



Revista de Matemática: Teoría y Aplicaciones
ISSN: 1409-2433
mta.cimpa@ucr.ac.cr
Universidad de Costa Rica
Costa Rica

Rosa Alvarado, Welman

Modelo geoestadístico espacio-temporal del crimen en el salvador: análisis estructural y predictivo

Revista de Matemática: Teoría y Aplicaciones, vol. 18, núm. 2, 2011, pp. 325-342

Universidad de Costa Rica

San José, Costa Rica

Disponible en: <http://www.redalyc.org/articulo.oa?id=45326930007>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

MODELO GEOESTADÍSTICO
ESPAZIO-TEMPORAL DEL CRIMEN EN EL
SALVADOR: ANÁLISIS ESTRUCTURAL Y
PREDICTIVO

GEOSTATISTICAL SPATIO-TIME MODEL OF
CRIME IN EL SALVADOR: STRUCTURAL AND
PREDICTIVE ANALYSIS

WELMAN ROSA ALVARADO*

*Received: 23 Feb 2010; Revised: 10 Jun 2011;
Accepted: 24 Jun 2011*

*FEDECREDITO, 25 Av. Norte y 23 Calle Poniente, Edificio Macario Armando Rosales Rosa, San Salvador, El Salvador. E-Mail: welman_16@hotmail.com

Resumen

En la actualidad, estudiar un fenómeno espacial y espacio-temporal requiere de la búsqueda de herramientas estadísticas que permitan analizar la dependencia de espacio, tiempo e interacciones. La ciencia que aborda este tipo de temas es la denominada Geoestadística cuya finalidad es predecir fenómenos espaciales. Esta ciencia es considerada el pilar para la modelización de fenómenos que involucran las interacciones de espacio y tiempo. En los últimos 10 años la geoestadística ha tenido gran aplicación en áreas como la geología, la edafología, tratamiento de imágenes, la epidemiología, la agronomía, la ecología, economía, etc. En esta investigación se aplica para construir un mapa predictivo de la criminalidad en El Salvador; para ello, se estudia la variabilidad conjunta del espacio y tiempo para predecir o generar escenarios delincuenciales: focalizar áreas geográficas de inseguridad, determinar grupos vulnerables a sufrir hechos delictivos, incentivar la formulación de políticas públicas y facilitar la toma de decisiones en el tema de inseguridad.

Palabras clave: Geoestadística; Espacio-Temporal; No-Separable; Crimen; Estructural; Variabilidad; Autocorrelación.

Abstract

Today, to study a geospatial and spatio-temporal phenomena requires searching statistical tools that enable the analysis of the dependency of space, time and interactions. The science that studies this kind of subjects is the Geoestatics which the goal is to predict spatial phenomenon. This science is considered the base for modeling phenomena that involves interactions between space and time. In the past 10 years, the Geostatistic had seen a great development in areas like the geology, soils, remote sensing, epidemiology, agriculture, ecology, economy, etc. In this research, the geostatistic had been apply to build a predictive map about crime in El Salvador; for that the variability of space and time together is studied to generate crime scenarios: crime hot spots are determined, crime vulnerable groups are identified, to improve political decisions and facilitate to decision makers about the insecurity in the country.

Keywords: Geostatistics; Spatio-Time; Non-Separable; Crime; Structural; Variability; Autocorrelation.

Mathematics Subject Classification: 86A32.

1 Introducción

En los últimos años la comunidad científica ha venido dándole un interés primordial en la modelización geoestadística de fenómenos, construyendo nuevas familias de estructuras espacio-temporales válidas para un mayor tratamiento a muchas disciplinas como medioambiente, epidemiología, geología, medicina o geofísica en el estudio de la variabilidad espacial y espacio-temporal. Gran parte de este esfuerzo viene encaminado en la búsqueda de nuevas herramientas que permiten estudiar fenómenos donde se incluya interacciones entre el espacio y tiempo para estudiar la variabilidad de un fenómeno de una forma separada. Uno de esos esfuerzos ha sido posible gracias a las primeras aproximaciones en la construcción de covarianzas espacio-temporales (Kyriakidis y Journel, 1999) que se basan en la extensión natural de métodos espaciales o temporales y una dimensión espacio-temporal. Por otro lado, en lo que respecta a la construcción de modelos donde involucre el análisis de variabilidad con tipo de estructuras espacio-temporal, De Cesaré et al. (2001) llevó esa extensión hacia un contexto no separable, es decir, construir modelos basados en covarianzas espacio-temporales estacionarias en las que asumen isotropía en espacio y tiempo.

El objetivo de esta investigación es aplicar la geoestadística espacio-temporal al crimen en El Salvador para estudiar la variabilidad conjunta entre el espacio y tiempo; modelar el fenómeno a partir del estimador Cokriging y otro mediante el modelo de covarianza espacial y espacio-temporal vía producto suma (De Iaco, Myers y Posa, 2001) para generar un escenario delictual del homicidio en el país.

2 Manejo del proceso espacial y espacio-temporal

La geoestadística es una rama de la estadística que trata de fenómenos espaciales. Su objetivo es la estimación, predicción y simulación de dichos fenómenos. Esta rama ofrece una forma de describir la continuidad espacial, que es un rasgo distintivo esencial de muchos fenómenos naturales, y proporciona adaptaciones de las técnicas clásicas de regresión para tomar ventaja de esta continuidad (Berlanga, 1970).

