{V}\int_V (Z(x) - Z_V)^2 dx\right\} \tag{3}$$

donde Z_V es una variable regularizada igual al valor medio de Z en el volumen V, es decir

$$Z_V = \frac{1}{V}\int_V Z(y)dy \tag{4}$$

Proponemos como medida de autocorrelación a σ_V^2, donde V es una bola de diámetro h y evaluamos su comportamiento en función de h. Es decir, $\sigma_V^2 = \sigma_V^2(h)$.

Por otra parte, el valor medio del semivariograma dentro de un volumen V está definido por el promedio de $\gamma(x-y)$ cuando x e y recorren el volumen V

$$\bar{\gamma}_{VV} = \frac{1}{V^2}\int_V\int_V \gamma(x-y)dxdy \tag{5}$$

A partir de las definiciones de las ecuaciones (3) y (5) se puede demostrar que[3]

$$\sigma_V^2 = \bar{\gamma}_{VV} \tag{6}$$

Para obtener una expresión del semivariograma como función de la nueva medida de autocorrelación, nos limitamos al caso unidimensional. Si se desea caracterizar la estructura espacial de datos distribuidos en más dimensiones, se pueden estimar varios semivariogramas unidimensionales a lo largo de las direcciones de interés. En el caso unidimensional, el volumen V queda reducido a un segmento de longitud h y

$$\sigma_V^2(h) = \frac{1}{h^2}\int_{\frac{-h}{2}}^{\frac{h}{2}}\int_{\frac{-h}{2}}^{\frac{h}{2}} \gamma(u-w)du\,dw \tag{7}$$

que mediante una sustitución de variables resulta

$$\sigma_V^2(h) = \frac{1}{h^2}\int_0^h\int_0^h \gamma(x-y)dxdy \tag{8}$$

Recordemos que γ es función de la distancia $|x-y|$. Definimos las siguientes primitivas

$$\Gamma(t)=\int_0^t \gamma(u)du \qquad G(t)=\int_0^t \Gamma(u)du \tag{9}$$

Entonces

$$\begin{aligned}\sigma_V^2(h) &= \frac{1}{h^2}\int_0^h\int_0^h \gamma(|x-y|)dxdy = \frac{1}{h^2}\int_0^h dx\left[\int_0^x \gamma(x-y)dy+\int_x^h \gamma(y-x)dy\right]= \\ &= \frac{1}{h^2}\int_0^h dx\left[\int_0^x \gamma(u)du+\int_0^{h-x}\gamma(u)du\right]=\frac{1}{h^2}\int_0^h dx\,[\Gamma(x)+\Gamma(h-x)]= \\ &= \frac{1}{h^2}\left[\int_0^h \Gamma(x)dx-\int_h^0 \Gamma(u)du\right]=\frac{1}{h^2}\left[\int_0^h \Gamma(x)dx+\int_0^h \Gamma(u)du\right]=\frac{2}{h^2}G(h)\end{aligned} \tag{10}$$

Finalmente obtenemos

$$G(h)=\frac{h^2}{2}\sigma_V^2(h) \tag{11}$$

Derivando $G(h)$ respecto de h

$$\Gamma(h)=\frac{h^2}{2}\frac{d\sigma_V^2}{dh}+h\sigma_V^2 \tag{12}$$

Derivando por segunda vez se obtiene la relación entre el semivariograma y la nueva medida de autocorrelación σ_V^2

$$\gamma(h)=\frac{h^2}{2}\frac{d^2\sigma_V^2}{dh^2}+2h\frac{d\sigma_V^2}{dh}+\sigma_V^2 \tag{13}$$

Puesto que $\sigma_V^2(h)$ es un valor promedio del semivariograma dentro de una región (es una función integral), es de esperar que presente una variabilidad espacial mucho más atenuada que la de las funciones de autocorrelación tradicionales.

Nuestros estimadores de autocorrelación: LV y SLV

Varianza local (LV) es un estimador insesgado de la nueva función de autocorrelación $\sigma_V^2(h)$ definida en la ecuación (3). Un estimador de la varianza dentro de una bola de diámetro h alrededor de un punto x_i está dado por

$$\hat{\sigma}_{V_i}^2(h)=\frac{1}{m_i}\sum_{j\in I_i}[Z(x_j)-\langle Z(x_i)\rangle]^2 \tag{14}$$

donde $I_i=\left\{j \text{ que satisfacen } |x_j-x_i|\le\frac{h}{2}\right\}$, (m_i+1) es la cantidad de x_j que satisfacen esta condición y $\langle Z(x_i)\rangle$ es el promedio de Z en la bola centrada en x_i dado por

$$\langle Z(x_i)\rangle=\frac{1}{m_i+1}\sum_{j\in I_i}Z(x_j) \tag{15}$$

Reescribiendo la ecuación (14) tenemos que

$$\hat{\sigma}_{V_i}^2(h)=\frac{1}{m_i}\left[\sum_{j\in I_i}Z^2(x_j)-\frac{1}{m_i+1}\left(\sum_{j\in I_i}Z(x_j)\right)^2\right] \tag{16}$$

