

**UNIVERSIDAD DE EL SALVADOR  
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA  
ESCUELA DE MATEMÁTICA  
MAESTRÍA EN ESTADÍSTICA**



**Universidad de El Salvador**

*Hacia la libertad por la cultura*

**“MODELO GEOESTADÍSTICO ESPACIO-TEMPORAL DEL CRIMEN EN EL SALVADOR: ANÁLISIS ESTRUCTURAL Y PREDICTIVO”**

**PRESENTADO POR:**

LIC. WELMAN DEL CARMEN ROSA ALVARADO

**TESIS DE POSGRADO PARA OPTAR AL TÍTULO DE:**

MAESTRO EN ESTADÍSTICA

**ASESOR EXTERNO:**

DR. ADOLFO HERNÁNDEZ ESTRADA.

**ASESOR LOCAL:**

DR. JOSÉ NERYS FUNES TORRES

JUNIO DE 2011

CIUDAD UNIVERSITARIA, SAN SALVADOR, EL SALVADOR, C.A.

**AUTORIDADES DE LA UNIVERSIDAD DE EL SALVADOR**

---

**RECTOR**

Máster. Rufino Antonio Quezada Sánchez

**VICE-RECTOR ACADÉMICO**

Máster. Miguel Ángel Pérez Ramos

**VICE-RECTOR ADMINISTRATIVO**

Máster Óscar Noé Navarrete

**SECRETARIO GENERAL**

Licenciado. Douglas Vladimir Alfaro Chávez.

**AUTORIDADES DE LA FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA**

---

**DECANO**

Dr. Rafael Antonio Gómez Escoto

**VICE-DECANO**

MSc. Marta Noemí Martínez Hernández

**SECRETARIO**

Licda. María Trinidad Trigueros de Castro

**ESCUELA DE MATEMÁTICA**

---

**DIRECTOR**

Msc. Carlos Mauricio Canjura

**COORDINADOR DE LA MAESTRIA EN ESTADÍSTICA**

Dr. José Nerys Funes Torres

**TRABAJO DE GRADUACIÓN:**

“MODELO GEOESTADÍSTICO ESPACIO-TEMPORAL DEL CRIMEN EN EL SALVADOR: ANÁLISIS ESTRUCTURAL Y PREDICTIVO”.

**PRESENTADO POR:**

WELMAN DEL CARMEN ROSA ALVARADO

**PARA OPTAR AL TITULO DE:**

MAESTRO EN ESTADÍSTICA.

**ASESOR:**

DR. ADOLFO HERNÁNDEZ ESTRADA: \_\_\_\_\_

DR. JOSÉ NERYS FUNES TORRES: \_\_\_\_\_

CIUDAD UNIVERSITARIA, JUNIO DE 2011.

## DEDICATORIA

A mi existencialismo...

*Welman Rosa.*

Contenido	Pág.
<b>Introducción</b> .....	1
<b>Capítulo I</b> .....	3
<b>Introducción A La Geoestadística Espacio-Temporal</b> .....	3
<b>1.1 Definición Geoestadística Clásica</b> .....	4
<b>1.2 Manejo de Proceso Espacial y Espacio-Temporal</b> .....	4
1.2.1 Análisis Estructural.....	4
1.2.2 Predictor Kriging en el contexto Espacio-Temporal.....	16
<b>1.3 Modelando una Regionalización Espacio Temporal</b> .....	18
1.3.1 Modelo Espacio Temporal de RF.....	20
1.3.2 Modelo Espacio Temporal Múltiple RF/TS.....	21
<b>1.4 Cokriging en el Modelo Espacio Temporal</b> .....	23
<b>1.5 Modelos de Covarianza Espacio Temporal</b> .....	25
1.5.1 Modelos de covarianza separable y no separables.....	26
1.5.2 Modelo Lineal de correogionalización: Variograma Producto-Suma Espacio Temporal.....	26
<b>Capítulo II:</b> .....	28
<b>Aplicación de la geoestadística espacial al crimen en el salvador</b> .....	28
<b>2.1 Descripción de Datos Espaciales</b> .....	28
<b>2.2 Análisis de Variabilidad Espacial del Crimen en El Salvador</b> .....	30
2.2.1 Análisis de Primer Orden: Análisis Exploratorio de Datos Espaciales.....	30
2.2.2 Análisis de Segundo Orden: Estructura y Correlación Espacial del Homicidio.....	35
<b>2.3 Predicción Espacial del Crimen: Kriging Disyutivo y Probabilístico</b> .....	47
<b>Capítulo III:</b> .....	52
<b>Aplicación De La Geoestadística Espacio-Temporal Al Crimen En El Salvador</b> .....	52
<b>3.1 Predicción Espacio-Temporal: Enfoque Cokriging</b> .....	53
3.1.1 Análisis Estructural.....	54
3.1.2 Predicción: Enfoque Cokriging.....	56
<b>3.2 Predicción Espacio-Temporal: Enfoque Modelo Lineal de Correogionalización-Producto-Suma</b> .....	58
3.2.1 El Modelo de Covarianza No-Separable Producto-Suma.....	58
3.2.2 Análisis Estructural Espacio-Temporal.....	63
3.2.3 Clasificación Espacial de Homicidios en El Salvador.....	66
3.2.4 Estimación de Parámetros de Modelos Espacio-Temporal.....	67
<b>Conclusiones</b> .....	80
<b>Referencias Bibliograficas</b> .....	82

## Introducción

En los últimos años la comunidad científica ha venido dándole un interés en la modelización geoestadística de fenómenos, construyendo nuevas familias de estructuras espacio-temporales válidos para darle un mayor tratamiento a muchas disciplinas como medioambiente, epidemiología, geología, medicina, geofísica, entre otros, en el estudio de la variabilidad espacial y espacio-temporal. Gran parte de este esfuerzo viene encaminado en la búsqueda de nuevas herramientas que permitan estudiar fenómenos donde incluya interacciones entre el espacio y tiempo para estudiar su variabilidad 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. También, 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 estacionarios en las que asumen isotropía en el espacio y tiempo.

Conceptualmente, 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).

Estudios recientes de la geoestadística han tocado a fondo la inclusión de dos componentes esenciales para la descripción de un proceso estocástico regionalizado y/o correogionalizado como lo son: el tiempo y la interacción espacio-tiempo. Esta inclusión ha sido el esfuerzo por parte de la comunidad científica para el manejo de este tipo de procesos, haciendo que hasta el momento la construcción de modelos covariográficos sea la nueva generación para el estudio geoestadístico espacio-temporal de cualquier fenómeno de la realidad. Esta trascendencia sobre el estudio de modelos covariográficos han surgido por el hecho de que algunas técnicas actuales que la geoestadística nos ofrece no son suficiente para darle un mejor tratamiento al fenómeno que se esté investigando. Kyriakidis y Journel (1999) propusieron tres puntos de vista para estudiar un proceso espacio-temporal: 1) un enfoque que involucre un modelo de función aleatoria simple integrado por los componentes espacio y tiempo; 2) un enfoque que involucre vectores de funciones aleatorias espaciales y 3) un enfoque que involucre vectores de series de tiempo. Sin embargo, estos tres enfoques no consideran una distribución multivariante espacio-temporal. Por ejemplo, cuando se desea estudiar en meteorología el pronóstico de las precipitaciones, presión, nubosidad y temperatura de una forma espacial y espacio-temporal. Es por ello, que el desarrollo de tales modelos van encaminados en la selección de modelos de covarianza espacio-temporal

separable y no separables (Phaedon C. Kyriakidis y André G. Journel., 1999) como el modelo lineal de correogionalización introducido por De Cesaré (2001).

En geoestadística existe un proceso importante y equivalente al que se presenta en estadística clásica para lograr predecir un fenómeno. Tal procedimiento está conformado por dos fases importantes: el análisis estructural y la predicción. En el análisis estructural el propósito es la caracterización de la estructura espacial y espacio-temporal de un fenómeno regionalizado. Es el proceso en el marco del cual conocemos si el proceso estocástico que se está analizando presenta un patrón de tipo aleatorio, agregado o uniforme. También, estima a partir de modelos matemáticos teóricos existentes en la literatura una función que refleje la correlación espacial y espacio-temporal de la variable regionalizada mediante la adopción razonada de la hipótesis más adecuada acerca de su variabilidad.

En la predicción lo que se pretende es interpolar el fenómeno ya regionalizado de una forma espacial y espacio-temporal, partiendo del modelo que fue identificado en el análisis estructural. Éste modelo por lo general se auxilia de la función de la función de covarianza o semivariograma, la cual, a pesar de ser una herramienta para describir el fenómeno, es capaz de predecir una variable medida en localizaciones situadas a distancias conocidas, pero que no han sido muestreadas.

Todos estos aspectos serán abordados en el presente trabajo de graduación que está estructurado en tres capítulos: introducción a la geoestadística espacio-temporal, aplicación de la geoestadística espacial al crimen en El Salvador y aplicación de la geoestadística espacio-temporal al crimen en El Salvador.

En el primer capítulo se presenta la definición de la geoestadística clásica, una introducción al manejo del proceso espacial y espacio-temporal como el análisis estructural y el predictor kriging. También, se describen brevemente algunos modelos de regionalización espacio-temporal como el modelo espacio-temporal de RF y múltiple RF/TS. Finalmente, una introducción a los nuevos modelos de covarianza espacio-temporal como los modelos de covarianza separable y no separable y el modelo lineal de correogionalización producto-suma.

En el segundo capítulo se desarrolla una aplicación de la geoestadística espacial al crimen en El Salvador. Disponiendo de una data de homicidios ocurridos en El Salvador desde el año 2003 hasta el 2008 por municipio. El objetivo de esta aplicación es analizar el fenómeno del crimen en El Salvador mediante un análisis estructural y predictivo de una forma espacial.

En el tercer capítulo se aplica la teoría de la geoestadística espacio-temporal al crimen en El Salvador. Se analiza el fenómeno mediante dos enfoques importantes para la generación de escenarios delictuales de los homicidios ocurridos en el país. El primer enfoque se considera el estimador lineal insesgado Cokriging. El segundo enfoque considera los modelos de covarianza no-separables como el producto-suma, donde este tipo de modelos a pesar de ser nuevos en la actualidad y dado que están todavía en debates por la comunidad científica, se realiza la aplicación considerando tres componentes como el espacio, el tiempo y sus interacciones espacio-temporal.

## CAPITULO I

### INTRODUCCIÓN A LA GEOESTADÍSTICA ESPACIO-TEMPORAL

#### **Introducción.**

En este capítulo se da una introducción a la geoestadística espacio-temporal como una herramienta para el estudio de fenómenos espaciales y espacio-temporales, que hoy en la actualidad están siendo abordados por la comunidad científica. El capítulo está estructurado en tres partes, una donde describe parte de la geoestadística espacial, otra donde enmarca la diversidad de métodos geoestadísticos espacio-temporal y la última que involucra el estudio de modelos de covarianza no-separables para el tratamiento espacio-temporal de un fenómeno regionalizado y/o corregionalizado. La primera, da a conocer las técnicas para el análisis estructural y predictivo de un fenómeno espacial, en el cual, para el análisis estructural se mencionan dos estadísticos primordiales para la caracterización de un proceso de puntos como el análisis del vecino más próximo, así como también la función de K de Ripley, cuyo objetivo es conocer si dicho proceso presenta un patrón de tipo aleatorio, uniforme y agregado. También, se analiza la parte de variabilidad espacial, considerando para tal caso, el estudio del test de I de Moran y de la función de covarianza o semivariograma. En el caso del semivariograma la variabilidad espacial es tratada para conocer si el proceso presenta una estacionariedad en cuanto a la separación espacial con que sucede tal fenómeno, para que luego de esto, poder caracterizarlo mediante una serie de modelos teóricos existentes en la literatura, los más comunes son: el esférico, exponencial y gaussiano. Modelos que describen el comportamiento de un proceso estocástico para permitir de esta manera, interpolar el fenómeno ya regionalizado con el estimador lineal insesgadokriging.

Por otro lado, en lo que respecta a la parte dos contenida en este capítulo, hay dos vertientes para el tratamiento de los fenómenos donde involucra además del componente espacial, el tiempo. En la primera vertiente menciona métodos espacio-temporales donde se exige que un fenómeno regionalizado espacio-temporal sea estudiado mediante un modelo único, que por lo general se descompone en dos componentes: una para modelar la tendencia mediante la media de la variabilidad espacio-temporal de la función aleatoria, y un componente residual estacionario que modele la mayor frecuencia de las fluctuaciones en torno a la tendencia en el espacio y tiempo. Este tipo de modelos son explicados por métodos deterministas y estocásticos, que por lo general en este capítulo se describe únicamente métodos estocásticos. Mientras tanto, la segunda vertiente tiene que ver con modelos donde el fenómeno se analiza a partir de funciones aleatorias múltiples RFs y un número finito  $n$  de funciones aleatorias correlacionadas espacialmente TS. Para modelar este tipo de métodos en la literatura geoestadística existen métodos o algoritmos que permiten describir este tipo de modelos regionalizados espacio-temporales, tal es el caso de la simulación goestadística secuencial gaussiana, donde a partir de los datos experimentales y del condicionamiento de un estimador como el kriging ordinario, poder simular una función aleatoria en otra virtual. Existe otra vertiente que tiene que ver con el estimador kriging simple, y es el caso del enfoque cokriging, cuyo propósito es interpolar un fenómeno regionalizado partiendo de la



correlación existente entre dos o más variables regionalizadas, considerando en tal caso, el tiempo.

Finalmente, se presenta la última parte, la cual tiene que ver con el estudio del modelo lineal de correogionalización producto-suma. Este tipo de modelos parte del estudio de funciones de covarianza separables y no-separables, donde en la actualidad se le ha dado mayor énfasis al tratamiento de fenómenos espacio-temporales con la función de covarianza no-separable vía producto-suma. Esto es, por el hecho de involucrar además de la componente espacial y temporal, otra como su interacción, de manera que el fenómeno puedan ser tratados no-separadamente para conocer la diferencia espacio-temporal entre el espacio y tiempo, así como también el espacio-temporal conjunto.

### **1.1 Definición Geoestadística Clásica.**

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).

### **1.2 Manejo de Proceso Espacial y Espacio-Temporal.**

Para el manejo de los procesos estocásticos espacio temporales es importante considerar dos aspectos importantes como el análisis estructural y la estimación para el análisis geoestadístico; el primero, se refiere al análisis estructural, el cual, consiste en modelos autorizados para el semivariograma para la identificación de modelos teóricos existentes en la teoría matemática en el modelado de fenómenos; el segundo, a la estimación o predicción mediante el estimador insesgado denominado Kriging (Berlanga y J. Obregon, 1970).

#### **1.2.1 Análisis Estructural.**

El análisis estructural es una de las partes más importantes de la geoestadística puesto que se encarga de la caracterización de la estructura espacial de una localidad o fenómeno regionalizado. Es el proceso en el marco del cual se obtiene un modelo geoestadístico para la función aleatoria que se estudia (Díaz, Martín. 20002).

Una manera más precisa de poder reflejar la estructura espacial de un fenómeno con naturaleza discreta es la representación cartográfica de todos los elementos del mismo en una región geográfica concreta. Para analizar esta estructura espacial es importante contrastar diversos supuestos que presentan tales procesos; es decir, si el análisis de dependencia espacial está dado mediante un análisis de correlación en base a diferentes herramientas como la covarianza o el semivariograma, es necesario antes de llegar a la conclusión de que el proceso presenta una autocorrelación espacial, analizar el proceso con técnicas estadísticas de distancias como el análisis del vecino más próximo y K-de-Ripley, el cual se analizan con más detalle en los siguientes

apartados. Este tipo de análisis de distancia es denominado como un análisis de primer orden y de segundo orden, y no hay que confundirlo con el análisis de estacionariedad de un proceso estacionario. Es más bien visto como un análisis de estructura del proceso espacial y de dependencia espacial.

### **Análisis del Vecino Más Próximo (NNA).**

El análisis del vecino más próximo (NNA) es una herramienta esencial para el análisis de distancia. El índice del vecino más próximo fue desarrollado por dos botánicos en el año de 1950 ( Clark y Evans, 1954) sobre todo para el trabajo en el terreno, pero que actualmente está siendo utilizado en diversos campos para una variedad amplia de problemas. Su objetivo es inferir por medio de él qué patrón espacial tendrá un proceso espacial, es decir, a partir de la hipótesis nula de aleatoriedad contrastar si el proceso presenta una estructura aleatoria, uniforme o agregada.

El índice del vecino más próximo es un cociente de dos medidas que compara la distancia entre los puntos más cercanos y la media distancia aleatoria. La primera media se expresa de la manera siguiente:

$$\text{Distancia del Vecino más Procimo} = d(NN) = \sum_{i=1}^N \frac{\text{Min}(d_{ij})}{N}$$

donde  $\text{Min}(d_{ij})$  es la distancia entre cada punto y su vecino más cercano y N es el número de puntos en la distribución. Así, se calcula la distancia de un solo punto a cada otro punto y la distancia más pequeña es seleccionada. Entonces, se toma el próximo punto y se calcula la distancia a todo el resto de los puntos (incluyendo el primer punto medido) con el vecino más próximo seleccionado y agrega la primera distancia mínima. Este proceso se repite hasta que todos los puntos hayan tenido su vecino más próximo seleccionado. La suma total de la distancia mínima es entonces dividida por N, el tamaño de la muestra de puntos, para producir así una media distancia mínima.

La segunda medida resumen es la distancia vecina más próxima prevista si la distribución de todos los puntos es completa y espacialmente aleatoria. Esta es la media distancia aleatoria (o la media aleatoria de la distancia más próxima). Está definida así:

$$\text{Media Distancia Aleatoria} = d(\text{ran}) = \frac{1}{2} \sqrt{\frac{A}{N}}$$

donde A es el área de la región y N es el número de eventos o incidentes. Puesto que A esta definida por el cuadrado de la unidad de medida (ejemplo, miles de cuadrados, miles de kilómetros, etc.), ésta produce una media aleatoria de la distancia en la misma unidad (es decir, miles, metros, etc.). Así tenemos que el índice del vecino más próximo es el cociente de estas dos medidas, de la distancia vecina más cercana observada y la media distancia aleatoria:

$$\text{Índice del Vecino más Próximo} = NNI = \frac{d(NN)}{d(ran)}$$

El resultado de este test determina la dependencia espacial de la distribución en base al valor del índice resultante; es decir, si el NNI es igual a 1, la población es aleatoria, un  $NNI > 1$  indica no aleatoriedad de la distribución con tendencia a la dispersión y si  $NNI < 1$  indica no aleatoriedad de la distribución con tendencia a la agregación.

El test de significancia del índice del vecino más próximo se hace mediante la prueba de una normal Z como el cociente de la diferencia de la distancia vecina más próxima observada y la media distancia aleatoria entre el error estándar de la media distancia aleatoria, así:

$$SE_{d(ran)} = \sqrt{\frac{(4 - \pi)A}{4\pi N^2}} \approx \frac{0.26136}{\sqrt{\frac{N^2}{A}}}$$

Con A siendo el área de la región y N el número de puntos.

#### **K-de-Ripley.**

Otro estadístico de análisis de distancia y esencialmente para el análisis de segundo es el llamado "K de Ripley (Ripley, 1976; Bailey y Gattrell, 1995). Se trata de 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. El supuesto con el que parte este estadístico se basa en que si un conjunto de puntos o eventos se distribuyen aleatoriamente, por ejemplo, desde una función Poisson, con una determinada densidad  $\lambda$ , el número esperado de puntos en un círculo de radio  $r$ , es igual a  $\lambda\pi r^2$ ; estimándose así la desviación respecto a la aleatoriedad mediante la función K de Ripley. Esta función K es una función de distribución acumulada, ya que a cada distancia  $r$ , todos los pares de puntos separados por una distancia menor que  $r$  se usan para estimar el valor de la función.

Como se ha mencionado anteriormente en el análisis del vecino más próximo sirve para contrastar si el proceso observado puede definirse como la realización de un proceso Poisson. Este test como el K de Ripley se basa de igual manera. Para ello, se construye una K teórica de tipo CSR (Completa Aleatoriedad Espacial) y se compara con la función  $K(r)$  de la función observada. El valor teórico esperado de la función K bajo el supuesto de CSR es igual a  $\pi r^2$ . De tal manera que:  $K(r) > \pi r^2$ , indica agregación,  $K(r) < \pi r^2$ , indica uniformidad y  $K(r) = \pi r^2$ , el proceso muestra aleatoriedad.

El test se deberá completarse con la construcción de intervalos de confianza, empleando para ello el método de Monte Carlo, que consiste en realizar una simulación de un número elevado de patrones de Poisson (con igual intensidad y en un área del mismo tamaño que el patrón observado). Este procedimiento es para contrarrestar el problema de que en las áreas de estudio

en que cuentan con límites definidos, se pueden producir variaciones estocásticas que causen valores mayores o menores que  $r^2$ , incluso para patrones de Poisson.

Ahora bien si A es el área de la región de interés, el número esperado de los acontecimientos en la región está dado por:

$$K(r) = \frac{1}{\lambda} E\{\text{número de eventos adicionales dentro de la distancia } h \text{ de un evento arbitrario}\}$$

Donde el estimador “naive” de  $K(r)$  es sólo el promedio empírico, dado por:

$$K(r) = \frac{1}{\hat{\lambda}N} \sum_{i=1}^N \sum_{i \neq j} I(d_{ij} \leq r)$$

Donde

$d_{ij}$  es igual a la distancia el  $i$ -ésimo y  $j$ -ésimo punto.

$$I(d_{ij} \leq r) = \begin{cases} 1 & \text{si } d_{ij} \leq r \\ 0 & \text{en otro caso} \end{cases}, y$$

$\hat{\lambda} = N/|A|$ , la intensidad estimada.

Para obtener una comprensión intuitiva de la función K de Repley, considera la idea de visitar un evento particular y la construcción de cada vez más círculos concéntricos de radio  $r$  en todo el evento. Dentro de cada círculo, se cuenta el número de eventos dentro de la distancia correspondiente de un evento particular. Esto debe repetirse para cada caso, con el número total de eventos acumulados dentro de una distancia determinada  $r$  calculada para muchos valores de distancia para da una estimación de  $K(r)$ .

Para tener en cuenta este problema, cada par de distancia en la función K de Repley empírica es ponderada por los pesos de la circunferencia con centro  $s_i$  y que pasa por  $s_j$  que cae dentro de la región de estudio A. Estos pesos están definidos por:

$$w_{ij} = \frac{a_{ij}}{2\pi d_{ij}}$$

Donde  $a_{ij}$  es la longitud del arco del círculo definido anteriormente clasificado en A. Con estos pesos, la función K de Repley cuenta los efectos de borde mediante el estimador siguiente:

$$K(r) = \frac{1}{\hat{\lambda}N} \sum_{i=1}^N \sum_{i \neq j} w_{ij}^{-1} I(d_{ij} \leq r)$$

## Análisis de correlación.

Los estadísticos definidos anteriormente son esenciales para el análisis estructural especialmente en lo que respecta al análisis de distancia de segundo orden. El análisis de correlación tiene que ver también con este tipo de análisis, con la ventaja que por medio de este análisis se conocerá si el proceso espacial presenta autocorrelación o dependencia espacial. Esta correlación espacial 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.

Existen varias técnicas de análisis de correlación espacial, dependiendo del tipo de variable y la dimensión de la variabilidad objeto de investigación. Las más utilizadas para el análisis de variables cuantitativas son aquellas que aplican diferentes test como: I de Moran, C de Geary, y la expresión gráfica conocida por Correlograma (Anselin, L., 1995). Y una especialmente a través del desarrollo de la Geoestadística como la técnica del semivariograma.

