

# Análisis exploratorio de variables regionalizadas con métodos funcionales

## Exploratory Analysis of Regionalized Variables with Functional Methods

RAMÓN GIRALDO<sup>a</sup>

UNIVERSIDAD NACIONAL DE COLOMBIA, FACULTAD DE CIENCIAS, DEPARTAMENTO DE  
ESTADÍSTICA, BOGOTÁ

---

### Resumen

Se muestra cómo las estadísticas descriptivas funcionales y el análisis en componentes principales funcional (ACPF) pueden emplearse en la evaluación empírica del supuesto de estacionariedad considerado en la modelación de variables regionalizadas. Se toma como ejemplo información georreferenciada correspondiente a mediciones de profundidad recogidas en 114 sitios de la Ciénaga Grande de Santa Marta, Colombia.

**Palabras clave:** análisis de datos funcionales, análisis en componentes principales funcional, estacionariedad.

### Abstract

It is shown how summary statistics of functional data and functional principal components analysis (FPCA) can be used to evaluate the stationarity assumption considered in modeling of regionalized variables. As an example is taken georeferenced information of depth measured at 114 locations at Ciénaga Grande de Santa Marta, Colombia.

**Key words:** Functional data analysis, Functional principal components analysis, Stationarity.

## 1. Introducción

Desde el trabajo pionero de Deville (1974) y más recientemente con el de Ramsay & Dalzell (1991), la comunidad estadística ha estado interesada en el análisis de datos funcionales (ADF). Se han propuesto versiones funcionales para métodos

---

<sup>a</sup>Profesor asociado. E-mail: rgiraldoh@unal.edu.co

estadísticos tradicionales como, entre otros, regresión (Cardot et al. 1999), análisis de varianza (Cuevas et al. 2004, Delicado 2007), modelo lineal generalizado (Escabias et al. 2004) o componentes principales (Pezulli & Silverman 1993). Los conceptos básicos del ADF y algunas de las metodologías antes mencionadas se encuentran en Ramsay & Silverman (2005).

No obstante, la tecnología moderna para la adquisición de información en tiempo real a menudo produce datos que pueden considerarse directamente como funcionales; es posible también obtener una cantidad finita y por lo tanto incompleta de información respecto a una función. Por ejemplo, cuando en las estaciones meteorológicas se colectan datos de temperatura diariamente, a pesar de que esta característica varía de manera continua en función del tiempo. En este último caso tiene sentido (debido a la continuidad) acudir al ajuste de curvas (Ramsay 1998) para obtener la observación funcional. Generalmente esta etapa se cumple usando métodos de suavizado y no paramétricos (Simonoff 1996). Debido a la continuidad espacial propia de la información geoestadística (Cressie 1993), surge de manera natural la posibilidad de adaptar las técnicas funcionales en este contexto.

En un análisis geoestadístico la fase exploratoria se lleva a cabo a través de gráficos que permiten inspeccionar la estacionariedad de la variable regionalizada estudiada (Isaaks & Srivastava 1987). Básicamente se construyen dispersogramas para establecer la presencia de tendencia y de autocorrelación. Un supuesto fundamental asumido es que la asociación espacial es función de la distancia y no de la posición, es decir que la correlación entre dos sitios depende de la distancia entre ellos y no de su ubicación dentro del área considerada (Samper & Carrera 1993).

En este artículo se muestra, a través de un estudio de caso, cómo las técnicas de análisis de datos funcionales permiten identificar tendencia en media y varianza y explorar la estructura de autocorrelación inherente a un conjunto de datos medidos en una región con continuidad espacial. Para exponer el procedimiento, se toma como ejemplo información georreferenciada sobre profundidad ( $m$ ) medida en 114 sitios del estuario Ciénaga Grande de Santa Marta, localizado en la costa norte de Colombia (figura 1). En la sección 3 se explica cómo el dato real de cada sitio (profundidad ( $m$ )) puede convertirse en funcional haciendo uso de la continuidad espacial y de métodos de suavizado. En la sección 4 se ilustra de qué manera esta adaptación permite cumplir con el objetivo exploratorio mencionado.

## 2. Teoría

En este apartado se presentan los conceptos que permiten contextualizar el trabajo y que se emplean en las secciones 3 y 4. En primera instancia se define, en el marco de la estadística espacial, qué es una variable regionalizada y cuáles son los supuestos que sobre ella deben estudiarse para lograr un apropiado uso de las herramientas de predicción espacial. Posteriormente se introducen los elementos básicos del análisis de datos funcionales y específicamente se hace referencia a algunas estadísticas descriptivas y al análisis en componentes principales.

