

ESTIMACION DEL HORIZONTE DE AUTOCORRELACION ESPACIAL: EL CASO DE LAS VARIABLES FISICO - QUIMICAS ESTUDIADAS EN LA CIENAGA GRANDE DE SANTA MARTA, COLOMBIA

Luis Hernando Hurtado Tobón y María Dolly García González

RESUMEN

El horizonte de autocorrelación espacial, entendido como la distancia mínima entre puntos a partir de la cual la correlación espacial se hace cero, es un parámetro de gran utilidad en el diseño de muestras que tienen como propósito el estudio del comportamiento espacial de una variable. En este trabajo se discute en detalle la estimación del horizonte de autocorrelación por la vía de los correlogramas de bandas de confianza y se hacen estimaciones para un conjunto de variables físico - químicas estudiadas en la Ciénaga Grande de Santa Marta, Norte de Colombia. Los resultados de la aplicación muestran que para cuatro de las variables incluidas en el estudio (Nitrito, Nitrato, suma de Nitrito más Nitrato y concentración de Oxígeno) no es posible analizar su comportamiento espacial debido a que el horizonte de autocorrelación es menor que la distancia mínima entre los puntos que formaron la muestra. Para las otras variables los horizontes de autocorrelación espacial en metros fueron: Amonio 4900, Salinidad 4600, Silicato 3900, Ortofosfato 2700 y Clorofila 2700. A través de la aplicación se nota además la distorsión que puede producirse en la estimación del horizonte de autocorrelación si previamente no se corrige la tendencia espacial que tiene la variable.

PALABRAS CLAVE: *Horizonte de autocorrelación espacial, Correlogramas de bandas de confianza, Índice de Moran.*

ABSTRACT

The spatial autocorrelation range: estimations of physical-chemical variables studied at the Ciénaga Grande de Santa Marta, Northern Colombia. The spatial autocorrelation range, understood here as the minimum distance between points at which the spatial correlation becomes zero, is a very useful parameter to design samples that have as goal the study of the spatial behavior of a variable. In this paper the estimation of the autocorrelation range by means of the correlograms of confidence bands are discussed in detail, and estimations for a set of physical-chemical variables studied at the Ciénaga Grande de Santa Marta, Northern Colombia, are made. The results of the application show that for four of the variables included in the study (Nitrite, Nitrate, the addition of Nitrite and Nitrate, and Oxygen concentration) it is not possible to analyze their spatial behavior because the autocorrelation range is less than the minimum distance between the points that constituted the sample. The spatial autocorrelation range for the other variables were, 4900 m for ammonia, 4600 m for salinity, 3900 m for silicate, 2700 m for orthophosphate and 2700 m for chlorophyll. Through the application it can be further noticed the distortion that may result in estimating the ranges if the spatial trend is not previously removed.

KEY WORDS: *Spatial autocorrelation range, Correlograms of confidence bands, Moran index.*

Una forma de estudiar variables aleatorias continuas que están autocorrelacionadas en el espacio es a través de la técnica del Kriging (Cressie, 1993). Para la utilización de esta técnica se parte de información sobre una variable en n localizaciones de una región y se predicen sus valores en un número de puntos suficientemente grande, de tal forma que se pueda visualizar su comportamiento a través de toda la región en estudio.

Un problema por resolver antes de aplicar la técnica del Kriging, es la ubicación de las n localizaciones donde se tomará la muestra que será base de la interpolación espacial o predicción. ¿Cuál debe ser el valor de n ?, ¿Cuál es la máxima distancia que puede haber entre las localizaciones vecinas para que exista autocorrelación espacial?, son preguntas a resolver inicialmente.

Para los procesos isotrópicos o sea aquellos donde la autocorrelación espacial depende solamente de la distancia entre los puntos sin importar la dirección, una solución a los anteriores interrogantes se obtiene con el procedimiento que se describe a continuación; siempre que se cumplan los supuestos de que existe autocorrelación espacial, la cual es función decreciente de la distancia entre las localizaciones, y que además la correlación entre la variable Z evaluada en el punto x_1 , $Z(x_1)$, y la misma variable Z en el punto x_2 , $Z(x_2)$, es cero cuando la distancia h entre los puntos x_1 y x_2 es mayor de h_0 , con $h_0 > 0$. El procedimiento mencionado es el siguiente:

Paso 1. Hallar h_0 a partir de la estructura de autocorrelación de la variable Z .