Entonces, como estimador global de la función $\sigma_V^2(h)$ hacemos un promedio entre los estimadores de la varianza en bolas alrededor de cada uno de los n puntos medidos x_i. Si realizamos un promedio pesado con los factores m_i nos queda

$$\mathrm{LV}(h) = \frac{1}{\sum_{i=1}^{n} m_i} \sum_{i=1}^{n} m_i \hat{\sigma}_{V_i}^2(h) = \frac{1}{\sum_{i=1}^{n} m_i} \sum_{i=1}^{n} \left[\sum_{j\in I_i} Z^2(x_j) - \frac{1}{m_i+1} \left(\sum_{j\in I_i} Z(x_j) \right)^2 \right] \tag{17}$$

Semivariograma desde la varianza local (SLV) es un estimador del semivariograma obtenido a partir de $\mathrm{LV}(h)$ y de la ecuación (13)

$$\mathrm{SLV}(h) = \frac{h^2}{2}\mathrm{LV}''(h) + 2h\mathrm{LV}'(h) + \mathrm{LV}(h) \tag{18}$$

El cálculo numérico de las derivadas de LV con respecto a h presenta problemas de inestabilidad. Estos son mayores para el caso de la derivada segunda, los cuales son magnificados al estar multiplicados por el cuadrado de la distancia. Para disminuir estos efectos, en el cálculo de las derivadas de LV aplicamos promedios sobre intervalos móviles. El algoritmo utilizado para el cálculo de las derivadas, válido para datos equiespaciados, es

$$f'(x_i) = \frac{f(x_{i+2}) - f(x_{i+1}) + f(x_{i-1}) - f(x_{i-2})}{x_{i+1} - x_{i-1}} \tag{19}$$

En el cálculo de la segunda derivada aplicamos la ecuación (19) a la derivada ya suavizada mediante los promedios móviles.

Otros estimadores de la autocorrelación

Semivariograma clásico (CSV) es el estimador de $\gamma(h)$ propuesto por Matheron[3] en 1962

$$\mathrm{CSV}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i+h)]^2 \tag{20}$$

donde $N(h)$ es el número de pares de datos separados por una distancia h. Este estimador es insesgado, pero no es robusto ni resistente a valores extremos. Además es impreciso para distancias medias y largas debido a que $N(h)$ disminuye con h.

Moving window 1 (MW1) es una medida de autocorrelación integral propuesta por Li y Lake[4] que se define como el momento $(d-1)$ del semivariograma dentro de una ventana de medida h, siendo d la dimensión del espacio

$$\gamma_N(h) = \frac{1}{\int_0^h \zeta^{d-1} d\zeta} \int_0^h \zeta^{d-1} \gamma(\zeta) d\zeta \tag{21}$$

Dichos autores también proponen un estimador de $\gamma_N(h)$ que es insesgado, $\mathrm{MW1}(h)$, definido como[4]

$$MW1(h) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{1}{2m_i'} \sum_{j\in K_i} [Z(x_i) - Z(x_j)]^2 \right\} \tag{22}$$

donde $K_i = \{j$ que satisfacen $0 < |x_i - x_j| \leq h\}$, m_i' es la cantidad de datos que satisfacen esta condición y n el número total de mediciones. Al considerar una mayor cantidad de datos en cada estimación, MW1 es menos afectado por valores extremos, lo cual lo hace más robusto y resistente que los semivariogramas clásicos.

Moving window 2 (MW2): existe una relación[4] entre el semivariograma tradicional $\gamma(h)$ y $\gamma_N(h)$

$$\gamma(h) = \gamma_N(h) + \frac{h}{d}\gamma'_N(h) \tag{23}$$

Entonces, basándonos en MW1 y la ecuación (23) podemos definir MW2, un nuevo estimador de semivariograma, como

$$\mathrm{MW2}(h) = \mathrm{MW1}(h) + \frac{h}{d}\mathrm{MW1}'(h) \tag{24}$$

Para el cálculo de MW1$'$ se utilizan diferencias finitas centradas. En este trabajo se analiza el caso $d = 1$.

Modelos de funciones de autocorrelación

Existen diferentes funciones que se han propuesto como modelos teóricos del semivariograma. Estas funciones deben satisfacer ciertas restricciones[3]:

- ser nula en el origen,
- ser condicionalmente definida negativa,
- tener un crecimiento inferior al cuadrado de la distancia.

Dos de los modelos teóricos más utilizados son el modelo esférico (ec.(25)) y el modelo exponencial (ec.(26))

$$\gamma(h) = \begin{cases} \frac{S}{2}\left(3\frac{h}{a} - (\frac{h}{a})^3\right) & h \le a \\ S & h > a \end{cases} \tag{25}$$

$$\gamma(h) = S\left(1 - e^{-\frac{h}{a}}\right) \tag{26}$$

En ambos casos, el parámetro a es la distancia de autocorrelación o rango y S la meseta, igual a la varianza de los datos.

Para lograr un modelo de LV se pueden seguir dos métodos. El primero es, a partir de un modelo de semivariograma, obtener el correspondiente modelo de LV mediante las ecuaciones (9) y (11). Así, si representamos al semivariograma mediante el modelo exponencial de la ecuación (26), tenemos que

$$\gamma(h)=S\left(1-e^{-\frac{h}{a}}\right)\Longrightarrow\Gamma(h)=S\left(h-a+ae^{-\frac{h}{a}}\right)\Longrightarrow G(h)=S\left(\frac{(h-a)^2}{2}+a^2\left(\frac{1}{2}-e^{-\frac{h}{a}}\right)\right)$$
$$\Longrightarrow\sigma_V^2(h)=S\left(\left(1-\frac{a}{h}\right)^2+\frac{2a^2}{h^2}\left(\frac{1}{2}-e^{-\frac{h}{a}}\right)\right) \tag{27}$$

El segundo método es ajustar LV a una función que sea más sencilla de parametrizar. Por ejemplo, se puede recurrir a un modelo esférico (ec.(25)) o un exponencial (ec.(26)). Los parámetros de estos modelos se pueden determinar realizando un ajuste mediante una técnica de optimización.