Para analizar la correlación espacial es importante conocer la matriz de pesos espacial, la cual es una generalización de la matriz de interacciones o contigüidades (conocida habitualmente como matriz de pesos, ponderaciones, retardos o contactos espaciales). Esta matriz de pesos generalizada ( $W$ ), en lugar de utilizar las ponderaciones binarias  $\delta_{ij}$  como cuantificación del concepto de vínculo, recoge el efecto de la región  $i$  sobre la región  $j$  a través de una ponderación  $w_{ij}$ , de forma que  $W = [w_{ij}]$ . Su interpretación es la siguiente: si  $w_{ij} = 0$  sería indicativo de ausencia de autocorrelación espacial entre las observaciones  $i, j$ ; si  $w_{ij} \neq 0$  sería indicativo de existencia de una interacción espacial entre las observaciones  $i, j$ , que podría ser expresada como simple contigüidad binaria (teniendo una frontera común), como contigüidad de distancia, a través de centroides, o como función inversa de la distancia simple o cuadrática.

En ese sentido, es como la matriz de ponderaciones espaciales hace posible la conexión entre el valor de una variable en un punto del espacio geográfico y las observaciones de dicha variable en otros puntos del área de estudio. En el contexto de series de tiempo o cronológicas, esto sería posible mediante un operador retardo que desfasara la variable uno o más períodos en el tiempo. Por ejemplo, la expresión:  $y_{t-s} = B^s y_t$  presenta la variable "y" desplazada "s" períodos en el pasado en términos del operador retardo B aplicado a dicha variable y.

Ahora bien, en el contexto espacial, las cosas no son tan sencillas, dadas las muchas direcciones en que el desfase puede tener lugar. Siguiendo el criterio de simple contigüidad binaria, el valor de  $y$  en  $(i, j)$ ,  $y_{ij}$  podría ser desplazado o retardado de varias maneras:

- Criterio de contigüidad torre, los desplazamientos espaciales (spatialshifts) serían:  
 $y_{i-1,j}; y_{i,j-1}; y_{i+1,j}; y_{i,j+1}$
- Criterio de contigüidad del alfil, los desplazamientos serían:  
 $y_{i-1,j-1}; y_{i+1,j-1}; y_{i+1,j+1}; y_{i-1,j+1}$

- Con el criterio de la reina, el número de posibles localizaciones consideradas como desplazamientos o “shifts” aumentarían hasta el total de 8.

En la tabla 1.1 se muestra el diagrama de los retardos espaciales en una red regular.

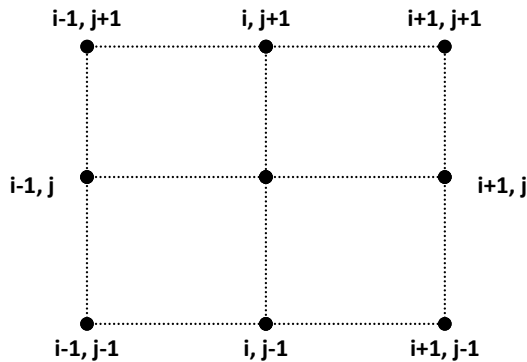


Tabla 1.1: Retardos espaciales en una red regular.

La mayoría de situaciones reales, existen regiones que por lo general presentan una forma geográfica no regular, por ello, lo descrito anteriormente no pueden utilizarse los criterios definidos, por el hecho de existir eventos en espacios geográficos irregulares. En tal sentido, esto puede resolverse mediante el concepto de retardo espacial que consiste en un promedio ponderado de los valores de una variable en las regiones vecinas, con ponderaciones fijas y dadas de forma exógena. Así el retardo espacial estaría conformado por la suma ponderada de todos los valores de una variable que pertenecen al mismo criterio de contigüidad (torre/alfil/reina), en vez de tomar individualmente cada uno de ellos. Esto se obtiene multiplicando las observaciones en cuestión ( $y_i$ ) por sus correspondientes pesos de la matriz de ponderaciones  $W$  del modo siguiente:

$$B^s y_t = \sum_j w_{ij} y_i \quad ; \quad \forall j \in J_i$$

Donde

$B^s$  es el operador retardo asociado con el criterio de contigüidad  $s$

$j$ : es el subíndice correspondiente al conjunto  $J_i$  de unidades relacionadas con  $i$ , según el criterio de contigüidad  $s$

$w_{ij}$  ponderaciones espaciales.

Lo anterior es importante para la elaboración de correlogramas espaciales, necesarios para la aplicación de los test de autocorrelación espacial que son analizados mediante el contraste de autocorrelación espacial global y el contraste de autocorrelación espacial local. A continuación se describen:

## Contraste de Autocorrelación Espacial Global.

El propósito es contrastar la hipótesis de que una variable se encuentre distribuida de forma totalmente aleatoria en un sistema espacial o, si por el contrario, existe algún tipo de asociación significativa de valores similares o distintos entre regiones vecinas. Los estadísticos que realizan este tipo de contraste son: el I de Moran (1948) y C de Geary (1954).

### Test I de Moran

Este test fue formulado como una función de una variable ( $y$ ), considerada en los puntos del espacio  $(i, j)$ , en desviaciones a la media, y los elementos de la matriz binaria de interacciones espaciales  $[\delta_{ij}]$ . Esta expresión inicial de Moran podría ser generalizada, sustituyendo la matriz de interacciones por la más general matriz de pesos espaciales  $w_{ij}$ , de la siguiente manera:

$$I = \frac{N \sum_{(2)} (y_i - \bar{y})(y_j - \bar{y})}{S_0 \sum_{i=1}^N (y_i - \bar{y})^2}$$

Donde

$w_{ij}$ : es el elemento de la matriz de pesos espaciales correspondiente al par  $(i, j)$ .

$S_0 = \sum_i \sum_j w_{ij} = \sum_{(2)} w_{ij}$ , es decir, la suma de los pesos espaciales.

$\bar{y}$ : es el valor medio o esperado de la variable  $y$

$N$ : número de observaciones o tamaño muestral.

Respecto a la distribución del contraste  $I$ , cuando el tamaño muestral es suficientemente amplio, la expresión estandarizada del test  $I$  se distribuye como una normal tipificada,  $N(0,1)$ . Es así como el proceso inferencial utiliza valores estandarizados ( $z$ ) de cada uno de ellos, obtenidos, a través del cociente entre la diferencia del valor inicial y la media teórica, y la desviación típica teórica, como:

$$z_I = \frac{I - E[I]}{SD[I]}$$

Donde

$E[I]$ : es la media teórica del estadístico  $I$ .

$SD[I]$ : desviación típica del estadístico  $I$ .

### Test de C de Geary

Este test es parecido al test  $I$  de Moran, ya que es función de la variable ( $y$ ) y los elementos de la matriz de interacciones espaciales  $[\delta_{ij}]$ , puede ser también generalizado sustituyendo esta matriz por la más general de pesos espaciales  $w_{ij}$ , de la siguiente manera:

$$c = \frac{(N - 1) \sum_{(2)} w_{ij} (y_i - \bar{y})^2}{2S_0 \sum_{i=1}^N (y_i - \bar{y})^2}$$

Considera el cuadrado de las diferencias entre los valores de (y) en dos puntos (i,j), a diferencia del contraste I. También, en este caso, puede afirmarse que el valor estandarizado de c ( $z_c$ ) se distribuye como una normal tipificada, N(0,1) asintótica, lo cual, se obtiene a partir de la expresión de los momentos de c de la manera siguiente:

$$z_c = \frac{c - E[c]}{SD[c]}$$

Donde

$E[c]$ : es la media teórica del estadístico c.

$SD[c]$ : desviación típica del estadístico c.

En la tabla 1.2 se muestra la interpretación de los valores estandarizados de los estadísticos de autocorrelación espacial global.

Test	Hipótesisnula	Hipótesisalternativa (z significativo)	
	(z no significativo)	z > 0	z < 0
<b>I de Moran</b>	No autocorrelación espacial	Autocorrelaciónespacial POSITIVA	Autocorrelaciónespacial NEGATIVA
<b>c de Geary</b>	No autocorrelación espacial	Autocorrelaciónespacial NEGATIVA	Autocorrelaciónespacial POSITIVA

Tabla 1.2: Interpretación de los valores estandarizados para los test de I de Moran y C de Geary.



### Contraste de autocorrelación espacial local.

El contraste de autocorrelación espacial global no es capaz de detectar la inestabilidad espacial de ciertas estructuras locales de asociación. Esto es así porque los test están centrados en el análisis de dependencia general propia de todas las unidades del espacio geográfico. En ese sentido, es necesario definir un contraste de autocorrelación espacial local que permita definir así concentraciones. Para ello, es importante plantearse dos puntos de vista sobre el problema de la dependencia espacial local como:

- Existe la posibilidad de que, en un espacio dado, no se detecte la presencia de autocorrelación espacial global en el proceso espacial de cualquier fenómeno aunque, de hecho, existan pequeños “clusters” espaciales en los que el proceso espacial experimenta una concentración importante.
- Existe también la posibilidad de que, habiéndose detectado dependencia a nivel global en un proceso espacial, no todas las regiones del espacio considerado contribuyan con igual peso en el indicador global, es decir, que coexistan unas zonas en las que el proceso se distribuya de forma aleatoria con una importante contribución a la dependencia existente.

Dados estos dos puntos de vista para dar respuesta al problema de la dependencia espacial es importante definir dos grupos de contrastes de asociación local que indiquen hasta qué punto una región se encuentra rodeada por otras con valores altos o bajos de una variable determinada: el primer grupo, la familia de estadísticos  $G_i$  de Getis y Ord; segundo, los indicadores locales de asociación espacial (LISA) (Anselin, L., 1995). Ambos tipos de contrastes parten de la hipótesis nula de ausencia de autocorrelación espacial global, aunque los indicadores LISA serán también capaces de responder a la segunda cuestión planteada anteriormente, demostrando la presencia de regiones con una participación en el estadístico global muy superior a la media (“outliers”). Para nuestro propósito se estudiarán los indicadores LISA como el local de Moran que a continuación se describe:

### Test I de Moran Local

El indicador I de Moran Local puede definirse como un indicador de dependencia local basado en el test I de Moran,  $I_i$ , de la siguiente manera:

$$I = \frac{N \sum_{(2)} z_i z_j}{S_0 \sum_{i=1}^N z_i^2}$$

Siendo  $\sum_{(2)} = \sum_i \sum_j$

De este modo, se cumple que el sumatorio de valores del estadístico local  $I_i$  es equivalente al test global I, siendo el factor de proporcionalidad  $\gamma = S_0 m_2$

Donde

$m_2 = \sum_i \frac{z_i^2}{N}$ , el momento de segundo orden la variable  $z_i$ .

Por la equivalencia de este indicador local con el global, el test I local de Moran podría también expresarse de la siguiente manera:

$$I = \frac{z_i}{m_2} \sum_j w_{ij} z_j$$

En la tabla 1.3 se muestra la interpretación de los valores estandarizados de los estadísticos de autocorrelación espacial local.

Contraste de Autocorrelación Local		Característica de la distribución	
Z(Local Moran, $I_i$ )	No rechazo $H_0$ $H_0$ : no autocorrelación espacial	No existencia de autocorrelación espacial	
	Rechazo $H_0$ $H_0$ : no autocorrelación espacial	$Z(I_i) > 0$	Existencia de un cluster alrededor de la región $i$ donde se concentran valores similares y elevados de la variable $x$ .
		$Z(I_i) < 0$	Existencia de un cluster de valores disimiles de la variable $x$ , donde la región $i$ concentra valores elevados de $x$ mientras que sus regiones vecinas muestran la situación contraria.

Tabla 1.3: Interpretación de los valores estandarizados para los test de I de Moran Local.

Con los indicadores descritos anteriormente se ha llegado al nivel de poder realizar un análisis de distancia de primer y segundo orden para conocer la estructura y correlación espacial del proceso. Ahora lo que resta es estudiar la forma en que puede ser ajustado el proceso, es decir, luego de haber inferido en el rechazo de la hipótesis nula de aleatoriedad en el proceso, conocer qué modelo teórico se ajusta a la distribución de los datos espaciales. En ese sentido, se abordará mediante la herramienta denominada 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 la distancia entre aquellos.

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 una distancia  $h$ . La función  $\gamma(h)$  se denomina semivariograma y su expresión es la siguiente:

$$\gamma(h) = \frac{1}{2N_p(h)} \sum_{i=1}^{N_p(h)} [Z(x_i) - Z(x_i + h)]^2$$

Donde

$N_p(h)$  es el número de pares a la distancia  $h$ .

$h$  es el incremento.

$Z(x_i)$  son los valores experimentales.

$x_i$  localizaciones donde son medidos los valores  $Z(x_i)$

Tal como se observa en la figura 1.1 se muestra el modelo gráfico de la función semivariograma tiene una serie de componentes que permiten describir ciertos aspectos de la variabilidad espacial de la población objeto de estudio como la meseta/umbral, efecto pepita y rango. Además de mostrar los tres modelos teóricos ajustados por su curva correspondiente (los más conocidos son: esférico, exponencial y gaussiano).

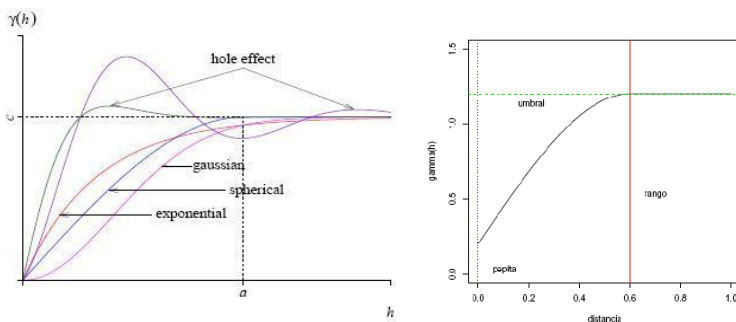


Figura 1.1: Representación Gráfica de algunos modelos de semivariograma y sus componentes.

El primer parámetro es el efecto pepita, el cual representa la discontinuidad puntual del semivariograma en el origen. El segundo es la meseta, que representa el límite del semivariograma cuando la distancia  $h$  tiende al infinito, generalmente los semivariogramas que tienen meseta finita cumplen con la hipótesis de estacionariedad fuerte, mientras que cuando ocurre lo contrario, el semivariograma define un fenómeno natural que cumple sólo con la hipótesis intrínseca. Si se interpreta la pepita como un error en las mediciones, esto explica porque se

sugiere que en un modelo que explique bien la realidad, “la pepita no debe representar más del 50% de la meseta. Finalmente el tercero, que corresponde al rango, es decir, a la distancia a partir de la cual dos observaciones son independientes, y se interpreta como la zona de influencia. Entre más pequeño sea el rango, más cerca se está del modelo de independencia espacial. El rango no siempre aparece de manera explícita en la fórmula del semivariograma (Díaz, Martín. 2002).

Teniendo en cuenta el significado que tienen los parámetros, se describe a continuación de una forma breve los modelos acotados para el análisis estructural.

**a) Modelo Lineal con Meseta:** En una dimensión los bloques son simplemente líneas. Supongamos que estos son de longitud  $a$  y la distancia entre sus centros es  $h$ . Sus solapamientos es entonces  $a - h$ , el cual, cuando se expresa como una proporción de  $a$  es  $1 - (h/a)$ , siendo  $h \leq a$ .

El semivariograma se expresa como:

$$\gamma(h) = \begin{cases} s(h/a) & \text{para } 0 \leq h \leq a \\ S & \text{para } h \geq a \end{cases}$$

**b) Modelo Circular:** En dos dimensiones, los bloques son discos de diámetro  $a$ . El área de intersección de éstos, separados a una distancia  $h$  de sus centros es:  $A = \frac{a^2}{2} \cos^{-1} \left( \frac{h}{a} \right) - \frac{h}{2} \sqrt{a^2 - h^2}$ ; para  $h \leq a$ . Su semivariograma es:

$$\gamma(h) = \begin{cases} S \left\{ 1 - \frac{2}{\pi} \cos^{-1} \left( \frac{h}{a} \right) - \frac{2h}{\pi a} \sqrt{1 - \left( \frac{h}{a} \right)^2} \right\} & \text{para } 0 \leq h \leq a \\ S & \text{para } h \geq a \end{cases}$$

$$\text{gradiente} = 4S/\pi a$$

Este modelo no ha encontrado mucha aplicación en las ciencias de la tierra.

**c) Modelo Esférico:** Por analogía podemos derivar este modelo considerando el solapamiento de los volúmenes de dos esferas de diámetro  $a$  y  $h$  la distancia que separa sus centros. El volumen de intersección es:  $V = \frac{\pi 2a^3}{4} - a^2 h + \frac{h^3}{3}$ ; para  $h \leq a$ . Su semivariograma es:

$$\gamma(h) = \begin{cases} \frac{S}{2} \left\{ 3 \left( \frac{h}{a} \right) - \left( \frac{h}{a} \right)^3 \right\} & \text{para } 0 \leq h \leq a \\ S & \text{para } h > a \end{cases}$$

$$\text{gradiente} = 3S/2a$$

**d) Modelo Exponencial:** Si el solapamiento de los bloques varía su tamaño de forma aleatoria, entonces el semivariograma resulta exponencial. En el caso isotrópico el semivariograma es:

$$\gamma(h) = S \left( 1 - e^{-\frac{h}{r}} \right) \text{ para } h \geq 0, \text{ gradiente} = S/r$$

Se considera como rango efectivo  $a = 3r$

Los procesos autorregresivos de primer orden y de Markov producen modelos exponenciales.

**e) Modelo Gaussiano:** Al igual que en el modelo exponencial, la dependencia espacial se desvanece solo en una distancia que tiende a infinito. El principal distintivo de este modelo es su forma parabólica cerca al origen.

$$\gamma(h) = S \left( 1 - e^{-\left(\frac{h}{r}\right)^2} \right) \text{ para } h \geq 0$$

Donde  $r$  es un parámetro no lineal que determina la escala espacial de la variación, como en el caso exponencial. El rango efectivo se considera  $a = \sqrt{3}r$ , que corresponde al valor  $0.95S$  del variograma.

### 1.2.2 Predictor Kriging en el contexto 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 (Ibáñez, 2003). Estos métodos de Kriging han sido ampliamente estudiados y aplicados en el contexto de la estadística espacial, y espacio temporal (Shekhar Srinivasan, 2003).

Es por ello que se hace necesario definir procesos espacio temporales que puedan ser caracterizados mediante funciones aleatorias  $Z(x, t)$  que varían en el espacio, es decir, dependen de la posición  $x \in R^3$  y dependen del tiempo  $t$ . Dicho de otra manera, a cada punto del espacio-tiempo  $(x, t)$  le hacemos corresponder una variable aleatoria  $Z$ .

Según la forma del estimador kriging se puede clasificar en tres grandes grupos. El primer grupo se divide en lineales (Simple, Ordinario, Universal y Residual) y no lineales (Disyuntivo, Indicador y Probabilístico); el segundo grupo, su soporte de medición puede ser puntual y en bloques; y el tercer grupo, paramétrico (multigaussiano, Disyuntivo y Lognormal) y no paramétrico (Simple, Ordinario, Universal y Residual) (Martín A. Díaz, 1969).

Se describe a continuación de una forma resumida el primer y tercer grupo, como sigue:

**a) Kriging Lineal con valores esperados conocidos: Kriging Simple.**

Considera la media conocida y constante a través de todo el dominio espacio temporal. Es decir que  $m(x, t) = m \forall (u, t) \in R^2 \times R^1$ . En este caso el estimador será de la forma:

$$\hat{Z}(x_0, t_0) = \sum_{i=1}^n \alpha_i Z(x_i, t_i) + \left(1 - \sum_{i=1}^n \alpha_i\right) m$$

Este estimador siempre será insesgado (Ibáñez, 2003), por lo tanto no necesitaremos imponer la restricción de que  $\sum_{i=1}^n \alpha_i = 1$ .

**b) Kriging lineal con valor esperado estacionario pero desconocido: Kriging Ordinario.**

Este estimador se utiliza cuando desconocemos el valor de la media, sabemos que no es constante en todo el área de estudio, pero podemos considerarla constante localmente, es decir, considerarla constante ( $= m_0$ ) en una vecindad centrada en el punto en el que queremos predecir. El dominio de estacionariedad de la media se restringe en este caso a una vecindad local, centrada en la localización  $(x_0, t_0)$  en la que queremos predecir.

Sea  $N(x_0, t_0)$  el conjunto de vecinos a  $(x_0, t_0)$  en el que consideramos la media constante. Si suponemos que  $N(x_0, t_0) = \{(x_i, t_i) : n_0 \leq i \leq n_1\}$ , el predictor kriging (1) quedará:

$$\hat{Z}(x_0, t_0) = m_0 + \sum_{i=n_0}^{n_1} \alpha_i (Z(x_i, t_i) - m_0) = \sum_{i=n_0}^{n_1} \alpha_i Z(x_i, t_i) + \left(1 - \sum_{i=1}^n \alpha_i\right) m_0$$

(Ibáñez, 2003), para que el estimador sea insesgado exigiremos que  $\sum_{i=1}^n \alpha_i = 1$ , y entonces el estimador kriging quedará:

$$\hat{Z}(x_0, t_0) = \sum_{i=n_0}^{n_1} \alpha_i Z(x_i, t_i)$$

Por lo tanto, el único objetivo es la predicción, no necesitaremos estimar la media del proceso.

**c) Kriging lineal en presencia de tendencia: Kriging Universal.**

Considera la media desconocida, y permite que varíe dentro de cada vecindad situada en el área de estudio.

La tendencia se modeliza como una combinación de funciones  $f_k(x, t)$  de las coordenadas, o de variables explicativas asociadas con los datos en  $(x, t)$ :  $m(x_0, t_0) = \sum_{k=0}^K B_k f_k(x_0, t_0)$ , siendo  $f_k(\cdot, \cdot)$  funciones conocidas, y  $B_k$  nuevas constantes a estimar.

En este caso, el estimador kriging será de la forma:

$$\hat{Z}(x_0, t_0) = \sum_{i=1}^n \alpha_i Z(x_i, t_i) + \sum_{k=0}^K B_k \left( f_k(x_0, t_0) - \sum_{i=1}^n \alpha_i f_k(x_i, t_i) \right)$$

(Ibáñez, 2003), al trabajar con Kriging universal, vamos a buscar un predictor uniformemente insesgado, es decir, vamos a exigir que:  $\sum_{i=1}^n \alpha_i f_k(x_i, t_i) = f_k(x_0, t_0)$  ya que entonces podremos expresar el estimador kriging, como función únicamente de los datos, es decir:

$$\hat{\gamma}(x_0, t_0) = \sum_{i=1}^n \alpha_i Z(x_i, t_i)$$

### 1.3 Modelando una Regionalización Espacio Temporal.

Existen un gran número de fenómenos cuyo estudio de manera natural requiere de un marco espacio-temporal. En ese sentido, por ejemplo, de las variables meteorológica o la evolución de un contaminante en un acuífero, etc. Históricamente, se han desarrollado dos vertientes principales usando un enfoque estadístico, una que se encarga de la modelación de fenómenos temporales, que son las series de tiempo o cronológicas, y la otra estudia los procesos que varían sólo en el espacio, de la cual se ha encargado la estadística espacial y en particular la geoestadística.

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. Por ejemplo, los homicidios registrados en ciertas localidades de El Salvador en  $t$  instantes de tiempo (días, semanas, meses y años). La variable aleatoria RV  $Z(u, t)$  está plenamente caracterizada por su función de distribución acumulada (fda) que da la probabilidad de que la variable  $Z$  en una localidad  $u$  en el espacio y un instante de tiempo de  $t$  no sea mayor a cualquier umbral  $z$ .

$$F(u, t, z) = Prob\{Z(u, t) \leq z\}, \forall z, (u, t) \in D \times T$$

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.