Actualmente, la geoestadística es un conjunto de técnicas usadas para analizar y predecir valores de una propiedad distribuida en espacio y tiempo. En contraposición con la estadística clásica, tales valores no se consideran independientes, por el contrario se supone de manera implícita

que están correlacionados unos con otros, es decir, que existe una dependencia espacial.

En todo trabajo o estudio geoestadístico tiene que llevarse a cabo en tres fases:

- a) **Análisis Exploratorio de Datos Espaciales:** aquí se estudian los datos muestrales sin tener en cuenta su distribución geográfica. Sería una etapa de aplicación de la estadística. Se comprueba la consistencia de los datos geográficos y se identifica la distribución de los que provienen.
- b) **Análisis Estructural:** esta etapa se encarga de la caracterización de la estructura espacial y espacio-temporal de una propiedad o fenómeno regionalizado. Es el proceso en el marco del cual se obtiene un modelo geoestadístico para la función aleatoria que se estudia. En pocas palabras consiste en estimar y modelar una función que refleje la correlación espacial y espacio-temporal de la variable regionalizada a partir de la adopción razonada de la hipótesis más adecuada acerca de su variabilidad. Esta variabilidad se realiza mediante un análisis de patrón de localizaciones de la población objeto de estudio, que permite definir el tipo de estructura espacial y espacio-temporal (en su caso) que organiza al conjunto de localizaciones; y un análisis de dependencia o autocorrelación espacial y espacio-temporal de la distribución, que permite medir el grado de similitud de una variable regionalizada de tipo cuantitativa entre puntos o eventos vecinos sobre un área muestral de estudio.
- c) **Predicción:** Estimación de la variable en los puntos muestrales, considerando la estructura de correlación espacial relacionada e integrando la información obtenida de forma directa en los puntos muestrales, así como la conseguida indirectamente en forma de tendencias conocidas.

En dicho proceso de estudio geoestadístico es importante también, el conocer el tipo de dato geográfico para su tratamiento. Generalmente son dos tipos de datos geográficos, un tipo de dato coordenado que consiste en un punto o evento georreferenciado y un dato geoestadístico o campo escalar que significa la intensidad o frecuencia en una zona geográfica.

2.1 Análisis estructural y correlación

A continuación se presentan algunos estadísticos existentes en la teoría geoestadística para el análisis estructural y correlación:

a) **Análisis estructural con dato coordenado:**

Índice del vecino más próximo $NNI = d(NN)/d(ran)$ (Clark y Evans, 1954), es un cociente de dos medidas que compara la distancia entre los puntos más cercanos y la media distancia aleatoria. Si el índice resulta ser igual a 1, indica que la distribución se ajusta a una de Poisson; si resulta ser mayor que 1, existe tendencia a la dispersión y si es menor que 1, existe tendencia a la agregación.

b) **Análisis estructural con dato mixto coordenado y geoestadístico y/o campo escalar:**

La función K de Ripley (Ripley, 1976; Bailey y Gatrell, 1995), es un operador que permite establecer el tipo, la intensidad y el rango del patrón espacial a través del análisis de las distancias existentes entre todos los puntos. Este estadístico contrasta la hipótesis de aleatoriedad (esta hipótesis bajo el supuesto de Completa Aleatorización Espacial es igual a πr^2 , entonces, si: $K > \pi r^2$ indica agregación, $K < \pi r^2$ indica uniformidad y $K = \pi r^2$, el proceso muestra aleatoriedad).

c) **Correlación:**

La correlación se establece mediante el análisis de cómo una región cuantitativa (intensidad, frecuencia) varía de acuerdo con los ejes de coordenadas X e Y , que miden la localización de cada uno de los valores de dicha localidad y que son variables independientes. En ese sentido, la descripción de la correlación sólo es posible si se analiza el fenómeno con el tipo de dato geoestadístico y/o campo escalar. El Test I de Moran (Anselin, L. 1995), establece el tipo, la intensidad y el rango de patrón espacial, midiendo el grado de autocorrelación de una distribución, y su interpretación es la siguiente: si el I es mayor a 0, y con tendencia a +1, la autocorrelación será positiva; si el I es igual a 0, entonces no existe autocorrelación; y si el I resulta ser negativo, con tendencia a -1, la autocorrelación será negativa.

d) **Semivariograma**

Una herramienta que cuantifica la semivarianza que existe en una distribución, midiendo el grado de correlación existente entre los valores de la variable en cada punto y distancia entre ellos. El

propósito primordial del semivariograma, integrado en las técnicas geoestadísticas, va más allá de la determinación de la correlación, esto porque al cuantificar la relación de una variable medida en una serie de puntos, puede conocer/predecir esa misma variable medida en puntos situados a distancias conocidas, pero que no han sido muestreados. El variograma se define como la media aritmética de todos los cuadrados de las diferencias entre pares de valores experimentales separados por una distancia h . La función $\gamma(h)$ se denomina semivariograma y su expresión es la siguiente: $\gamma(h) = (1/2N_p(h)) \sum_{i=1}^{N_p(h)} [Z(x_i - Z(x_i + h)]^2$, donde $N_p(h)$ es el número de pares a la distancia h , h es el incremento, $Z(x_i)$ son los valores experimentales y x_i localizaciones donde son medidos los valores $Z(x_i)$.