Paso 2. Verificar que las localizaciones x_i estén distribuidas en la región R de tal forma que el círculo de centro x_i y radio h_0 , para cada i , contenga varias localizaciones.

Cuando el proceso no es isotrópico (hay anisotropía) el procedimiento debe modificarse levemente considerando que existen dos valores de h_0 , esto es h_0' y h_0'' y en lugar de un círculo se asocia a cada localización una elipse cuyos ejes $2h_0'$ y $2h_0''$ que se intersectan en x_i .

La constante h_0 es un parámetro de la distribución espacial de la variable que se llama el horizonte de autocorrelación espacial y corresponde a la distancia mínima entre puntos a partir de la cual la autocorrelación espacial es cero.

Así como el tamaño de una muestra depende de la varianza de la variable, el número de localizaciones es una función del horizonte de autocorrelación espacial; en ambos casos se requiere de una información de la cual generalmente no se dispone y lo acostumbrado es acudir a una muestra piloto.

Una situación concreta donde se requiere determinar n localizaciones para analizar el comportamiento espacial de un conjunto de variables, se presenta en un estudio sobre la Ciénaga Grande de Santa Marta (CGSM) en el Norte de Colombia donde se dispone de una muestra tomada por el INVEMAR en 115 estaciones ubicadas a espacios regulares a lo largo y ancho de la Ciénaga con información sobre nueve (9) parámetros físico-químicos. Esta muestra se puede considerar como piloto para estimar horizontes de autocorrelación espacial que permitan posteriormente el diseño de muestreos fijando distancias mínimas entre las localizaciones de acuerdo con los horizontes de autocorrelación de las variables que interesa analizar.

Las técnicas utilizadas para la estimación de los horizontes de autocorrelación espacial fueron en este caso el Índice de Moran (Moran, 1950) y la construcción de correlogramas de bandas de confianza.

Los resultados de la aplicación muestran que para las variables Nitrito, Nitrato, suma de Nitrito más Nitrato y concentración de Oxígeno, no es posible estimar los horizontes de autocorrelación con esta muestra piloto ya que la distancia a la cual fueron ubicadas las 115 localizaciones hace que la correlación global sea cero, según el Índice de Moran en la variable Nitrito, y después de remover la tendencia en las otras tres variables, por lo que no se cumple uno de los supuestos para aplicar el procedimiento propuesto. Para los demás parámetros los horizontes de autocorrelación espacial en metros son los siguientes: Salinidad 4600, Ortofosfato 2700, Silicato 3900, Amonio 4900 y Clorofila 2700.

La información utilizada

La muestra que se considera como piloto está formada por 115 puntos, ubicados a través de la CGSM a espacios regulares, con distancias entre ellos que varían en un intervalo de 1480 m como mínimo y 32000 m como máximo. La ubicación de los 115 puntos de la muestra piloto se ilustra en la Figura 1.