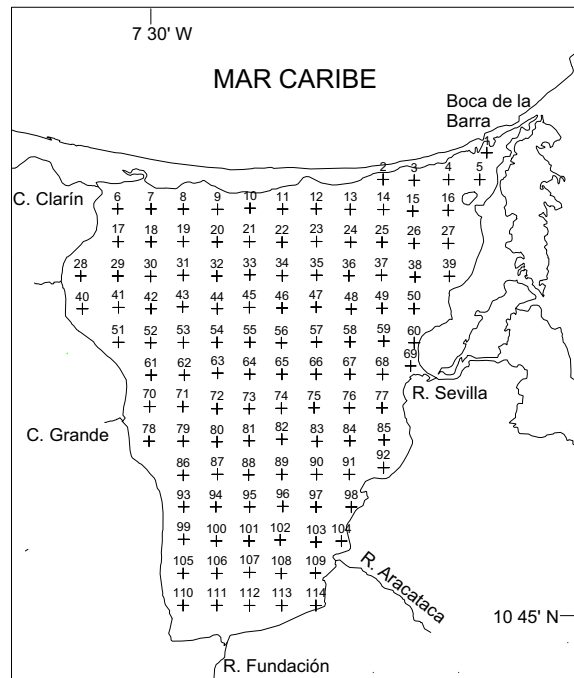


FIGURA 1: Región de estudio y localización de los sitios de muestreo.

## 2.1. Estadística espacial, variable regionalizada y estacionariedad

Estadística espacial es la reunión de un conjunto de técnicas apropiadas para el análisis de datos que corresponden a la medición de variables aleatorias en diversos sitios (puntos del espacio o agregaciones espaciales) de una región. De manera más formal se dice que esta misma trata con realizaciones de un proceso estocástico (campo aleatorio)  $\{Z(s) : s \in D \subset R^d\}$ , en el que  $s$  representa una ubicación en el espacio euclidiano  $d$  dimensional y  $Z(s)$  es una variable aleatoria en la ubicación  $s$ . Dicha disciplina se subdivide en tres grandes áreas (geoestadística, patrones puntuales y datos de áreas). La pertinencia de cada una de ellas está asociada a las características del conjunto  $D$  de índices del proceso estocástico de interés. Cuando las ubicaciones  $s$  provienen de un conjunto  $D$  continuo y son seleccionadas a juicio del investigador ( $D$  fijo), la información se enmarca dentro de la teoría geoestadística (Cressie 1993). En geología y minería se denomina como variable regionalizada al proceso estocástico así obtenido (Journel & Huijbregts 1978, Isaaks & Srivastava 1987). Este mismo se define como estacionario si la función de distribución del vector aleatorio  $\vec{Z}(s) = [Z(s_1), \dots, Z(s_n)]^t$  es idéntica a la del vector  $\vec{Z}(s+h) = [Z(s_1+h), \dots, Z(s_n+h)]^t$  para cualquier  $h$ . El proceso  $\{Z(s) : s \in D \subset R^d\}$  es estacionario fuerte si:

1.  $E(Z(s)) = m, \forall s \in D \subset \mathbb{R}^d$ , con  $m \in \mathbb{R}$ ,
2.  $COV[Z(s), Z(s+h)] = C(h) < \infty$ .

Las dos condiciones anteriores implican que la media y la varianza son constantes en la región y que la covarianza depende solo de la distancia entre los sitios y no de su posición dentro del área de estudio. Un supuesto muy usado en la práctica del análisis geoestadístico es el de estacionariedad débil que se basa en los siguientes requisitos:

1.  $E(Z(s) - Z(s+h)) = k, \forall (s, s+h) \in D \subset \mathbb{R}^d$ , con  $m \in \mathbb{R}$ ,
2.  $V[Z(s) - Z(s+h)] = 2\gamma(h)$ .

A  $2\gamma(h)$  se le denomina variograma y es la función comúnmente empleada para hacer estimación de la autocorrelación espacial.

Para la correcta práctica del análisis geoestadístico es de vital importancia juzgar la hipótesis de estacionariedad, puesto que la selección del método de predicción kriging (Christakos 2000) está directamente asociada con la propiedad de media constante. Cuando se tienen datos muestrales este requerimiento es estudiado de manera empírica usando dispersogramas de los valores de la variable de interés contra las coordenadas de medición (Cressie 1993). La suposición referente a la covarianza o al variograma es difícil de probar, incluso de forma descriptiva, y no se evalúa en la mayoría de trabajos aplicados.