El primer método sirve cuando necesitamos conocer el semivariograma a partir del LV, ya que quedan garantizados los requerimientos teóricos de la forma funcional del semivariograma. Por otro lado, el segundo método puede ser de mayor utilidad en aquellos casos en donde no se necesita el semivariograma sino sólo una medida de la autocorrelación.

DATOS Y RESULTADOS

Se compara el comportamiento de los cinco estimadores de autocorrelación (CSV, MW1, MW2, LV y SLV) con conjuntos de datos sintéticos y con mediciones de permeabilidad en muestras de un pozo de la cuenca de Neuquén en la Patagonia Argentina. Gran parte de los datos utilizados son los mismos que se analizaron anteriormente[7] para comparar el comportamiento de distintos estimadores punto a punto con MW1 y MW2.

Datos sintéticos

Se analiza el comportamiento de los cinco estimadores en diferentes estructuras geológicas que pueden encontrarse en rocas-reservorio: una estructura cíclica, una escalonada y varias con distribución normal.

Estructura cíclica. Se genera una estructura cíclica adicionando un ruido gaussiano a una onda cuadrada (Figura 1a). Esta estructura puede representar dos capas de porosidades 10 % y 30 % que se repiten cíclicamente. El ruido se genera con distribuciones N(10,1) y N(30,3), respectivamente. La estructura también podría representar distribuciones de permeabilidades luego de haber aplicado algún tipo de transformación no lineal, por ejemplo logarítmica. El comportamiento de los estimadores se observa en la Figura 1b (CSV, MW2 y SLV) y la Figura 1c (MW1 y LV).

En la Figura 1b, la periodicidad de la estructura es claramente reflejada por los tres estimadores de semivariograma. SLV y MW2 lo hacen en forma más atenuada que CSV a medida que la distancia de separación h aumenta. Para h muy grandes, CSV se distorsiona debido a la escasez de pares de datos con los que se efectúa la estimación. En cambio, MW2 y SLV no sufren este fenómeno debido a que utilizan una mayor cantidad de datos en cada estimación y, en este caso, conservan las características de robustez de MW1 y LV, respectivamente. En general, se observa que los semivariogramas dependen principalmente de los valores medios locales (originados con la onda cuadrada) y son apenas afectados por el ruido gaussiano.

En la Figura 1c se muestra que MW1 y LV tienen un comportamiento mucho más suave que MW2 y SLV. Sus pequeñas oscilaciones se atenúan rápidamente, tendiendo a la varianza de los datos. En ambos estimadores es fácil establecer que la autocorrelación entre los datos se extiende a una distancia de 0,15, poco mayor que un semiperíodo, como así también el valor de la varianza ($\sigma^2 = 100$). Por otro lado, se destaca que LV es más suave y estable que MW1.

Estructura escalonada. Se genera una estructura compuesta por tres escalones con un ruido gaussiano adicional (Figura 2a). Los datos podrían representar un campo de porosidades (o del logaritmo de las permeabilidades) con una distribución N(10,1) para el primer escalón, $N(20,2)$ para el segundo y $N(30,3)$ para el tercero. El comportamiento de los estimadores se observa en la Figura 2b (CSV, MW2 y SLV) y la Figura 2c (MW1 y LV).

En la Figura 2b, CSV tiene un comportamiento creciente. Esto se debe a que cuanto más separados están los pares de valores, su diferencia es mayor en promedio (ec.(20)). MW2 también presenta una tendencia creciente, pero cuando la distancia es mayor que los dos primeros escalones, decrece hasta el valor de la varianza. SLV muestra claramente tres etapas: en las dos primeras crece en forma cada vez más acentuada y en la última decrece. La pequeña fluctuación que se observa entre los dos primeros tramos se debe principalmente al término de la segunda derivada de LV que interviene en la estimación de SLV (ec.(18)).

La tendencia creciente de los tres estimadores resalta la naturaleza no estacionaria de los datos, mientras que el cambio de pendiente del último tramo se debe a los efectos de borde.

En la Figura 2c, MW1 y LV muestran una pendiente positiva para todas las distancias y son muy similares entre sí. LV manifiesta más claramente la existencia de tres etapas. Nótese que este fenómeno se refleja en la mayor variabilidad de SLV (Figura 2b).