Puntos de muestreo

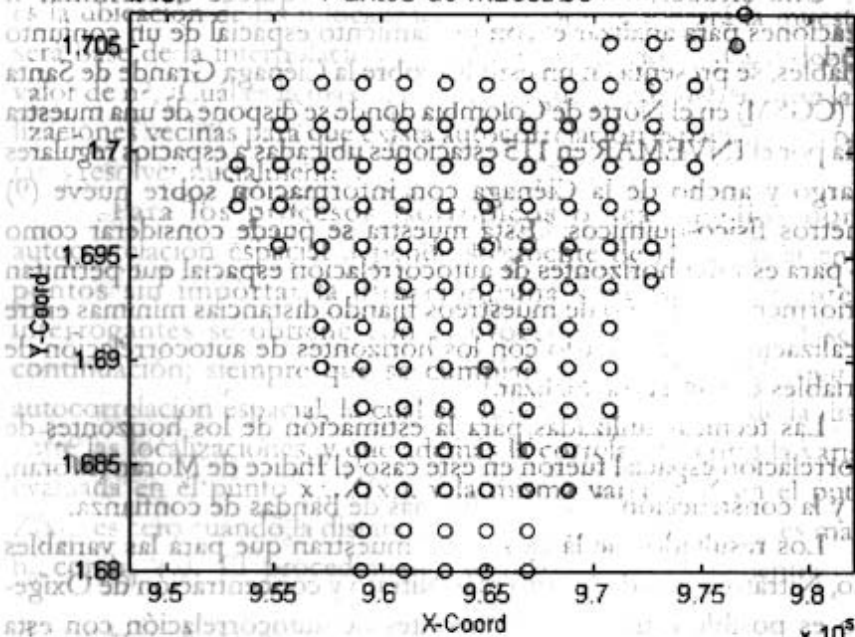


Figura 1. Ubicación de los puntos de muestreo en la Ciénaga Grande de Santa Marta para la muestra piloto

En cada una de las localizaciones se registró información sobre las siguientes variables: Salinidad, Nitrito, Ortofosfato, Silicato, Nitrato, Amonio, concentración de Oxígeno, Clorofila y suma de Nitrito y Nitrato. El diseño de esta muestra piloto y la toma de la información fueron realizados por investigadores del INVEMAR el 16 de Diciembre de 1995 como una actividad dentro de un programa de investigación sobre la CGSM.

La autocorrelación espacial

La autocorrelación espacial se analiza desde dos puntos de vista, en términos globales y como función de la distancia entre las localizaciones.

La autocorrelación global. La autocorrelación global en cada una de las variables se evalúa por medio del Índice de Moran, el cual es esencialmente un coeficiente de correlación ponderado por un factor que expresa la relación espacial entre las localizaciones. Formalmente, para una variable Z , de la cual se tienen valores en n localizaciones x_i ($i=1,2,\dots,n$) el Índice de Moran I se obtiene a través de

$$I = \frac{n}{n-1} \sum_{i=1}^n \sum_{j=1}^n \frac{w_{ij} (z(t_i) - \bar{z}) (z(t_j) - \bar{z})}{ws_i^2} \quad \text{donde } w_{ij} \text{ es el}$$

conjunto de ponderadores, $w = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$ y s_z^2 la varianza estimada de Z en las x_i localizaciones (Kern, 1997).

Considerando el Índice I como una variable aleatoria (es un estimador), se tiene que

$$E[I] = -\frac{1}{n-1}$$

Si se supone además que los valores de Z están asociados a las n localizaciones x_i solamente por azar, la varianza de I está dada por

$$V(I) = \frac{as_1 + bs_2 + cw^2}{(n-1)^2 (n-2) (n-3)w^2}$$

donde,

$$a = (n-1) [n^3 - (3+b_2)(n^2-2)],$$

$$b = n(n-1)(2b_2-n),$$

$$c = (n-1)(3n-6b_2) - (n-2)(n-3)$$

$$s_1 = \sum_{i=1}^n \sum_{j=1}^n (w_{ij}^2 + w_{ij} w_{ji})$$

$$s_2 = \sum_{i=1}^n w_i^2 + 2w_i \cdot w_j + w_j^2$$

Quando se supone que los valores de Z en los sitios $x_i, i = 1, 2, \dots, n$, corresponden a una muestra de variables independientes, Normales y con idéntica distribución, la varianza de I se obtiene por medio de

$$V(I) = \frac{n(n-1)(ns_1 - s_2) + (2n-4)w^2}{(n+1)(n-1)^2 w^2}$$

Con cualquiera de los dos supuestos sobre el comportamiento de Z, el estimador I tiene distribución asintóticamente Normal. La anterior afirmación y las formas que tienen el valor esperado y las varianzas de I se encuentran en Cliff y Ord (1981).