## 2.2. Análisis de datos funcionales (ADF)

En un número creciente de campos aplicados las observaciones tomadas son curvas o imágenes. Estas son ejemplos de funciones puesto que la intensidad medida es disponible en cada punto de un segmento de línea o de una porción de un plano o volumen. En el ADF la unidad básica de información es la función completa, más que un conjunto de valores (Ramsay & Dalzell 1991). En el contexto multivariado los datos provienen de la observación de la familia aleatoria  $\{X(t_j)\}_{j=1, \dots, J}$ . De otro lado, en análisis funcional se asume que estos mismos proceden de una familia continua  $\chi = \{X(t); t \in T\}$  (Ferraty & Vieu 2006). Dos definiciones importantes para establecer diferencias entre los contextos real y funcional son (Ferraty & Vieu 2006):

**Definición 1.** Una variable aleatoria  $\chi$  se llama variable funcional (*v.f.*) si toma valores en un espacio infinito dimensional (espacio funcional). Una observación  $\chi$  de  $\chi$  se llama un dato funcional.

**Definición 2.** Un conjunto de datos funcionales  $\chi_1, \dots, \chi_n$  es la observación de  $n$  variables funcionales  $\chi_1, \dots, \chi_n$  con igual distribución que  $\chi$ .

Sea  $T = [a, b] \subseteq \mathbb{R}$ . Normalmente se asume que se tienen elementos de

$$L_2(T) = \left\{ f : T \rightarrow \mathbb{R}, \text{ tal que } \int_T f(t)^2 dt < \infty \right\}$$

$L_2(T)$  con el producto interno usual  $\langle f, g \rangle = \int_T f(t)g(t) dt$  es un espacio Euclidiano (He et al. 2000).

### 2.2.1. Estadísticas descriptivas en ADF

Las estadísticas descriptivas univariadas y bivariadas clásicas se aplican igualmente cuando se tienen datos funcionales. Dado un conjunto de datos funcionales  $\chi_1, \dots, \chi_n$ , definidos en  $t \in T \subset \mathbb{R}$ , las correspondientes funciones descriptivas están dadas por las siguientes expresiones (Ramsay & Silverman 1997):

- Media:  $\bar{\chi}(t) = n^{-1} \sum_{i=1}^n \chi_i(t)$ .
- Varianza:  $Var(\chi(t)) = (n-1)^{-1} \sum_{i=1}^n (\chi_i(t) - \bar{\chi}(t))^2$ .
- Desviación estándar:  $D.E.(\chi(t)) = \sqrt{Var(\chi(t))}$ .
- Covarianza:  $Cov(\chi(t_1), \chi(t_2)) = (n-1)^{-1} \sum_{i=1}^n (\chi_i(t_1) - \bar{\chi}(t_1))(\chi_i(t_2) - \bar{\chi}(t_2))$ .
- Correlación:  $Corr(\chi(t_1), \chi(t_2)) = \frac{Cov(\chi(t_1), \chi(t_2))}{\sqrt{Var(\chi(t_1))Var(\chi(t_2))}}$ .

### 2.2.2. Análisis en componentes principales funcional (ACPF)

Los objetivos del ACPF son los mismos del análisis en componentes principales clásico (ACP), es decir encontrar los modos dominantes de variación y conocer cuántos de estos son necesarios para lograr una aproximación satisfactoria a los datos originales. El ACP puede definirse en términos del siguiente procedimiento iterativo (Ramsay & Silverman 2005):

1. Se encuentra el vector de pesos  $\xi = (\xi_{11}, \dots, \xi_{p1})^t$  para el cual los valores (*scores*) del componente principal  $f_{i1} = \sum_j \xi_{j1} x_{ij}$  maximizan  $\sum_i f_{i1}^2$  sujeto a  $\sum_j \xi_{j1}^2 = \|\xi_1\|^2 = 1$ .
2. Se lleva a cabo una segunda y hasta  $p$  subsecuentes etapas. En la  $m$ -ésima etapa se calcula un nuevo vector de pesos  $\xi_m$ , con componentes  $\xi_{jm}$ , tal que  $\sum_j f_{jm}^2$  es máxima, sujeto a las restricciones  $\|\xi_m\|^2 = 1$  y  $\sum_j \xi_{jk} \xi_{jm} = \xi_k^t \xi_m = 0, k < m$ .

En el caso del ACPF, los valores de las variables se reemplazan por valores de funciones  $\chi_i(t)$ , tal que el índice discreto  $j$  en el contexto multivariado, descrito arriba, se sustituye por un índice continuo  $t$ . Las sumas sobre  $j$  se reemplazan por integrales sobre  $t$ . Por consiguiente el ACPF se encuentra como sigue:

1. Se halla la función de pesos  $\xi_1(t)$  que maximiza  $\sum_i f_{i1}^2$  sujeto a  $\int_T \xi_1^2(t) dt = \|\xi_1\|^2 = 1$ , con  $f_{i1} = \int_T \chi_i(t) \xi_1(t) dt$ .

