

PROPUESTA DE UN INDICADOR COMO VARIABLE AUXILIAR EN EL ANÁLISIS COKRIGING

RAMÓN GIRALDO H.*

Resumen

Se propone un indicador que facilita la estimación del modelo lineal de correogionalización (MLC), necesario para la aplicación de la técnica cokriging. A través de un estudio de caso se comparan varianzas de predicción obtenidas con las técnicas kriging y cokriging empleando como variable auxiliar el indicador propuesto. Se muestra, para la información considerada, que el método cokriging resulta preferible en términos de costos y de precisión de las predicciones.

Palabras claves: Indicador $IGC_i(P)$, kriging, cokriging, varianza de predicción.

Abstract

It is proposed an indicator that facilitates the estimation of the linear model of coreogionalization necessary for the application of the cokriging method. Through a case study, the prediction variances, obtained by kriging and cokriging methods with the proposed indicator as auxiliary variable, are compared. It is shown, with the considered information, that the cokriging method is preferable in terms of costs and precision of the predictions.

Key words: $IGC_i(P)$ Indicator, kriging, cokriging, prediction variance

*Profesor Asistente, Departamento de Matemáticas y Estadística, Universidad Nacional de Colombia. email: rgiraldo@matematicas.unal.edu.co

1. Introducción.

Cuando se requiere llevar a cabo un análisis geoestadístico puede optarse por aplicar los métodos de predicción espacial kriging o cokriging (Cressie, 1993). La escogencia de una u otra técnica depende de si existe información auxiliar relacionada, desde un punto de vista espacial, con la variable de interés y de los costos de muestreo de la variable primaria. En los casos donde hay suficiente información muestral de variables auxiliares y los costos de muestreo de la variable de interés son altos, se recomienda la aplicación del método cokriging (Isaaks & Srivastava, 1989). No obstante cuando el número de variables auxiliares se incrementa, el ajuste del modelo lineal de correogionalización necesario para la aplicación del método cokriging suele ser dispendioso; puesto que es necesario que todos los modelos de semivarianza y de semivarianza cruzada sean estimados con la misma combinación de modelos teóricos (Bogaert et al., 1995), lo cual resulta complejo desde el punto de vista práctico. En este trabajo se propone el uso del indicador $IGC_i(P)$ (Giraldo, 2002) como método para reducir la dimensionalidad de las variables auxiliares, de forma tal que se simplifique el proceso de ajuste del modelo lineal de correogionalización y por consiguiente la aplicación del método cokriging. El procedimiento es ilustrado a través de un estudio de caso.

2. Marco Teórico.

En este acápite se hace la definición del indicador que se empleará como variable auxiliar en el análisis cokriging y se presentan los conceptos básicos de la teoría geoestadística, específicamente en lo referente a las técnicas kriging y cokriging.

2.1. Indicador $IGC_i(P)$

Suponga que se tienen datos de P variables binarias medidas en n sitios de una región. La estructura de esta información se presenta a continuación:

Sitio	X	Y	V_1	$V_2 \cdots$	V_P
1	x_1	y_1	v_{11}	$v_{12} \cdots$	v_{1P}
2	x_2	y_2	v_{21}	$v_{22} \cdots$	v_{2P}
3	x_3	y_3	v_{31}	$v_{32} \cdots$	v_{3P}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n	x_n	y_n	v_{n1}	$v_{n2} \cdots$	v_{nP}

Donde X y Y representan las coordenadas de ubicación geográfica (grados, planas o cartesianas), V_1, \dots, V_P , son P variables binarias con v_{ij} que indica éxito (1) o fracaso (0) respecto a la característica j en el sitio i . Éxito y fracaso son definidos a conveniencia. Por ejemplo éxito puede indicar la presencia de una especie, de un contaminante o puede indicar que el valor medido de una variable cuantitativa es mayor que el de su mediana muestral. La estructura de datos arriba descrita puede presentarse en una matriz de la siguiente forma:

$$\begin{pmatrix} n_{11} & n_{12} \cdots & n_{1P} \\ n_{21} & n_{22} \cdots & n_{2P} \\ \vdots & \vdots & \vdots \\ n_{n1} & n_{n2} \cdots & n_{nP} \end{pmatrix} n_{ij} = \begin{cases} 1 & \text{Éxito} \\ 0 & \text{Fracaso} \end{cases} \quad (1)$$