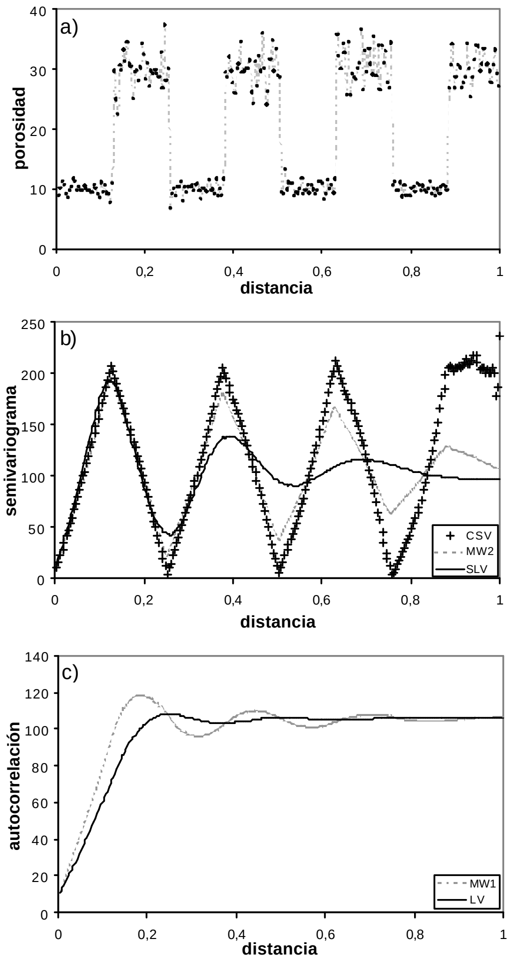

Figura 1. a) Estructura cíclica sintética; b) CSV, MW2 y SLV en función de la distancia de separación adimensional; c) MW1 y LV en función de la distancia de separación

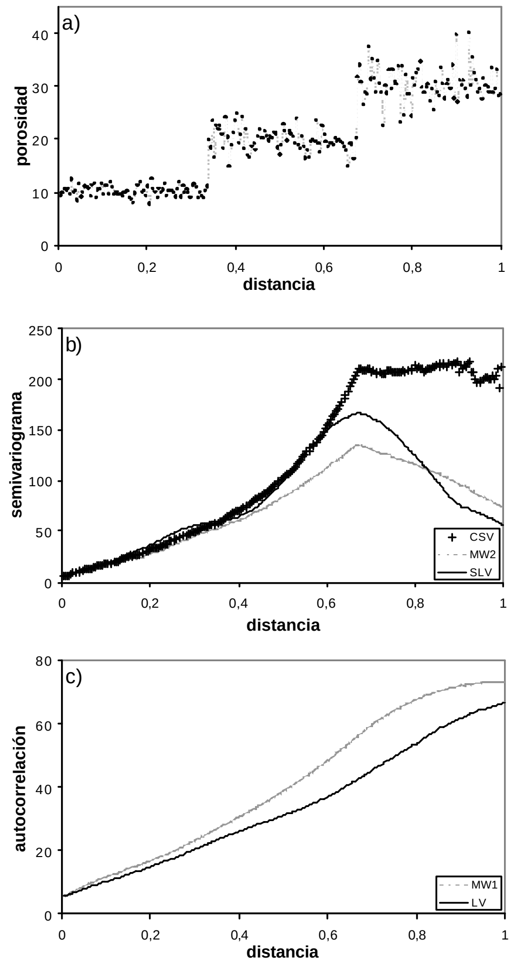

Figura 2. a) Estructura escalonada sintética; b) CSV, MW2 y SLV en función de la distancia de separación adimensional; c) MW1 y LV en función de la distancia de separación

Estructura con distribución normal. Utilizando una de las rutinas de simulación gaussiana secuencial de Deutsch y Journel[5] se genera un campo sintético unidimensional (Figura 3a). Este campo puede representar valores ya normalizados de un campo de permeabilidad, porosidad o cualquier otra propiedad de interés. Los datos generados presentan una distribución normal $N(0,1)$ y una estructura espacial modelada mediante un semivariograma exponencial (ec.(26)) con distancia de correlación $a = 0,05$ y meseta unidad (Figura 3b). El comportamiento de los estimadores se muestra en las Figuras 3b y 3c.

En la Figura 3b se representa CSV, MW2 y SLV, junto al modelo exponencial teórico. Se observa que el comportamiento de los semivariogramas se aparta del modelo teórico. Esto se debe a que el cálculo de los semivariogramas es muy sensible a las características particulares de cada conjunto de datos. MW2 es más robusto a largas distancias, tendiendo suavemente a la varianza de los datos. Los otros dos estimadores, CSV y SLV, presentan oscilaciones mayores.