Esta función de semivariograma tiene una serie de componentes que permiten describir ciertos aspectos de la variabilidad espacial y espacio-temporal (en su caso) de la población objeto de estudio como la meseta/umbral, efecto pepita y rango. El primero mide la discontinuidad en el origen; el segundo, mide el valor máximo de variabilidad (sill); y el último, mide el área de influencia de la correlación (alcance o rango).

En la literatura existe una serie de modelos teóricos admisibles o autorizados de semivariograma que se usan en la práctica, algunos de ellos son: el modelo esférico, exponencial y gaussiano (J. Berlanga and J. Obregón (1970)).

2.2 Predicción espacial y espacio-temporal

Dentro del contexto de la Geoestadística se conoce con el nombre de kriging a una familia de algoritmos de regresión por mínimos cuadrados generalizados que, a partir de un conjunto de observaciones $Z(x_i, t_i)_{i \in 1, \dots, n}$, proporcionan el predictor lineal óptimo para la variable Z en una nueva posición (x_0, t_0) . Por lo general, a estos algoritmos se les conoce con el nombre de kriging en reconocimiento del trabajo realizado en este campo por Danie Krige (1950). Estos métodos de Kriging han sido ampliamente estudiados y aplicados en el contexto de la estadística espacial, y espacio-temporal. Según la forma del estimador (Martín A. Diaz, 1969) el estimador lineal con valores esperados conocidos llamado Kriging Simple, tiene la forma de $\hat{Z}(x_0, t_0) = \sum_{i=1}^n \alpha_i Z(x_i, t_i) + (1 - \sum_{i=1}^n \alpha_i)m$. Este estimador siempre será insesgado, por lo tanto no necesitaremos imponer

la restricción de que $\sum_{i=1}^n \alpha_i = 1$.

Modelando una regionalización espacio-temporal

Consideremos a D un dominio en el espacio, y a T un dominio en el tiempo con $D \subseteq R^d$ y $T \subseteq R^1$. Sin pérdida de generalidad, asumamos que $d = 2$; así tenemos que R^2 es de dos dimensiones (horizontal) en espacio. Y la variable aleatoria (*RV*) $Z(u, t)$ una variable que puede tomar una serie de valores (realizaciones) en cualquier lugar en el espacio $u \in D$ y el instante de tiempo $t \in T$, de acuerdo a una distribución de probabilidad. Una función aleatoria espacio temporal (*RF*) $Z(u, t), (u, t) \in D \times T$ está definida como un conjunto generalmente dependientes *RVs* $Z(u, t)$, una para cada lugar en el espacio $u \in D$ y un instante de tiempo $t \in T$. Una realización de la función aleatoria *RF* $Z(u, t)$ (donde $u = (x, y, z)$) es una colección de las realizaciones de sus componentes *RVs*. Esta función aleatoria o campo aleatorio $Z(u, t)$ es estrictamente estacionario si su distribución de probabilidad es invariante ante traslaciones $(h, \tau) \in D \times T$, es decir, si dados dos vectores cualesquiera *RVs* $Z(u_1, t_1), \dots, Z(u_N, t_T)$ y $Z(u_1 + h, t_1 + \tau), \dots, Z(u_N + h, t_T + \tau)$ tienen la misma función de distribución multivariante, cualquier traslación del vector $(h, \tau) \in D \times T$, es decir, una estacionariedad débil o intrínseca. Una condición menos exigente que la estacionariedad estricta, es la estacionariedad de segundo orden o estacionario en sentido amplio si se cumple que $E[Z(u, t)] = m(u, t) = m$ (cte) $\forall (u, t) \in D \times T$ y la función de covarianza espacio-tiempo $C_z(u, t; u', t')$ se supone que depende sólo de los retardos de la distribución espacial y temporal $h = u - u'$ y $\tau = t - t'$, es decir, $E[Z(u, t) - m][Z(u', t') - m] = C_z(h; \tau)$.

Existen dos importantes puntos de vistas conceptuales para el modelado de distribuciones espacio-temporal a través de herramientas estadísticas para incluir una dimensión temporal. Primero, exige estudiarlo desde un punto de vista espacio temporal mediante un modelo único *RF* $Z(u, t)$, por lo general se descompone en dos componentes: una componente para modelar la tendencia mediante la media de la variabilidad espaciotemporal del proceso $Z(u, t)$, y un componente residual estacionario que modele la mayor frecuencia de las fluctuaciones en torno a la tendencia en el espacio y tiempo. Segundo, exige estudiarlo desde un punto de vista espacio temporal mediante un modelo que contenga múltiples vectores *RFs* o vectores de *TS*. Dos subclases de modelos que puedan analizar el espacio temporal del proceso. La primera subclase que trate el espaciotemporal *RF* $Z(u, t)$ como una colección de un finito número de T de correlación espacio tem-

poral RFs $Z(u)$, mientras que los modelos en la segunda subclase vea la RF $Z(u, t)$ como una colección de un finito numero N de correlación espacial TS $Z(t)$.

