

EL CONCEPTO GEOESTADÍSTICO DE VARIABLE REGIONALIZADA

Jorge A. Rodríguez G.*

La Geoestadística puede considerarse como una disciplina que se ocupa del análisis estadístico de variables espacialmente distribuidas; en el ámbito de la geografía venezolana los métodos y conceptos que constituyen a la Geoestadística son prácticamente poco conocidos; términos como semivarianza, variograma o técnica kriging son usuales en Geoestadística y el significado de los mismos solamente es pertinente en el contexto espacial si se aplican a variables geográficas que se comporten como *variables regionalizadas*.

Se conviene en que una variable geográfica o espacial es una variable regionalizada si cumple con los siguientes requisitos:

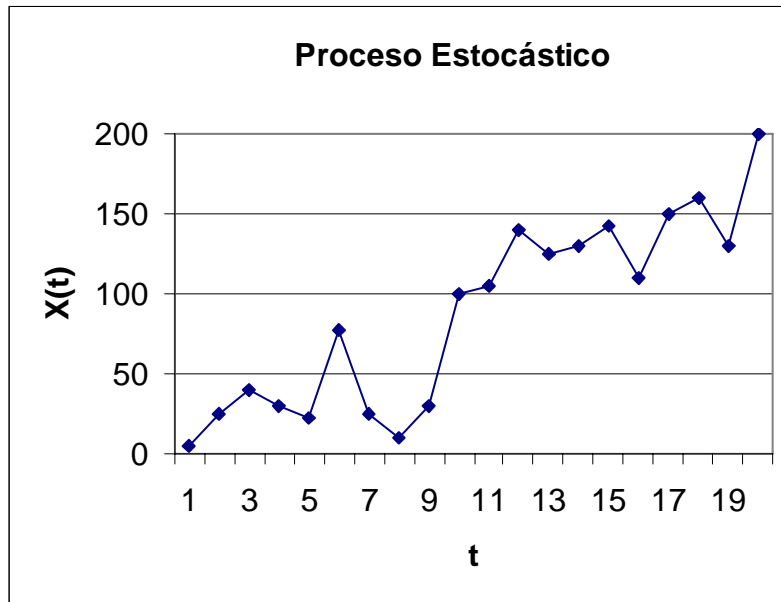
1. Es continua pero no matematizable
2. Posee variación local aleatoria
3. Posee variación regional o conjunta no aleatoria

Los requisitos (2) y (3) significan que la variable regionalizada deba asumirse como una variable estocástica; se considera que una variable es estocástica si ella resulta de la combinación de factores

* Profesor de la Escuela de Geografía de la Universidad Central de Venezuela

deterministas y de factores aleatorios. La naturaleza aleatoria está prescrita en el requisito (2), mientras que la naturaleza determinista está dada por la exigencia (3). En la literatura estadística especializada se acepta que si una variable geográfica cumple con la condición (3) entonces existe entre los datos autocorrelación espacial o que la variable geográfica posee una estructura de autocovarianza o de autocorrelación. Finalmente, si la variable geográfica posee una estructura de autocorrelación se concluye que ella forma parte o es el resultado de un Proceso Espacial Estocástico.

El concepto de Proceso Espacial Estocástico es análogo al Proceso Estocástico que se estudia en Análisis de Series de Tiempo, y, particularmente en el estudio que se realiza en Climatología de las variables meteorológicas, tal como se muestra en el siguiente ejemplo.



El proceso $X(t)$ tiene dos componentes: uno de carácter general –el cual es de tipo lineal tendencial– y el otro de carácter probabilístico. Se afirma que el proceso está correlacionado linealmente con la variable tiempo t pero además que cada dato individual es independiente del precedente en el sentido de que no es factible precisar de manera exacta qué valor ocurrirá en el próximo evento. Una descomposición lineal del proceso $X(t)$ vendría dada según la siguiente relación:

$$X(t) = XD(t) + XP(t) \quad (1)$$

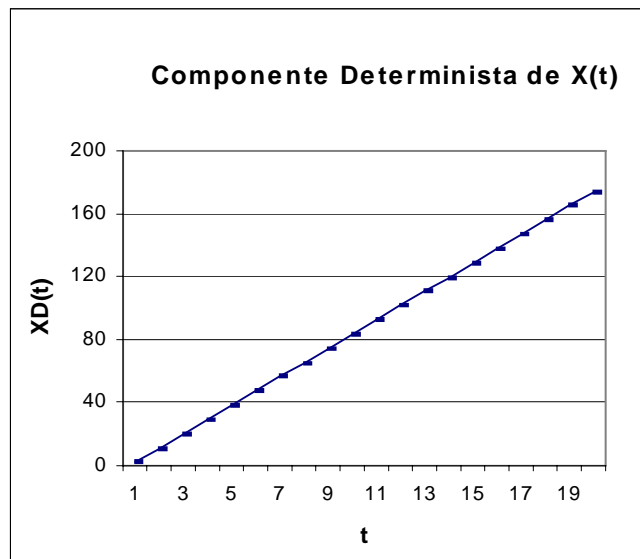
donde,

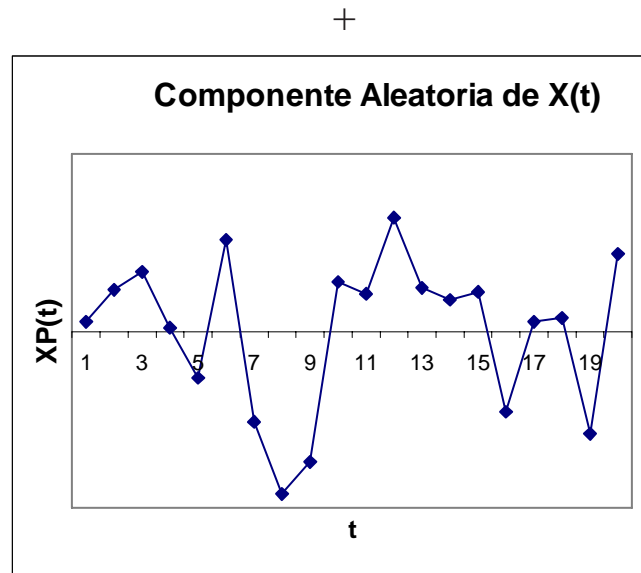
$X(t)$ = Proceso estocástico (temporal)

$XD(t)$ = componente determinista del proceso $X(t)$

$XP(t)$ = componente aleatorio del proceso $X(t)$

Una desagregación gráfica de $X(t)$ en sus componentes determinista y aleatoria podría ser la siguiente:





Éstas dos ilustraciones son las representaciones gráficas de los sumandos de la ecuación (1), pudiendo observarse la estructura lineal tendencial del sumando o componente determinista y la estructura aleatoria del componente o sumando probabilista del proceso estocástico $X(t)$.

Ahora bien, todo punto del proceso estocástico "temporal" requiere de un par de coordenadas: una para representar la variable tiempo cronológico –la variable t – y la segunda para representar la cantidad meteorológica –la variable $X(t)$. En ese mismo orden de ideas, todo punto que pertenece a un proceso estocástico espacial requiere de una terna de coordenadas: dos para representar la localización espacial y una para cuantificar la cantidad geográfica correspondiente. Por ejemplo, si se desea obtener la representación cartográfica de la hipsometría de un área determinada, cada punto de muestreo requiere los datos de latitud y longitud para su localización geográfica o geodésica, y como tercera cantidad la altitud o cota del punto en cuestión.

Si se conviene en que las cantidades georeferenciales de un lugar se simbolizan con la letra \mathbf{x} , la cantidad de la variable geográfica pudiera ser, por ejemplo, $\mathbf{Z}(\mathbf{x})$. Ello significa que la simbología matemática de toda cantidad o variable geográfica estaría dada por el par ordenado $[\mathbf{x}, \mathbf{Z}(\mathbf{x})]$. Esto implica que el proceso estocástico espacial (acrónimo, PEE) de una superficie geográfica dada está constituida por un conjunto finito, pero relativamente grande, de puntos de coordenadas $[\mathbf{x}, \mathbf{Z}(\mathbf{x})]$; es decir, la representación matemática del proceso estocástico espacial es $\{ [\mathbf{x}, \mathbf{Z}(\mathbf{x})] \}$, donde el par de llaves $\{ \}$ equivalen al término matemático de conjunto.