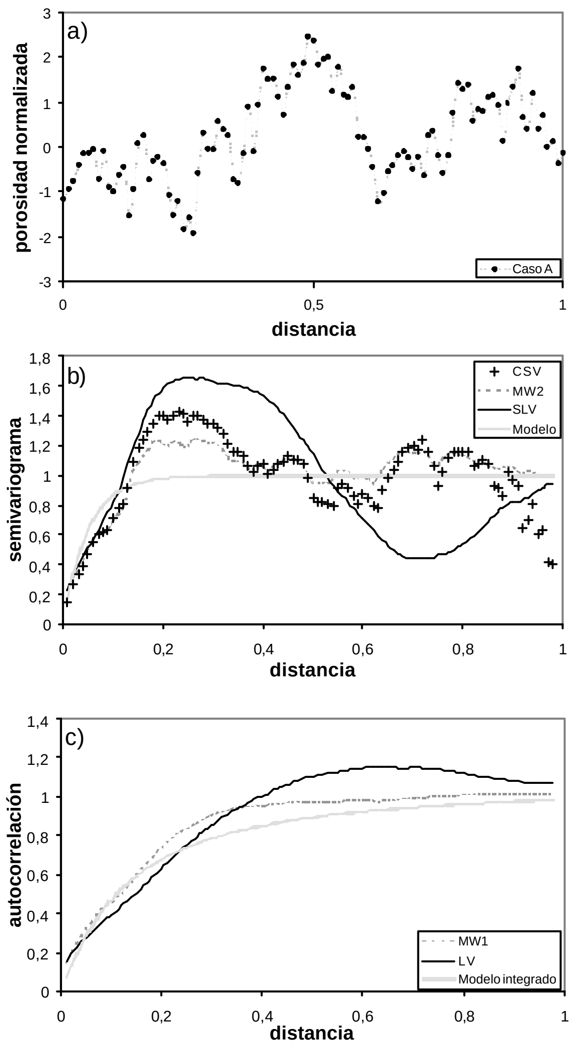

Figura 3. a) Estructura sintética con distribución de porosidad normalizada; b) CSV, MW2 y SLV en función de la distancia de separación adimensional y comparación con el modelo exponencial; c) MW1 y LV en función de la distancia de separación y comparación con el modelo de LV integrado del exponencial (ec.(27))

En la Figura 3c se muestran MW1 y LV. Ambos son mucho más suaves que los semivariogramas y tienen un comportamiento cualitativamente similar. Así como en la Figura 3b se comparan los semivariogramas con el modelo exponencial que generó el campo sintético de la Figura 3a, aquí comparamos LV con el correspondiente modelo integrado del exponencial, dado por la ecuación (27) con $a = 0,05$ y $S = 1$. Este procedimiento es el primer método descrito en el apartado "Modelos de funciones de autocorrelación". Se puede observar que LV respeta mejor al modelo que los semivariogramas.

Análisis de la robustez de los estimadores. Evaluamos el comportamiento de los cinco estimadores de autocorrelación con varios conjuntos de datos que tienen la misma distribución y estructura espacial que los datos del caso anterior. Estos campos son distintas realizaciones generadas con la rutina de simulación gaussiana secuencial de Deutsch y Journel, todas ellas utilizando la misma distribución $N(0,1)$ y el mismo semivariograma. El objetivo es poder analizar la robustez de los estimadores, es decir si sus comportamientos dependen sólo de las propiedades estadísticas de los datos analizados o si se ven afectados por las características particulares de cada realización.

Si bien analizamos numerosas realizaciones, por simplicidad mostramos en este trabajo los resultados con sólo tres de ellas a modo de ejemplo, ya que todas mostraban un comportamiento cualitativo similar. En la Figura 4a se representan dos nuevas realizaciones (que llamaremos Caso B y Caso C) y en las Figuras 4b-f se comparan los comportamientos de cada uno de los estimadores con estas dos nuevas realizaciones y también con la del punto anterior (que llamaremos Caso A).