Modelos de covarianza espacio-temporal

Los primeros intentos en construir funciones de covarianza espacio temporal, hicieron uso de los modelos separables con suposiciones demasiados simplistas sobre la naturaleza de la variabilidad espacio-temporal, ya sea combinando la covarianza espacial y temporal de una manera de suma o producto. Esta construcción de modelos separables ignora las interacciones espacio-temporales, y afirman que la covarianza espacio-temporal presenta una completa independencia. Pocos procesos observados se comportan de esta manera, y un esfuerzo considerable se ha hecho en la búsqueda de alternativas de representaciones no separables.

El desarrollo de las funciones de covarianzas espacio temporal comenzó con los modelos métricos, donde la función de covarianza se expresaba en términos de una combinación lineal de los retardos espaciales y temporales. Posteriormente se propuso el modelo producto, expresando la covarianza espacio temporal como producto de las varianzas individuales espacial y temporal (Cressie and Huang 1999). Finalmente, se introdujo el modelo lineal, donde las covarianzas espacio temporal se expresan como suma de las varianzas tanto para el componente espacial como temporal. Cressie y Huang desarrollaron un conjunto de funciones de covarianzas espacio temporal no separable válido a través de la transformada de Fourier de la función de covarianza unidimensional (Cressie y Huang 1999). Gneiting desarrolló este enfoque para una representación de Fourier libre (Geniting 2001). Estos acontecimientos dieron un paso importante en la geostadística espacio-temporal, pero no fue hasta la contribución de De Iaco que esta forma de estacionariedad, de funciones covarianza no separables se generalizó para su aplicación (De Iaco, Myers y Posas 2003).

El modelo de covarianza producto-suma de De Iaco (2001) permite la combinación lineal de estructuras de covarianza arbitrariamente complejas (incluyendo zona de anisotropía y geométrica) en el espacio y tiempo, con una completa interacción. La representación del modelo producto-suma que incorpora Cressie y Huang ofrece nuevas familias de funciones de covarianza.

Modelo Lineal de corregionalización: variograma producto-suma espacio temporal

Estudiar un fenómeno donde involucre una distribución multivariante espacio-temporal, por ejemplo, varias variables podrían medirse en localizaciones o lugares dados y para diferentes momentos o intantes de tiempo. En años recientes muchos autores han venido estudiando el Modelo Lineal de Corregionalización (LCM) para un análisis geoestadístico multivariante en el contexto espacial, este modelo no se ha utilizado para procesos espacio-temporales. Pero el esfuerzo a sido encaminado a que el variograma marginal para aun modelo espacio-temporal pueda ser extendido al caso multivariante. En particular el LCM es extendido para aplicaciones espacio temporal, donde el variograma básico espacio temporal es modelado como un modelo producto-suma (De Iaco, Myers, y Posa, 2001). En este caso, cada uno de los variogramas en el MLC está escrito en términos de marginales en el espacio-tiempo y pueden ser fácilmente obtenidas. Sin embargo, un modelo espacio-temporal mas general con MLC, como el que integra el modelo producto y el producto-suma pueden ser usados.

Sea $(RF) \{Z(u, t), (u, t) \in D \times T\}$ una función aleatoria espacio-temporal estacionaria de segundo orden, donde D representa el espacio n -dimensional y T representa la dimensión del tiempo. La función $C(h, \tau)$ definida anteriormente debe ser definida positiva, a fin de que su covarianza sea válida, es decir, en el marco de la predicción krigada, el cálculo de la inversa de la matriz de covarianza, necesaria para la obtención del mejor predictor lineal e insesgado. Esto es, para cualquier $(r_1; q_1), \dots, (r_m; q_m)$, cualquier valor real a_1, \dots, a_m , y cualquier intero positivo m C debe satisfacer que $\sum_{i=1}^m \sum_{j=1}^m C(r_i - r_j; q_i - q_j) \geq 0$.

De Cesare (2001) introduce el desarrollo de modelos de covarianza producto-suma ver De Iaco, Myers, y Posa, 2001):

$$C(h, \tau) = k_1 C_x(h) C_t(\tau) + k_2 C_x(h) + k_3 C_t(\tau)$$

donde C_x y C_t son respectivamente los modelos de covarianza espacial y temporal validos. Así, la condición para que estas covarianzas sean válidas es $k_1 > 0, k_2 \geq 0$ y $k_3 \geq 0$.

En términos de semivariograma usando estacionariedad, la expresión anterior puede ser expresada por

$$\gamma(h, \tau) = (k_2 + k_1 C_t(0)) \gamma_x(h) + (k_3 + k_1 C_x(0)) \gamma_t(\tau) - k_1 \gamma(h) \gamma_t(\tau)$$

donde γ_x y γ_t son los modelos de semivariograma espacial y temporal, mientras que $C_x(0)$ y $C_t(0)$ son los valores *sill*. De Cesare (2001) definió las relaciones entre el espacio-temporal y los semivariogramas marginales como:

$$\begin{aligned}\gamma(h, 0) &= (k_2 + k_1 C_t(0)) \gamma_x(h) = k_x \gamma_x(h) \\ \gamma(0, \tau) &= (k_3 + k_1 C_x(0)) \gamma_t(\tau) = k_t \gamma_t(\tau)\end{aligned}$$