Consideremos una discretización de tres dimensiones (3D) espacio-tiempo con dominio  $DxT$ , dentro de un conjunto de  $N \times T$  puntos, donde  $N$  y  $T$  son el número de puntos en el espacio y dominio temporal, respectivamente.

Para cualquier conjunto de  $NT$  puntos en  $DxT$  corresponde al vector de  $NT$  RVs  $\{Z(u_1, t_1), \dots, Z(u_N, t_N)\}$ , que se caracteriza por la variable  $NT$  o  $NT$ -puntos fda:

$$F(u_1, t_1, \dots, u_N, t_N; z_{11}, \dots, z_{NT}) = Prob\{Z(u_1, t_1) \leq z_{11}, \dots, Z(u_N, t_N) \leq z_{NT}\}$$

El conjunto de todas esas  $NT$  Variables de la función de distribución acumulativa, para cualquier entero positivo  $N, T$  y para cualquier elección de espacio-tiempo de lugares, constituye el espacio temporal de la RF  $Z(u, t)$ .

La inferencia de esta distribución espacio temporal requiere de repetidas realizaciones de la componente RVs en cada lugar del espacio-tiempo  $(u, t) \in DxT$ , el cual, nunca se dispone en la práctica. El paradigma clásico de la inferencia consiste en poner en común los pares de mediciones separadas por el mismo espacio-tiempo  $(h, \tau) \in DxT$ , con  $h = u - u'$  y  $\tau = t - t'$ , y utilizar estos pares como un conjunto de repeticiones. Esta relación corresponde al modelo de decisión (no una hipótesis) de dos puntos espacio-tiempo de estacionariedad.

Un campo aleatorio  $Z(u, t)$  es estrictamente estacionario si su distribución de probabilidad es invariante ante traslaciones  $(h, \tau) \in DxT$ , 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  $(h, \tau) \in DxT$ , es decir:

$$F(u_1, t_1, \dots, u_N, t_T; z_{11}, \dots, z_{NT}) = F(u_1 + h, t_1 + \tau, \dots, u_N + h, t_T + \tau; z_{11}, \dots, z_{NT}) \\ \forall u_1, t_1, \dots, u_N, t_T, \text{ and } (h, \tau) \in (DxT)$$

Una condición menos exigente que la estacionariedad estricta, es la estacionariedad de segundo orden.

Un proceso aleatorio espacio-temporal  $Z(u, t)$  es estacionario de segundo orden, o estacionario en sentido amplio si se cumple que:

$$E\{Z(u, t)\} = m(u, t) = m(cte) \forall (u, t) \in DxT$$

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,

$$E\{[Z(u, t) - m][Z(u', t') - m]\} = C_z(h; \tau)$$



En muchos casos, sin embargo, los procesos naturales que actúan sobre una gran región y/o espacio de tiempo están mejor representados por la no estacionariedad espaciotemporal de los modelos RF. La no estacionariedad normalmente se limita a la componente media (tendencia), la cual es dependiente de la ubicación espacial  $u$ , o el instante de tiempo  $t$ , o ambos. Por ejemplo, en el fenómeno del crimen, la realización del proceso estocástico se volverá no estacionario a medida que el fenómeno vaya a la tendencia, de manera que no se estaría cumpliendo el supuesto de estacionariedad; también, al tomar en cuenta la variabilidad existente en el fenómeno, dejará de ser constante cuando no dependa de la distancia entre las localidades. Generando así un proceso estocástico no estacionario.

Existen dos importantes puntos de vistas conceptuales para el modelado de distribuciones espacio temporal a través de herramientas estadísticas para incluir una dimensión temporal, y que se examinan a continuación:

- a) Primero exige estudiarlo desde un punto de vista espaciotemporal 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.
- b) 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 temporal 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  $Z(t)$ .

### 1.3.1 Modelo Espacio Temporal de RF.

Se considera un solo modelo espacio temporal RF  $Z(u, t)$ , por lo general cuando la función aleatoria espacio temporal RF se descompone en dos componentes. Un componente media  $m(u, t)$  para modelar la tendencia, que es un promedio de variabilidad de  $Z(u, t)$ , y un componente residual  $R(u, t)$  modelando en torno a las fluctuaciones de la tendencia en el espacio y el tiempo. Las dos clases de modelos se pueden definir dependiendo de si la media se considera como determinística o estocástica. Para nuestro propósito describiremos únicamente la clase de modelos estocásticos.

La media  $M(u, t)$  es estocástica, la descomposición de  $Z(u, t)$  se escribe como:

$$Z(u, t) = M(u, t) + R(u, t) \forall (u, t) \in D \times T$$

Donde  $M(u, t)$  es un espacio temporal de RF no estacionario modelando la distribución espacio y tiempo de la media del proceso, con  $E\{M(u, t)\} = m(u, t)$ , y  $R(u, t)$  es media cero estacionario de RF modelando en el espacio y tiempo en torno a las fluctuaciones de  $M(u, t)$ . Por lo general  $R(u, t)$  sigue un modelo independiente de (o al menos no correlacionado)  $M(u, t), \forall(u, t) \in DxT$ .

Los modelos de tendencia como los de tipo determinística se consideran aleatorios si los coeficientes  $b_{kl}$  que multiplican las funciones deterministas  $f_{kl}(u, t)$  son modelos que se basan en variables aleatorias, es decir,

$$M(u, t) = \sum_{l=0}^L \sum_{k=0}^K B_{kl} f_{kl}(u, t), \forall(u, t) \in DxT$$

Con

$$E[B_{kl}] = b_{kl}, \quad k = 0, \dots, K, l = 0, \dots, L$$

Cuando el valor esperado de los coeficientes aleatorios  $B_{kl}$ , resulta ser el valor de la tendencia aleatoria  $M(u, t)$ , puede ser estimado de los z-datos a través de el procedimiento de regresión generalizada, similar a kriging simple.

Los modelos donde participa la componente media estocástica son generalmente expresados como:

$$Z(u, t) = M(u, t) + R(u, t) = M_u(t) + R(u, t), \quad \forall(u, t) \in DxT \text{ (Phaedon C. Kyriakidis, 1999)}$$

Una tendencia temporal determinística  $m_u(t)$  en la localidad  $u_\alpha$  es en primer lugar una función de t ajustada mediante el procedimiento de mínimos cuadrados, independientemente de otros lugares o localidades, utilizando funciones básicas lineales o trigonométricas de t como en la expresión de la tendencia determinística  $m(u, t)$ . La estimación de los parámetros, que es el estimador  $\hat{B}_{kl}, k = 0, \dots, K, l = 0, \dots, L$  son entonces regionalizados en el espacio, que permite una reconstrucción de la función aleatoria espacio temporal RF  $M(u, t)$  en localidades  $u$  y  $t$  instante de tiempo no muestreadas. Notemos que estos parámetros pueden ser transversalmente correlacionados, ya que, derivan de los mismos z-datos, y tal correlación cruzada debe tenerse en cuenta, cuando los parámetros de la tendencia temporal son regionalizados en el espacio.

### 1.3.2 Modelo Espacio Temporal Múltiple RF/TS.

Esta clase de modelos espacio temporales desde un punto de vista de un proceso espacio temporal  $Z(u, t)$  como un conjunto de funciones aleatorias correlacionadas RFs, o un conjunto de funciones correlacionadas TS, que en función del dominio (espacio o tiempo) es más densamente informado.

**a) Vector de RFs.**

La función aleatoria espacio temporal RF  $\{Z(u, t), u, t \in D \times T\}$  esta modelada como una colección de series de tiempo correlacionadas espacialmente en RFs. La siguiente representación es adoptada para el proceso espacio temporal  $Z(u, t)$  (ver figura 1 (a)). Por ejemplo, el registro diario, semanal, mensual y anual de homicidios en El Salvador desagregado por municipio, cantón, etc.

$$\begin{aligned} Z(u, t) = \mathbf{Z}(u) = [Z_t(u)] &= [Z_{t_1}(u), Z_{t_2}(u), \dots, Z_{t_T}(u)]' & (1) \\ &= [Z_1(u), Z_2(u), \dots, Z_T(u)]' \end{aligned}$$

es decir, T RFs son considerados, uno para cada instantes de tiempo  $t_i, i = 1, \dots, T$ . Los mapas de distribución espaciales pueden ser construidos solo para los T instantes de tiempo  $\{t_1, t_2, \dots, t_T\}$ , y no es posible interpolar en tiempo sin algún modelo adicional.

Asumiendo estacionariedad, la matriz de valores de la covarianza  $\mathbf{C}(h)$  entre los datos para dos instantes separados por el mismo retardo  $h$ , se expresa como:

$$\mathbf{C}(h) = \begin{bmatrix} C_{11}(h) \dots & C_{1j}(h) \dots & C_{1T}(h) \\ \vdots & \vdots & \vdots \\ C_{i1}(h) \dots & C_{ij}(h) \dots & C_{iT}(h) \\ & \vdots & \vdots \\ C_{T1}(h) \dots & C_{Tj}(h) \dots & C_{TT}(h) \end{bmatrix}$$

Donde

$$C_{ij}(h) = Cov\{Z_i(u), Z_j(u + h)\} \cong \frac{1}{n} \sum_{\alpha=1}^n [Z(u_\alpha, t_i) - m_i] [Z(u_\alpha + h, t_j) - m_j]$$

Es la covarianza cruzada entre los datos en el instante de  $t_i$  y  $t_j$ , con  $m_i = (1/n) \sum_{\alpha=1}^n Z(u_\alpha, t_i)$  siendo una estimación de la media de la función aleatoria RF  $Z_i(u)$  en el instante  $t_i$ . Para  $h = 0$ , la matriz  $\mathbf{C}(0)$  es la matriz estacionaria de varianza cruzada del conjunto T de datos, por un instante de tiempo  $t_i$ .

**b) Vector de TS.**

La función aleatoria espacio temporal RF  $Z(u, t)$  es ahora modelada por un número finito  $n$  de funciones aleatorias correlacionadas espacialmente TS (ver figura 1.2) con la siguiente representación:

$$\begin{aligned} Z(u, t) = \mathbf{Z}(t) = [Z_u(t)] &= [Z_{u_1}(t), Z_{u_2}(t), \dots, Z_{u_n}(t)]' & (2) \quad (6) \\ &= [Z_1(t), Z_2(t), \dots, Z_n(t)]' \end{aligned}$$

Es decir, son consideradas  $n$  series de tiempo, una para cada localización  $u_\alpha$ . La construcción de TS es posible sólo en las  $n$  localizaciones  $\{u_1, u_2, \dots, u_n\}$ .

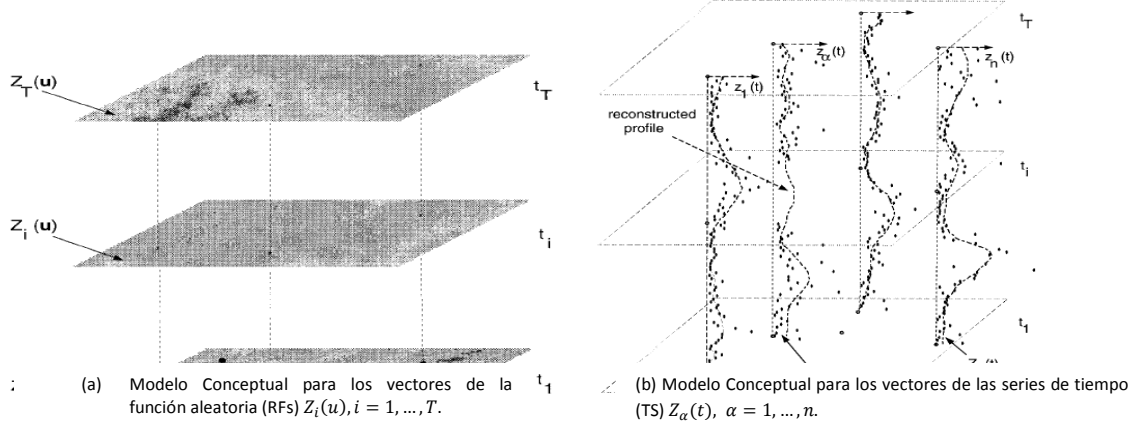


Figura 1.2: Representación gráfica de los dos modelos conceptuales espacio-temporales.

#### 1.4 Cokriging en el Modelo Espacio Temporal.

Existen interesantes enlaces entre el enfoque de considerar una única función aleatoria espacio temporal RF y un enfoque de considerar múltiples vectores de RFs o Ts. Estos vínculos se han investigado de una forma parcial para el caso de modelos de covarianza espacio temporal separable de tipo (1)(Phaedon C. Kyriakidis, 1999).

Consideremos, por ejemplo,  $T$  espacios RFs  $\{Z_i(u), i = 1, \dots, T\}$  correlacionados a través del tiempo. En el caso no-isotópico, es decir, niveles de atributos medidos y  $T$  instante de tiempo  $\{t_1, t_2, \dots, t_T\}$  en diferentes  $n_i \neq n$  localidades  $u_{\alpha_i}, \alpha_i = 1, \dots, n_i$ , el estimador simple de Cokriging (SCK)  $\hat{Z}_k(u)$  para  $Z(u, t_k)$  desconocida en la localidad  $u \in Dyt_k$  instantes de tiempo se escribe como:

$$\hat{Z}_k(u) - m_k = \sum_{i=1}^T \sum_{\alpha_i=1}^{n_i} \lambda_{\alpha_i} [Z_i(u_{\alpha_i}) - m_i] \quad (3)$$

donde  $m_i$  es la media estacionaria espacial de la función aleatoria RF  $Z_i(u)$  definida para los  $t_i$  instantes de tiempo. Generalmente, sólo los datos más cercanos a los puntos  $(u, t_k) \in DxT$  siendo estimados son considerados, en cada caso  $n_i$  se sustituye por  $n_i(u, t)$  con  $n_i(u, t) < n_i$ . El número de datos conservados y el tamaño de la búsqueda de una zona no tienen porque ser los mismos para todos los instantes de tiempo  $t_i, i = 1, \dots, T$ .

Los pesos de SCK  $\lambda_{\alpha_i}(u, t_k)$  son obtenidos a partir del sistema de SCK:

$$\sum_{j=1}^T \sum_{\beta_i=1}^{n_j} \lambda_{\beta_i}(u, t_k) C_{ij}(u_{\alpha_i} - u_{\beta_i}) = C_{ik}(u_{\alpha_i} - u), \quad \alpha_i = 1, \dots, n_i, i = 1, \dots, T \quad (4)$$

Denotado por el vector de ponderaciones de SCK para los  $n_i$  datos en el  $t_i$ -ésimo instante de tiempo como  $\mathbf{A}_i(u, t_k)$ , la representación vectorial del estimador SCK anterior es

$$\hat{Z}_k(u) = \sum_{i=1}^T [\mathbf{A}_i(u, t_k)]' [\mathbf{Z}_i(u) - \mathbf{m}_i] + m_k$$

donde

- $\mathbf{A}_i(u, t_k) = [\lambda_i, \dots, \lambda_{n_i}]'$  es el vector  $n_i \times 1$  de los pesos cokriging asignado a los  $z_i$  datos en las localidades  $u_{\alpha_i}$ ,  $\alpha_i = 1, \dots, n_i$ ,
- $\mathbf{Z}_i(u) = [Z_1(u_1), \dots, Z_i(u_{n_i})]'$ , y
- $\mathbf{m}_i$  es el vector  $n_i \times 1$  de las medias estacionaria  $m_i$  (constante) de las funciones aleatorias espaciales RFs  $\mathbf{Z}_i(u)$ ,  $i = 1, \dots, T$ .

El vector T de los pesos cokriging  $\mathbf{A}_i(u, t_k)$  son obtenidos de la solución del único sistema (4), expresado en la forma matricial como:

$$\begin{bmatrix} \mathbf{K}_{11} & \cdots & \mathbf{K}_{1T} \\ \vdots & \ddots & \vdots \\ \mathbf{K}_{T1} & \cdots & \mathbf{K}_{TT} \end{bmatrix} \times \begin{bmatrix} \mathbf{A}_1(u, t_k) \\ \vdots \\ \mathbf{A}_T(u, t_k) \end{bmatrix} = \begin{bmatrix} \mathbf{k}_{11} \\ \vdots \\ \mathbf{k}_{Tk} \end{bmatrix} \quad (5)$$

donde

- $\mathbf{K}_{ij} = [C_{ij}(u_{\alpha_i} - u_{\beta_j})]$  es la matriz  $n_i \times n_j$  de autocovarianza y autocovarianza cruzada entre los  $n_i$  y  $n_j$  datos  $z_i(u_{\alpha_i})$ ,  $\alpha_i = 1, \dots, n_i$  y  $z_j(u_{\beta_j})$ ,  $\beta_j = 1, \dots, n_j$  para los instantes de tiempo  $t_i$  y  $t_j$ , respectivamente, y
- $\mathbf{k}_{t1} = [C_{ik}(u_1 - u), \dots, C_{ik}(u_{n_i} - u)]'$  es el vector  $n_i \times 1$  de los datos desconocidos autocovarianza y covarianza cruzada.

En virtud de la decisión del modelo de estacionariedad espacio temporal de segundo orden, la covarianza cruzada  $\mathbf{K}_{ij}$  entre las mediciones en cualquier par de instantes de tiempo  $t_i$  y  $t_j$  no depende de estos instantes, pero sólo en el retardo temporal  $\tau = t_i - t_j$  separados. Por lo tanto, la siguiente relación es válida:

$$\mathbf{K}_{ij} = \mathbf{K}_{\tau_{ij}} = [C(h_{\alpha\beta}, \tau_{ij})], \quad \forall h_{\alpha\beta} \in D, \quad \forall \tau_{ij} = t_i - t_j \in T$$

donde

$C(h_{\alpha\beta}, \tau_{ij})$  es la matriz de dato a dato covarianza espacio temporal.

El anterior sistema de SCK es el mismo que el sistema generalizado SK (Simple Kriging) en un contexto espacio temporal, que puede escribirse como

$$\begin{bmatrix} C(u_\alpha - u_\beta, t_1 - t_1) & \cdots & C(u_\alpha - u_\beta, t_1 - t_T) \\ \vdots & \ddots & \vdots \\ C(u_\alpha - u_\beta, t_T - t_1) & \cdots & C(u_\alpha - u_\beta, t_T - t_T) \end{bmatrix} \begin{bmatrix} \lambda_{\alpha 1}(u, t_k) \\ \vdots \\ \lambda_{\alpha T}(u, t_k) \end{bmatrix} = \begin{bmatrix} C(u_\alpha - u, t_1 - t_k) \\ \vdots \\ C(u_\alpha - u, t_1 - t_k) \end{bmatrix} \quad (6)$$

Por lo tanto, en virtud de la estacionariedad espacio temporal de segundo orden, los sistemas (4) y (5) son equivalentes, y el estimador (3) vuelve a él estimador Simple Kriging (SK) en el contexto espacio temporal, es decir,

$$Z_k(u) - m_k = \sum_{i=1}^T \sum_{\alpha_i=1}^{n_i} \lambda_{\alpha_i}(u, t_k) [Z_i(u_{\alpha_i}) - m_k]$$

Notemos que, por simetría, de forma similar puede sacarse conclusiones para el caso de  $n$  correlaciones espaciales TS  $Z_\alpha(u)$ ,  $\alpha = 1, \dots, n$  definida en las localidades  $\{u_1, \dots, u_n\}$ .

Asumiendo estacionariedad, la matriz de valores de covarianza  $\mathbf{C}(\tau)$  entre los datos de localidades separados por algún retardo temporal  $\tau$  puede escribirse como:

$$\mathbf{C}(\tau) = \begin{bmatrix} C_{11}(\tau) \dots & C_{1\beta}(\tau) \dots & C_{1n}(\tau) \\ \vdots & \vdots & \vdots \\ C_{\alpha 1}(\tau) \dots & C_{\alpha\beta}(\tau) \dots & C_{\alpha n}(\tau) \\ \vdots & \vdots & \vdots \\ C_{n1}(\tau) \dots & C_{n\beta}(\tau) \dots & C_{nn}(\tau) \end{bmatrix}$$

Donde

$$C_{\alpha\beta}(\tau) = Cov\{Z_\alpha(\tau), Z_\beta(t + \tau)\} \cong \frac{1}{T} \sum_{i=1}^T [Z(u_\alpha, t_i) - m_\alpha.] [Z(u_\beta, t_i + \tau) - m_\beta.]$$

La matriz de valores de covarianza cruzada entre los datos de las localidades  $u_\alpha$  y  $u_\beta$  para el retardo temporal  $\tau$ , con  $m_\alpha = (1/T) \sum_{i=1}^T Z(u_\alpha, t_i)$  siendo una estimación de la media de la serie temporal  $Z(u_\alpha)$  en la localidad  $u_\alpha$ . Para  $\tau = 0$ , la matriz  $\mathbf{C}(0)$  es la matriz estacionaria varianza covarianza entre los  $n$  series de tiempo.

### 1.5 Modelos de Covarianza Espacio Temporal.

Analizar un fenómeno donde existe un componente espacial y otro temporal, requiere de modelos covariográficos o semivariograma válidos para lograr así estimaciones más óptimas.

Hasta ahora solo se han discutido modelos de semivariograma en términos de coordenadas en espacio y tiempo. La combinación de estructuras de autocorrelación espacial y temporal a obtener una forma completa para regionalización ha sido un área activa de investigaciones en la década pasada, revisiones se pueden encontrar en (Cressie y Huang 1999, Kyriakidis y Journel 1999). Estrategias para combinar las estructuras de autocorrelación espacio temporal (donde

generalmente se debaten en forma de covarianza) se pueden dividir en clases de modelos separables y no separables.

### **1.5.1 Modelos de covarianza separable y no separables.**

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 inversión de Fourier de la función de covarianza unidimensional (Cressie y Huang 1999). Gneiting desarrolló este enfoque para una representación de Fourier libre (Gneiting 2001). Estos acontecimientos dieron un paso importante en la geoestadística espacio-temporal, pero no fue hasta la contribución de Delaco que esta forma de estacionariedad, de funciones covarianza no separables se generalizó para su aplicación (De Iaco, Myers y Posa 2003).

El modelo de covarianza producto-suma de Delaco (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.

### **1.5.2 Modelo Lineal de correogionalización: Variograma Producto-Suma Espacio Temporal.**

Como hemos mencionado en la sección 3, gran parte del desarrollo de los modelos de regionalización espacio temporal involucra tres puntos de vistas conceptuales (Kyriakidis y Journel 1999), ellas son: i) Enfoque que supone un único modelo de función aleatoria integrando los componentes espacial y temporal; ii) Enfoque en donde involucra vectores de funciones aleatorias espacial; iii) Enfoque en donde involucra vectores de series de tiempo.

Sin embargo, ninguno de ellos involucra la distribución multivariante espacio temporal (excepción del enfoque cokriging). Por ejemplo, varias variables podrían medirse en localizaciones o lugares dados y para diferentes momentos, resultando una información multivariante espacio temporal. Aunque muchos autores han utilizado 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.

El objetivo aquí es mostrar cómo el variograma marginal para un modelo espacio temporal puede 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 LCM está escrito en términos de marginales en el espacio-tiempo y pueden ser fácilmente obtenidas. Sin embargo, un modelo espacio-temporal más general con LCM, como el que integra el modelo producto y el producto-suma pueden ser usados.