Se define el número de éxitos por sitio como:

$$\sum_{j=1}^P n_{ij} = n_{i.}, \quad \forall i, i = 1, \dots, n \quad (2)$$

y el número de éxitos en la característica j como:

$$\sum_{i=1}^n n_{ij} = n_{.j}, \quad \forall j, j = 1, \dots, P \quad (3)$$

Además sea:

$$K_i = \sum_{j=1}^P \delta_{ij}, \quad \text{donde } \delta_{ij} = \begin{cases} j & \text{si } n_{ij} = 1 \\ n_{i.} - P & \text{si } n_{ij} = 0 \end{cases} \quad (4)$$

El indicador $IGC_i(P)$ en el sitio i -ésimo se calcula como:

$$IGC_i(P) = P(P-2) + n_{i.} + K_i - h(P, n_{i.}) \quad (5)$$

donde:

$$h(P, n_{i.}) = \begin{cases} -2P, & \text{si } n_{i.} = 0 \\ 0, & \text{si } n_{i.} = 1 \\ (a_m - b_m P) & \text{si } m = n_{i.} - 2 \text{ y } n_{i.} = 2, 3, \dots, P \end{cases} \quad (6)$$

con:

$$\begin{aligned} a_0 &= 0, & b_0 &= -1, \\ a_m &= a_{m+1} + m(m+1) & b_m &= b_{m-1} + (m-1) \text{ y } m \text{ entero finito.} \end{aligned}$$

Se puede observar en la ecuación (5) que el indicador da prioridad a la detección del número de éxitos a través de n_i , y en función de éstos pondera, por medio de K_i , para identificar las posiciones de los éxitos dentro de la sucesión. Con las otras dos funciones involucradas en el cálculo del indicador se logra que, para cualquier P , su valor mínimo sea cero, y que los demás valores tengan incrementos unitarios hasta llegar al máximo. Las propiedades de la distribución del indicador y algunas guías para su interpretación en contextos prácticos puede obtenerse en Giraldo (2002).

2.2. Análisis Estructural y Kriging Ordinario.

Una variable regionalizada es un proceso estocástico $\{Z(x) : x \in D \subset R^m\}$ definido sobre un dominio D del espacio m -dimensional R^m . La variable regionalizada es estacionaria de segundo orden si cumple (Díaz-Francés, 1993):

- a) $E[Z(x)] = k, k \in R, \forall x \in D \subset R^m$
- b) $COV[Z(x), Z(x+h)] = C(h) < \infty$

Generalmente se trabaja sólo con la hipótesis que pide que los incrementos $[Z(x+h) - Z(x)]$ sean estacionarios, esto es (Clark, 1979):

- a) $E[Z(x+h) - Z(x)] = 0$
- b) $V[Z(x+h) - Z(x)] = 2\gamma(h)$

$\gamma(h)$ se conoce como la función de semivarianza y caracteriza las propiedades de dependencia espacial del proceso (Robertson, 1987). El estimador de momentos de la función de semivarianza es el semivariograma experimental (Wackernagel, 1995):

$$\gamma(h) = \frac{\sum^{n(h)} ((Z(x+h) - Z(x))^2)}{2n(h)} \quad (7)$$

donde $Z(x)$ y $Z(x+h)$ son variables en sitios separados por una distancia h y $n(h)$ es el número de parejas de puntos que se encuentran separadas por la distancia h . En Samper & Carrera (1990) se presentan modelos teóricos de comportamiento del semivariograma experimental (esférico, exponencial, gaussiano, entre otros) y se discuten métodos de ajuste de los mismos.

Una vez ajustado un modelo teórico al semivariograma experimental, es posible realizar la predicción en puntos no muestreados empleando, entre otros,

el método kriging ordinario. Cada predicción es calculada de los n datos muestrales $Z(x_i)$ como:

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i), \text{ con } \sum_{i=1}^n \lambda_i = 1 \quad (8)$$

$Z^*(x_0)$ es insesgado y los λ_i se obtienen de tal manera que se minimice la varianza del error de predicción $E(Z(x_0) - (\hat{Z})(x_0))^2$.