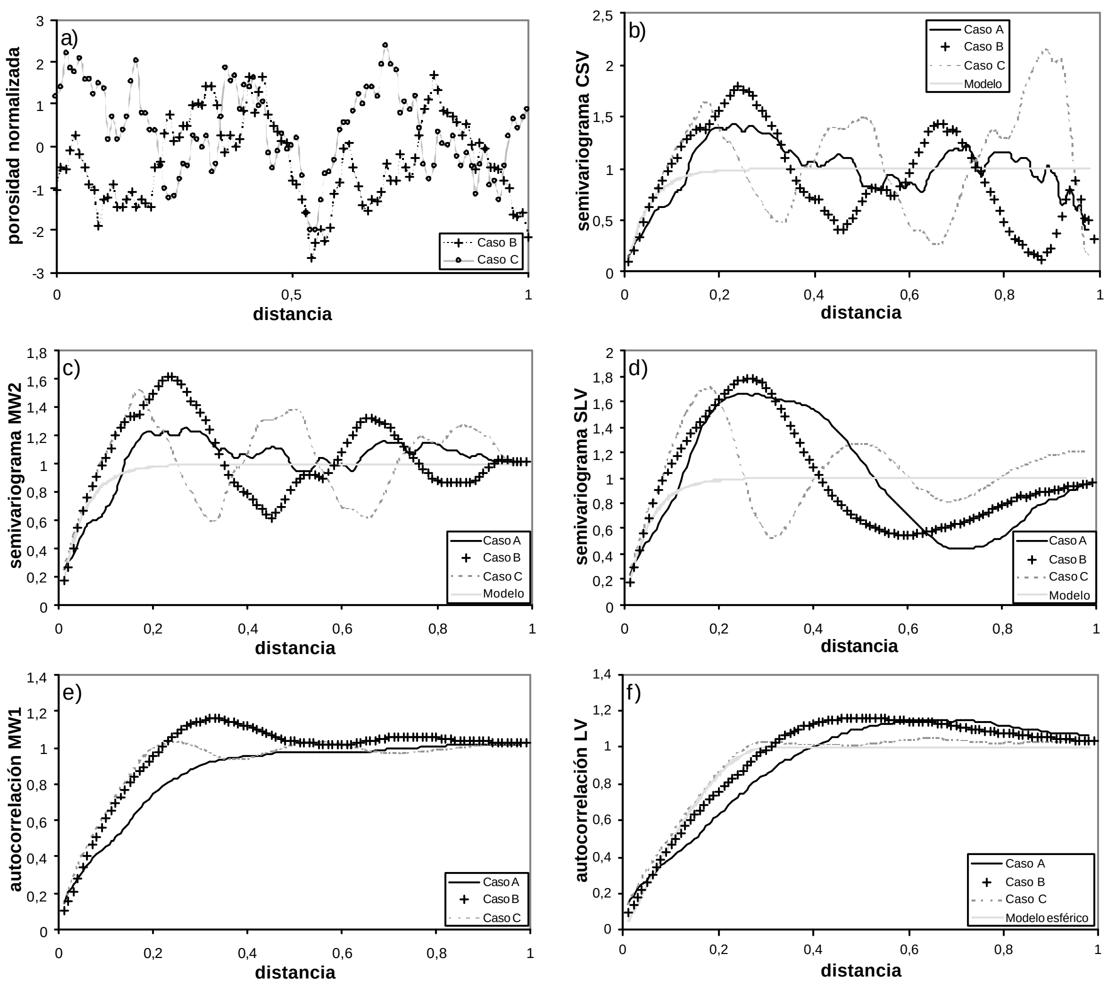

Figura 4. a) Dos campos sintéticos con distribución de porosidad normalizada, generados con la misma estructura espacial que el Caso A de la Figura 3a; b) Comparación del CSV aplicado a los tres casos sintéticos con el modelo teórico de semivariograma; c) idem b) con el estimador MW2; d) Idem b) con el SLV; e) Desempeño del MW1 en los tres campos sintéticos; f) Desempeño del LV en los tres campos sintéticos y ajuste mediante un modelo esférico (ec.(25))

Se puede notar que el comportamiento de MW2 (Figura 4c) es ligeramente más robusto que el de CSV (Figura 4b) y SLV (Figura 4d), los cuales presentan mayores oscilaciones. Por otro lado, analizando el comportamiento de cada uno de los estimadores, las curvas resultantes en los distintos casos no se asemejan entre sí y menos aún al modelo teórico. Esto es un claro indicio de la fuerte dependencia de estos estimadores con las irregularidades particulares de cada conjunto de datos para distancias medias y grandes. Sin embargo, para distancias del orden del rango ($a = 0,05$) tienen comportamientos similares.

En cuanto a las Figuras 4e y 4f correspondientes a MW1 y LV, se destaca la robustez de ambos estimadores, si bien presentan leves diferencias en la caracterización de los distintos conjuntos de datos. En la Figura 4f se modela la autocorrelación LV usando el modelo esférico de la ecuación (25) con parámetros $a = 0,3$ y $S = 1$. Esto ejemplifica el segundo método descrito en el apartado "Modelos de funciones de autocorrelación", obteniendo una descripción de LV mediante un modelo teórico más sencillo y confiable.

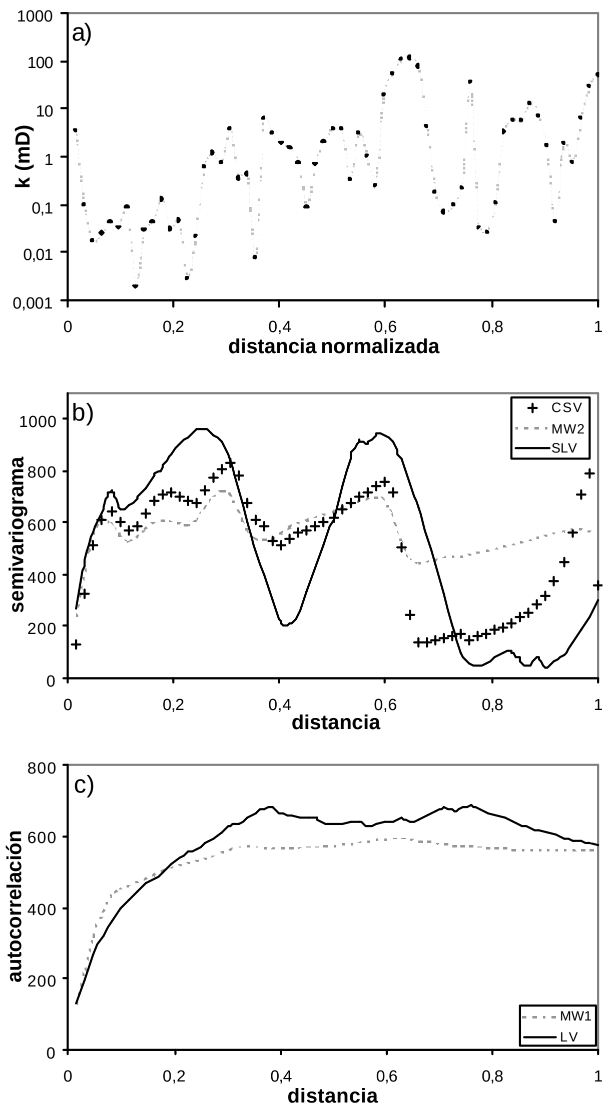

Figura 5. a) Mediciones de permeabilidad de un pozo real, en escala logarítmica y distancia adimensional; b) CSV, MW2 y SLV en función de la distancia de separación adimensional; c) MW1 y LV en función de la distancia de separación adimensional

Datos reales

Los datos analizados constituyen una muestra de 64 mediciones de permeabilidad sobre testigos rocosos tomados a diferentes profundidades. Estos testigos fueron obtenidos de un pozo de la cuenca de Neuquén, Argentina. Los datos se muestran en la Figura 5a. Los valores extremos de permeabilidad son de 0,002 mD y 116 mD y la media aritmética es de 10,3 mD. El comportamiento de los estimadores se observa en la Figura 5b (CSV, MW2 y SLV) y la Figura 5c (MW1 y LV).