### **El modelo LCM Espacio-Temporal**

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

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

Donde  $C_x(h)$  y  $C_t(\tau)$  son los modelos de covarianza tanto para la parte espacial y temporal,  $k_1 > 0, k_2 \geq 0, k_3 \geq 0$ .



## CAPITULO II

### APLICACIÓN DE LA GEOESTADÍSTICA ESPACIAL AL CRIMEN EN EL SALVADOR.

#### Introducción

En este capítulo se aplica la teórica geoestadística al crimen en El Salvador considerando el componente espacial. El desarrollo del capítulo incluye un bagaje de técnicas y herramientas geoestadística que existen en nuestro medio para abordar el fenómeno de la criminalidad mediante el enfoque espacial. El capítulo está estructurado por dos partes esenciales en lo que respecta al estudio geoestadístico, la parte estructural y predictiva. En la parte de análisis estructural involucra el estudio del fenómeno regionalizado del crimen mediante dos puntos de vistas, el primero, cuando se analiza el proceso estocástico para conocer qué tipo de patrón esta describiendo tal proceso y el segundo, conocer la variabilidad propia del fenómeno para lograr así una interpolación geoestadística de la criminalidad en el país. Se estudian los estadísticos como el índice del vecino más próximo y la función K de Ripley para contrastar que el fenómeno de la criminalidad presenta un patrón de tipo aleatorio, uniforme y agregado; esto, para tener la evidencia de la existencia de concentraciones de regiones donde presentan una mayor y menor concentración de la incidencia del homicidio. Por otro lado, se analiza dentro del análisis estructural, la caracterización del fenómeno, es decir, estudiar la correlación y variabilidad espacial mediante el contraste de I de Moran y función de covarianza o semivariograma. En la parte predictiva, se tiene que, partiendo del análisis estructural ser capaz de generar por medio de un modelo teórico una superficie o mapa delictual del fenómeno del crimen, es decir, interpolar a partir de un estimador como el kriging las intensidades de homicidios ocurridos en los municipios que conforman El Salvador.

#### 2.1 Descripción de Datos Espaciales.

En el siguiente apartado se analiza brevemente la descripción de los datos espaciales de homicidio en El Salvador desde el año de 2003 hasta el año de 2008. El objetivo de esta descripción es tener claro la forma de datos estadísticos con el cual se partirá para la realización del análisis de la variabilidad espacial y estimación del fenómeno del crimen. Existen varios formatos de datos espaciales que generalmente se utilizan y que a continuación se describen:

- a) **Datos geoestadísticos:** está referido a la distribución continua que presenta el fenómeno del crimen sobre una región objeto de análisis (a nivel de municipio), es decir, cuando se toman los datos espaciales en un número finito de localizaciones, permitiendo definir una superficie continua de variación. En este caso, el dato espacial que se ha considerado para el análisis geoestadístico ha sido la frecuencia o intensidad de homicidios ocurridos en los diferentes municipios que conforman el país. Ese dato espacial está representado por el centroide de dicho municipio para realizar el análisis estructural y predictivo del crimen.

- b) **Dato coordinado:** son observaciones puntuales o eventos discretos vinculados al área de estudio, donde cada ocurrencia o hecho delictivo es un coordinado. Un conjunto de coordinados se visualiza como una nube de puntos y sirve para conocer la distribución del fenómeno del crimen a partir de la técnica de análisis de distancia dentro del análisis estructural como el análisis del vecino más próximo (NNA). Los datos que se disponen son mayores al 85% del total de homicidios ocurridos en promedio desde el año de 2004 y 2008 en el país, aunque no son el 100% del universo, si son lo suficientemente grandes para la realización de este tipo de análisis.
- c) **Campo escalar:** Son datos espaciales de frecuencia referenciados sobre unidades de superficie. Estas frecuencias espaciales suelen encontrarse espacialmente representados de igual manera que los datos geoestadísticos, es decir, para el caso de las frecuencias, cada polígono o municipio su centroide representara la cantidad de homicidios o acumulación de eventos.

El tipo de análisis a realizar bajo estos tipos de formato de datos espaciales viene dado según la naturaleza del fenómeno y del análisis que se pretenda abordar con las técnicas existentes en nuestro medio. Por ello, es importante tener claro qué tipo de formato utilizar para lograr los resultados esperados, así como también las herramientas de software para su aplicación.

Actualmente existe una variedad de software para el tratamiento de estos tipos de formatos de datos espaciales. A continuación se mencionan los más esenciales para la generación de los resultados:

- **CrimeStat:** especialista en realización de análisis de distancia y estructural como de primer y segundo orden. Para el análisis de primer orden, la generación los índices como NNA y K-Ripley se requiere del formato de datos coordinados y campo escalar, respectivamente, aunque es importante mencionar que los dos pueden ser calculados por el tipo de formato de dato coordinado. Para el análisis de segundo orden, en lo que respecta a la estructura y correlación del crimen, se necesitará del formato de datos espaciales de campo escalar o dato geoestadístico para la generación de los Índices de Moran y de C de Geary, así como también, el correlograma para dichos índices.
- **ARCGIS:** especialista para el análisis de segundo orden en lo que respecta a la generación de semivariograma, así como también, en la predicción del crimen a través del estimador de Kriging Simple (Disyuntivo).

## **2.2 Análisis de Variabilidad Espacial del Crimen en El Salvador.**

Analizar la variabilidad espacial del fenómeno de la criminalidad en El Salvador requiere en primer lugar de describir el número de eventos ocurridos por unidad de área en cualquier localidad; en segundo lugar, requiere de dar respuesta a la interrogante de cuál será la distribución que tendrán esos eventos ocurridos en el país. Para ello, es importante tener presente que la observación de dichos eventos como un primer paso consiste en visualizar la regularidad e irregularidad de la distribución, iniciando con el supuesto por el cual esa distribución es regular cuando las observaciones se distribuyen con tendencia a la homogeneidad en toda el área de estudio, en este caso, los 21,040km cuadrados del país.

En tal sentido, esa variabilidad espacial es determinada por dos tipos de análisis como el análisis de distancia de primer y segundo orden. El primero consiste en analizar el patrón de localizaciones en donde ha ocurrido un homicidio, lo que permitirá definir el tipo de estructura espacial que organiza a la intensidad global en la ocurrencia del fenómeno del crimen, es decir, conociendo la descripción del proceso de eventos u ocurrencias en localidades, ser capaz de inferir en si el fenómeno presenta una estructura aleatoria, uniforme y agregada a distancias territoriales largas o pequeñas. El segundo, tiene que ver con la autocorrelación espacial de la distribución, lo que permitirá definir cómo varía la intensidad o frecuencia del homicidio en función de su localización, es decir, la clave primordial se encuentra en la medida de correlación espacial, el cual mide el grado de similitud para una variable de tipo cuantitativa entre puntos vecinos sobre el área muestral. Este análisis solo es posible cuando se calcula esta medida de autocorrelación espacial como lo es el “variograma” (o semivariograma).

El procedimiento para conocer la estructura y correlación es necesario realizarlo ya que de lo contrario a la hora de poder definir un modelo teórico al fenómeno del crimen para su estimación se caería en el error de no conocer a ciencia cierta su patrón espacial, es decir, no habría forma de modelar el fenómeno cuando el patrón de coordenados y campo escalar (dato geoestadístico) presenta aleatoriedad en el área de estudio.

A continuación se analiza el fenómeno del crimen de forma espacial a partir de estas premisas definidas anteriormente

### **2.2.1 Análisis de Primer Orden: Análisis Exploratorio de Datos Espaciales.**

Darle solución al problema de la criminalidad implica necesariamente analizar la relación entre la distribución espacial observada de las evidencias del hecho donde ocurrió el evento o crimen, las cuales fueron recogidas en los instrumentos de levantamiento de la escena del hecho. Empero, este problema puede ser expresado en las siguientes interrogantes:

- ¿Qué tuvo lugar aquí?
- ¿Cuándo tuvo lugar?

- ¿Cómo tuvo lugar?
- ¿Por qué tuvo lugar aquí, en este momento y de esta manera, y no en otro lugar, en otro momento y de otra manera?

Al expresarse estas interrogantes dan la pauta de buscar una solución lo suficientemente razonable, en el sentido de caracterizar el fenómeno mediante una relación en base a distribuciones mediante el análisis de vecindad, estacionariedad, repulsión o independencia, distancia métrica, entre otros. Por ejemplo, para el caso de vecindad, cuando ocurre un homicidio en cierta localidad, se establecerá una vecindad espacial en la localización de los demás eventos mediante la atracción. Permitiendo reconocer si lo que sucede en una determinada localidad, es la causa de lo que sucede en las localidades vecinas.

En ese sentido, el análisis de primer orden, ayudará en gran medida a caracterizar la intensidad del proceso espacial, es decir, determinar el número de acontecimientos por área muestral. Para ello, se realiza un análisis exploratorio de datos espaciales para conocer la distribución de los datos, utilizando las más importantes gráficas como el histograma y el de tendencia. El primero ayuda a conocer la distribución de frecuencias de los datos muestrales, presentando así, medidas de tendencia central, dispersión y de forma. El segundo, permite conocer si el proceso espacial cumple los supuestos de estacionariedad en el proceso espacial, para luego así, tener una luz para esclarecer si tiene sentido o no aplicar una interpolación o predicción del fenómeno del crimen como el estimador de Kriging.

Es importante recalcar que los datos de intensidades de homicidios en El Salvador para darle el tratamiento respectivo del análisis exploratorio de datos espacial, tiene que pasar con un análisis previo a ello, y es el análisis de datos atípicos, con el objetivo de no caer en malas estimaciones a la hora de abordar los cálculos. También, mencionar que el tipo de formato de datos para este análisis de primer orden ha sido el de dato geoestadístico, es decir, utilizar el centroide de cada municipio del país para representar la frecuencia o intensidad de ocurrencia de la variable homicidio.

En la figura 2.1 se muestra el resultado obtenido mediante la librería Geoestadística de ArcGis para el cálculo del histograma, y tal como se observa en dicho histograma existe una evidente anormalidad en los datos de homicidios desde el año 2003 hasta el año 2008, reflejando que la mayor parte de la ocurrencia del homicidio en El Salvador se desplaza hacia la parte izquierda del país (área centro-sur). Este comportamiento es natural, debido a que la mayoría de municipios con mayor población e industrialización están en el centro y sur del país. Se mira además, la medida de tendencia central como la media, donde en el período de análisis el año que experimentó menor intensidad en la ocurrencia fue para el año 2007, 12 homicidios en promedio. La forma de la figura del histograma es evidente la forma de una distribución exponencial para todos los años considerados en el análisis. Aunque este resultado es observatorio, es clave para llegar a definir en los datos de homicidios una estructura mediante una distribución que logren acomodarse o

ajustarse a ellos, y que solo es posible al buscar un método de transformación para darle el respectivo análisis.

En la literatura clásica de la estadística, muestra una diversidad de métodos para lograr llevar una serie de datos a la normalidad, el software arcgis proporciona tres métodos de aproximación como el directo, lineal y núcleo gaussiano (gaussiankernel), y luego así, ajustar dichos datos geoestadísticos a un modelo teórico. En nuestro caso se ha considerado el método de transformación puntuación normal con aproximación del núcleo gaussiano. Este procedimiento es hasta cierto punto importante para lograr realizar de una manera óptima los resultados que se obtengan en el análisis de segundo orden que en los siguientes apartados se analiza.

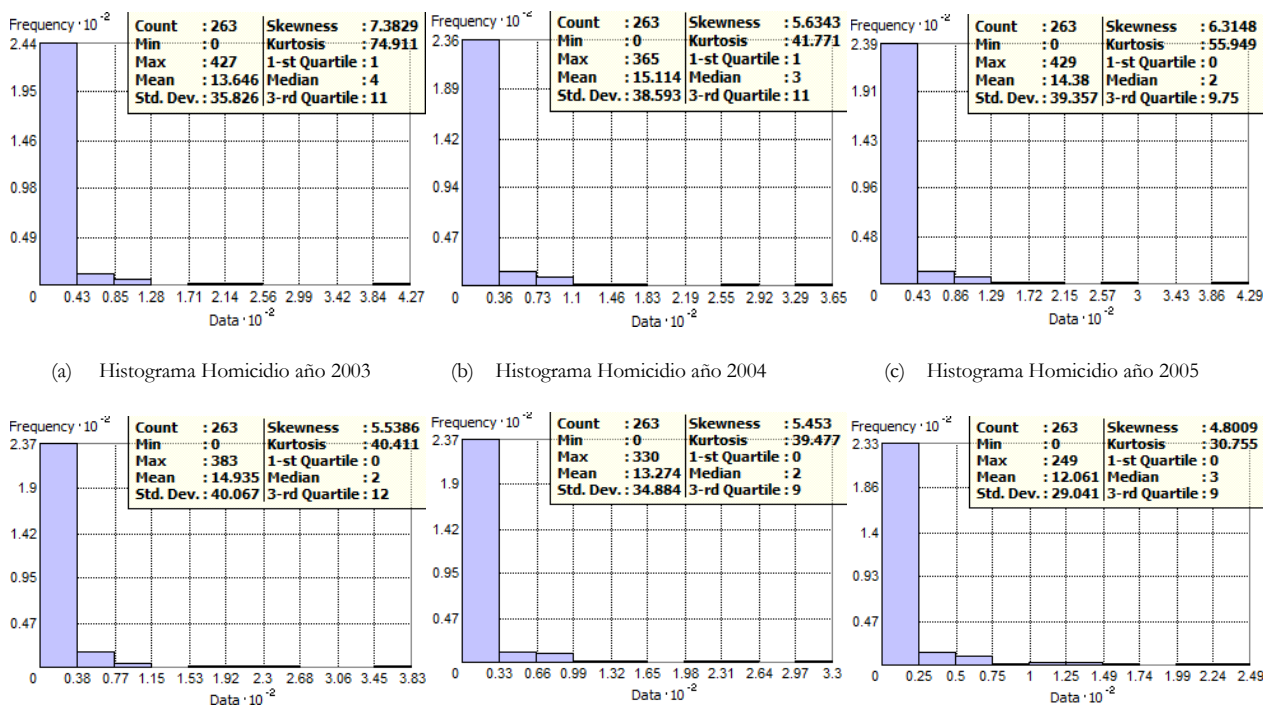


Figura 2.1: Histograma de la ocurrencia de homicidio en El Salvador desde 2003 hasta 2008

Analizar la tendencia del homicidio en el período comprendido del año 2003 hasta el 2008 es evidente que el proceso espacial muestra un claro comportamiento estacionario, esto tomando al menos un proceso estacionario en sentido débil y fuerte. Este último es el que más se observa a simple vista en los gráficos de tendencia (ver tabla 2.2), en donde se mira que en ningún año no se ha experimentado en todo el conjunto de datos geoestadísticos un patrón de tendencia respecto a la media, esto si se observa las líneas de color azul y verde, que representan los dos planos XY y YZ, respectivamente. Se observaría tendencia si existiera una fluctuación en dichas líneas sobre los planos, pero no existe esa evidencia en todos los gráficos para los datos espaciales del crimen. Infiriendo de esta manera que el proceso espacial del fenómeno del crimen puede ser modelado

con algún modelo teórico existente en la literatura que en los siguientes apartados se abordará con más detalle.

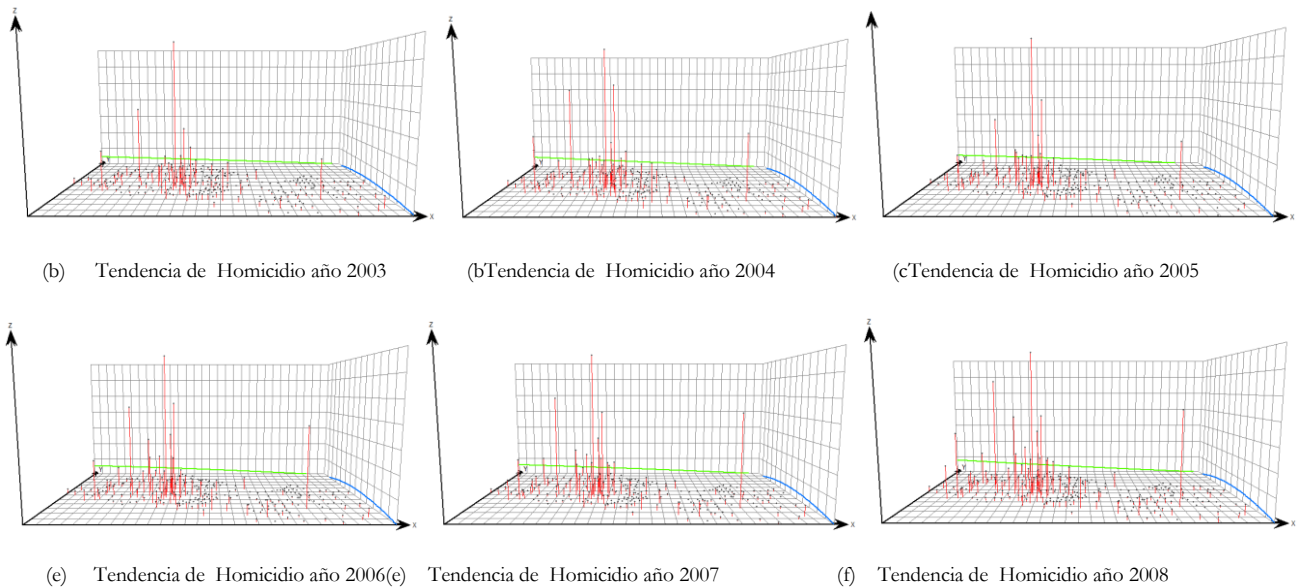


Figura 2.2: Gráfico de Tendencia de homicidio en El Salvador desde 2003 hasta 2008

Uno de los objetivos como se ha mencionado anteriormente acerca del análisis de primer orden es conocer la distribución de los datos (coordenados y/o campo escalar) para plantear de esta manera una probabilidad espacial del fenómeno del crimen, es decir, conocer el número esperado de sucesos por unidad muestral según el tipo del proceso espacial. En tal sentido, partiendo de las evidencias identificadas en lo anterior, y de la hipótesis ( $H_0$ ) que suele asociarse al patrón espacial con tendencia a la aleatoriedad. Se puede inferir visualmente que el fenómeno del crimen presenta un patrón con tendencia a la no aleatoriedad, esto debido a que al observar la intensidad con la que ocurren los homicidios en todos los municipios del país, su comportamiento es concentrado en cierta medida en varias áreas del territorio, lo cual da una noción de orden en cuanto a la distribución de los datos geostadísticos del homicidio. Este orden existente territorialmente asume la intensidad de un proceso espacial en términos de probabilidad, haciendo entonces, que esa distribución espacialmente ordenada sea más probable para sólo algunas localizaciones, mientras que en una serie espacialmente desordenada, la consecuencia u ocurrencia del homicidio puede aparecer en cualquier lugar del espacio. Es así como se genera una entropía de distribución espacial cuando puede ser medida mediante el grado de orden existente en una región o localidad y por tanto, con la mayor o menor probabilidad en la localidad de un evento u homicidio en cierta localización dada. Ahora bien, si se parte de esta entropía, es decir, del valor medio de lo inesperable, permite caracterizar el grado de aleatorización que acontece a un determinado evento. Esto es, cuanto mayor sea la probabilidad de ocurrencia de un homicidio (grado de certeza), la entropía tenderá a cero (siendo una evidencia de regularidad con tendencia a la agregación concentrada). Por el contrario, si la probabilidad de ocurrencia del homicidio es nula, entonces, la entropía (se define como el grado de ordenamiento de la estructura, es decir, es

una medida de organización, mejor dicho, de desorganización donde el máximo grado de entropía equivale al mayor grado de desorden o lo que es lo mismo, sin estructura), será alta (define la aleatorización de los datos estudiados).

Con lo analizado anteriormente se puede llegar a definir un principio importante en cuanto a la ocurrencia del homicidio, ya que en términos de entropía es claro que los datos geoestadísticos observados desde 2003 hasta el 2008 están en condiciones de orden espacial, y el problema de variabilidad espacial tiene respuesta, es decir, si se observa la distribución de las intensidades de homicidios en el período de análisis (tabla 2.2), se mira la clara evidencia de una distribución de baja entropía. Caso contrario, se estaría en una incertidumbre al momento de averiguar por qué ciertos eventos observados aparecen allí donde han aparecido.

Para confirmar este supuesto analizado, se analiza esta agregación mediante la medición del grado de agrupación de un coordenado/dato geoestadístico con una elipse que agrupe el 95% de todas las observaciones. Esta prueba sólo aporta una descripción espacial de cómo se encuentra agrupada la distribución.

En la tabla 2.1 se muestra la desviación estándar de la elipse para la intensidad de homicidios ocurridos en los 262 municipios de El Salvador. Tal como se observa, la elipse presenta un tamaño menor, es decir, cuanto menor sea la longitud de los ejes y el área, más concentrada será la serie y por tanto, habrá menor tendencia a la dispersión.

<b>Año</b>	<b>Longitud eje X (m)</b>	<b>Longitud eje Y (m)</b>	<b>Área elipse (m<sup>2</sup>)</b>
2003	72722.11 m	24795.41 m	5664840569.84 sq m
2004	71187.45 m	24678.49 m	5519146039.14 sq m
2005	66758.86 m	21857.51 m	4584157921.55 sq m
2006	69111.93 m	21050.29 m	4570471190.87 sq m
2007	71997.02 m	21707.04 m	4909813330.13 sq m
2008	72282.66 m	23196.17 m	5267448442.52 sq m

Tabla 2.1: Desviación estándar de la Elipse para la intensidad de Homicidios en El Salvador.

Hemos analizado parte importante en cuanto a la variabilidad espacial del crimen considerando el análisis de primer orden, y en resumen se ha llegado a la conclusión que al analizar el fenómeno con medidas descriptivas el fenómeno o proceso espacial presenta una distribución en los datos con tendencia a la agregación.

## **2.2.2 Análisis de Segundo Orden: Estructura y Correlación Espacial del Homicidio**

El análisis de variabilidad espacial se requiere como hemos mencionado anteriormente del análisis de patrón de localizaciones de la población objeto de estudio, y de la autocorrelación espacial de la distribución para llegar a definir el tipo de estructura que presentan las localizaciones o eventos/campo escalar y conocer como varían los valores de la variable homicidio en función de su localización.

### **Análisis Estructural**

Como se observó en el análisis de primer orden, el fenómeno del crimen presenta una distribución evidentemente agregada producto de la observación y análisis a las medidas descriptivas y superficiales del área de estudio. En este apartado se pretende confirmar completamente ese supuesto de agregación que presentan las intensidades de homicidios en El Salvador.

Para ello, se considera necesario abordar dicho análisis mediante el análisis de segundo orden, el cual permite estimar la forma de la variación de cualquier variable regional sobre el área investigada a una o varias escalas seleccionadas, de tal manera que se cuantifique esa variación espacial en distintas direcciones del espacio, es decir, conocer si existe evidencia de agregación en los eventos de homicidios y saber si existe una correlación entre las intensidades de homicidio ocurridos en diversas regiones del país.

Es importante tener presente que el patrón espacial siempre caracteriza a una determinada distribución, y el paso firme para poder realizar esta conjetura es mediante el rastreo del patrón de puntos o eventos/campo escalar. Esto porque el comportamiento espacial sólo es visto a partir de la forma o estructura que presentan dichos datos, y luego así, equiparlo en base a una distribución teórica como la de Poisson, donde al observar que una región del plano tiene la misma probabilidad de contener un punto o evento, se estaría convirtiendo este supuesto como aleatorio (hipótesis nula de la variabilidad espacial), la cual es conocida como CSR (complete spatial randomness). Caso contrario, el fenómeno del crimen presentaría un patrón espacial uniforme o agregado.