Llamando $S_e(I)$ a la desviación estándar de I los resultados anteriores permiten llegar a que, para valores de n suficientemente grandes, el estadístico $I/S_e(I)$ tiene una distribución aproximada a la Normal estándar y se puede utilizar para probar la hipótesis nula de que la variable Z se distribuye espacialmente en forma aleatoria. Para la prueba de esta hipótesis se utiliza como forma alterna la construcción de intervalos de confianza para I y como criterio de prueba se rechaza la hipótesis nula cuando el intervalo construido no contiene el cero.

La aplicación de lo expuesto, sobre el Índice de Moran y la autocorrelación global, a las nueve (9) variables estudiadas en la CGSM, tomando como ponderador el inverso de la distancia entre localizaciones, se presenta en la Tabla 1.

Como se puede observar en la Tabla 1, solamente se acepta la hipótesis nula (el intervalo contiene el cero) en la variable Nitrito, es decir que según la muestra tomada en la CGSM en esta variable no se presenta autocorrelación espacial global. En otros términos, el resultado permite afirmar que las distancias a las cuales estuvieron ubicadas las

estaciones de muestreo para la información procesada fueron mayores que el horizonte de autocorrelación espacial de la variable Nitrito. Así que un estudio sobre el comportamiento espacial del Nitrito requiere ubicar los puntos de muestreo a una distancia menor de 1480 m que fue la distancia mínima en este caso.

Tabla 1. Estimación del Índice de Moran para las nueve (9) variables físico-químicas estudiadas en la CGSM.

Nombre de la variable	Estimación del Índice I	Intervalo de confianza bajo Aleatoriedad		Intervalo de confianza bajo Normalidad	
<i>Salinidad</i>	0.1675	0.1485	0.1865	0.1484	0.1866
<i>Nitrito</i>	0.0088	- 0.0102	0.0278	- 0.0102	0.0278
<i>Ortofosfato</i>	0.0816	0.0634	0.0990	0.0627	0.1007
<i>Silicato</i>	0.1665	0.1478	0.1852	0.1477	0.1853
<i>Nitrato</i>	0.0426	0.0259	0.0594	0.0239	0.0614
<i>Amonio</i>	0.0564	0.0673	0.1040	0.0665	0.1048
<i>Can. Oxígeno</i>	0.7027	0.0461	0.0945	0.0460	0.0945
<i>Clorofila</i>	0.3391	0.1106	0.1496	0.1106	0.1495
<i>Nitrito y Nitrato</i>	0.0685	0.0507	0.0865	0.0494	0.0878

La autocorrelación como función de la distancia

Para estudiar la forma como la autocorrelación espacial depende de la distancia entre localizaciones empezamos por separar la variable de análisis $Z(x)$ en dos componentes: una que se refiere a la variación a gran escala o tendencia espacial $T(x)$ y la otra que incluye la variación espacial en pequeña escala $e(x)$, así que:

$$Z(x) = T(x) + e(x),$$

donde $e(x)$ se considera una variable aleatoria para la cual $E[e(x)] = 0$ y representa las variaciones alrededor de la media de $Z(x)$.

La forma como varía la autocorrelación en función de la distancia entre las localizaciones, en pequeña escala, permite ubicar un punto a partir del cual la autocorrelación espacial se hace cero, es decir el horizonte de autocorrelación espacial. Esta relación entre distancia y autocorrelación se puede analizar por medio de un tipo especial de correlograma que se construye para cada variable luego de remover la tendencia, es decir un correlograma para la componente aleatoria $e(x)$.

Para la construcción del mencionado correlograma se empieza por un agrupamiento de las distancias entre localizaciones en intervalos de clase y con los valores que toma la variable y sus coordenadas en cada clase se construyen intervalos de confianza para el Índice de Moran, es decir que en cada marca de clase se tiene un intervalo de confianza para I.

Este procedimiento puede representarse gráficamente en un plano cartesiano, ubicando las marcas de clase en el eje horizontal y haciendo corresponder a cada marca de clase un segmento paralelo al eje vertical y de longitud proporcional al intervalo de confianza respectivo. El gráfico así construido se llama un correlograma de bandas de confianza y el punto de intersección, con el eje horizontal, del primer intervalo de confianza que contenga el cero es una estimación del horizonte de autocorrelación espacial para la variable que se está analizando.