Combinando $\gamma(h, \tau)$, $\gamma(h, 0)$ y $\gamma(0, \tau)$ con $C(0, 0) = k_1 C_x(0) C_t(0) + k_2 C_x(0) + k_3 C_t(0)$ resolvemos para k_1 , k_2 y k_3 :

$$\begin{aligned}k_1 &= \frac{k_x C_x(0) + k_t C_t(0) - C(0, 0)}{C_x(0) C_t(0)} \\ k_2 &= \frac{C(0, 0) - k_t C_t(0)}{C_x(0)} \\ k_3 &= \frac{C(0, 0) - k_x C_x(0)}{C_t(0)} \\ k &= \frac{k_1}{k_x k_t} = \frac{k_x C_x(0) + k_t C_t(0)}{k_x C_x(0) \times k_t C_t(0)}\end{aligned}$$

donde $k_x C_x(0) k_t C_t(0)$ y $C(0, 0)$ son los valores *sill* de $\gamma(h, 0)$, $\gamma(0, t)$ y $\gamma(h, t)$. Según De Iaco, la condición de definición de positividad se garantiza cuando

$$0 < k \leq \frac{1}{\max\{\text{sill}(\gamma(h, 0)), \text{sill}(\gamma(0, \tau))\}} .$$

3 Planteamiento del problema geoestadístico

Se tienen las intensidades de homicidios en los 262 municipios de El Salvador del año 2003 y 2008 y se dispone de dos tipos de datos (coordenado y campo escalar) para darle solución al problema. Entonces, la configuración de los homicidios ocurridos en el país es la siguiente: cada hecho delictivo es georeferenciado a partir del punto coordenado (según la proyección geográfica que se esté utilizando), estas coordenadas espaciales se dan como (x, y) . Las intensidades serán el conjunto de esos eventos de homicidios ocurridos en todo el territorio de El Salvador, tomando como punto geográfico el centroide para cada municipio o lo que es lo mismo un dato de tipo campo escalar. En ese sentido, para predecir un valor no muestrado en la zona geográfica o municipio en particular, por ejemplo, el municipio con coordenada $x = -89.1254, y = 14.1542$. La ecuación

tomará la forma de $Z(s) = \mu + \epsilon(s)$ donde $s = (x, y)$ es la localidad, ejemplo, un homicidio ocurrido en el municipio “z” con ubicación geográfica $s = (-89.1254, 14.1542)$; y $Z(s)$ es el valor que toma la localidad, ejemplo, la intensidad de homicidios ocurridos en un municipio “z” igual a 50, su expresión queda así: $Z(-89.1254, 14.1542) = 50$.

El predictor toma la forma como una suma de pesos de las observaciones o datos, así: $\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i Z(s_i)$. Donde $Z(s_i)$ es el valor medido en el municipio i -ésimo, por ejemplo, $Z(-89.1254, 14.1542) = 50$, λ_i es el peso desconocido para el valor medido en el municipio i -ésimo, s_0 es la predicción del municipio, por ejemplo, $(-89.1142, 14.4587)$, y $N = 262$ el número de observaciones o valores medidos.

El problema es: Predecir espacialmente y espacio-temporalmente (en su caso) un valor muestreado o no observado en el área de estudio, por ejemplo, predecir una intensidad de homicidios en una región en donde no se ha observado ningún evento en la ocurrencia del homicidio.

4 Metodología y herramientas

Metodología

La metodología que se utilizó fue la siguiente:

Fase 1: Análisis Estructural Espacial y Espacio-Temporal: aquí se hizo uso del análisis de segundo orden para determinar la caracterización de la estructura del fenómeno del crimen en El Salvador. Esto mediante el análisis del vecino más próximo (considerando el dato ordenado), el estadístico K de Repley (considerando el tipo mixto de dato), el análisis de correlación con los Test de I de Moran para contrastar la hipótesis de aleatoriedad en el proceso estocástico del fenómeno del crimen, y finalmente, el estudio de la función de semi-variograma.

Fase 2: Predicción o Simulación Geoestadística Espacial y Espacio-Temporal: se realizó una interpolar del fenómeno del homicidio tanto espacial como espacio-temporal. Para la creación del escenario delictual se utilizó el estimador Cokriging en un contexto espacio-temporal. En lo que respecta al escenario delictual del homicidio para el año 2009 en El Salvador fue necesario realizar una simulación secuencial gaussiana geoestadística espacio-temporal considerando

el modelo lineal de corregrionalización basado en el modelo de covarianza no separable producto-suma De Cesare (2001).

Herramientas

En lo que respecta al análisis exploratorio de datos espacio-temporal, estimación de parámetros del semivariograma y generación del escenario delictual del homicidio se utilizaron: -lenguaje de programación R con paquetes maptools, spdep, rgdal, geoR, sp, foreing, RColorBrewer y classInt.

Por otro lado, para la transformación anamorfosis/tipo gaussiano (indispensable para la simulación secuencial gaussiana) se utilizaron los paquetes Gslib, Nscore, Backtr y gaussim; y para el escenario espacio-temporal considerando el enfoque cokriging se utilizó el paquete ARCGIS9.2.