En la Figura 5b se puede observar que todos los estimadores de semivariograma presentan fuertes fluctuaciones. El más robusto es MW2, mientras que SLV es inestable. Esto no se debe solamente a la gran variabilidad de los datos. La presencia de valores extremos (*outliers*) afecta en gran medida a SLV, ya que en su estimación interviene el cómputo de la derivada segunda de LV, la cual se torna numéricamente inestable (ec.(19)).

En la Figura 5c se puede apreciar que MW1 y LV muestran un aspecto cualitativamente similar, aunque, en este caso, MW1 es más suave que LV. El comportamiento irregular se debe a los valores extremos que inducen a la inestabilidad de los estimadores.

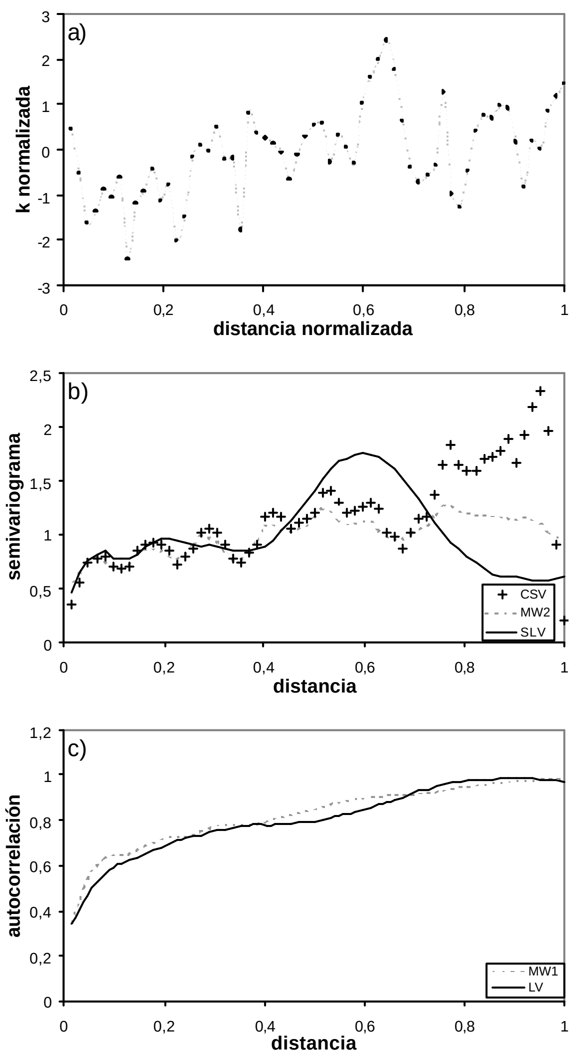

Figura 6. a) Mediciones de permeabilidad de un pozo real, luego de ser normalizadas; b) CSV, MW2 y SLV en función de la distancia de separación adimensional; c) MW1 y LV en función de la distancia de separación adimensional

Debido a la gran variabilidad de los valores de permeabilidad, este conjunto de datos fue normalizado (Figura 6a) mediante la técnica de *normal score*[5]. Se aplican los estimadores a los datos transformados; CSV, MW2 y SLV se muestran en la Figura 6b y MW1 y LV en la Figura 6c.

En la Figura 6b se puede ver el mejor comportamiento con los datos normalizados. Tanto CSV como SLV se vuelven más estables, SLV en mayor medida. MW2 no se ve muy afectado por la normalización.

En la Figura 6c MW1 y LV son muy suaves debido a que los datos tienen una distribución normalizada.

DISCUSIÓN

Tiempo de cálculo de los estimadores

En todos los casos estudiados, el método más rápido para determinar el semivariograma es aplicando el estimador punto a punto CSV. La estimación de SLV consume aproximadamente tres veces más tiempo. MW2 es aún más lento, demandando cerca de diez veces más que CSV. En el cálculo de SLV y MW2, lo más costoso es la estimación de LV y MW1. La derivación de SLV y MW2 a partir de LV y MW1, respectivamente, es prácticamente inmediata.

Comportamiento de los estimadores

En general podemos decir que LV y MW1 tienen un comportamiento similar. Son buenos estimadores de la autocorrelación espacial, muy robustos, resistentes y confiables para cualquier distancia. La estabilidad de estos estimadores integrales se debe a que utilizan una gran cantidad de datos y consideran un valor promedio dentro de una región para cada estimación. Esto hace que las fluctuaciones que presentan (debido a la heterogeneidad de los datos) sean muy atenuadas y disminuyan a medida que aumenta la distancia de separación. Su principal característica es que pierden sensibilidad a las particularidades del conjunto de datos que pretenden describir, aunque LV es, en general, más sensible que MW1. Esto puede ser una gran ventaja de los estimadores integrales, ya que permite obtener medidas estadísticas y de autocorrelación muy confiables a partir de un conjunto particular de datos.