La construcción de correlogramas para las ocho (8) variables que tienen autocorrelación global diferente de cero produce los resultados que se resumen en la Tabla 2, donde E_1 y E_8 representan respectivamente los extremos inferior y superior de cada intervalo de confianza y solamente se registran los intervalos que corresponden a las primeras cuatro (4) marcas de clase de las 15 en que fueron agrupadas las distancias entre las 115 localizaciones.

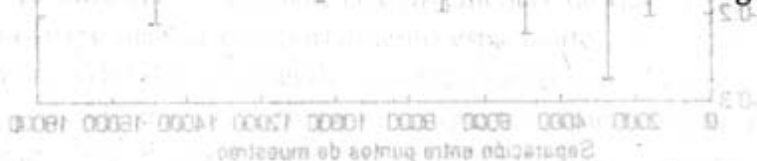
Tabla 2. Estimación de los horizontes de autocorrelación espacial para ocho (8) variables físico-químicas estudiadas en la CGSM.

Nombre de la variable	Intervalos de confianza para I								Horizonte de Autocorrelación (metros)
	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	
	0.29	0.63	0.06	0.31	-0.05	0.20	-0.09	0.09	4600
Ortofosfato	0.05	0.42	-0.01	0.37	-0.06	0.16	-0.28	0.14	2700
Silicato	0.30	0.68	0.12	0.51	-0.04	0.19	-0.29	0.14	3900
Nitrato	-0.18	0.17	-0.25	0.09	-0.10	0.10	-0.20	0.19	
Amonio	0.28	0.65	0.12	0.50	0.02	0.25	-0.31	0.11	4900
Can. Oxígeno	0.02	0.41	-0.39	0.08	-0.14	0.14	-0.08	0.50	
Clorofila	0.21	0.58	0.04	0.35	-0.10	0.15	-0.38	0.06	2700
Nitrito y Nitroso	-0.17	0.19	-0.27	0.10	-0.12	0.07	-0.17	0.12	

Además de las estimaciones de los horizontes de autocorrelación espacial para cinco de las ocho variables estudiadas, que se muestran en la última columna de la Tabla 2, otro resultado para destacar es lo que se refiere al comportamiento del Nitrato, concentración de Oxígeno y la suma de Nitrito y Nitrato. Aunque en la Tabla 1 estas mismas variables mostraron autocorrelación global positiva, luego de remover la tendencia para construir la Tabla 2 se observa que el primer intervalo contiene el cero, es decir que desaparece la autocorrelación. Este comportamiento se puede apreciar mejor comparando el correlograma de la variable en su forma original con el correspondiente a los residuales de la misma variable después de remover la tendencia. Para esta comparación en el caso del Nitrato se construyeron las Figuras 2 y 3 utilizando el software GEOPATH.

En forma similar se puede visualizar lo que ocurre con la variable concentración de Oxígeno antes y después de removerle la tendencia. Con la variable suma de Nitrito y Nitrato la situación es similar, por supuesto como una consecuencia de lo que pasa con el Nitrato y además porque el otro elemento de esta suma que es el Nitrito no tiene autocorrelación global positiva.

La aplicación de la técnica de los correlogramas de bandas de confianza a la Información obtenida a través del "área de barrida" de Diciembre 16 de 1995, sobre parámetros físico-químicos en la CGSM, permite mostrar que para cuatro de las nueve variables estudiadas allí no es posible evaluar su comportamiento espacial ya que sus horizontes de autocorrelación son menores de 1480 m que fue la distancia mínima entre localizaciones según la forma como se diseñó la muestra; mientras que el mínimo valor del horizonte de autocorrelación en las otras cinco variables es de 2700 m correspondiente al ortofosfato y la clorofila, así que para estudiar espacialmente solo estas cinco variables habría bastado con una muestra de 56 estaciones localizadas como en la Figura 4.