Los λ_i se calculan resolviendo el siguiente sistema de ecuaciones (Clark, 1979):

$$\begin{pmatrix} \gamma_{11} & \cdot & \cdot & \cdot & \gamma_{1n} & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \gamma_{ij} & \cdot & \cdot & \cdot \\ \gamma_{n1} & \cdot & \cdot & \cdot & \gamma_{nn} & 1 \\ 1 & \cdot & \cdot & \cdot & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \cdot \\ \cdot \\ \cdot \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma_{10} \\ \cdot \\ \cdot \\ \cdot \\ \gamma_{n0} \\ 1 \end{pmatrix} \quad (9)$$

en donde γ_{ij} es igual a $\gamma(h)$, calculado para la distancia entre los puntos x_i , x_j , μ es la mitad del multiplicador de Lagrange empleado para la restricción dada por la condición de insesgamiento y γ_{io} es igual a $\gamma(h)$ calculado para la distancia entre la i -ésima observación muestral y el punto donde se desea hacer la predicción. La varianza de predicción en cada punto es calculada por (Cressie, 1993):

$$\sigma_0^2 = \sum_{i=1}^n \lambda_i \gamma_{io} + \mu \quad (10)$$

en donde γ_{io} y μ se interpretan igual que en el párrafo anterior y los (λ_i) son los valores encontrados al resolver (9).

2.3. MLC y Cokriging

Si se tienen dos variables regionalizadas $Z_{v_1}(x)$ y $Z_{v_2}(x)$ tomadas en cada uno de los puntos, el estimador de momentos de la función de semivarianza cruzada está dado por el semivariograma cruzado (Bogaert et al., 1995):

$$\gamma_{v_1 v_2}(h) = \frac{1}{2n(h)} \sum^{n(h)} \{Z_{v_1}(x+h) - Z_{v_1}(x)\} \{Z_{v_2}(x+h) - Z_{v_2}(x)\} \quad (11)$$

donde $n(h)$ es el número de parejas de datos que se encuentran a una distancia h .

Un modelo lineal de correogionalización (MLC) asume que todos los semivariogramas simples (7) y cruzados (11) pueden expresarse como una combinación lineal de los mismos modelos teóricos (Isaaks & Srivastava, 1989).

En el caso de considerar sólo dos variables se tiene que:

$$\begin{aligned}\gamma_{v_1}(h) &= \alpha_0\gamma_0(h) + \dots + \alpha_m\gamma_m(h) \\ \gamma_{v_2}(h) &= \beta_0\gamma_0(h) + \dots + \beta_m\gamma_m(h) \\ \gamma_{v_1v_2}(h) &= \delta_0\gamma_0(h) + \dots + \delta_m\gamma_m(h)\end{aligned}\tag{12}$$

donde $\gamma_{v_1}(h)$ y $\gamma_{v_2}(h)$ son los semivariogramas simples y $\gamma_{v_1v_2}(h)$ es el semivariograma cruzado, $\gamma_0(h), \gamma_1(h), \dots, \gamma_m(h)$ son los modelos teóricos de semivarianza y α_i, β_i y δ_i , con $i = 1, 2, \dots, m$, son constantes.

Matricialmente:

$$\begin{aligned}\Gamma(h) &= \begin{pmatrix} \gamma_{v_1}(h) & \gamma_{v_1v_2}(h) \\ \gamma_{v_1v_2}(h) & \gamma_{v_2}(h) \end{pmatrix} = \sum_{s=0}^m B_s \gamma_s(h), \quad \text{donde} \\ B_s &= \begin{pmatrix} \alpha_s & \delta_s \\ \delta_s & \beta_s \end{pmatrix} \quad \gamma_s(h) = \begin{pmatrix} \gamma_s(h) & 0 \\ 0 & \gamma_s(h) \end{pmatrix}\end{aligned}\tag{13}$$

A $\Gamma(h)$ en la ecuación (13) se le denomina *matriz de correogionalización*.