2. Se realiza una segunda y hasta  $p$  nuevas etapas. En la  $m$ -ésima etapa, se calcula un nuevo vector de pesos  $\xi_m(t)$  y un nuevo componente principal tal que  $\sum_j f_{jm}^2$  se maximiza sujeto a las restricciones  $\|\xi_m\|^2 = 1$  y  $\int_T \xi_k(t)\xi_m(t) dt = 0$ .

En ACP clásico los componentes se obtienen resolviendo la ecuación propia  $V\xi = \rho\xi$ , donde  $V$  es una matriz de covarianzas o de correlación,  $\xi$  es un vector propio de  $V$  y  $\rho$  es un valor propio de  $V$ . La versión funcional (Ramsay & Silverman 1997) es análoga tomando  $\int_T \nu(s,t)\xi(t) dt = \rho\xi(t)$  como ecuación propia, donde  $\nu(s,t) = \sum_{i=1}^n \chi_i(s)\chi_i(t)$  es la función de covarianza ( $\chi_i(t)$  ha sido centrada),  $\rho$  es un valor propio y  $\xi(t)$  es una función propia de la función de covarianza  $\nu(s,t)$ . En el contexto funcional en general se trabaja con la función de covarianzas y no con la de correlación, porque los valores de las funciones están en la misma escala.

### 3. Método

Cuando se tienen datos de una variable en  $n$  sitios de una región con continuidad espacial, como los usados dentro del análisis geoestadístico (Cressie 1993), implícitamente en cada uno de ellos hay una observación funcional. Suponga que se calculan las distancias euclidianas  $d_{ij}$  del sitio  $i$ ,  $i = 1, \dots, n$ , a cada uno de los  $j$  restantes,  $j = 1, \dots, n$ , y que estas se ordenan en el eje de las abscisas. Si a cada  $d_{ij}$  se le asigna como ordenada el valor medido de la característica de interés en el sitio  $j$ , se tienen medidas puntuales de una observación funcional para el sitio  $i$  en el intervalo  $[0, h]$ , donde  $h$  es la distancia máxima entre este mismo y sus vecinos. Empleando técnicas de suavizado (Simonoff 1996), los valores así encontrados pueden convertirse en un dato funcional. En la figura 2 se muestran los dispersogramas de profundidad contra distancia en cuatro estaciones de muestreo de la CGSM y funciones (líneas continuas) halladas mediante *B-splines* (Simonoff 1996) en cada uno de ellos. Es claro que la distancia máxima entre un sitio y sus vecinos fluctúa dependiendo de su ubicación dentro del sistema y que por ende las funciones ajustadas tienen dominios distintos. Por ejemplo, en las estaciones de muestreo 56 y 82, ubicadas en el centro de la CGSM (figura 1), las distancias máximas están alrededor de 17 km y 21 km, respectivamente (figura 2), mientras que en las estaciones 1 y 114, localizadas en los extremos norte y sur (figura 1), estas son cercanas a los 30 km (figura 2).

El procedimiento descrito se realizó en cada uno de los 114 sitios de muestreo (figura 1), obteniéndose así igual número de funciones en el intervalo  $[0, 17000]$  (figura 3). El valor  $h = 17000$  corresponde al mínimo de las distancias máximas entre parejas de sitios y fue considerado como extremo del soporte de todas las funciones para homogeneizar los análisis (para tener igual número de datos en todas las distancias). Cada curva expresa el cambio esperado (suavizado) de profundidad en la medida en que hay alejamiento del correspondiente sitio. Con base en este conjunto de datos funcionales se calcularon las medidas descriptivas dadas en la sección 2.2.1 y se aplicó el ACPF de la sección 2.2.2, con el propósito de identificar tendencia espacial en media y varianza, de establecer el rango de auto-

correlación y de evaluar empíricamente si esta última función es homogénea dentro de la CGSM. Los gráficos de las funciones propias del ACPF se presentan como perturbaciones positivas y negativas de la media funcional (media funcional más o menos una constante por la función propia)(Ramsay & Silverman 2005). Con base en un *biplot* y en gráficos de contornos de los *scores* resultantes del ACPF, se hace la interpretación práctica desde el punto de vista espacial. Los análisis se llevaron a cabo usando las librerías *fda* y *geoR* del software R (R Development Core Team 2005).