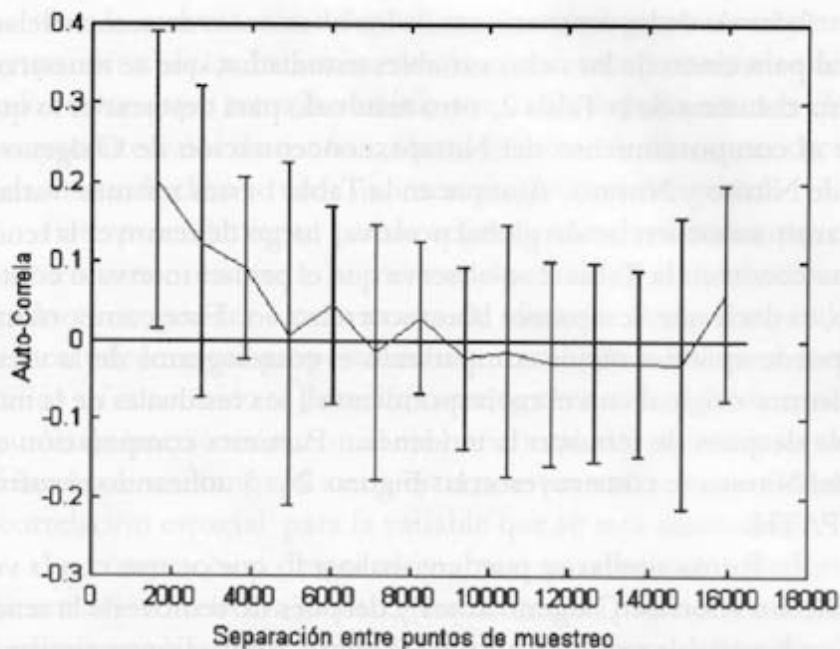


Figura 2. Correlograma para la variable Nitrato en su forma original.

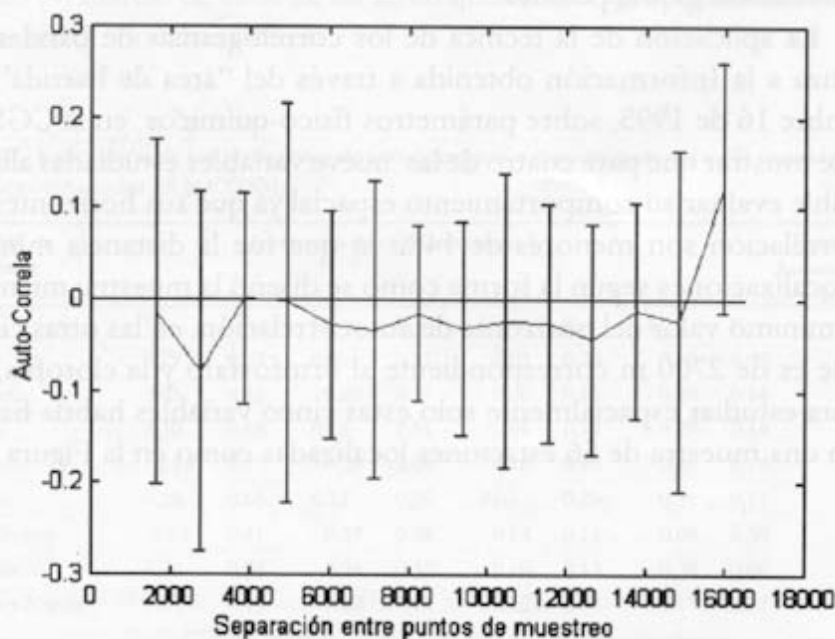


Figura 3. Correlograma de los residuos de la variable Nitrato luego de remover la tendencia.