5 Análisis de resultados

Descripción de datos y proyección geográfica

Para la aplicación se utilizaron las intensidades de homicidios ocurridos en El Salvador en el período de 2003 hasta 2008 de los 262 municipios. La intensidad fue georreferenciada mediante el punto centroide correspondiente a la forma geográfica de cada municipio. Esta información fue llevada al Sistema de Información Geográfico para darle su respectivo tratamiento de la información. Este procesamiento de información fue realizado por separado: uno dirigido a datos de tipo coordenados y otro de tipo geostadístico. En el primer tipo de datos no fue necesario el realizar una transformación de proyección geográfica, trabajándose de esta manera con la proyección NAD27 (Norte América Datum 1927). Para el segundo tipo de dato si fue necesario realizar una transformación de proyección geográfica, convirtiendo la proyección NAD27 a una proyección UTM (zona 16 Norte) con Datum WGS1984.

5.1 Distribución de los datos

En la Figura 1 se muestra la distribución de los datos de las intensidades de homicidios ocurridos en El Salvador, y tal como se observa, el promedio de homicidios durante los seis años fue de 13.96. Las mayores intensidades se observan en los municipios de San Salvador, ubicado en la zona central del país, Santa Ana, en la zona occidental, Sonsonate y Acajutla,

en la zona sur-occidente y San Miguel, ubicado en la zona oriental del país. Al analizar esta distribución territorial, el fenómeno presenta una variabilidad en cuanto a la ocurrencia del homicidio en municipios cuando se alejan de la mayor intensidad promedio de homicidios, provocando así una heterogeneidad geográfica. Esto se observa en los gráficos de distribución de tales intensidades vista desde una zona Este (East) y una zona Norte (North). En estos dos tipos de gráficos está bien evidenciada la concentración de homicidios territorialmente, es decir, los municipios que se encuentran dentro de la franja o paleta de color rojo, su intensidad es mayor en la ocurrencia del homicidio, mientras que los municipios que experimentan menor intensidad en la ocurrencia del homicidio su color es menos intenso. Se observa además el histograma de los datos, y efectivamente, su distribución se corre hacia la izquierda, generando de esta manera una distribución asimétrica en los datos.

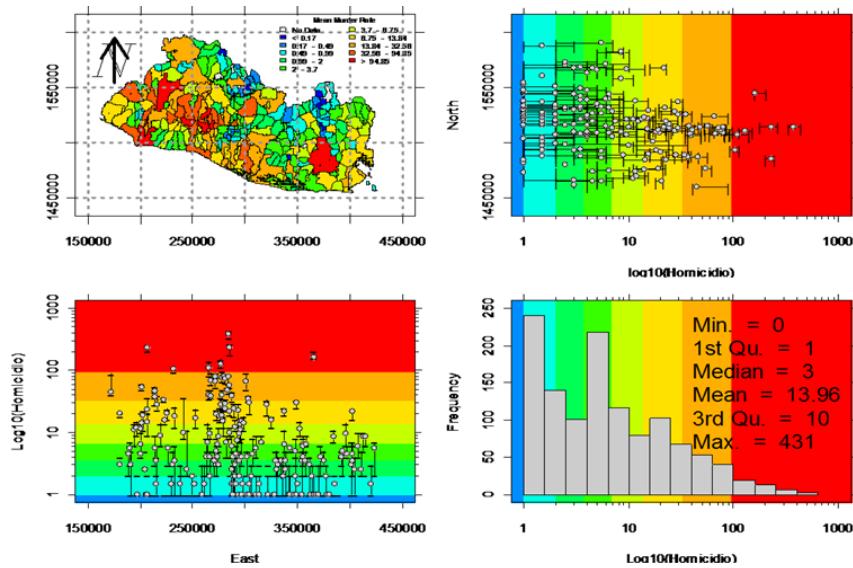


Figura 1: Distribucion de los datos.

5.2 Análisis estructural y correlación

Como el objetivo nuestro es generar un escenario delictual del homicidio en El Salvador, se necesita grandemente de un análisis estructural que pueda por medio de ello saber si el proceso estocástico espacio-temporal del crimen en el país presenta un patrón de tipo agregado y de dependen-

cia o autocorrelación espacio-temporal. Esto para tener evidencia de la existencia de grupos de municipios que presentan una baja, media y alta intensidad en la ocurrencia del homicidio.

En el **Cuadro 1** se muestra el índice del vecino más próximo (NNI) y el test de I de Moran. Se puede inferir a partir del análisis del vecino más próximo que el fenómeno del crimen en El Salvador presenta un patrón no aleatorizado, es decir, existen la evidencia de conformación de grupos. También, al analizar el índice de Moran con un nivel de significancia del 5% se rechaza la hipótesis nula (hipótesis de aleatoriedad), infiriendo de esta manera que hay presencia de dependencia o autocorrelación espacial en la distribución de homicidios en El Salvador. Mencionar además que se obtuvieron resultados similares en cuanto al estadístico de K de Repley, donde se rechazó de igual manera la hipótesis de aleatoriedad.

En la Figura 2 se muestra la clasificación espacial de la intensidad de homicidios ocurridos en El Salvador durante el año de 2003 hasta 2008 y el estadístico de autocorrelación de I de Moran visto desde una perspectiva espacio-temporal (gráfico Scatterplot). Tal como se observa, el indicador de local de autocorrelación espacial es significativo, ya que tiene un valor de 0.54, infiriendo de esta manera que el fenómeno presenta una dependencia espacial y con tendencia al agrupamiento. Al analizar el mapa de clasificación se mira que el grupo que presenta una mayor intensidad están distribuidos en el primer cuadrante del gráfico del estadístico de *I* de Moran, y así sucesivamente en el resto de cuadrantes del plano cartesiano.