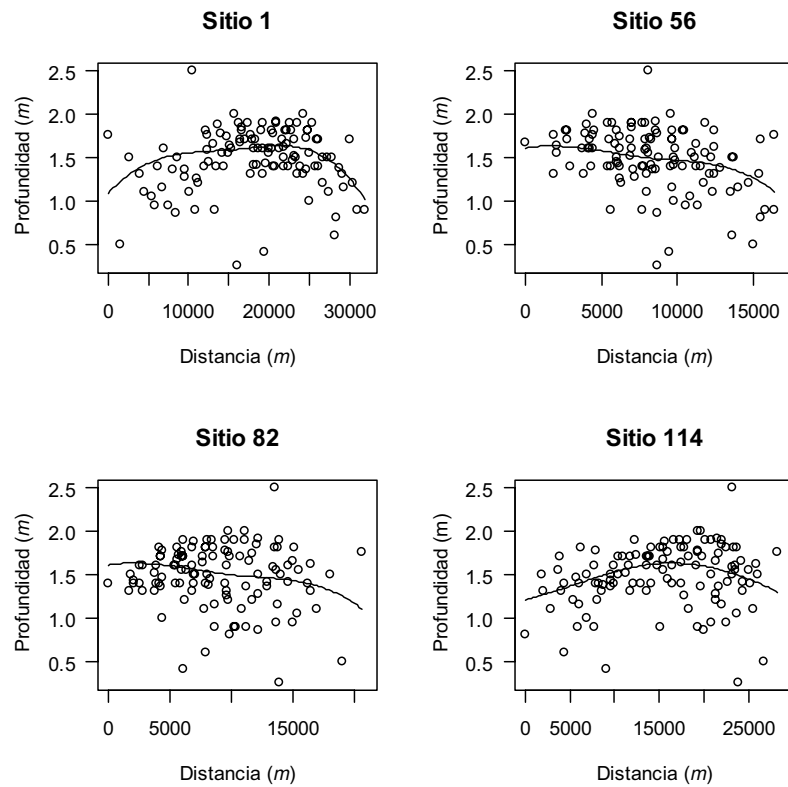


FIGURA 2: Valores de profundidad en función de la distancia, en cuatro sitios de muestreo de la Ciénaga Grande de Santa Marta. Las líneas continuas representan observaciones funcionales obtenidas por ajustes mediante *B-splines*.

#### 4. Resultados y discusión

Las medidas resúmenes globales de profundidad (tabla 1) hacen pensar que la CGSM es un sistema somero (promedio alrededor de 1.5 *m*) y homogéneo (C. V. < 30%). Sin embargo, a pesar de la poca variabilidad, las funciones descriptivas calculadas (figura 3) muestran que puede haber zonas con comportamientos

diferenciales (además permiten darse una idea global del radio de homogeneidad). El gráfico de la función media (figura 3, izquierda) presenta una pendiente muy pequeña antes de 10 *km* (aproximadamente) con magnitudes próximas al promedio global (tabla 1). Después de los 10 *km* predominan las curvas con valores bajos (figura 3), haciendo que la función media tienda a disminuir. Esto insinúa, desde el punto de vista práctico, que puede estarse presentando un fenómeno de estacionariedad local, es decir que hay zonas del sistema donde es razonable esperar un valor promedio alrededor de 1.47 *m*, pero que hay otras, especialmente de las fronteras del mismo, de baja batimetría. Es claro que la profundidad es menor en algunos de los extremos de la región de estudio porque las curvas tienden a disminuir para distancias grandes, es decir para aquellas con los sitios más alejados (los que están ubicados en el borde). El valor agregado de este resultado no es la identificación de las zonas con características similares, puesto que eso puede detectarse directamente con los datos medidos (Giraldo et al. 2000), sino la identificación global del alcance dentro de estas. Este patrón general se describe de manera detallada con los resultados del ACPF.

TABLA 1: Medidas descriptivas de datos de profundidad (*m*) tomados en marzo de 1997 en la Ciénaga Grande de Santa Marta, Colombia. C. V.: Coeficiente de Variación (%).

Medida	Profundidad ( <i>m</i> )
Media	1.47
Mínimo	0.25
Máximo	2.50
C. V.	24.1

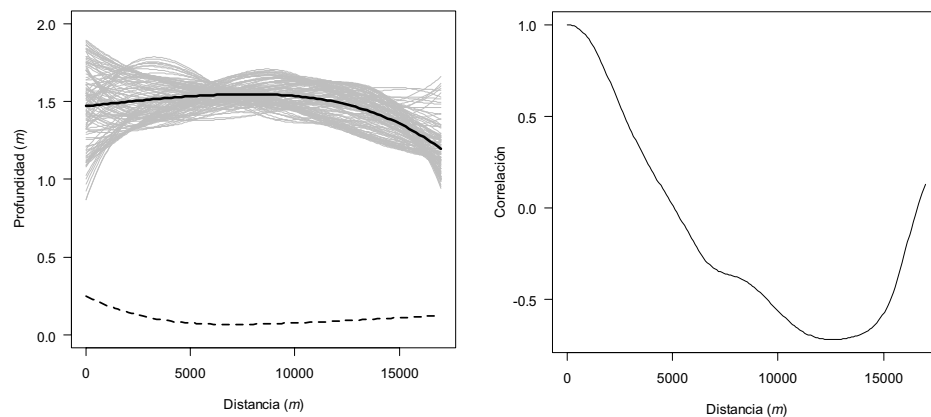


FIGURA 3: Izquierda: curvas de profundidad en función de distancias entre sitios (líneas continuas claras), función media (línea continua oscura) y función de desviación estándar (línea punteada). Derecha. función de autocorrelación.