Los test que realizan esta inferencia de patrón espacial del crimen son: el análisis del vecino más próximo (NNA) y el K-de Ripley. El primero, parte de la función de distribución acumulativa que informa acerca del patrón espacial observado como una función de distancia, el cual, establece la comparación entre las distancias de eventos próximos y las distancias esperadas entre dichos eventos según el tipo de condición, es decir, si el índice de vecino más próximo (NNI) resulta ser igual a 1, indica que las distancias reales y las esperadas coinciden, de tal manera que la distribución se ajusta a una de Poisson; si el NN1 es mayor a 1, existe tendencia a la dispersión y si el NN1 es menor a 1, existe tendencia a la agregación. El segundo indicador, 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 test consiste en calcular la función K del patrón de eventos

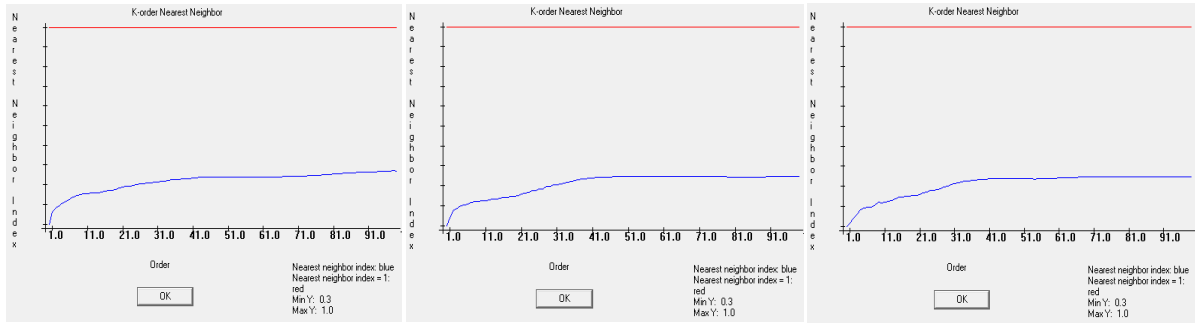


observados en localidades y compara luego a una función K teórica de un patrón de Poisson de la misma intensidad. El software CrimeStatIII permite visualizar el correlograma de la función K de Repley para diferentes intervalos de confianza. En dicho gráfico, la función observada esta representada con color azul y se compara con un número de distribuciones simuladas (color verde y rojo) mediante Monte Carlo, que representa la hipótesis de aleatoriedad.

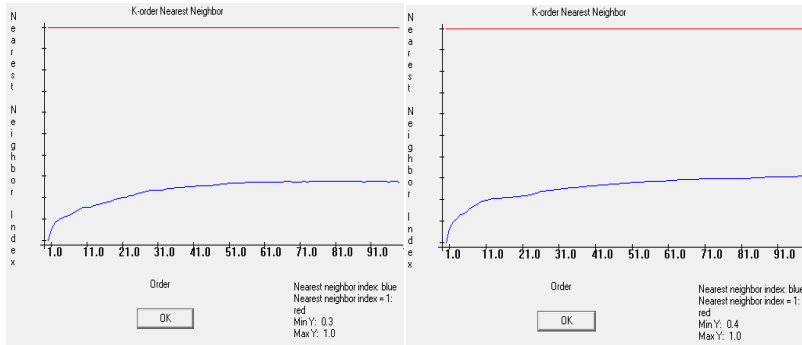
En la tabla 2.2 se presentan los resultados del NN1 para los años de 2004 hasta 2008. Se ha considerado para este test los datos de eventos, y tal como se observa, la incidencia del homicidio está representada por un patrón no aleatorio con tendencia a la agregación. Es claro que este indicador diera semejante resultado, ya que cuando se analizó anteriormente la forma de las intensidades de homicidios, estos presentaban una tendencia a la concentración, dando así una evidencia plausible del rechazo del patrón de aleatorio. La distancia estándar para los cinco años ronda en promedio casi el kilómetro medio, con un índice NNI cercano al 0.4. Este resultado es reflejado en los gráficos de K-orden del índice del vecino más próximo, el cual se muestra en la figura 2.3. En dichos gráficos se observa la clara evidencia de que el proceso espacial del homicidio mediante un orden de 45 presenta tendencia al agrupamiento. Se observaría una tendencia a la dispersión si la línea azul que representa la intensidad de homicidios saliera de la línea del supuesto de aleatorización (color rojo), pero que en los cinco años el comportamiento es similar. Concluyendo así, que la distribución de los eventos de homicidios en presentan un patrón espacial no aleatorio con tendencia a la concentración.

<b>Año</b>	<b>DistanciaEstándar</b>	<b>NN Índex</b>	<b>Z</b>	<b>P-valor</b>	<b>Evidencia de Cluster</b>
2004	1435.03 m	0.3473	-62.6711	0.0001	Si
2005	1195.86 m	0.3136	-78.1591	0.0001	Si
2006	1457.49 m	0.3471	-63.3121	0.0001	Si
2007	1199.36 m	0.3316	-70.4641	0.0001	Si
2008	1449.03 m	0.3668	-53.9149	0.0001	Si

Tabla 2.2: Análisis del Vecino más Próximo (NNA).



(a) K-Orden Índice Vecino más Próximo 2004 (b) K-Orden Índice Vecino más Próximo 2005 (c) K-Orden Índice Vecino más Próximo 2006



(d) K-Orden Índice Vecino más Próximo 2007 (e) K-Orden Índice Vecino más Próximo 2008

Figura 2.3: Gráfico de K-orden del Índice del Vecino más Próximo

En la figura 2.4 se muestra el test de K-Ripley de la intensidad de homicidios ocurridos en El Salvador desde el año de 2003 hasta el 2008. Se tiene que la distribución observada será aleatoria si su recorrido se sitúa entre la menor (función roja) y la mayor (función verde). Si el recorrido de la función observada se sitúa por encima de la condición de aleatoriedad (0,0), se concluye que la distribución observada tenderá a la concentración o agrupamiento. Mientras que si se sitúa bajo la condición de aleatoriedad, la distribución observada de los homicidios tenderá a la dispersión.

Tal como se observa en los correlogramas para cada año, existe una evidencia de agrupamiento en los seis años considerados en el análisis, notando que a una distancia de aproximadamente un kilómetro, se mira claro que no existe un dato geoestadístico que pueda relacionarse, produciendo así una aleatoriedad, ya que la distribución es acumulativa y por ende a medida que existe una zona de influencia (distancia), la distribución observada se aleja de la teórica, concluyendo de esta manera, que el proceso espacial muestra una no aleatoriedad con tendencia al agrupamiento.

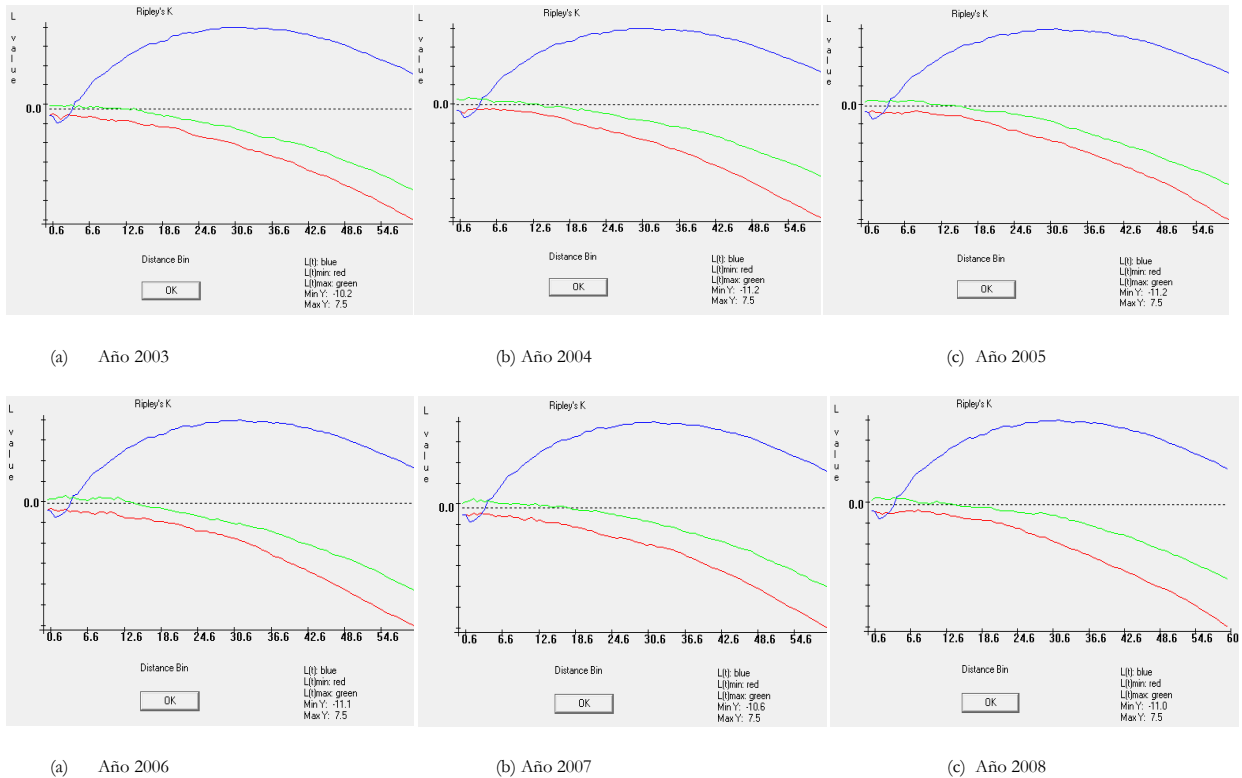


Figura 2.4: Correlograma del Test K-Ripley de la intensidad de homicidios en El Salvador.

### Análisis de la Correlación.

La caracterización espacial de una distribución se realiza mediante la estimación de la autocorrelación de los distintos valores que adopta la variable dependiente en distintas localidades del espacio, permitiendo así, establecer una medida del grado de similitud de esos valores entre localizaciones vecinas. Esta perspectiva global de cómo están distribuidos estos valores es contrastada mediante el análisis de correlación, el cual tiene como objetivo contrastar la hipótesis de que una variable se encuentre distribuida de forma totalmente aleatoria en un sistema espacial o, si por el contrario, existe algún tipo de asociación significativa de valores similares o distintos entre localidades vecinas.

Tal como se analizó anteriormente en base al análisis del vecino más próximo, se concluyó que la estructura espacial del homicidio en El Salvador presentaba un patrón no aleatorio con tendencia a la agregación, y que era necesario conocer las similitudes que existe entre los diferentes municipios que conforman el país, es decir, inferir si toda el área de estudio (o país) presenta una dependencia o autocorrelación espacial. Esta descripción de la correlación sólo es posible si se cuenta con un tipo formato de dato geoestadístico (y/o campo escalar de intensidades), para que así, se pueda analizar el proceso mediante las funciones (correlogramas y semivariograma) o a través de una serie de índices como el Índice de Moran y C de Geary. Estos test y la información

que aporta primordialmente el correlograma de Moran y el análisis de Semivariograma, se establece el tipo, la intensidad y el rango de patrón espacial a través del análisis de la ocurrencia del homicidio en los diferentes municipios del país.

Es importante luego de haber contrastado el patrón espacial del fenómeno del crimen, analizar la autocorrelación espacial cuando resulta ser positiva o negativa. Por ello, es que a partir del axioma por el cual la presencia de autocorrelación (del signo que sea) en la distribución de homicidios ocurridos en el país, invalida el supuesto de independencia espacial (la hipótesis de aleatoriedad), es decir, si la autocorrelación es positiva, indica que las localidades vecinas muestran una intensidad similar del homicidio, lo que implica que la ocurrencia del homicidio, observada en cierta localidad o municipio, dependerá de la intensidad de los homicidios que ocurran en las localidades vecinas que resulten ser significativamente similares. En el caso de la autocorrelación negativa, se contrastaría que las localidades o municipios tendrán muy diferentes intensidades en la ocurrencia del homicidio, es decir, que en la mayoría de los casos existirán municipios con una mayor intensidad y que sus vecinos experimenten una intensidad menor en la ocurrencia.

Como se ha mencionado anteriormente, los dos índices que contrastan la estructura espacial del crimen, son el  $I$  de Moran y el  $C$  de Moran; el primero, mide 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. El segundo, al igual que  $I$  de Moran, mide el grado de autocorrelación, y advirtiendo que este coeficiente se parece en la forma al estadístico de deDurbin y Watson, establecido para contrastar la hipótesis alternativa de existencia de un proceso autorregresivo de primer orden en los residuos de una regresión. Su interpretación es la siguiente: si valores de  $c$  que tiendan a 0 ( $c < 1$ ) indica autocorrelación positiva, si  $c$  tiende a 2 ( $c > 1$ ) corresponde con autocorrelación negativa, y si el valor de  $c$  es igual a 1, entonces habrá una ausencia de autocorrelación.

Para el cálculo de este test existe una variedad de software. En nuestro caso los resultados son obtenidos mediante el empleo de CrimeStat, los cuales son válidos. También, el software proporciona el correlograma de la función de Moran, la función para la distribución observada aparece en color azul y la compara con un número de distribuciones teóricas (color verde y rojo) simuladas (se realizaron 100 simulaciones para cada año) que representan la hipótesis de aleatoriedad. La distribución observada será aleatoria si su recorrido se sitúa entre la menor (función en rojo) y la mayor (función en verde). Si el recorrido de la función observada se sitúa por encima de la condición de aleatoriedad (0,0) se concluye que la distribución observada, tendrá una naturaleza agregada. Mientras que si se sitúa bajo la condición de aleatoriedad, la distribución observada tiende a la dispersión.

En la tabla 2.3 se muestra el I de Moran Global para las intensidades de homicidios ocurridos en los años de 2003 a 2009 comparado con los mismos indicadores pero ajustado a distancias pequeñas. Tal como se observa el I de Moran, con un nivel de significancia del 5% se rechaza la hipótesis nula (hipótesis de aleatoriedad), infiriendo que hay presencia de dependencia o autocorrelación espacial en la distribución de homicidios en El Salvador. Se realizó este mismo contraste mediante el test de C de Geary y los resultados fueron los mismos, es decir, la distribución de homicidios ocurridos en los años de 2003 hasta 2007 tiene dependencia espacial.

Año	Cantidad	Original Índice de Moran	Z	P-valor	Ajustado Índice de Moran	Z	P-valor
2003	3684	0.057892	8.459647	0.0001	0.050408	8.394413	0.0001
2004	3990	0.068560	9.922043	0.0001	0.059687	9.856699	0.0001
2005	3782	0.081236	11.659812	0.0001	0.070967	11.607621	0.0001
2006	3928	0.077652	11.168438	0.0001	0.067980	11.143917	0.0001
2007	3497	0.075466	10.868819	0.0001	0.065320	10.731062	0.0001
2008	3179	0.075605	10.887842	0.0001	0.066545	10.921198	0.0001

Tabla 2.3: Índice de Moran versus Índice de Moran Ajustado a distancia pequeña.

Una vez que se ha llegado a la conclusión de que el proceso espacial presenta una estructura de dependencia o autocorrelación espacial, es importante confirmar este resultado mediante el correlograma para las intensidades de homicidios ocurridas en el mismo período de análisis.

En la figura 2.5 se presentan los correlogramas de I de Moran para los años de 2003 hasta 2008, y tal como se muestra, la distribución de homicidios ocurridos en el año de 2003, demuestra una naturaleza no aleatoria, esto porque la línea azul que representa la intensidad de homicidios observados en tal año, está por encima de la condición de aleatoriedad (0,0) y de las distribuciones teóricas simuladas mediante Monte Carlo (color verde y rojo). Este comportamiento es evidente en el resto de años. Concluyendo de esta manera que el proceso espacial del homicidio en El Salvador además de presentar una dependencia espacial, su patrón sigue siendo con tendencia al agrupamiento o concentración.

## a) Correlograma de I de Moran.

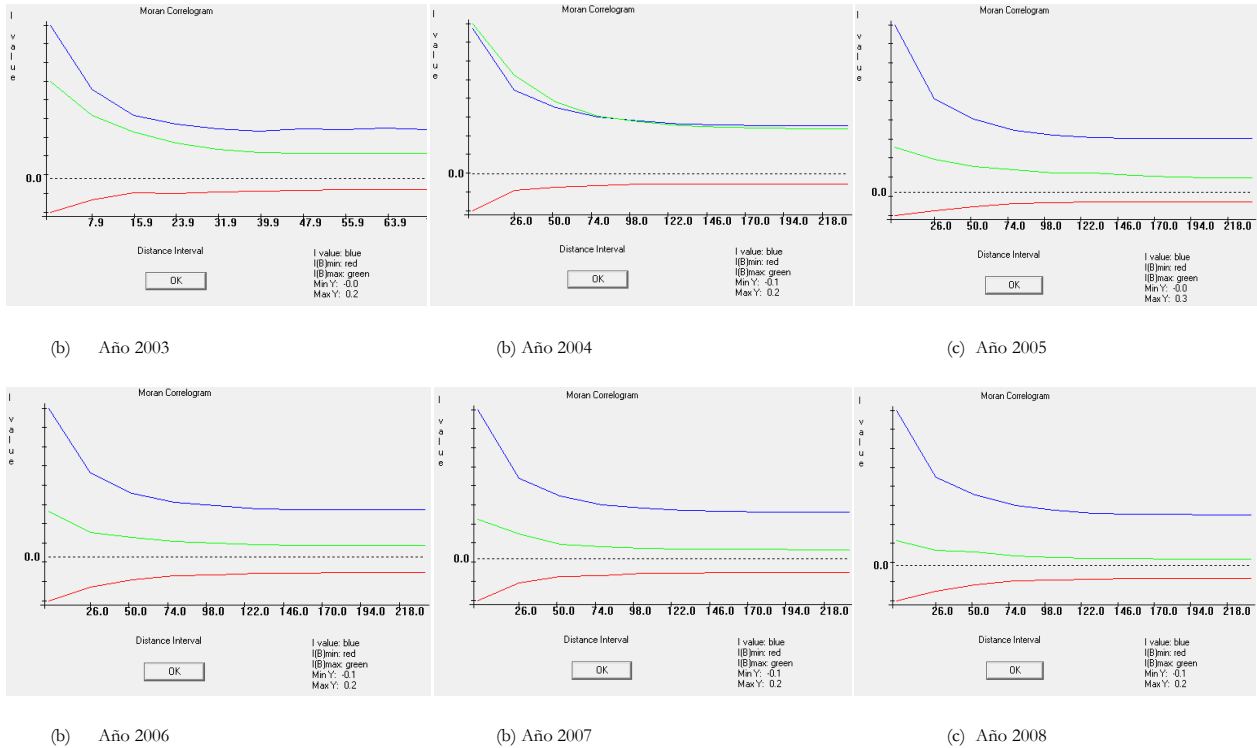


Figura 2.5: Correlograma de I de Moran de la intensidad de homicidios ocurridos desde 2003 hasta 2008.

## Contraste de I de Moran Local

El contraste de I de Moran Local al igual que el Global lleva al mismo objetivo de poder inferir sobre si el proceso espacial del crimen presenta una estructura de autocorrelación o dependencia espacial. El propósito como se mencionó en el test I de Moran Global permitió conocer que tanta dependencia espacial existía en las unidades muestrales (en este caso las intensidades de homicidios ocurridas en los diferentes municipios del país). El test I de Moran Local es capaz de contrastar la presencia de dependencia espacial por unidad muestral, es decir, saber si cierto municipio presenta una dependencia en cuanto a los vecinos que están próximos a él. Esta dependencia espacial puede ser negativa o positiva, cuando resulta ser positiva, hay un cluster o grupo alrededor de una unidad muestral o municipio, donde se concentran valores similares y elevados de la intensidad de homicidios; otro es, cuando la dependencia espacial es negativa, entonces el comportamiento será de existir un cluster o grupo de intensidades diferentes, donde un municipio específico concentra una intensidad mayor, mientras que en otros municipios aledaños o vecinos a él muestran la situación contraria.

El contraste se realizó para las intensidades de homicidios ocurridos desde el año 2003 hasta el 2008, y se constato que efectivamente el test I de Moran Local es capaz de identificar aquellos

municipios que presentan una mayor intensidad en la ocurrencia el homicidio, además de conocer su dependencia espacial, si es positiva o negativa. En el anexo A1 se muestra el test I de Moran Local para el período de análisis, en dicho contraste se observó que la mayor incidencia de homicidios ocurre en municipios que están en la zona metropolitana, también en aquellos que están en la zona sur-centro y sur oriente del país. Son 14 municipios que se han identificado como más violentos en todo el período de análisis, los cuales son: Santa Ana, San Juan Opico, Quezaltepeque, Apopa, San Martín, Ciudad Delgado, Colón, San Salvador, Mejicanos, Santa Tecla, Soyapango, Ilopango, San Marcos y San Miguel.

Una parte muy importante que hay que recalcar es que estos municipios a pesar de ser los más violentos son los que hacen que municipios vecinos a ellos tengan el mismo comportamiento en intensidad de homicidios, aunque mencionando el caso más atípico como San Miguel, donde presenta una dependencia espacial negativa en todo el período de análisis. Por ejemplo, en el año de 2008 la intensidad que experimentó este municipio fue de 155 homicidios, y su valor z-test fue de -15.290152, un valor significativo para inferir de la existencia de un cluster o grupo de intensidades diferentes, concentrando este municipio de San Miguel una intensidad mayor, mientras que el resto de municipios que colindan a él muestran situaciones diferentes.

En resumen, el comportamiento del fenómeno del crimen en El Salvador presenta una estructura con patrón de agrupamiento, y que su autocorrelación o dependencia espacial es muy evidente tanto de forma global como local, en donde se puede identificar por medio del test I de Moran Local, cuáles son aquellos municipios que presentan una mayor intensidad en la ocurrencia del homicidio y además de darnos información de si un municipio específico que presento una dependencia espacial, cómo éste influye para que otros aunque no presentan una intensidad mayor sean parte del agrupamiento producto del municipio que reflejó una mayor intensidad.

### **Semivariograma.**

En esta sección se analiza el semivariograma como herramienta para la búsqueda de una función analítica que represente adecuadamente el proceso espacial del fenómeno del crimen en El Salvador. Esa búsqueda requiere de la estimación de dicha función, la cual está caracterizada por tres importantes parámetros: uno para medir la discontinuidad en el origen o existencia de efecto pepita (nugget), otro para medir el valor máximo de variabilidad (sill), y el área de influencia de la correlación llamado alcance (range).

En la literatura existe una serie de modelos teóricos admisibles o autorizados de variogramas que se usan en la práctica y son: modelo de efecto pepita puro, esférico, exponencial, gaussiano y potencia. Los cuales están definidos mediante dos características importantes como: comportamiento del variograma en el origen, que puede ser lineal, parabólico y con efecto de pepita; y la presencia o ausencia de variabilidad.

Uno de los objetivos más primordiales del semivariograma es determinar la correlación espacial, es también, capaz de cuantificar la correlación de una variable regionalizada medida en serie de puntos (en este caso, los eventos/intensidades de homicidios en los municipios del país), es decir, puede conocer/predecir esa intensidad medida a distancias conocidas, pero que no ha sido muestreada. Esta función también responde a la pregunta ¿Cómo de parecidos son los puntos en el espacio a medida que estos se encuentran más alejados? Siendo importante que cumpla tres principios fundamentales: que el número de puntos muestrales para estimar un variograma en condiciones de isotropía (el proceso es invariante a la rotación) debe ser 100. Si existe anisotropía, los datos deben ser bastante más abundantes; que el número de pares de datos espaciales requeridos en el cálculo de cada punto del variograma debe estar, al menos, entre 30 y 50; y el último, ajustar el semivariograma mediante un modelo teórico, debido a que el empírico carece de función matemática precisa. Para nuestro caso, la asignación del variograma teórico para analizar el fenómeno del crimen, es mediante tres modelos: el esférico, exponencial y gaussiano.