Año	NN Index	Evidencia de Cluster	Índice de Moran	Dependencia Espacial
2003	—		0.057892	Sí
2004	0.347	Sí	0.06856	Sí
2005	0.314	Sí	0.081236	Sí
2006	0.347	Sí	0.077652	Sí
2007	0.332	Sí	0.075466	Sí
2008	0.367	Sí	0.075605	Sí

Tabla 1: Análisis del vecino más próximo y test de Moran.

5.3 Semiovriograma

La generación de un escenario delictual del homicidio en El Salvador para el año 2009 se realizó a partir de dos enfoques: un enfoque mediante el es-

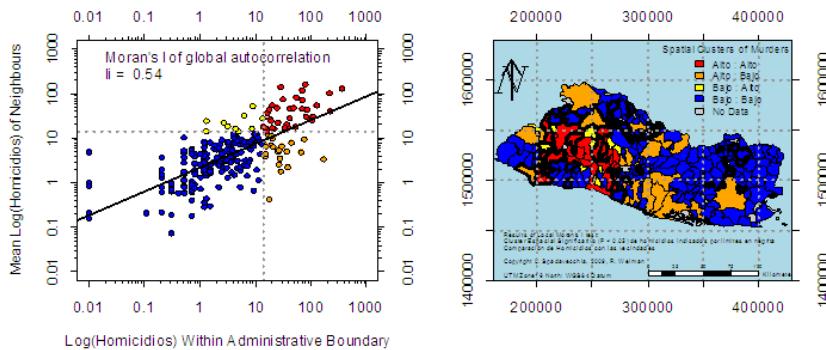


Figura 2: Clasificación espacial del homicidio en El Salvador.

timador Cokriging y otro mediante el Modelo Lineal de Corregionalización Producto-Suma.

En la Figura 3 se muestra el cálculo del semivariograma y los parámetros de los modelos teóricos: esférico, exponencial y gaussiano. En tal resultado se tiene que al analizar estos tres modelos, los más idóneos son el modelo exponencial y esférico, uno por la mayor variabilidad explicada en los datos de homicidios y otro por no presentar una discontinuidad o efecto pepita nula (ver parámetro de ecuación correspondiente). Según las estimaciones de la función de semivariograma, se observa que el modelo esférico no sobrepasa el 50% del valor de la meseta. Con esto, se llega a la conclusión que para estimar las intensidades de homicidios se utilizará cualquiera de los dos modelos, aunque el que se utilizará será el modelo esférico por no haber experimentado un efecto pepita nula.

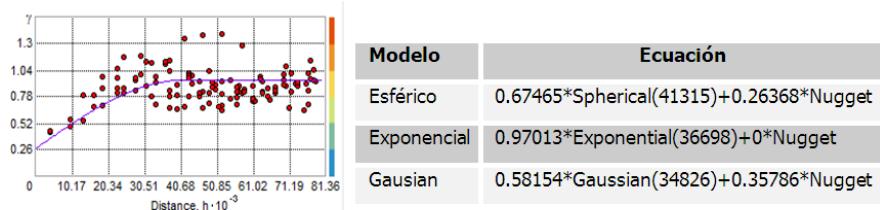


Figura 3: Semivariograma y modelos teóricos.

En la Figura 4 se presenta el cálculo de la función de covarianza no separable vía modelo lineal de corregionalización producto-suma. Se mues-

tra la estimación del modelo de covarianza válido, y tal como se observa en grafico de semivariograma, la separación espacial, su autocorrelación es evidente en cuanto comparamos las intensidades de homicidios a medida que la distancia se incrementa, mostrando de esta manera una estacionariedad espacial en los datos. El modelo que ajusta a los datos de homicidios es el exponencial. Por otro lado, se observa un efecto pepita o nugget para la parte temporal.

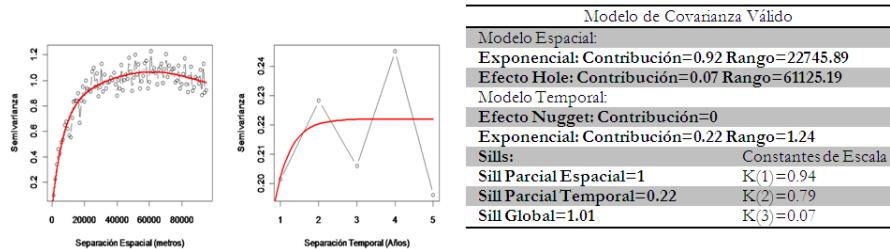


Figura 4: Semivariograma y modelo de covarianza no separable vía producto-suma.

5.4 Predicción y simulación geoestadística

La generación de un escenario delictual se realizó mediante el enfoque de Cokriging y una simulación geoestadística espacio-temporal (basado en una simulación secuencial gaussiana). En lo que respecta a la simulación se tuvo que realizar una transformación anamorfosis a la función aleatoria (considerada la intensidad de homicidios en El Salvador durante los seis años) como un requisito indispensable para este tipo de simulación vía modelo de covarianza no separable producto-suma (School of GeoSciences, Institute of Atmospheric and Environmental Science, University of Edinburgh, UK). De manera que las estimaciones o simulaciones obtenidas se tuvieron que transformar nuevamente para volverlas a su escala original utilizando la transformación inversa gaussiana.