En el gráfico de la función de desviación estándar (figura 3) se observa que esta toma su máximo en cero (desviación estándar alrededor de 0.25 *m* de profundi-



dad) y que disminuye hasta lograr valores próximos a cero (desviaciones estándar alrededor de 0 m de profundidad) para distancias entre 5 y 10 km. Este resultado, a pesar de su apariencia, no demuestra que la variabilidad cambia en función de la distancia. Más bien refleja la presencia de observaciones extremas en el conjunto original de datos. El rango mayor de 2 m (ver máximo y mínimo, tabla 1) sugiere que hay por lo menos dos sitios con condiciones de profundidad muy distintas. Cuando se emplean métodos de suavizado, hay mayor influencia de los datos en los extremos del soporte que en el centro donde hay más información para llevar a cabo el ajuste (Simonoff 1996) y por ello los datos atípicos afectan más la curva resultante si están al comienzo o al final del dominio de la función. Los valores mínimos y máximos de profundidad (dados en las estaciones 9 y 12, respectivamente, figura 1) aparecen, por construcción, en todos los dispersogramas calculados (ver por ejemplo los máximos y los mínimos de los dispersogramas de la figura 2). Sin embargo, por la razón anterior, su influencia en el cálculo de las funciones no es la misma. Las curvas de los sitios 9 y 12 (figura 3) o las de los vecinos a estos (los ubicados hacia la zona norte-centro) tendrán puntos de partida muy distintos (alejados) de las restantes y a consecuencia de esto la variabilidad funcional aumenta en las proximidades del origen.

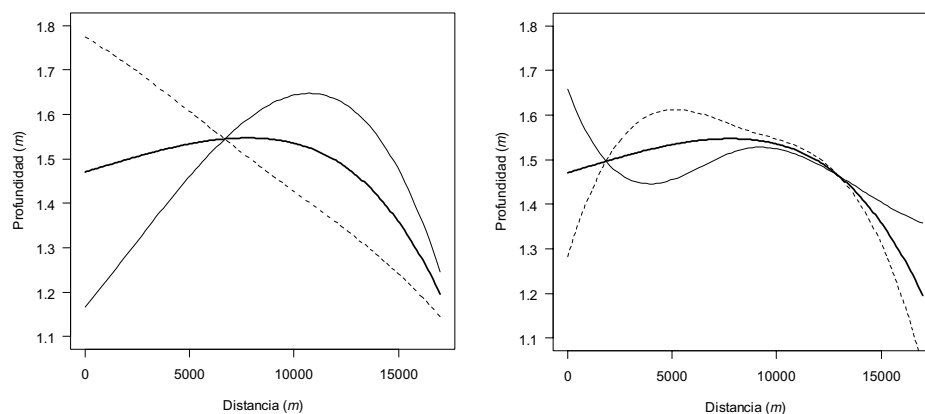


FIGURA 4: Funciones propias (*eigenfunctions*) del primer y segundo componente principal funcional (izquierda y derecha, respectivamente), graficadas como perturbaciones positivas (líneas continuas delgadas) y negativas (líneas punteadas) de la media funcional (línea continua gruesa).

La función de autocorrelación (figura 3) muestra un decrecimiento muy rápido en términos de la distancia. Toma el valor cero alrededor de los 5 km, lo cual implica que en general sitios separados más de 5 km no tienen asociación espacial. Cabe anotar que este es un punto de referencia global para la autocorrelación. Como se mostrará adelante, un estudio local (punto a punto) permite detectar relaciones espaciales de mayor distancia. De todas formas el rango estimado resulta mucho menor al que se obtiene a través del cálculo de la función de semivarianza (Giraldo et al. 2000). Una posible explicación de la diferencia entre las dos estimaciones

es la no consideración de valores extremos en el ajuste del semivariograma. La discrepancia entre los dos procedimientos podría evaluarse a través de un estudio de simulación.