El método de predicción espacial *cokriging* consiste en hacer predicción espacial de una variable con base en su información y en la de algunas variables auxiliares correlacionadas espacialmente con ella. El predictor cokriging tiene la siguiente expresión cuando se considera una sola variable auxiliar:

$$\hat{Z}_{v_i}^*(x_0) = \sum_{i=1}^{n_1} a_i Z_{v_1}(x_i) + \sum_{j=1}^{n_2} b_j Z_{v_2}(x_j)\tag{14}$$

El lado izquierdo de la igualdad anterior representa la predicción de la variable de interés en el punto x_0 no muestreado. $Z_{v_1}(x_i)$ con $i = 1, 2, \dots, n_1$, representa los valores observados de la variable primaria. Así mismo, los $Z_{v_2}(x_j)$ con $j = 1, 2, \dots, n_2$, representan los valores observados de la variable auxiliar. a_i , y b_j con $i = 1, 2, \dots, n_1$ y $j = 1, 2, \dots, n_2$ respectivamente, representan los pesos o ponderaciones de las observaciones de las variables primaria y auxiliar y se estiman con base en el MLC ajustado a los semivariogramas simples y cruzados (ecuación 13). Los pesos a_i y b_j se estiman con un proceso análogo

al descrito para el método kriging; es decir, éstos serán los que minimizan la varianza de predicción, sujeta a la restricción de que el predictor sea insesgado.

La estimación de los parámetros se obtiene resolviendo el siguiente sistema de ecuaciones (Isaaks y Srivastava, 1989):

$$\Upsilon_{ij} \lambda = \Upsilon_{i0} \quad (15)$$

donde

$$\Upsilon_{ij} = \begin{pmatrix} \gamma_{v_1}(1,1) & \cdots & \gamma_{v_1}(n,1) & \gamma_{v_1v_2}(1,1) & \cdots & \gamma_{v_1v_2}(m,1) & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{v_1}(1,n) & \cdots & \gamma_{v_1}(n,n) & \gamma_{v_1v_2}(1,n) & \cdots & \gamma_{v_1v_2}(m,n) & 1 & 0 \\ \gamma_{v_1v_2}(1,1) & \cdots & \gamma_{v_1v_2}(n,1) & \gamma_{v_2}(1,1) & \cdots & \gamma_{v_2}(m,1) & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma_{v_1v_2}(1,m) & \cdots & \gamma_{v_1v_2}(n,m) & \gamma_{v_2}(1,m) & \cdots & \gamma_{v_2}(m,m) & 0 & 1 \\ 1 & \cdots & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & 0 \end{pmatrix}$$

$$\lambda = \begin{pmatrix} a_1 \\ \vdots \\ a_n \\ b_1 \\ \vdots \\ b_m \\ \mu_1 \\ \mu_2 \end{pmatrix} \quad \Upsilon_{i0} = \begin{pmatrix} \gamma_{v_1}(0,1) \\ \vdots \\ \gamma_{v_1}(0,n) \\ \gamma_{v_1v_2}(0,1) \\ \vdots \\ \gamma_{v_1v_2}(0,m) \\ 1 \\ 0 \end{pmatrix}$$

La matriz Υ_{ij} contiene los valores de las funciones de semivarianza y de semivarianza cruzada, calculadas por medio del *MLC* estimado, para todas las distancias entre las parejas de puntos consideradas. Las dos últimas filas de dicha matriz corresponden a la restricción de insesgamiento del predictor. a_i y b_j con $i = 1, 2, \dots, n$ y $j = 1, 2, \dots, m$, son los parámetros a estimar, μ_1 y μ_2 son los multiplicadores de Lagrange empleados para la restricción de insesgamiento. El vector Υ_{i0} contiene los valores de las funciones de semivarianza y semivarianza cruzada, evaluados por medio del *MLC* estimado, para las distancias entre los sitios de muestreo (de ambas variables) y el sitio donde se desea hacer la predicción. Las dos últimas filas del vector están asociadas a la condición de insesgamiento. La correspondiente varianza de predicción del

método cokriging se calcula como (Bogaert et al., 1995):