Los estimadores del semivariograma son menos estables que los anteriores. MW2 es el más robusto de los tres aquí analizados, mientras que SLV es un buen estimador para distancias cortas, al igual que CSV.

La mayor sensibilidad de LV frente a MW1 se convierte en desventaja para SLV. Este último se torna inestable para datos muy heterogéneos que presentan valores extremos. La pequeña sensibilidad de LV se amplifica en la estimación de SLV debido al término de la derivada segunda de LV multiplicada por la distancia de separación al cuadrado (ec.(18)).

CONCLUSIONES

Se proponen dos nuevos estimadores de autocorrelación espacial. El primero, LV (Varianza Local), está basado en la varianza de dispersión y es un estimador del valor medio del semivariograma dentro de una región. El segundo, SLV (Semivariograma desde la Varianza Local), está relacionado con las derivadas del primero y es un estimador del semivariograma.

Para datos reales y sintéticos, se comparan sus comportamientos con el del estimador clásico de Matheron (CSV) y con el de los dos estimadores integrales propuestos por Li y Lake (MW1 y MW2).

Las conclusiones son:

1) El nuevo estimador integral LV es estable a largas distancias en contraposición a CSV, cuya precisión disminuye a medida que la distancia de separación aumenta. Comparte esta característica con MW1.
2) A partir de LV y MW1 se puede definir claramente una distancia de autocorrelación.
3) Los estimadores integrales son ideales para ser ajustados mediante un modelo teórico a diferencia de los estimadores de semivariograma (CSV, MW2, SLV) cuyo ajuste suele ser muy artesanal y subjetivo debido a las grandes fluctuaciones que pueden presentar para datos heterogéneos.
4) LV cuenta con algunas características que lo hacen más eficaz que MW1. Es una medida de la varianza de dispersión en regiones de diferentes tamaños y su tiempo de cálculo es tres veces menor que el de MW1.
5) SLV describe eficazmente la variabilidad a corta distancia, pero se torna más inestable que MW2 para datos heterogéneos a grandes distancias de separación. Su única ventaja frente a MW2 es que su tiempo de cálculo es tres veces menor.
6) Los estimadores integrales de autocorrelación LV y MW1 brindan una descripción más confiable acerca de la estructura espacial de los datos analizados, mientras que los semivariogramas CSV, MW2 y SLV son fuertemente influenciados por los valores puntuales que presenta el conjunto particular de datos.

Lista de símbolos

CSV	semivariograma clásico, definido por Matheron
d	dimensión del espacio euclideano
$E\{\}$	operador de esperanza
G	primitiva de Γ ((ec.(9))
h	distancia de separación
k	permeabilidad
m_i, m'_i	número de datos en una ventana móvil
MW1	estimador integral *moving window* 1
MW2	estimador de semivariograma *moving window* 2
$N(h)$	cantidad de pares de datos separados por una distancia h
n	número total de datos
$N(a,b)$	distribución normal, donde a es el valor medio y b la desviación estándar
LV	estimador integral de varianza local
SLV	estimador de semivariograma desde la varianza local
V	región del espacio muestral
Z	variable aleatoria
Z_V	variable regularizada (valor medio de Z en V)

Letras griegas

γ	semivariograma
γ_N	medida de autocorrelación integral (ec.(21))
Γ	primitiva de γ ((ec.(9))
σ_V^2	varianza de dispersión

AGRADECIMIENTOS

Este trabajo fue financiado con subsidios de la Universidad de Buenos Aires, de la Agencia Nacional de Promoción Científica y Tecnológica y del Consejo Nacional de Investigaciones Científicas y Técnicas de la Argentina. G.B. Savioli y M.S. Bidner son investigadoras de esta última institución.

REFERENCIAS

1 E.A. Breitenbach, "Reservoir simulation: state of the art", *Journal of Petroleum Technology*, Vol. **43**, N° 9, pp. 1033–1036, (1991).

2 E.A. Darderes, O.M. Sorarrain y M.S. Bidner, "Simulación numérica del flujo unidimensional de un gas real a través de un medio poroso", *Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería*, Vol. **4**, N° 2, pp. 151–174, (1988).

3 F. Samper Calvete y J. Carrera Ramírez, "*Geoestadística: Aplicaciones a la hidrología subterránea*", Centro Internacional de Métodos Numéricos en Ingeniería, Barcelona, España, (1990).

4 D. Li y L.W. Lake, "A moving window semivariance estimator", *Water Resources Research*, Vol. **30**, N° 5, pp. 1479–1489, (1994).

5 C.V. Deutsch y A.G. Journel, "*GSLIB- Geostatistical software library and user's guide*", New York University Press, New York, (1992).

6 J.L. Jensen, L.W. Lake, P.W.M. Corbett y D.J. Goggin, "*Statistics for petroleum engineers and geoscientists*", Prentice Hall PTR, New Jersey, (1997).

7 G.B. Savioli, A.F. Saccomano y M.S. Bidner, "Comparison of classical and new autocorrelation estimators", *IV World Congress on Computational Mechanics*, Buenos Aires, Argentina, 29 de junio al 2 de julio de 1998, publicado en "*Computational mechanics, new trends and applications*", E. Oñate and S.R. Idelsohn (Eds.), CIMNE, Barcelona, España, (1998).