En lo referente al ACPF, el primer componente principal funcional (CPF1) explicó el 68.2% y el segundo (CPF2) el 16.1% de la variabilidad en la profundidad (considerada de manera funcional). De acuerdo con las funciones de peso halladas (figura 4), el CPF1 explica la relación entre sitios distantes hasta 7 km (mediana escala) y el CPF2 la asociación hasta 2 km (pequeña escala). Sitios de muestreo con valores (*scores*) positivos en el CPF1 tenderán a ser someros y aquellos con valores negativos en este mismo tendrán mayor batimetría. En ambos casos en presencia de correlación espacial moderada. De otro lado, respecto al CPF2, podrá decirse que sitios con valores positivos en dicho componente, tienen valores de profundidad por encima del promedio y poca asociación espacial y que valores negativos de este corresponden a sitios someros de baja correlación espacial. El *biplot* resultante (figura 5) sugiere que el CPF2 está asociado a la variabilidad en las estaciones ubicadas en las zonas centro, centro-oriental y sur-oriental (estaciones de la parte superior del cuadrante uno) y que el CPF1 evidencia lo que ocurre en las zonas occidental, norte y sur (estaciones en los extremos del eje horizontal).

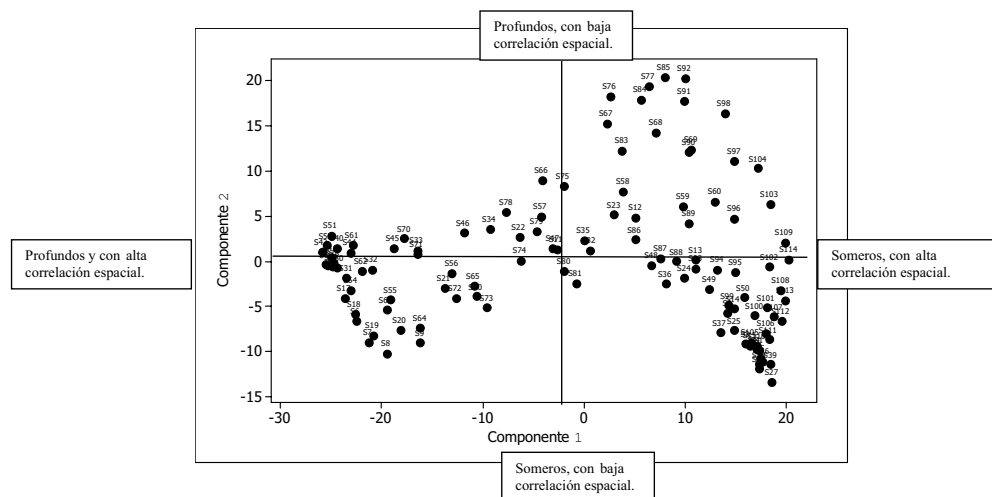


FIGURA 5: Valores (*scores*) de los sitios de muestreo sobre los dos primeros componentes principales hallados con base en datos funcionales de profundidad. Cada sitio es identificado por una etiqueta que contiene la letra *s* de sitio y un número correspondiente al dado en la figura 1.

Los gráficos de contornos de los *scores* correspondientes a los componentes principales funcionales 1 y 2 (figura 6) revelan que la profundidad está por encima del promedio en el costado occidental de la CGSM, puesto que en esta zona los valores del CPF1 son negativos. Además puede afirmarse que allí existe mayor dependencia espacial que en el resto del sistema, es decir, se espera encontrar estaciones de esta zona distantes hasta 7 km en las que de manera conjunta hay

valores altos de profundidad. Lo contrario ocurre en las zonas sur y nor-oriental donde las magnitudes de este componente son positivas.

Los *scores* correspondientes al CPF2 son especialmente interpretables, de acuerdo con la figura 5, en la zona centro-oriental. Se puede decir entonces que en esta subregión del área de estudio, el nivel de batimetría es alto, pero que hay poca asociación espacial, posiblemente debido a que sus estaciones están en un punto intermedio entre dos zonas someras (sur y nor-oriental).

En general el ACPF permite concluir que la autocorrelación espacial no depende de la posición, puesto que tanto en la zona somera como en la profunda hay correlación a mediana escala. El 16% explicado por el CPF2 corresponde a la relación a pequeña escala presente en las estaciones de transición entre áreas profundas y someras. Así mismo podría pensarse que el supuesto de estacionariedad fuerte no es válido, porque el nivel promedio global tiende a cambiar para distancias mayores de 10 km (distancia corta teniendo en cuenta que hay sitios separados cerca de 30 km) y porque parece haber diferencias entre los promedios de batimetría de la zona occidental y los de las zonas sur y nor-oriental del sistema.