$$\begin{aligned} \sigma_k^2 = & Cov(Z_{v_1}(x_0), Z_{v_1}(x_0)) + \mu_1 + \sum_{i=1}^n a_i Cov(Z_{v_1}(x_i), Z_{v_1}(x_0)) \\ & + \sum_{j=1}^m b_j Cov(Z_{v_2}(x_j), Z_{v_2}(x_0)) \end{aligned} \quad (16)$$

donde μ_1 es el multiplicador de Lagrange asociado a la condición de insesgamiento ($\sum_{i=1}^n a_i = 1$). $Cov(Z_{v_i}(x_k), Z_{v_i}(x_l)) = \sigma_{v_i}^2 - \gamma_{v_i v_i}(k, l)$ es la función de covarianza espacial de la variable i , $i = 1, 2$, evaluada para la distancia entre los sitios de muestreo k, l .

Una ventaja del método *cokriging* (ecuación 14) frente al *kriging* (ecuación 8) consiste en que cuando la variable auxiliar está ampliamente correlacionada con la variable de interés se puede obtener una disminución en la varianza de predicción (la varianza dada por la ecuación 16 es menor que la que se encuentra con la ecuación 9), no obstante dicha variable tenga menor densidad de muestreo. En situaciones en las que la variable objetivo tiene costos altos de muestreo se recomienda la aplicación de esta metodología (Bogaert et al., 1995).

Un problema práctico asociado al uso de cokriging es el ajuste del *MLC* cuando hay varias variables auxiliares, puesto que resulta dispendioso identificar la combinación de modelos teóricos y los parámetros de los mismos que permiten que todos los semivariogramas simples y cruzados sean ajustados. En esta situación una reducción de la dimensionalidad en las variables auxiliares, por ejemplo a través del indicador $IGC_i(P)$ antes descrito, puede servir para llevar el problema al caso de una sola variable auxiliar donde el ajuste del *MLC* resulta menos complejo.

3. Información utilizada y metodología

Para la realización del trabajo se tomó información de las variables profundidad (m), Secchi (cm), Salinidad, Oxígeno (mg/l) y Clorofila “a” (mg/l), medidas en un muestreo realizado en marzo de 1997 en el estuario Ciénaga Grande de Santa Marta ubicado en la costa norte de Colombia. La variable de interés para la aplicación del método cokriging es la clorofila “a”, dados sus altos costos de muestreo (insumos de laboratorio). Las restantes variables fueron consideradas como auxiliares (aunque no se muestra en el trabajo, los semivariogramas cruzados de dichas variables con la variable clorofila “a” muestran

fuertes estructuras de dependencia espacial entre ellas). Debido a la dificultad de ajuste del MLC para la aplicación del método cokriging (en este caso es necesario modelar 5 semivariogramas simples y 10 semivariogramas cruzados con la misma combinación de modelos teóricos) se decidió tomar la información de las variables auxiliares de forma “condensada” a través del indicador $IGC_i(4)$ (ecuación 5), calculado con base en la información dicotomizada de dichas variables. Para dicotomizarlas se compararon los valores observados con su correspondientes medianas muestrales.

Las 4 variables auxiliares fueron medidas en 114 sitios del ecosistema de estudio, por lo cual se calcularon 114 datos para la variable $IGC_i(4)$. La variable clorofila “a” fue medida en 107 sitios (en adelante se le denomina *CLO*). Para propósitos comparativos de las técnicas kriging y cokriging se considera también información de la variable clorofila “a” reducida a cerca de la mitad (54 datos) y en adelante se llamará *CLOI*.

Para estudiar la eficiencia de los métodos kriging y cokriging se compararon las estimaciones de las varianzas de predicción obtenidas al aplicar:

- Kriging sobre *CLO*.
- Kriging sobre *CLOI*.
- Cokriging sobre *CLOI* usando como variable auxiliar $IGC_i(4)$.

El ajuste del modelo lineal de correogionalización y el cálculo de las varianzas de predicción para los tres casos considerados fue realizado en el software The Spatial Interpolation of Agroclimatic Data (Bogaert et al., 1995).