Es necesario definir antes del cálculo del semivariograma una serie de consideraciones como el ángulo de dirección, tolerancia y bandwidth (ver figura 2.6). La cual, hacen que el área de estudio sea lo más representativa posible.

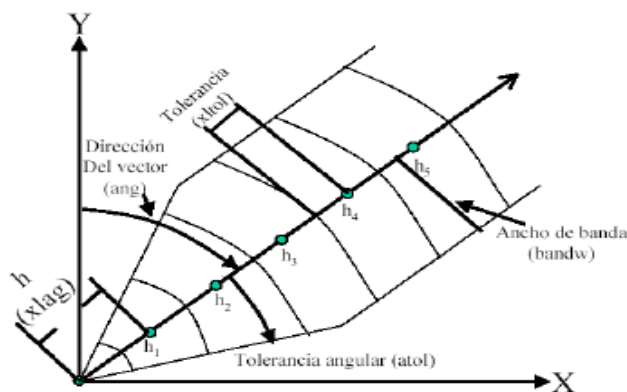


Figura 2.6: Consideración para el computo del variograma.

En la tabla 2.4 se muestran las estimaciones de los modelos de semivariograma ajustados mediante los modelos teóricos: esférico, exponencial y gaussiano. Estos resultados de las estimaciones son generados del software Arcgis. Se tiene que para el año de 2003 las consideraciones que se utilizaron para dicho cálculo fue el considerar un ángulo dirección 109, tolerancia 58 y bandwidth 6. Tal como se observa, el modelo teórico que mejor se ajusta es el exponencial, ya que el efecto pepita (nugget) no sobrepasa el 50% de la meseta (sill), la cual es evidente en la forma que tiene dicha función al mirar su semivariograma (ver figura 2.7), mostrando que el proceso espacial presenta una autocorrelación, la cual es lógico afirmarlo, puesto que en el análisis estructural se determinó ese patrón espacial del fenómeno del crimen.

En lo que respecta a la intensidad de homicidios ocurridos en el año 2004, de igual manera ocurre el mismo comportamiento de autocorrelación espacial, observando que el efecto nugget alcanzó



un valor del 0.065662 y una variabilidad espacial de 0.8224. Es claro afirmar que con ese valor de la variación estructural del fenómeno enmarca en sí, una relación espacial en cuanto a las intensidades ocurridas en los diferentes municipios del país (aumentando la dependencia espacial a medida que se incrementa la distancia). Considerando para este caso un ángulo de dirección de 110 grados, una tolerancia de 60 grados y un Bandwidth de 6. Valores semejantes en relación a los que se tomaron para las intensidades del año 2003. Por otro lado, al observar el modelo teórico para las intensidades del año 2005, es evidente que el modelo que mejor se ajusta al proceso espacial es el exponencial, esto porque su efecto pepita no sobrepasa el 50% de la meseta, la cual es igual a 0.8637.

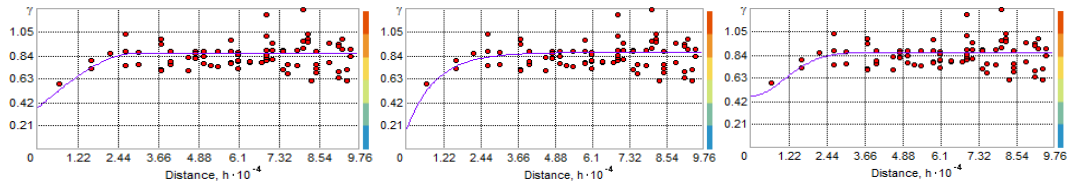
En comparación a los otros modelos como esférico y gaussiano, el que más tiene similitud al exponencial en cuanto a la variación espacial es el modelo esférico, existiendo de esta manera dos modelos que podrían estar describiendo el fenómeno del homicidio, aunque se dejará el exponencial por su mínimo efecto pepita. En cuanto al año 2006, la estimación de la función de semivariograma es más significativa en lo que respecta al modelo teórico exponencial, esto debido primeramente a la presencia de variabilidad en pequeñas distancias (menor al 25%) y segundo a que presenta un error experimental menor (nugget igual a 0.1091). Esto se mira en el correlograma de su semivariograma (ver figura 2.7), el cual la estructura espacial para las intensidades de homicidios en este año está bien definida, observando que la dependencia espacial es más fuerte a medida se incrementa la distancia.

En lo que respecta a los años de 2007 y 2008, la dependencia espacial es igualmente evidente tal cual se ha observado en los años anteriores. Este comportamiento es más que significativo puesto que en lo que concierne a la variabilidad espacial, a medida se incrementa la distancia el proceso sigue presentando una estacionariedad, es decir, el proceso es invariante a la traslación, justificando así, que es factible lograr interpolar el fenómeno del crimen a partir de los modelos de estimación que la teoría geoestadística ofrece como: el estimador de kriging, disyuntivo, etc.

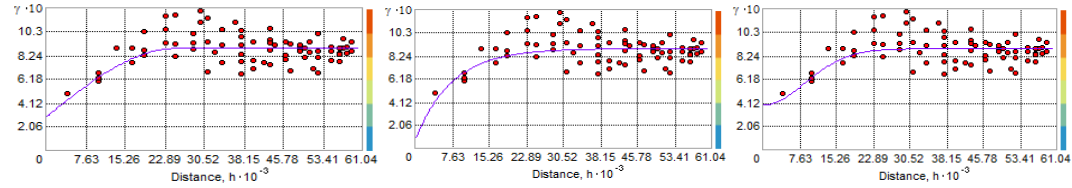
En resumen, se puede afirmar estadísticamente que las funciones de semivariograma para las intensidades de homicidios ocurridos en los años de 2003 hasta 2008, hay una evidencia de variabilidad y correlación espacial, permitiendo así, definir un modelo teórico que pueda describir dicho proceso espacial. El modelo que se utilizará para la interpolación del fenómeno del crimen será mediante el esférico y exponencial, aunque únicamente con utilizar uno de ellos bastaría.

<b>Año</b>	<b>Modelo</b>	<b>Ecuación</b>	<b>Rango</b>	<b>Intervalo</b>	<b>Distancia</b>
2003	Esférico	$0.5005 * \text{Spherical}(30277) + 0.36032 * \text{Nugget}$	30277.5	10814	9
	Exponencial	$0.71585 * \text{Exponential}(28480) + 0.15096 * \text{Nugget}$	28479.7	10814	9
	Gaussian	$0.40028 * \text{Gaussian}(27220) + 0.46204 * \text{Nugget}$	27219.8	10814	9
2004	Esférico	$0.59562 * \text{Spherical}(24991) + 0.28715 * \text{Nugget}$	24990.6	6697	9
	Exponencial	$0.82241 * \text{Exponential}(22622) + 0.065662 * \text{Nugget}$	22621.8	6697	9
	Gaussian	$0.49392 * \text{Gaussian}(21720) + 0.39013 * \text{Nugget}$	21719.7	6697	9
2005	Esférico	$0.63149 * \text{Spherical}(30756) + 0.21811 * \text{Nugget}$	30756.2	6788.8	12
	Exponencial	$0.76903 * \text{Exponential}(33612) + 0.094677 * \text{Nugget}$	33611.8	6788.8	12
	Gaussian	$0.52897 * \text{Gaussian}(27063) + 0.32317 * \text{Nugget}$	27062.6	6788.8	12
2006	Esférico	$0.58793 * \text{Spherical}(37232) + 0.27294 * \text{Nugget}$	37231.5	8516	11
	Exponencial	$0.76162 * \text{Exponential}(37195) + 0.10901 * \text{Nugget}$	37195.4	8516	11
	Gaussian	$0.49509 * \text{Gaussian}(32017) + 0.36686 * \text{Nugget}$	32016.6	8516	11
2007	Esférico	$0.58322 * \text{Spherical}(32243) + 0.28659 * \text{Nugget}$	32243.4	6697	12
	Exponencial	$0.72801 * \text{Exponential}(32797) + 0.15114 * \text{Nugget}$	32797.3	6697	12
	Gaussian	$0.49223 * \text{Gaussian}(27875) + 0.37916 * \text{Nugget}$	27875.2	6697	12
2008	Esférico	$0.57833 * \text{Spherical}(31991) + 0.27445 * \text{Nugget}$	31991.3	8516	11
	Exponencial	$0.772 * \text{Exponential}(30900) + 0.087767 * \text{Nugget}$	32797.3	8516	11
	Gaussian	$0.4796 * \text{Gaussian}(27905) + 0.37443 * \text{Nugget}$	27904.9	8516	11

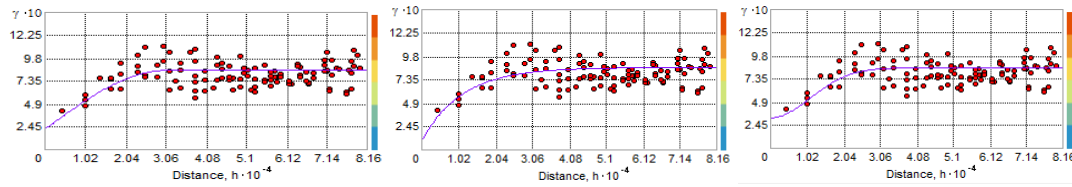
Tabla 2.4. Estimación de parámetros de los modelos teóricos esférico, exponencial y gaussiano.



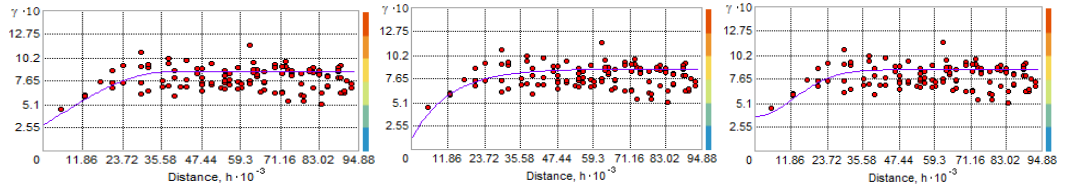
(c) Modelo Esférico Homicidio Año 2003 (b) Modelo Exponencial Homicidio Año 2003 (c) Modelo Gaussiano Homicidio Año 2003



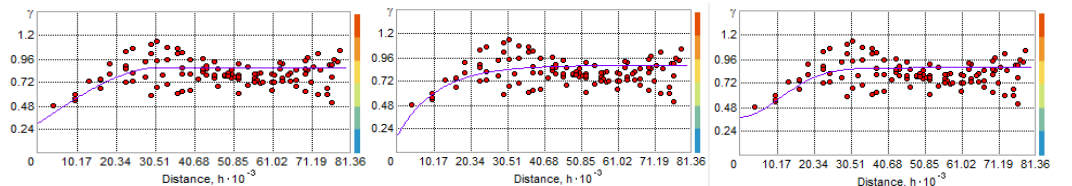
(c) Modelo Esférico Homicidio Año 2004 (b) Modelo Exponencial Homicidio Año 2004 (c) Modelo Gaussiano Homicidio Año 2004



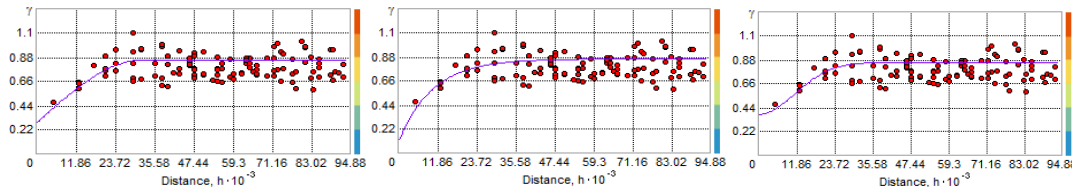
(a) Modelo Esférico Homicidio Año 2005 (b) Modelo Exponencial Homicidio Año 2005 (c) Modelo Gaussiano Homicidio Año 2005



(a) Modelo Esférico Homicidio Año 2006 (b) Modelo Exponencial Homicidio Año 2006 (c) Modelo Gaussiano Homicidio Año 2006



(a) Modelo Esférico Homicidio Año 2007 (b) Modelo Exponencial Homicidio Año 2007 (c) Modelo Gaussiano Homicidio Año 2007



(a) Modelo Esférico Homicidio Año 2008 (b) Modelo Exponencial Homicidio Año 2008 (c) Modelo Gaussiano Homicidio Año 2008

Figura 2.7: Semivariograma para los modelos teóricos esférico, exponencial y gaussiano.

### **2.3 Predicción Espacial del Crimen: Kriging Disyuntivo y Probabilístico.**

Predecir el crimen en El Salvador desde un enfoque espacial es un problema que puede ser abordado mediante técnicas que la Geoestadística nos ofrece como el estimador lineal insesgado denominado "kriging". El término kriging apareció alrededor de 1960 para nombrar una técnica creada en Francia por Matheron a partir de los trabajos de D. G. Krige quien fue el primero que hizo uso de la correlación espacial y de este estimador en el campo de la evaluación de yacimientos minerales.

Para lograr el objetivo de poder predecir las intensidades de homicidios en el país, es importante que se pueda realizar una serie de supuestos como: el análisis estructural y el de correlación. Sin haber pasado la prueba mediante estas herramientas difícilmente se podría abordar el estimador de kriging. Anteriormente en la sección 2.1.2 se llegó a la conclusión de que el proceso espacial del crimen en El Salvador pasaba esta prueba y que era posible pasar a la interpolación o predicción espacial del crimen.

En términos estadísticos se pueda advertir que algunos de los supuestos que se hacen para lograr implementar un estimador lineal insesgado como lo es el kriging, es necesario que puedan cumplirse para llegar al objetivo. En tal sentido, es natural que al estudiar un proceso espacial se tenga el conocimiento del concepto de estacionariedad, tanto del momento de primer orden como el de segundo orden de la función aleatoria (la covarianza o el variograma).

En lo que respecta al planteamiento del problema general radica en el conocimiento de si el proceso espacial presenta estacionariedad de segundo orden. Para ello, es importante que la función aleatoria tenga un valor esperado, una covarianza centrada y un variograma, donde al menos uno de estos dos momentos de segundo orden se supone conocido. Generalmente cuando solamente existe el variograma, entonces la función aleatoria (que representa el proceso espacial del crimen) se considera intrínseca (o momento de primer orden).

Teniendo presente lo anterior ahora es posible conocer este estimador según su forma, es decir, sea lineal como no lineal. La primera forma incluye los estimadores: Simple, Ordinario y Residual; la segunda, incluye los estimadores: Disyuntivo, Indicador y Probabilístico.

En el apartado anterior donde se realizó el análisis de distancia tanto de primero y segundo orden, los resultados que se obtuvieron hicieron que se logrará definir la forma del estimador para aplicarlo al proceso espacial del crimen. Se constató que en primera instancia el proceso presentaba una estacionariedad, en el sentido de no haber observado una tendencia en el proceso espacial, pero que era importante utilizar una transformación a los datos espaciales, ya que se reflejaba en el gráfico del histograma, donde la mayor parte de los datos se corren hacia la izquierda.

Uno de los estimadores que han sido propuestos para solventar este tipo de inconvenientes es el denominado disyuntivo, ya que por naturaleza, luego de haber realizado una transformación mediante un método en particular (sea directo, lineal y núcleo gaussiano) el estimador kriging toma la forma de no lineal. El método que se utilizó para la aplicación del estimador al fenómeno del crimen en El Salvador fue mediante el núcleo gaussiano, y el modelo teórico de semivariograma identificado en el apartado anterior fue el exponencial.

En la figura 2.8 se muestran los mapas de incidencia en la ocurrencia del homicidio en El Salvador desde el año de 2003 hasta 2008, y tal como se observa en los seis mapas la hipótesis nula de aleatoriedad como fue rechazada en los apartados anteriores, llegado a inferir que el proceso espacial presentaba una tendencia a la concentración, aquí en estos mapas es bien reflejado dicho patrón. Se mira la conformación de seis a siete concentraciones en todo el país, donde la mayoría de ellas están ubicadas en el área centro-sur y sur-oeste, mientras que en el área oriente es evidente la conformación de una a dos concentraciones al igual que en el área centro. Los municipios que se identificaron con una intensidad mayor mediante el test I de Moran Local, estos son los que conforman estas concentraciones. Tal es el caso de aquellos que conforman la concentración ubicada en el departamento de San Salvador y La Libertad; por ejemplo, en el año de 2008 los municipios que conforman la concentración de mayor incidencia de entre 123 a 249 homicidios son: San Salvador, Soyapango, Ciudad Delgado, Mejicanos, Ayutuxtepeque, Apopa, Nueva San Salvador (Santa Tecla), Nejapa, Quezaltepeque, San Juan Opico y Colon. Mientras que la concentración ubicada en el Departamento de Santa Ana están: Santa Ana y Coatepeque. Por otro lado, la concentración ubicada en el área sur-oeste del país, específicamente en el Departamento de Sonsonate se encuentran los municipios que experimentaron para este año una intensidad mayor como: Acajutla, Sonsonate y Nahuilingo. Si se analiza el área centro del país, se observa según el mapa, la unión de tres concentraciones, aunque la mayor concentración está ubicada en el centro-sur del país, específicamente en los Departamentos de San Vicente y La Paz, de los cuales, los municipios que experimentaron una mayor intensidad son: Zacatecoluca, San Luís de la Herradura, Tecoluca, San Vicente y Jiquilisco.

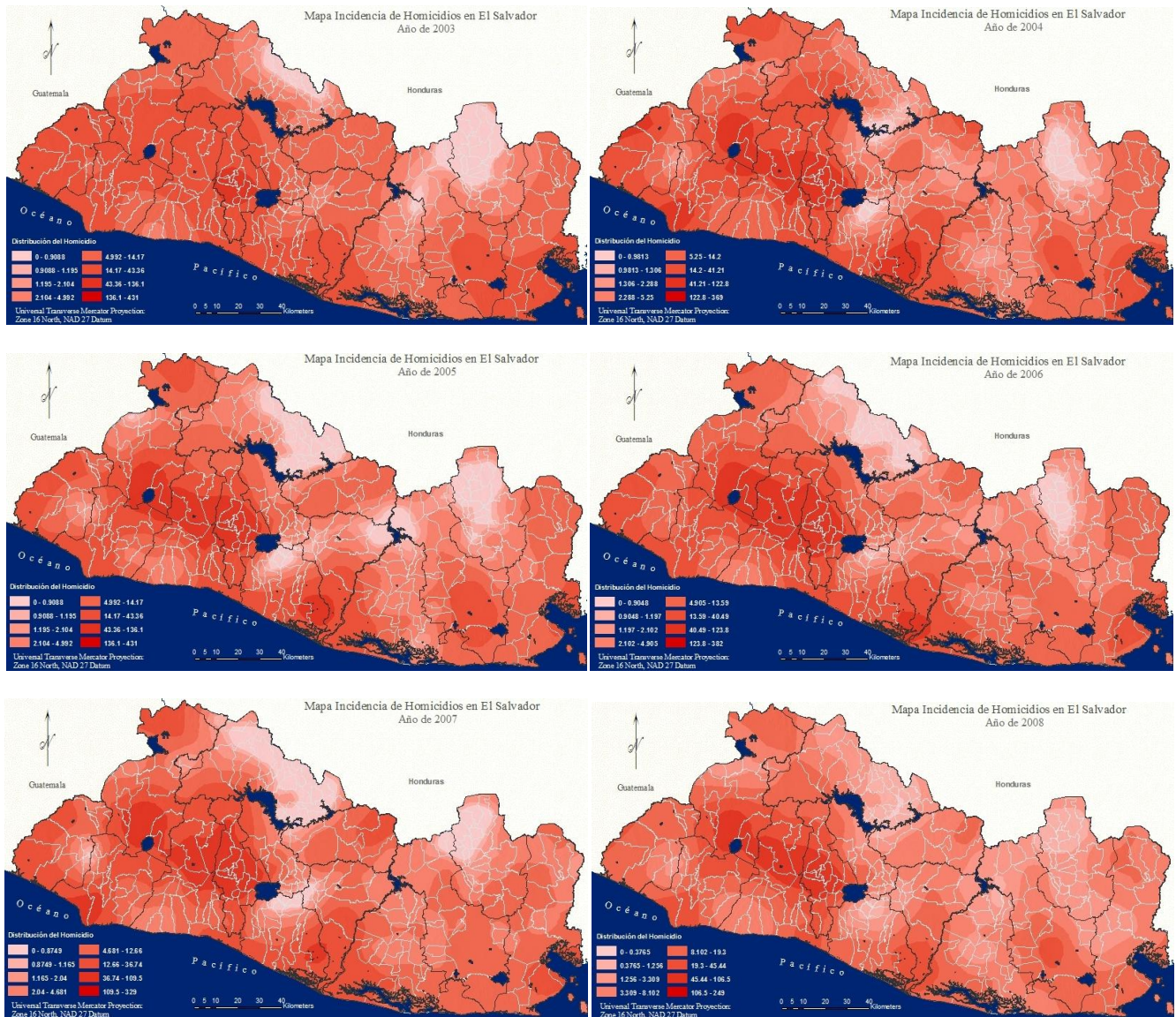


Figura 2.8: Mapas de Incidencia del Homicidio en El Salvador desde el año 2003 hasta 2008.

Finalmente está la concentración ubicada en la zona oriental del país, y aunque para este año la intensidad fue menor respecto al año 2008 (según el mapa) el comportamiento respecto a la conformación de las concentraciones es el mismo, esto por el grado de agrupamiento que existe en cuanto a los municipios aledaños a aquellos que presentan una mayor intensidad en la ocurrencia del homicidio. Por ejemplo, el municipio de San Miguel se observó anteriormente en el test I de Moran Local que su dependencia espacial era negativa, y que es reflejado en la concentración ubicada en el Departamento de San Miguel. Importante resultado, ya que con esto se puede confirmar en qué momento se podría tener una estructura de dependencia espacial negativa, es decir, municipios que presentan una mayor o menor intensidad y que sus aledaños experimentan una mayor o menor intensidad, el resultado será la existencia de agrupamiento respecto a estos municipios que experimentan tal situación.