Sin embargo, la literatura consultada reporta que el PEE no es fácilmente matematizable pero es factible de manejar estadísticamente si la variable regionalizada o el proceso estocástico es **homogéneo**. Se asume que el PEE es homogéneo si: (1) todos los puntos tienen el mismo valor esperado; (2) se cumple que la autocorrelación espacial existe en toda la región o área estudiada. Al satisfacer el PEE los dos requisitos mencionados se afirma que el PEE es estacionario de segundo orden.

Si todos los puntos tienen el mismo valor esperado ello significa que estadísticamente se cumplirá que: $\mathbf{E} [\mathbf{Z}(\mathbf{x})] = \mathbf{E} [\mathbf{Z}(\mathbf{x}+\mathbf{h})] = \boldsymbol{\mu}$; es decir, el valor esperado de la variable geográfica en el sitio con coordenadas georeferenciales \mathbf{x} es igual al valor esperado de la variable geográfica en el sitio con coordenadas georeferenciales $\mathbf{x} + \mathbf{h}$, o sea en el sitio que se encuentra a una distancia \mathbf{h} del lugar con coordenadas georeferenciales \mathbf{x} . Como el valor esperado es único, entonces tal valor esperado puede representarse mediante la constante $\boldsymbol{\mu}$.

La existencia de la autocorrelación o autocovarianza espacial en un área A dada se expresa estadísticamente como: $\mathbf{COV} [\mathbf{Z}(\mathbf{x}), \mathbf{Z}(\mathbf{x}+\mathbf{h})] = \mathbf{C}(\mathbf{h}) \forall \mathbf{x} \in A$, donde las siglas \mathbf{COV} o \mathbf{C} simbolizan el operador estadístico matemático de la covarianza. La expresión $\mathbf{C}(\mathbf{h})$ implica que la covarianza es una función de la distancia \mathbf{h} y, específicamente, que es una función inversa de \mathbf{h} , o sea que a medida que aumenta la

distancia \mathbf{h} entre dos puntos $\mathbf{Z}(\mathbf{x})$ y $\mathbf{Z}(\mathbf{x}+\mathbf{h})$ menor será el valor de $\mathbf{C}(\mathbf{h})$. En resumen, la condición de homogeneidad se traduce estadísticamente mediante las siguientes representaciones algebraicas:

$$\mathbf{E} [\mathbf{Z}(\mathbf{x})] = \mathbf{E} [\mathbf{Z}(\mathbf{x}+\mathbf{h})] = \boldsymbol{\mu} \quad \forall \mathbf{x} \in \mathbf{A} \quad (2)$$

$$\mathbf{COV} [\mathbf{Z}(\mathbf{x}), \mathbf{Z}(\mathbf{x}+\mathbf{h})] = \mathbf{C}(\mathbf{h}) \quad \forall \mathbf{x} \in \mathbf{A} \quad (3)$$

Si en la **ecuación (3)**, $\mathbf{h} = \mathbf{0}$, resultará que:

$$\mathbf{C} [\mathbf{Z}(\mathbf{x}), \mathbf{Z}(\mathbf{x})] = \mathbf{VAR}(\mathbf{x}) = \mathbf{V}(\mathbf{v}) = \sigma^2 \quad \forall \mathbf{x} \in \mathbf{A} \quad (4)$$

La interpretación de la **relación (4)** es que todos los puntos de un área \mathbf{A} , donde la variable regionalizada es homogénea, tienen la misma varianza. En consecuencia, para toda el área \mathbf{A} se tienen como constantes los parámetros $\boldsymbol{\mu}$ y σ^2 , mientras que la covarianza \mathbf{C} variará según varíe la distancia \mathbf{h} .

Cuando una variable geográfica es a la vez una variable regionalizada y una variable homogénea es posible calcularle lo que se denomina la **semivarianza**. Para comprender el significado de este concepto es pertinente analizar como se relacionan los operadores varianza y covarianza, lo cual se mostrará seguidamente.