4. Resultados y Discusión.

En primera instancia se calcularon los semivariogramas experimentales de las variables *CLO*, *CLOI* e $IGC_i(4)$ y el semivariograma experimental cruzado entre *CLOI* e $IGC_i(4)$. Sólo se incluye en el trabajo el semivariograma cruzado (Fig. 1). Este indica que las dos variables consideradas presentan correlación espacial inversa, es decir que valores altos de productividad biológica (alta clorofila “a”) están asociados con valores bajos del indicador $IGC_i(4)$ en zonas circundantes (incluso mayores a 10 km). Los valores bajos del indicador $IGC_i(4)$ están asociados a magnitudes por debajo de la mediana en las variables profundidad, Secchi, salinidad y oxígeno disuelto (ver interpretación del $IGC_i(P)$ en Giraldo (2002)). Lo anterior indica que zonas con alta biomasa fitoplanctónica estarán asociadas a baja profundidad, alta turbidez y a masas de agua con baja salinidad y bajo nivel de oxígeno (tal vez como consecuencia del consumo de éste durante las horas del día).

Una vez calculados los semivariogramas experimentales se ajustó el MLC entre las variables $CLOI$ e $IGC_i(4)$ (tabla 1), el cual incluye efecto pepita puro y un modelo esférico (para estudiar las propiedades de estos modelos ver Samper & Carrera (1990)). Con base en el MLC se realizaron las predicciones, de la variable $CLOI$, a través del método cokriging en 53 sitios de muestreo (en los que fue eliminada inicialmente la información) y se calculó la varianza de predicción máxima, mínima y promedio (tabla 2.)

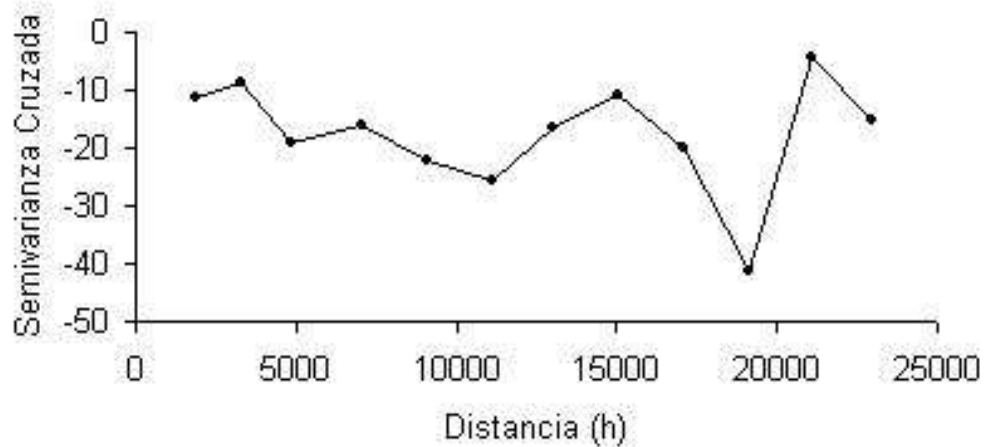


Figura 1. Semivariograma experimental cruzado de las variables clorofila "a" e $IGC_i(4)$. Información tomada en marzo de 1997 en la Ciénaga Grande de Santa Marta.

Tabla 1: Modelo de correionalización ajustado a los semivariogramas experimentales (simples y cruzado) de las variables $CLOI$ e $IGC_i(4)$. Información original medida en una muestreo realizado en marzo de 1997 en la Ciénaga Grande de Santa Marta.

Función de Semivarianza	Modelo Ajustado
$\gamma_{v_1}(h)$	$131.82 + 535.4$ Esférico (8000)
$\gamma_{v_2}(h)$	$3.89 + 9.59$ Esférico (8000)
$\gamma_{v_1 v_2}(h)$	$-1.18 - 18.70$ Esférico (8000)

Utilizando la información de $CLOI$ se llevó a cabo predicción en 53 sitios (en los que fue eliminada la información original de CLO) a través del método kriging y se calcularon nuevamente las varianzas de predicción máxima, mínima y promedio (tabla 2). Los resultados obtenidos en las varianzas de predicción

sobre *CLOI* mediante cokriging y kriging fueron comparados con las varianzas de predicción mínima y máxima obtenidas por Giraldo et al. (2001) para la variable *CLO* (tabla 2).