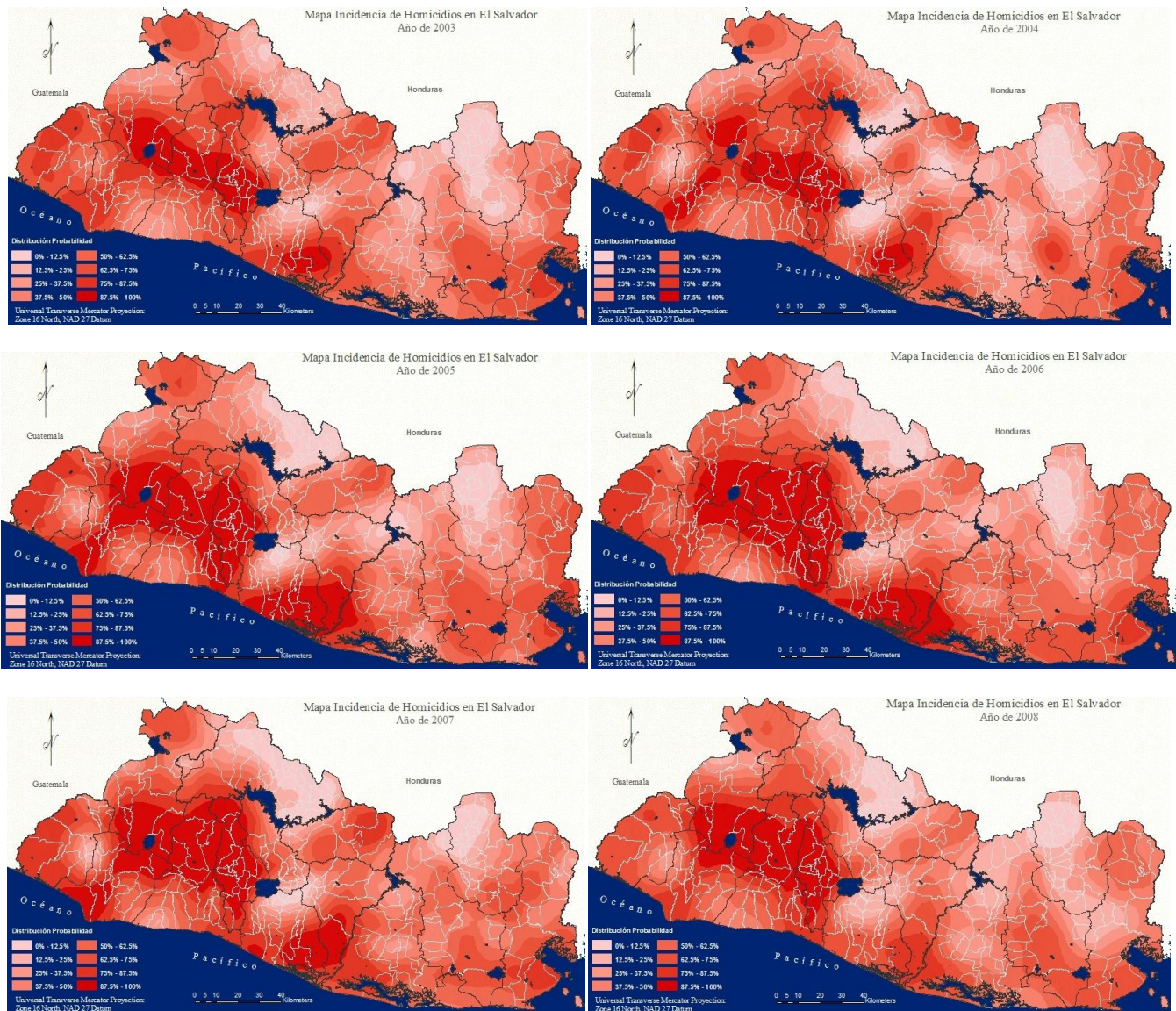
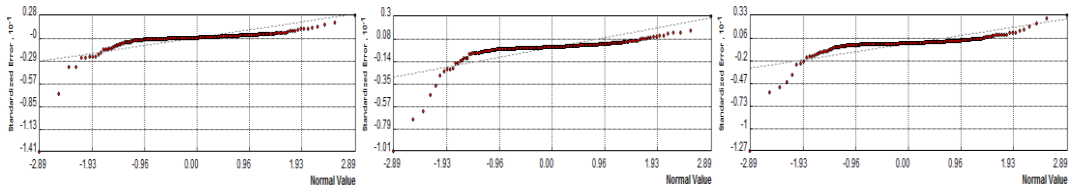
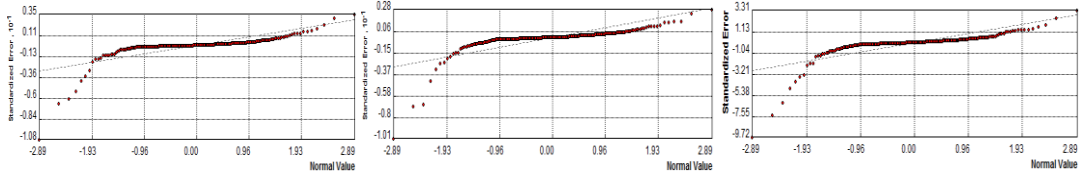


Figura 2.9: Mapas de Vulnerabilidad del Homicidio en El Salvador desde el año 2003 hasta 2008.

Finalmente, en la figura 2.10 se muestra el gráfico QQ de normalidad para los residuos, y tal como se observa, la estimación realizada a partir del estimador kriging disyuntivo y probabilístico cumple los supuestos de normalidad para los residuos, infiriendo un supuesto esencial en lo que respecta a la parte final de predicción espacial, que es la validación del modelo. Existen puntos o datos que se alejan considerablemente de la línea recta, pero que por la naturaleza con la que ocurren los homicidios en El Salvador permite que en la realidad existan municipios con mayor intensidad y otros con menor intensidad. Este comportamiento se genera similarmente para los seis años en los cuales se han en el análisis.



(a) Modelo Esférico Homicidio Año 2003 (b) Modelo Exponencial Homicidio Año 2004 (c) Modelo Gaussiano Homicidio Año 2005



(d) Modelo Esférico Homicidio Año 2006 (e) Modelo Exponencial Homicidio Año 2007 (f) Modelo Gaussiano Homicidio Año 2008

Figura 2.10: Gráfico QQ de normalidad



## CAPITULO III

### APLICACIÓN DE LA GEOESTADÍSTICA ESPACIO-TEMPORAL AL CRIMEN EN EL SALVADOR

#### Introducción

En este capítulo se presenta una aplicación de la geoestadística espacio-temporal al crimen en El Salvador, analizando el fenómeno mediante dos enfoques importantes para la generación de escenarios delictuales del homicidio en el país. En el primer enfoque tiene que ver con el estimador lineal insesgado Cokriging, lo cual es un estimador que incluye la correlación de varias variables aleatorias o una única variables medida en diferentes instantes de tiempo para la interpolación de un fenómeno, en este caso, corregionalizado. Para llegar a interpolar el crimen mediante este estimador es importante que se utilice la misma metodología para llegar a la predicción como el análisis estructural. Es por ello, que se aborda esta aplicación con este enfoque considerando cuatro años, es decir, las intensidades de homicidios ocurridos en los 262 municipios de El Salvador durante el año de 2005 hasta 2008. Se ha considerado tal período debido a que la herramienta Arcgis permite únicamente hasta un máximo de cuatro funciones aleatorias. Se describe el tipo de estructura, así como también, el análisis pertinente para llegar a la interpolación, de manera que se pueda generar un escenario delictual del homicidio para el año 2009.

El segundo enfoque tiene que ver con modelos de covarianza no-separables como el producto-suma. Este tipo de modelos a pesar de ser nuevos en la actualidad y por el hecho de estar todavía en debate por la comunidad científica, existen investigaciones que muestran lo potente que son para el modelaje de fenómenos donde se quiera analizar no solamente con dos componentes como el espacio y tiempo, sino como también, sus interacciones. Actualmente en países europeos como Inglaterra están utilizando estas herramientas para fenómenos geofísicos, donde se han llegado a resultados óptimos en cuanto a la interpolación espacio-temporal. Para nuestro caso, este tipo de enfoques se aplica a la criminalidad, y que tiene como objetivo primordial el estudio estructural y predictivo, utilizando una simulación gestadística como la secuencial gaussiano para su tratamiento espacio-temporal de dicho fenómeno.

Se estudia la criminalidad generando funciones aleatorias partiendo de la información temporal de seis años para su aplicación, considerando dos puntos de vistas para el modelaje de procesos espacio-temporales como: el modelo espacio-temporal múltiple RF/TS y Cokriging. Estos puntos de vistas pueden ser generados mediante una simulación geoestadística, utilizando un condicionamiento de la información, uno tiene que ver con los puntos muestrales que se han considerado (centroide de cada municipio del país) y otro con el condicionamiento como el estimador kriging ordinario. Se elige este estimador de kriging ordinario debido al hecho de desconocer la media para la estimación (no es constante en todo el área de estudio) de la criminalidad en el país. En ese sentido, se realiza una simulación secuencial gaussiana para simular la realidad de la incidencia de homicidios, generando de tal manera un escenario delictual del homicidio para el año 2009.

### 3.1 Predicción Espacio-Temporal: Enfoque Cokriging.

Una de las vertientes importantes y que se ha dedicado la geoestadística espacio-temporal ha sido en el planteamiento de definir procesos estocásticos espacio-temporal caracterizado mediante funciones aleatorias que varíen tanto en el espacio como el tiempo. Dicho de otra manera, a cada punto en el espacio-tiempo le hacemos corresponder una variable aleatoria.

En la práctica este tipo de planteamientos consiste en estimar valores en puntos donde no hemos muestreado, dados un conjunto finito, y con frecuencia reducido, de mediciones en espacio y en tiempo. En este caso, se aborda este tipo de planteamientos mediante el enfoque de Cokriging para la estimación de la criminalidad en El Salvador, donde se supone que cada coordenado o punto geográfico en el país  $x_j$  (centroide de cada municipio),  $Z(x_j, t)$  es una función aleatoria cuya media y covarianza dependen del punto  $x_j$ , es decir, como primer caso define  $M$  funciones aleatorias  $\{Z(x_1, t), Z(x_2, t), \dots, Z(x_M, t)\}$  correlacionadas y resulta eficiente cuando  $M \leq N$ ; y como segundo, caso se considera que para un tiempo  $t_i$  dado,  $Z(x, t_i)$  es una función aleatoria, entonces tenemos  $N$  funciones aleatorias espaciales  $\{Z(x, t_1), Z(x, t_2), \dots, Z(x, t_N)\}$  correlacionadas entre sí. De manera similar este enfoque se resuelve usando Cokriging y resulta eficiente cuando  $M \geq N$ , (Myers, 1988).

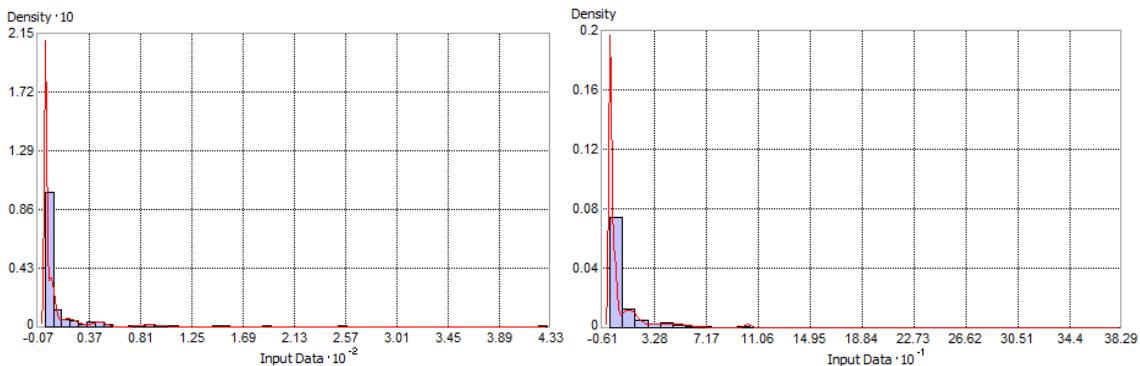
Dado lo anterior se define un conjunto de datos espacio-temporales considerando el caso dos, es decir,  $N$  funciones aleatorias espaciales (en este caso, las intensidades de homicidios ocurridos en los 262 municipios de El Salvador desde el año 2005 hasta 2008). El conjunto o set de datos se muestra en la tabla 3.1.

Coordenada X	Coordenada Y	Homicidios 2005	Homicidios 2006	Homicidios 2007	Homicidios 2008
40487	315929	47	43	41	24
41343	303084	4	3	0	3
...	...	...	...	...	...
44915	309434	12	24	12	13

Tabla 3.1: Descripción del conjunto de datos de homicidios desde el año 2005 hasta 2008.

Con dicho conjunto de datos será posible aplicar el enfoque Cokriging para predecir o generar un escenario delictual en El Salvador para el año 2009. En tal sentido, se utilizará la misma metodología que se utilizó en el capítulo anterior, cuando estimamos la criminalidad de una forma espacial mediante el estimador lineal óptimo insesgadokriging. Realizando así, un análisis estructural para estudiar la variabilidad espacial considerando los cuatro años y una estimación del fenómeno de la criminalidad bajo este enfoque de cokriging.

En la figura 3.1 se muestra el histograma para las intensidades de homicidios ocurridos en El Salvador desde el año 2005 hasta 2008, y tal como se observa, dicha distribución presenta una asimetría, es decir, gran parte de los datos están con mayor frecuencia en la parte izquierda y pocos datos muy extremos se corren a la cola derecha. En tal sentido, para poder realizar una estimación mediante cokriging será necesario realizar una transformación a los datos para lograr una estimación más óptima debido a la presencia de datos extremos. El método de transformación es el de punteo normal (normal score, Roger S. Bivand, 2008) con aproximación al método de núcleo gaussiano (kernelgaussian) para calcular así su densidad, vista con color rojo en el gráfico para el conjunto de datos. La herramienta que realiza esta generación de gráficos se encarga de transformar las intensidades de homicidios y de llevarlos nuevamente a la escala original mediante la transformación inversa (back-transformation).



(a) Estimación Densidad – Núcleo Gaussiano Año 2005 (b) Estimación Densidad – Núcleo Gaussiano Año 2006

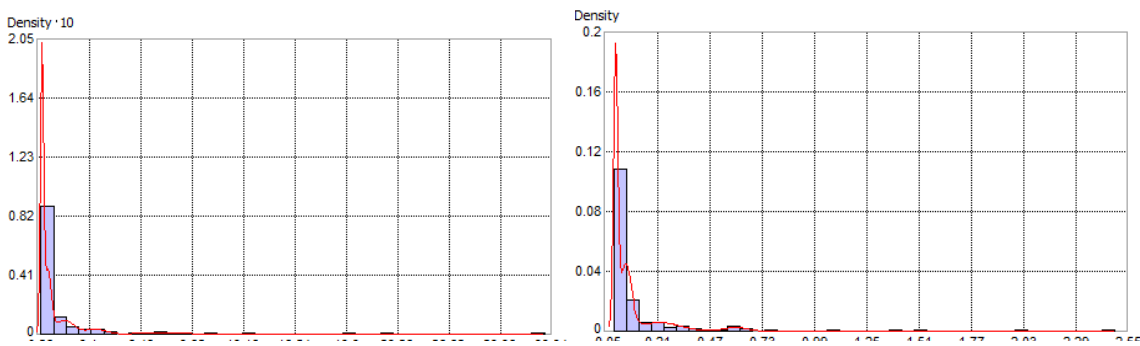


Figura 3.1: Transformación de Datos por medio del método del núcleo gaussiano.

### 3.1.1 Análisis Estructural.

El análisis estructural como se analizó en el capítulo 2, donde se analizaba la parte espacial, consistía en buscar un modelo teórico existente en la literatura geoestadística como el modelo esférico, exponencial y gaussiano para la estimación del fenómeno de la criminalidad, considerando un enfoque hasta cierto punto, un tipo de correlación espacial en cuanto a las

intensidades de homicidios ocurridos desde el año 2005 hasta 2008. La herramienta que permite identificar tales modelos teóricos es el gráfico del semivariograma.

Es de mencionar además, que al aplicar este enfoque de cokriging la generación de la función de semivariograma mostrará únicamente la parte espacial aunque internamente haya considerado el conjunto de datos para un período de tiempo determinado (en nuestro caso, cuatro años) considerado en el análisis.

En la figura 3.2 se muestra el semivariograma para el conjunto de datos de homicidios en El Salvador, considerando para efectos de análisis tres modelos teóricos, el esférico, exponencial y gaussiano. Se observa el semivariograma que representa el modelo esférico, donde su efecto pepita alcanza hasta un valor de 0.26, reflejando de tal manera un error muestral mínimo para las intensidades de homicidios. A medida que la distancia incrementa la dependencia espacial tiende a estabilizarse, infiriendo así el cumplimiento del supuesto de estacionariedad en la distribución de los datos. Por otro lado, en lo que respecta al modelo exponencial se muestra que a pesar de la existencia de una mayor variabilidad en los datos a medida su distancia incrementa ocurre un efecto pepita nulo, mostrando una discontinuidad en el origen en el semivariograma. Mientras tanto, en el semivariograma del modelo gaussiano se observa la mínima variabilidad a medida que incrementa la distancia, logrando de esta manera, una estabilización en cuanto a la parte espacial a medida que su distancia es mayor.

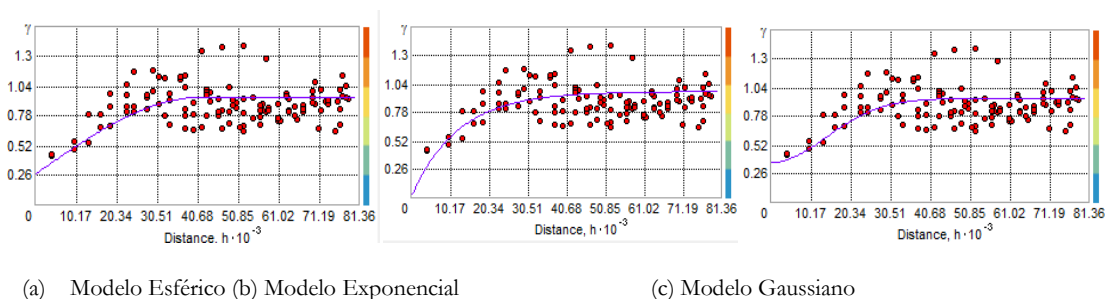


Figura 3.2: Semivariograma para los modelos teóricos Esférico, Exponencial y Gaussiano.

Al analizar estos tres modelos teóricos se puede llegar a la conclusión que con los tres modelos se puede interpolar las intensidades de homicidios, pero los más idóneos son el exponencial y el esférico, uno por la mayor variabilidad explicada en los datos de homicidios para los cuatro años y otro por no presentar una discontinuidad o efecto pepita nulo.

Las estimaciones de los parámetros de la función de semivariograma se muestra en la tabla 3.2, y se observa que el modelo esférico no sobrepasa en 50% del valor de la meseta, aunque como se ha mencionado, la variabilidad se estabiliza a una distancia de 80 kilometros, mientras que para el modelo exponencial la situación es muy diferente, su variabilidad se estabiliza a una distancia menor. Con esto, se llega a la conclusión que para estimar las intensidades de homicidios se utilizará cualquiera de los tres modelos teóricos, aunque el que se utilizará será modelo esférico por no haber experimentado un efecto pepita nulo.

Modelo	Ecuación	Rango	Intervalo	Distancia
Esférico	$0.67465 * \text{Spherical}(41315) + 0.26368 * \text{Nugget}$	41315	6697	12
Exponencial	$0.97013 * \text{Exponential}(36698) + 0 * \text{Nugget}$	36698	6697	12
Gaussian	$0.58154 * \text{Gaussian}(34826) + 0.35786 * \text{Nugget}$	34826	6697	12

Tabla 3.2: Estimación de los parámetros del semivariograma para los modelos teóricos.

### 3.1.2 Predicción: Enfoque Cokriging.

Una vez realizado el análisis estructural el paso siguiente es la predicción mediante el enfoque Cokriging. Para ello, fue necesario transformar los datos para aplicar el Cokriging al conjunto de datos. Dicha transformación que está contenida en el software Arcgis permite transformar los datos mediante el conteo normal aproximado por el núcleo gaussiano, esto, para eliminar el efecto de aquellas intensidades extremas en el conjunto de datos. Seguidamente los resultados de la interpolación o generación del escenario delictual correspondiente al año 2009 fue importante también, volver a transformar los datos utilizando la inversa de la transformada (back-transformation) para llegar a la escala original de los datos.

La presentación del escenario delictual para el año 2009 se muestra en la figura 3.3, el cual se observa la conformación de las mismas concentraciones de intensidades de homicidios en el país. Alrededor de cinco concentraciones distribuidas en todo el territorio de El Salvador: la concentración CH1 está conformada por los municipios de San Salvador, Santa Tecla, Antiguo Cuscatlán, Mejicanos, Ayutuxtepeque, San Marcos, San Martín, Apopa, Nejapa, Ciudad Delgado, Soyapango, Quezaltepeque, San Juan Opico, Colón, Ciudad Arce, Santa Ana, Coatepeque y Coatepeque. Esta concentración representa más del 90% de los municipios más violentos a nivel nacional donde se concentra una intensidad de homicidios con una franja de entre los 43 a más homicidios. En lo que respecta a la concentración CH2, los municipios que la conforman son: Acajutla y Sonsonate, donde presentan una intensidad mayor en comparación a los alrededores como Nahilingo, Izalco y San Antonio del Monte, el cual se encuentran en una franja de intensidad de entre 14 a 43 homicidios. Por otro lado, si se analiza la concentración CH3 ubicada en el área Sur-Centro del país, se observa que dicha concentración se desplaza territorialmente hacia el occidente y oriente del país. Municipios como Zacatecoluca, Tecoluca, San Vicente y Jiquilisco su escenario delictual en lo que respecta al año 2009 ronda entre los 14 a 50 homicidios.

Esta zona de influencia provoca que municipios aledaños como San Sebastián y San Estaban Catarina presenten una intensidad de homicidios de entre los 4 a 14 homicidios. Mientras tanto, al analizar la concentración CH4, se observa que hay una zona de influencia que proviene del área centro-norte del país, provocando que municipios que conforman esta concentración como Sensuntepeque, Ilobasco y Victoria experimenten una franja con intensidad entre 14 a 43, y 4 a 14 homicidios, respectivamente. Finalmente, la concentración CH5 ubicada en la zona oriental del país, específicamente en los municipios de San Miguel, Chirilagua, Chinameca y La Unión su escenario delictual es más intenso para los Municipios de San Miguel y La Unión, una franja de 43 a más homicidios; mientras que los que están aledaños a ellos, experimentando una franja de entre los 14 a 43 homicidios. En resumen, se puede concluir que estas estimaciones a partir de este enfoque de cokriging, explica la realidad en base a la correlación existente entre las intensidades ocurridas en los municipios de El Salvador, relacionando dichas intensidades en un período de cuatro años para lograr generar el escenario delictual en el país para el año 2009. Existe otro enfoque que va más allá de la estimación basada en correlacionar dichas intensidades, y es el modelo de covarianza no-separable vía producto-suma que en el siguiente apartado se analiza a profundidad. Dicho enfoque genera un escenario delictual partiendo de un modelo donde considera tres componentes importantes como la parte espacial, temporal y su interacción para la predicción espacio-temporal del crimen en El Salvador.

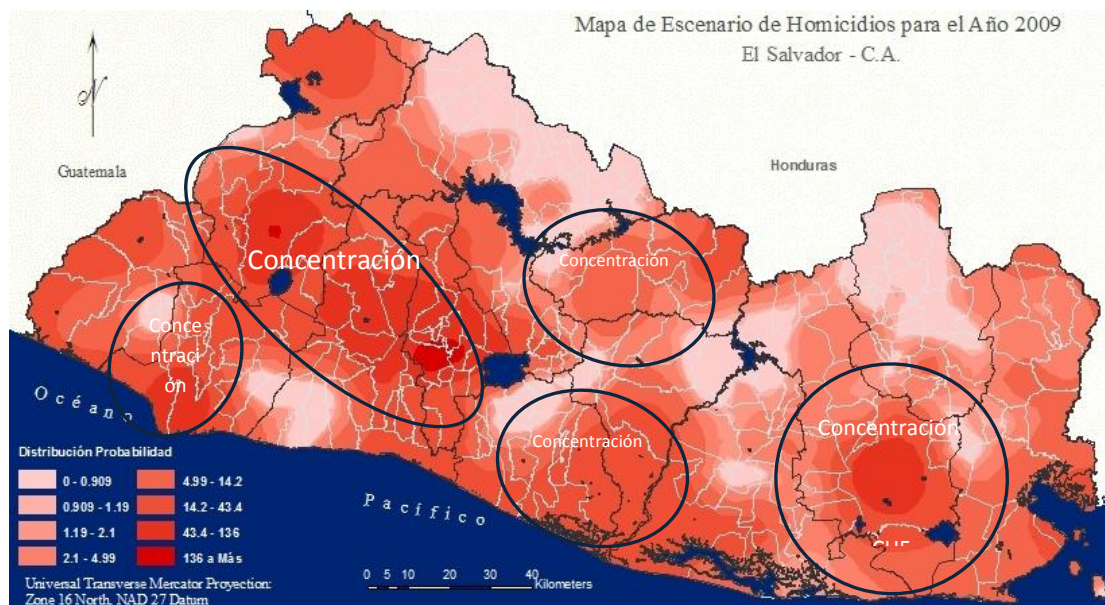


Figura 3.3: Escenario delictual del crimen en El Salvador para el año 2009.