Sean dos variables \mathbf{x} , \mathbf{y} ; puede demostrarse que:

$$\mathbf{VAR} (\mathbf{X} - \mathbf{Y}) = \mathbf{V}(\mathbf{X} - \mathbf{Y}) = \mathbf{V}(\mathbf{X}) + \mathbf{V}(\mathbf{Y}) - 2 \mathbf{COV}(\mathbf{X}, \mathbf{Y})$$

O, de modo más resumido,

$$\mathbf{V}(\mathbf{X} - \mathbf{Y}) = \mathbf{V}(\mathbf{X}) + \mathbf{V}(\mathbf{Y}) - 2 \mathbf{C}(\mathbf{X}, \mathbf{Y}) \quad (5)$$

Si se conviene en que $\mathbf{V}(\mathbf{X}) = \mathbf{V}(\mathbf{Y})$, entonces resulta que:

$$V(\mathbf{X}-\mathbf{Y}) = 2 V(\mathbf{X}) - C(\mathbf{X},\mathbf{Y}) \quad (6)$$

Al factorizar la **expresión (6)**, queda:

$$\frac{1}{2} V(\mathbf{X}-\mathbf{Y}) = V(\mathbf{X}) - C(\mathbf{X},\mathbf{Y}) \quad (7)$$

Pero también, $C(\mathbf{X},\mathbf{X}) = V(\mathbf{X}) = \sigma^2$. En consecuencia, al sustituir en (7):

$$\frac{1}{2} V(\mathbf{X}-\mathbf{Y}) = \sigma_x^2 - C(\mathbf{X},\mathbf{Y}) \quad (8)$$

Si ahora se traslada la ecuación (9) a la variable regionalizada y homogénea $Z(x)$, se tiene:

$$\frac{1}{2} [V(Z(x) = Z(x+h))] = \sigma^2 \quad (9)$$

donde,

$V[Z(x)] = V[Z(x+h)] = \sigma^2$, dado que la variable $Z(x)$ es homogénea.

Así mismo, $C[Z(x), Z(x+h)]$, como ya se mencionó, es una función de la **distancia h** y por ello es válido simbolizarla así: $C[Z(x), Z(x+h)] = C(h)$. Al término a la izquierda de la **relación (9)** se le ha denominado semivarianza y se le simboliza como $\gamma(h)$; por lo tanto, se concluye que:

$$\gamma(h) = \sigma^2 - C(h) \quad (10)$$

En un punto cualquiera de coordenadas georeferenciales \mathbf{x} del **área A**, se cumplirá que:

$C[\mathbf{Z}(\mathbf{x}), \mathbf{Z}(\mathbf{x}+\mathbf{h})] = C[\mathbf{Z}(\mathbf{x}), \mathbf{Z}(\mathbf{x})] = \sigma^2$, dado que $\mathbf{h} = \mathbf{0}$. Ello significa que en cada punto del **área A**:

$$C[\mathbf{Z}(\mathbf{x}), \mathbf{Z}(\mathbf{x})] = \sigma^2 = C(\mathbf{0})$$

Sustituyendo en (10):

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}) \quad (11)$$

En consecuencia, si $Z(x)$ es la representación de una variable geográfica regionalizada y homogénea, implica que la semivarianza de $Z(x)$ – $\gamma(\mathbf{h})$ – es una función de la **distancia h** entre los puntos muestreados o medidos en el **área A** y que $\gamma(\mathbf{h})$ puede ser cuantificada conociendo la varianza regional de $Z(x)$ – $C(\mathbf{0})$ – y la estructura de autocovarianza de $Z(x)$, es decir, $C(\mathbf{h})$.

Los términos de la **ecuación (11)** pueden ser llevados a un gráfico bidimensional conocido como **semivariograma**, en el cual la **variable h** se plotea en el eje horizontal mientras que en el eje vertical se representan $C(\mathbf{0})$ y $C(\mathbf{h})$. Mediante el semivariograma es que se describe gráficamente la estructura del Proceso Estocástico Espacial de la variable regionalizada y homogénea $Z(x)$ y por ello la literatura especializada en el tema le presta particular atención a dicho esquema gráfico.