Tabla 2: Varianzas de predicción mínima, máxima y promedio ($\mu g^2/I$) para las variables *CLO* y *CLOI* obtenidas con los métodos kriging y cokriging. Entre paréntesis se encuentra la pérdida en precisión respecto al método kriging sobre *CLO* al comparar varianzas máximas y la pérdida en precisión respecto al método cokriging sobre *CLOI* al comparar las varianzas promedio

MÉTODO	Varianza de predicción máxima	Varianza de predicción mínima	Varianza de predicción promedio
Kriging sobre CLO	379 (0%)	0	...
Kriging sobre CLOI	506 (33.5%)	0	194.2 (2.2%)
Cokriging sobre CLOI con base en información auxiliar de $IGC_i(4)$	488 (28.7%)	0	190.1 (0%)

Los resultados mostrados en la tabla 2 respecto a las varianzas de predicción máximas indican que una reducción en los sitios de muestreo a cerca de la mitad en la variable *CLO* ocasiona una pérdida de precisión del 33.5% cuando se aplica kriging y del 28.7% cuando se aplica cokriging. Esto revela que en situaciones en las que sea necesaria la reducción de sitios de muestreo, por ejemplo por razones de costos (pasar de 107 sitios a 54 en la variable *CLO* disminuye costos en cerca de \$ 2'000.000, es preferible el uso del predictor cokriging.

Una conclusión análoga se obtiene al comparar las varianzas de predicción promedio obtenidas con los métodos kriging y cokriging sobre la variable *CLOI*. En esta situación se pierde un 2.2% de precisión al emplear el método kriging (tabla 2).

A manera de conclusión se establece, para el caso considerado, que si es necesario hacer una disminución en el número de sitios de muestreo de la variable clorofila "a", la mejor alternativa para identificar su distribución espacial es la aplicación del método cokriging, puesto que éste proporciona una menor varianza de predicción que el método kriging. Así mismo se muestra que una alternativa para facilitar la aplicación del método cokriging con información de múltiples variables auxiliares es la reducción de éstas a través del indicador $IGC_i(P)$.

Agradecimientos. El autor agradece a INVEMAR y COLCIENCIAS. La información utilizada en este trabajo fue tomada del proyecto “Análisis y aplicación de técnicas geoestadísticas en la modelación de procesos estocásticos relacionados con variables ecológicas en ambientes estuarinos” cofinanciado por ambas instituciones.

Referencias

- [1] Bogaert, P., P. Mahau & F. Beckers. (1995). *The Spatial Interpolation of Agroclimatic Data. Cokriging Software and Source Code*. FAO, Rome.
- [2] Clark, I. (1979). *Practical Geostatistics*. Elsevier Publishing, New York.
- [3] Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley & Sons, New York.
- [4] Díaz- Francés, E. (1993). *Introducción a Conceptos Básicos de Geoestadística*. Memorias Seminario Estadística y Medio Ambiente. Centro de Investigación en Matemáticas, Guanajuato, México.
- [5] Giraldo, R. (2002). *Construcción de un Indicador para el Estudio Conjunto de la Distribución Espacial de Múltiples Variables Binarias*. Tesis de Maestría en Estadística. Universidad Nacional de Colombia (en preparación para sustentación).
- [6] Isaaks, E. & R. M. Srivastava.(1989). *Applied Geostatistics*. Oxford University Press, New York.
- [7] Robertson, G. P. (1987). *Geostatistics in Ecology: Interpolating with Known Variance*. *Ecology* **68(3)**, 744–748.
- [8] Samper, F. J. & J. Carrera. (1990). *Geoestadística. Aplicaciones a la Hidrogeología Subterránea*. Centro Internacional de Métodos Numéricos en Ingeniería. Universidad Politécnica de Cataluña, Barcelona.
- [9] Wackernagel. H. (1995). *Multivariate Geostatistics. An Introduction with Applications*. Springer-Verlag, Berlín.