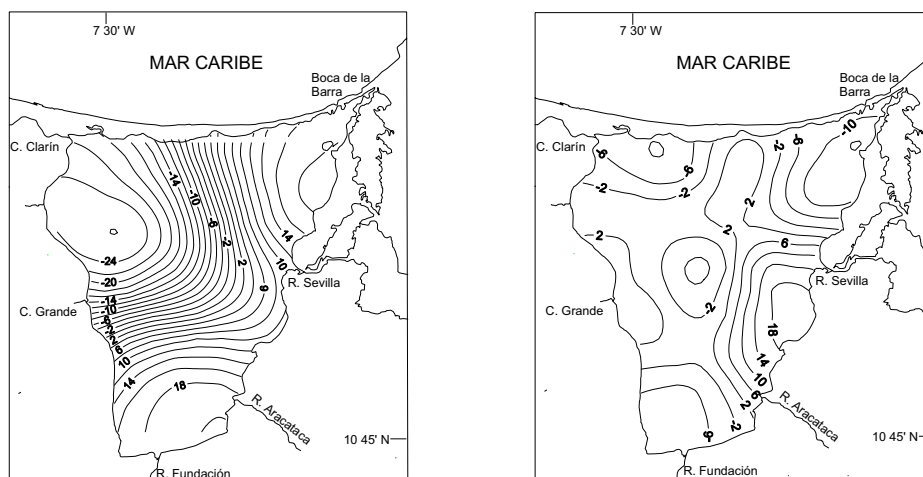


FIGURA 6: Gráficos de contornos de los *scores* de los componentes principales funcionales 1 (izquierda) y 2 (derecha).

## 5. Conclusión y perspectivas

La metodología empleada permite establecer de manera global cambios en media, varianza y correlación en una realización de un campo aleatorio y por consiguiente hace posible el estudio empírico de la estacionariedad del proceso de interés. Además hace factible, a través del uso del ACPF, la evaluación de la interacción local entre los valores medidos de la variable regionalizada considerada.

En el trabajo se presentó una aplicación del ADF dentro del contexto exploratorio geoestadístico. La extensión a las otras áreas de la estadística espacial y la formulación de teorías apropiadas para modelar datos funcionales con correlación espacial (cuando en cada sitio se observe directamente una función), son campos de investigación abiertos.

*Recibido: febrero de 2007*

*Aceptado: mayo de 2007*

## Referencias

- Cardot, H., Ferraty, F. & Sarda, P. (1999), 'Functional Linear Model', *Statistics and Probability Letters* **45**, 11–22.
- Christakos, G. (2000), *Modern Spatio Temporal Geostatistics*, Oxford University Press, New York.
- Cressie, N. (1993), *Statistic for Spatial Data*, John Wiley & Sons, New York.
- Cuevas, A., Febrero, M. & Fraiman, R. (2004), 'An ANOVA Test for Functional Data', *Computational Statistics and Data Analysis* **47**, 111–122.
- Delicado, P. (2007), 'Functional  $k$ -Sample Problem when Data are Density Functions', *Computational Statistics* .  
\*Published online, <http://www.Springerlink.com>
- Deville, J. (1974), 'Méthodes statistiques et numeriques de l'analyse harmonique', *Ann. Insee* **15**, 3–104.
- Escabias, M., Aguilera, A. & Valderrama, M. (2004), 'Principal Components Estimation of Functional Logistic Regression: Discussion of Two Different Approaches', *Journal of non Parametric Statistics* **16**(3-4), 365–384.
- Ferraty, F. & Vieu, P. (2006), *Non Parametric Functional Data Analysis. Theory and Practice*, Springer, New York.
- Giraldo, R., Troncoso, W., Mancera, J. & Méndez, N. (2000), 'Geoestadística: una herramienta para la modelación en estuarios', *Rev. Acad. Col. Cienc.* **24**(90), 60–72.
- He, G., Muller, G. & Wang, J. (2000), Extending Correlation and Regression from Multivariate to Functional Data, in M. Puri, ed., 'Asymptotics in Statistics and Probability', Brill Academic Publisher, Leiden, pp. 1–14.
- Isaaks, E. & Srivastava, M. (1987), *Applied Geostatistics*, Oxford University Press, New York.
- Journel, A. & Huijbregts, C. (1978), *Mining Geostatistics*, Academic Press, London.

- Pezulli, S. & Silverman, B. (1993), 'Some Properties of Smoothed Components Analysis for Functional Data', *Computational Statistics* **8**, 1–16.
- R Development Core Team (2005), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org>.
- Ramsay, J. (1998), 'Estimating Smooth Monotone Functions', *Journal Royal Statistical Society, Series B* **60**, 365–375.
- Ramsay, J. & Dalzell, C. (1991), 'Some Tools for Functional Data Analysis', *Journal Royal Statistical Society* **53**(3), 539–572.
- Ramsay, J. & Silverman, B. (1997), *Functional Data Analysis*, Springer, New York, USA.
- Ramsay, J. & Silverman, B. (2005), *Functional Data Analysis*, 2nd edn, Springer, New York, USA.
- Samper, F. & Carrera, J. (1993), *Geoestadística. Aplicaciones a la hidrogeología subterránea*, Centro Internacional de Métodos Numéricos en Ingeniería, UPC Barcelona.
- Simonoff, J. (1996), *Smoothing Methods in Statistics*, Springer.