En la Figura 5 se muestra el escenario delictivo de homicidio en El Salvador, al lado derecho se presenta el escenario de homicidios para el año 2009 (enfoque Cokriging) y al lado izquierdo el escenario de homicidios generado a partir del modelo lineal de correacionalización vía producto-suma (considerando el promedio de 100 simulaciones o realizaciones de las intensidades de homicidios). Tal como se observa en dichos mapas, el escenario muestra claramente las concentraciones de homicidios en las

zonas que se habían identificado en la clasificación espacial contrastado con el estadístico de I de Moran (ver figura 2). Son cinco concentraciones las que están distribuidas territorialmente en el país, donde la mayor incidencia está siendo más influenciada en la parte central y occidente del país, dejando una concentración en lo que respecta a la zona oriental. En la zona occidental los municipios más violentos son Acajutla, Sonsonate y Santa Ana. Por otro lado, en la zona central los municipios más violentos son San Salvador, Mejicanos, Ayutuxtepeque, San Marcos, San Martín, Soyapango, entre otros aledaños a ellos. Finalmente, la zona oriental los municipios que presentan una mayor intensidad en la ocurrencia del homicidio son las cabeceras departamentales de San Miguel y La Unión.

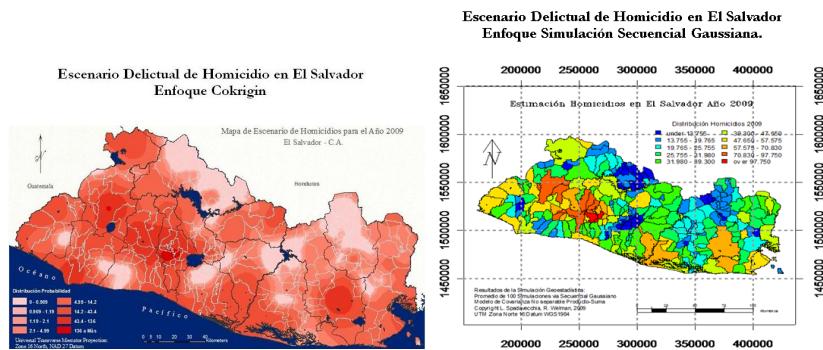


Figura 5: Escenario delictual de homicidios: enfoques Cokriging y MLC.

6 Conclusiones

El fenómeno del crimen en El Salvador en cuanto al análisis estructural se constató que presenta una estructura de tipo agregada y una dependencia espacial positiva muy fuerte. Se ha logrado identificar por medio del análisis de segundo orden qué zonas son las más violentas, esto mediante el análisis del estadístico de I de Moran como una herramienta para la clasificación espacio-temporal.

La predicción fue posible realizarla por el hecho de haber detectado una variabilidad espacial y espacio-temporal del crimen muy fuerte, esto mediante el gráfico de la función de semivariograma. Estimando así el fenómeno a partir del modelo teórico esférico en el enfoque Cokriging y una simulación secuencial gaussiana vía modelo de covarianza no separable producto-suma. Teniendo como resultado un escenario delictual del

homicidio para el año 2009.

Referencias

- [1] Anselin, L. (1995) “Local indicators of spatial association - LISA”, *Geographical Analysis* **27**(2): 93–115.
- [2] Castillejo, M. (2007) *Teoría Geoestadística aplicada al análisis de la Variabilidad Espacial Arqueológica Intra-Site*. Tesis Doctoral, Universitat Autònoma de Barcelona.
- [3] Cressie, N.; Wikle, C. (2002) “Space-time Kalman filter”, in: A. El-Shaarawi & W.W. Piegorsch (Eds.) *Encyclopedia of Environmetrics* , Volume 4: 2045–2049.
- [4] De la Cruz Rot, M. (2006) “Introducción al análisis de datos mapeados o algunas de las (muchas) cosas que puedo hacer si tengo coordenadas”, *Ecosistemas* **15**(3): 19–23,
- [5] Díaz, M.A. (2002) *Geoestadística Aplicada*. Instituto de Geofísica, UNAM, México, 131 págs.
- [6] Gneiting, T.; Genton, M.G.; Guttorp, P. (1999) “Geostatistical space-time models, stationarity, separability and full symmetry”, Technical Report No. 475, Department of Statistics, University of Washington, Seattle.
- [7] Huang, H.C.; Martínez, F.; Mateu, J.; Montes, F. (2007) “Model comparison and selection for stationary space-time models”, *Computational Statistics & Data Analysis* **51**(9): 4577–4596.
- [8] Berlanga, J.; Obregón, J. (1970) *Geoestadística*, apuntes: 44–115.
- [9] Kyriakidis, P.C.; Journel, A.G. (1999) “Geostatistical space-time models: a review”, *Mathematical Geology* **31**(6): 651–684.
- [10] De Iaco, S.; Myers, D.E.; Posa, D. (2003) “The linear coregionalization model y the product-sum space-time variogram”, *Mathematical Geology* **35**(1): 25–38