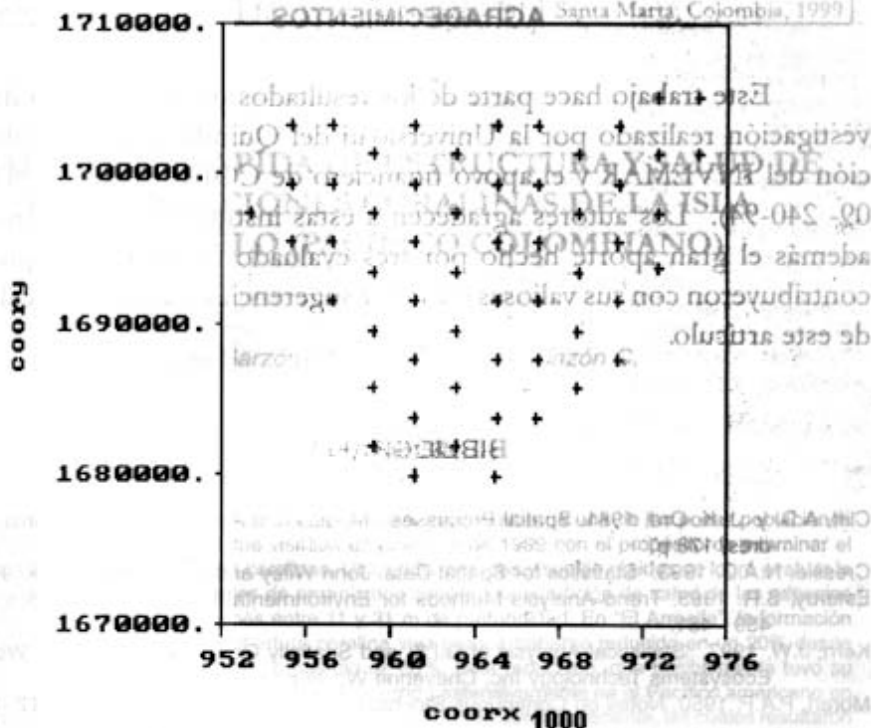


Figura 4. Relocalización de los puntos de muestreo

De otra parte se debe destacar la forma como la tendencia espacial puede enmascarar la verdadera dimensión de la autocorrelación espacial. Algunas variables como Nitrato y concentración de Oxígeno en su forma original muestran autocorrelación global positiva, pero una vez removida la tendencia la autocorrelación desaparece. Esto significa que se obtendrían estimaciones de los horizontes de autocorrelación espacial muy diferentes si en el análisis se omite el estudio y corrección de la tendencia.

Resulta claro, también, la conveniencia de que el diseño de muestras para analizar comportamiento espacial de variables se apoye en los valores de los horizontes de autocorrelación espacial de las variables que se pretende analizar, esto garantiza que la ubicación de los puntos de muestreo si permita detectar y aprovechar la autocorrelación espacial para los propósitos del estudio.

Este trabajo hace parte de los resultados de un proyecto de investigación realizado por la Universidad del Quindío con la colaboración del INVEMAR y el apoyo financiero de COLCIENCIAS (1113-09-240-94). Los autores agradecen a estas instituciones y reconocen además el gran aporte hecho por tres evaluadores anónimos quienes contribuyeron con sus valiosas críticas y sugerencias a mejorar la calidad de este artículo.

BIBLIOGRAFIA

Cliff, A.D. y J. K. Ord. 1981. Spatial Processes: Models and Applications. Page Bros. Londres, 178 p.
 Cressie, N.A.C. 1993. Statistics for Spatial Data. John Wiley and Sons. Nueva York, 900p.
 Esterby, S.R. 1993. Trend Analysis Methods for Environmental Data. Environmetrics, 4(4): 459 - 481.
 Kern, J.W. 1997. Statistical Analysis of G.I.S. and Spatially Correlated Field Data. Western Ecosystems Technology Inc. Cheyenne Wy, 86p.
 Moran, P.A.P. 1950. Notes on Continuous Stochastic Phenomena. Biometrika, 37: 17 - 23

FECHA DE RECEPCION: Mayo 27 de 1998 FECHA DE ACEPTACION: julio 22 de 1999

DIRECCION DE LOS AUTORES:

Departamento de Matemáticas, Universidad del Quindío, calle 12 Norte, Armenia, Quindío-Colombia. A.A. 460, E-mail: piesa@cocora.uniquindio.edu.co