### 3.2 Predicción Espacio-Temporal: Enfoque Modelo Lineal de Corregionalización - Producto-Suma.

#### 3.2.1 El Modelo de Covarianza No-Separable Producto-Suma.

En los últimos años el estudio de modelos de regionalización espacio-temporal ha sido el interés de los científicos para predecir fenómenos donde interactúe el espacio y tiempo, así como también sus interacciones. En un primer momento se estudiaron modelos de covarianza espacio-temporales como: el modelo métrico, producto y suma. Dichos modelos analizaban el fenómeno considerando el espacio y tiempo separadamente sin incluir interacciones que permitiera así conocer la variabilidad en conjunto (espacio-temporal). En ese sentido, fue como se propusieron otros modelos de covarianza donde incluyera sus interacciones, tal como los modelos de covarianza no-separables producto-suma incluir (De Iaco, Myers y Posa, 2001). Estas interacciones incluidas en el modelo provocan un efecto estructural en cuanto a la separación entre el espacio y tiempo (ver figura 3.4), en donde se muestra claramente tal efecto de diferencia sobre la combinación espacio-temporal en el modelo (b).

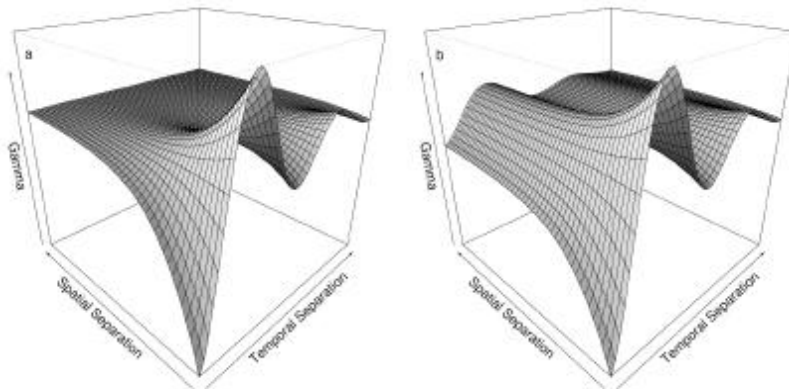


Figura 3. 4: Comparación de la representación del semivariograma espacio-temporal. a) Modelo Producto y b) Modelo Producto-Suma.

Para lograr implementar un modelo de covarianza no-separable como el producto-suma se requiere de estimar los parámetros que incluyen este tipo de modelos lineales de corregionalización. Así como en la estimación de la función de covarianza o variograma en donde se consideraba únicamente el componente espacial se estimaban los parámetros como el rango (range), efecto pepita (sill) y meseta. En este caso se tendrá que hacer lo mismo, pero tanto para el componente espacial y temporal, así como también su interacción espacio-temporal.

Por otro lado, para la estimación del modelo es importante también, estimar los coeficientes ( $k_1, k_2$  y  $k_3$ ) que acompañan las covarianzas para los componentes en el modelo lineal de corregionalización, estos coeficientes deben cumplir que  $k_1 > 0, k_2 \geq 0$  y  $k_3 \geq 0$  para que el modelo sea válido, es decir, en el marco de predecir de una manera espacio-temporal es

importante que el cálculo de la inversa de la matriz de covarianzas sea definida positiva, de lo contrario, se estaría ante un modelo no válido, y por consiguiente, no poder estimar la función de covarianza para un modelo vía producto-suma.

Como el objetivo es predecir la criminalidad en El Salvador, el conjunto de información que se tiene para tal propósito, hasta cierto punto, es excelente, debido a que se dispone de una gran cantidad de información de intensidades de homicidios a un período de seis años. Por lo que se espera que resulte un modelo de covarianza no-separable válido. Generalmente, para lograr esto, mediante una estimación o simulación se requiere de la aplicación de técnicas geoestadísticas donde incluya no solamente la parte espacial, sino que además la parte temporal. Estas herramientas son escasas para aplicar cierto fenómeno de la criminalidad, pero que se hace el esfuerzo en buscar algoritmos que permitan realizar este tipo de análisis. Dentro de este análisis tiene que existir la disciplina de iniciar con el análisis estructural para luego predecir, ya que obviando esto, se caería en el error de no llegar a estimar o predecir el fenómeno de una manera válida.

En ese sentido, este enfoque que se le da a la criminalidad en El Salvador, se inicia con el análisis estructural, es decir, realizar un análisis exploratorio de datos y un estudio exhaustivo en lo que respecta al cálculo de la función de covarianza o variograma mediante el modelo lineal de correogionalización producto-suma. Luego de ello, estimar o predecir el fenómeno mediante una simulación geoestadística.

La simulación geoestadística en nuestro caso será interpretada como la realización de una función aleatoria (considerada la intensidad de homicidios en El Salvador durante los seis años). Esta función puede ser caracterizada por una función de probabilidad y, en dependencia de su grado de estacionariedad, por la función de covarianzas o variograma. Entonces, la idea central de la simulación radica en la obtención de nuevas realizaciones artificiales de la función aleatoria, de manera tal que éstas simuladas reflejen las mismas propiedades estadísticas que se espera que posea la función aleatoria. Pero en la realidad no conocemos a simple vista dichas propiedades y lo único que podemos hacer es inferir a través de una sola realización o muestra de la función aleatoria, entonces lo que se hace es intentar obtener realizaciones simuladas que sean estadísticamente equivalentes a la muestra que se posee de la función aleatoria. Esta equivalencia tiene que ver en que todas las realizaciones tengan la misma distribución de probabilidad de la función aleatoria que se simula, conformándonos con que posean los mismos primeros momentos de primer y segundo orden que inferimos a partir de la muestra, en este caso, de las intensidades de homicidios ocurridos en El Salvador en los 262 municipios.

La mayoría se pregunta ¿Cómo es posible estimar la realidad “in situ”(desafortunadamente no se dispone de un conocimiento exacto de la realidad) correctamente, con suficiente detalle?

La respuesta es simular la realidad o la intensidad de homicidios en base a un modelo, por lo que la realidad y la simulación son variables diferentes de un mismo fenómeno.



Por ello, se pretende simular tal realidad existente en El Salvador como mediciones reales (intensidad de homicidios), es decir, una realización de la función aleatoria en los diferentes municipios del país consistirá en interpretar la distribución espacial y temporal de esa variable regionalizada como una realización particular de la función aleatoria.

Es importante saber además de esto, que la simulación puede ser condicionada a otros valores sean experimentales o estimaciones a partir de un estimador como el kriging sea simple o de tipo ordinario. En nuestro caso, tal condicionamiento para la simulación requeriría de una variable secundaria para que se logre una simulación de tipo correogionalizada, pero dado el caso que se dispone de una única variable y no de otra para relacionarla conjuntamente (como por ejemplo, la pobreza u otro tipo de variable relacionada con el fenómeno de la criminalidad) y generar simulaciones donde este condicionado otra información.

Ese tipo de condicionamiento a partir de un estimador de kriging se le conoce como KED (KrigingwithanExternalDrift), que generalmente se genera un modelo de regresión más complejo. Pero el proceso se vuelve simple cuando la segunda variable que se desea relacionar con la de interés, en este caso, la intensidad de homicidios, se logrará predecir o simular en todas las localizaciones o municipios del país.

En la literatura existen diferentes métodos de simulación de funciones aleatorias como el matricial, espectral, bandas rotantes, secuencial gaussiano, secuencial indicador, gaussiano truncado y recocido simulado. Todos ellos, excepto el secuencial indicador son de tipo condicionado y gaussiano.

Para nuestro caso, se utiliza el método secuencial gaussiano, el cual, la estimación se obtiene usando kriging ordinario aplicado a los datos estandarizados. Pero para esto, se requiere que el modelo del semivariograma calculado esté basado en datos transformados y cada vez que se obtengan valores estimados éstos serán transformados hacia atrás para volver a su escala original.

De lo anterior, se puede decir que para aplicación de la simulación secuencial gaussiano será importante realizar una exploración a los datos o intensidades de homicidios para conocer el tipo de distribución de probabilidad que tiene dichos datos. Esta aplicación se hace en dos tipos, una donde se calcula la función de semivariograma sin aplicarle ninguna transformación a los datos y otra utilizando un método de transformación de tipo gaussiana. Finalmente comparar resultados de dichas simulaciones.

Las herramientas que se utilizarán para la realización de esta simulación o generación de escenarios delictuales en El Salvador son las siguientes, dependiendo su aplicabilidad:

- a) Análisis Exploratorio de Datos Espacio-Temporal. Lenguaje de Programación R con paquetes mapprotools, spdep, rgdal, geoR, sp, foreign, RColorBrewer y classInt.
- b) Transformación anamorfosis de los datos: gslib, nscore y backtr.

- c) Estimación de parámetros de la función de covarianza: Lenguaje de Programación R con paquetes `mapproj`, `spdep`, `rgdal`, `geoR`, `sp`, `foreign`, `RColorBrewer` y `classInt`.
- d) Simulación geoestadística producto-suma: `gaussim`.
- e) Generación de Escenarios Delictuales: Lenguaje de Programación R con paquetes `mapproj`, `spdep`, `rgdal`, `geoR`, `sp`, `foreign`, `RColorBrewer` y `classInt`.

En la aplicación de la simulación secuencial gaussiano mediante el software `gussim` (paquete desarrollado por Edinburgh Geoestadística Espacio-Temporal) fue necesario definir una estructura de datos para tal utilización. La estructura de datos se muestra en la tabla 3.3 y especifica la variable de interés y la secundaria, además de las coordenadas y el tiempo. Ese tipo de estructura es la que se maneja para modelar modelos donde incluya el espacio y el tiempo.

OID	SID	East	North	Tiempo	Homicidio	Secundaria
1	1	405188.291	315421.868	1	87	0
2	1	405191.611	315420.388	2	87	0
3	1	405193.101	315421.128	3	47	0
...	...	...	...	...	...	...
1571	262	608412.411	268572.108	5	7	0
1572	262	608412.381	268569.938	6	2	0

Tabla 3.3: Estructura de Datos de Homicidio Espacio-Temporal en El Salvador.

Es importante además de definir la estructura de datos, saber el tipo de proyección geográfica a trabajar. En este caso, se tuvo que realizar un procesamiento del sistema de proyección para las coordenadas. Este procedimiento se generó en el software R para pasar de una proyección UTM (Universal Transverse Mercator) con NAD27 (Norte América Datum 1927) a una proyección UTM con Datum WGS1984. Ejemplo, transformar la coordenada (405188,315122) a (188522, 1544831). Esta transformación se decidió hacerlo por el hecho de que se trabajara con una proyección mundial y no una específica como la que generalmente se trabaja en El Salvador.

Posterior a ello fue necesario también generar una grilla o archivo grid (cuadrícula) para utilizarlo en la predicción o simulación. Este tipo de archivo igualmente como se transformó en R la proyección, fue posible crear dicho archivo en el paquete R, generando así una cuadrícula aleatoria tanto para realizar las simulaciones correspondiente al período de análisis (año 2003=1, 2004=2, ..., 2008=6) y otra **cuadrícula** o archivo grid para la simulación o predicción del año siguiente o de años siguientes como pueden ser 2009=7 y 2010=8. En la figura 3.5 se muestra el

mapa generado a partir de la simulación de la grid en El Salvador con proyección UTM Zona 16 Datum WGS84.

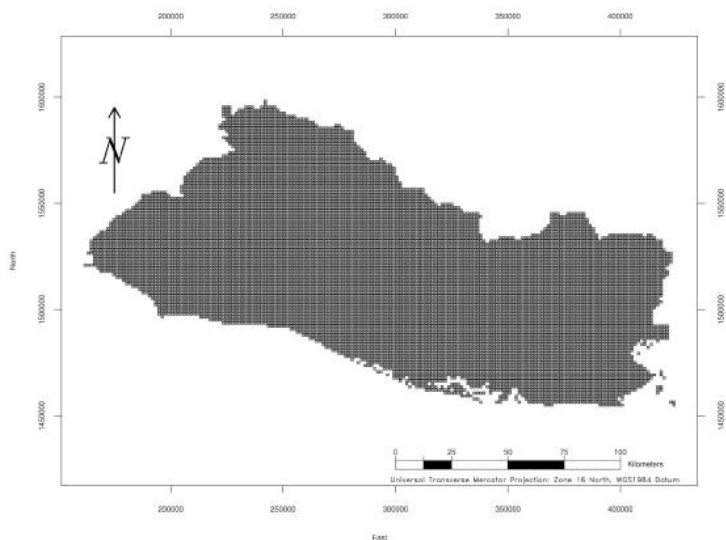


Figura 3.5: Simulación de la grid en El Salvador.

En la tabla 3.4 se muestra la estructura de datos para el archivo grid. Este archivo incluye las coordenadas y las dos variables, una de interés y otra secundaria. Como nuestro interés es simular la criminalidad en El Salvador y por el hecho de disponer información desde el año de 2003 hasta el 2008 es posible generar una simulación. Fuera conveniente además de incluir esta variable de interés, el considerar otra que esté grandemente correlacionada con ella como por ejemplo, la pobreza, niveles socio-económicos, narcotráfico, etc.

East	North	Tiempo	Secundaria
405188.291	315421.868	1	0
405191.611	315420.388	1	0
405193.101	315421.128	1	0
...	...	...	...
608412.411	268572.108	6	0
608412.381	268569.938	6	0

Tabla 3.4: Formato de la generación aleatoria del archivo grid.

### **3.2.2 Análisis Estructural Espacio-Temporal.**

Como el objetivo nuestro es generar un escenario delictual en El Salvador, se necesita grandemente de un análisis estructural que pueda por medio de ello saber cuál es la distribución de los datos espacio-temporales de las intensidades de homicidios y de conocer si existe una variabilidad espacio-temporal. En todo caso, el paso a seguir mediante este análisis requiere del estudio de su distribución en un primer momento, para luego así, pasar al análisis local de las intensidades de homicidios, es decir, contrastar por medio del Test de I de Moran (con el Scatterplot) si el proceso estocástico espacio-temporal del crimen en el país presenta una estructura de dependencia o autocorrelación espacio-temporal para tener la evidencia de poder conocer los grupos que presentan una baja, media y alta intensidad en la ocurrencia del homicidio en los diferentes municipios de El Salvador.

Importante además que se relacione las intensidades de homicidios con el coeficiente de variación, esto, para saber qué tipo de relación existente se encuentra, es decir, conocer si a mayor coeficiente de variación existe en un municipio en particular, cómo será su intensidad, si tendrá mayor o menor intensidad.

#### **Distribución de Datos Espacio-Temporal.**

Iniciamos el análisis mediante el mapa promedio de homicidios ocurridos en El Salvador durante el período de 2003 hasta 2008 (ver figura 3.6). Se muestra claramente la distribución de las intensidades de homicidios, donde 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, ubicado en la zona occidental del país, Sonsonate y Acajutla, ubicados en la zona sur-occidente y San Miguel, ubicado en la zona oriental del país. Dichos municipios son los que generan una zona de influencia para que otros municipios aledaños a ellos experimenten un comportamiento similar.

Es de considerar además que hay municipios que aunque estén aledaños a otros donde si existe una intensidad mayor, su comportamiento temporal es bajo. Es aquí donde al realizar un contraste local, se podría llegar a la conclusión que en tal efecto, el fenómeno es causado no solo por la naturaleza en la cual ocurren los homicidios, sino más bien, en la existencia de otros factores o variables que causan tal efecto.

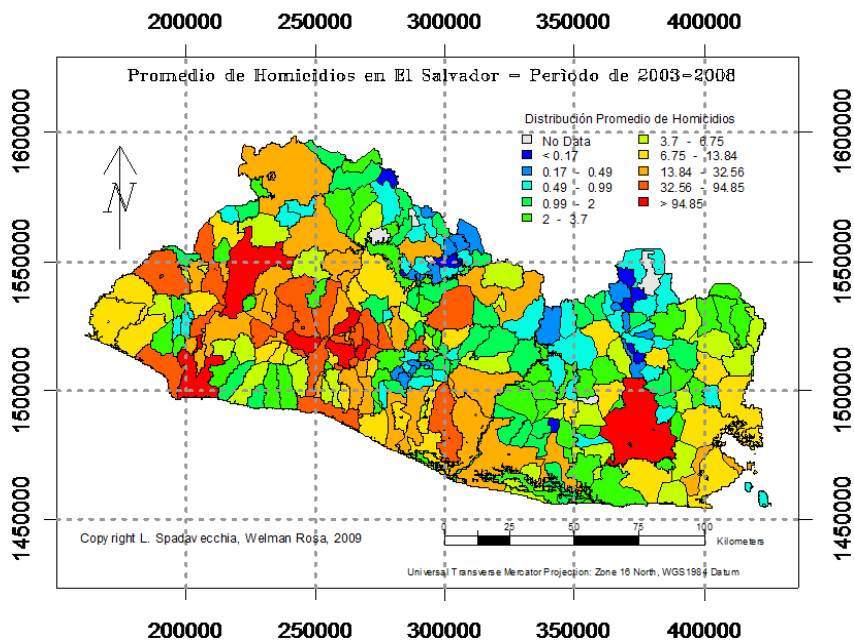


Figura 3.6: Promedio de Homicidios en El Salvador durante los 6 años (2003 a 2008).

En la figura 3.7 se muestra el coeficiente de variación para las intensidades de homicidios ocurridos en El Salvador, y tal como se observa, efectivamente, a mayor coeficiente de variación en cierto municipio menor es su intensidad en la ocurrencia del homicidio. Si analizamos los municipios antes mencionados como más violentos en el país, su coeficiente de variación es abajo del 20%. Mientras que aquellos que han experimentado un comportamiento de menos intensidad, provocan que su coeficiente sea mayor. Los municipios que experimentan un mayor coeficiente están más concentrados en la zona norte del país, específicamente en el departamento de Chalatenango y gran parte de municipios al norte del departamento de Morazán.

Por otro lado, al analizar su distribución territorial, el fenómeno presenta una variabilidad en cuanto a la ocurrencia del homicidio en municipios cuando se alejan de la mayor intensidad promedio homicidio, provocando de esta manera una heterogeneidad geográfica. Esto se observa claramente en los gráficos de distribución de tales intensidades vista desde una zona este y una zona norte (ver figura 3.8). Al analizar el gráfico con coordenada Este (East) la distribución de tales intensidades desde la más mínima hasta la más intensa, se observa que los municipios que están dentro de la franja roja son efectivamente los que se han identificado como más violentos. Mientras que los de menor intensidad caen en la franja de color menos intenso, un color celeste claro. Ahora bien, si se analiza desde una vista norte (North) el fenómeno no cambia en cuanto a la dirección que se analice, haciendo que los municipios con intensidad moderada se agrupen en una franja con color verde. Gran parte de estos municipios están agrupados dentro de franjas con color verde y amarillo. Este agrupamiento será contrastado mediante el Test de I de Moran que se analiza en el siguiente apartado.

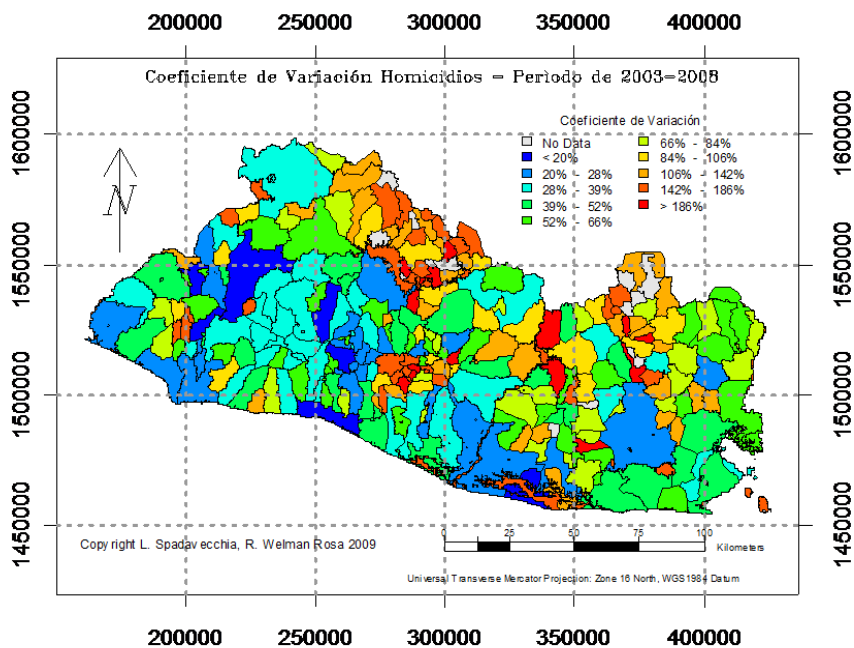


Figura 3.7: Coeficiente de Variación de los homicidios ocurridos en El Salvador en el período de 2003 a 2008.

Una vez analizado el comportamiento del homicidio en cuanto a la dirección Este y Norte será importante estudiar su distribución. En el gráfico donde se presenta el histograma se puede observar que el fenómeno de la criminalidad al transformarlo mediante el logaritmo su distribución no es tan normal como parece, pero da un síntoma de tendencia a la concentración de municipios con menor y mayor intensidad en la ocurrencia del homicidio. La máxima intensidad es de 431 homicidios y cae en el color rojo reflejado en la franja del gráfico. Se puede decir que la distribución es sensible a valores pequeños extremos y a otros con pequeña intensidad. Es aquí donde surge la necesidad de que a la hora de poder aplicar un modelo se tenga presente de buscar una transformación a los datos de manera tal que los resultados sean tal cual se han observado en la realidad. La distribución de probabilidad para este tipo de datos tiende a tener una distribución de tipo exponencial. Una idea clara para ir definiendo los posibles modelos teóricos existentes en la literatura para el cálculo de la función de covarianza o variograma.

Por otro lado, en cuanto al efecto que puede existir en lo que respecta a la separación espacial y temporal de las intensidades de homicidios, es importante que se logren identificar qué municipios son los que originan el efecto de agrupamiento, causando de esta manera lo que se le conoce como una continuidad espacial y temporal en dichas intensidades. Por ello, es necesario mediante este análisis conocer la heterogeneidad espacial en lo que se refiere a aquellos municipios que experimentan una mayor intensidad en la ocurrencia del homicidio. Tal es el caso del municipio de San Salvador, que generalmente el fenómeno está mayormente representado en zonas centrales y occidentales del país, es decir, su incidencia es originada de una forma tan

intensa que el resto de municipios que colindan en zona oriental. No obstante, hay que observar que el municipio que genera un efecto de separación en cuanto a la estacionariedad espacial de los datos es San Miguel. Este efecto se verá cuando se analice la función de semivariograma tanto para la parte espacial y temporal. Importante realizar este tipo de análisis ya que por medio de él es posible identificar a qué distancia se genera tal efecto denominado como “efecto hole”.

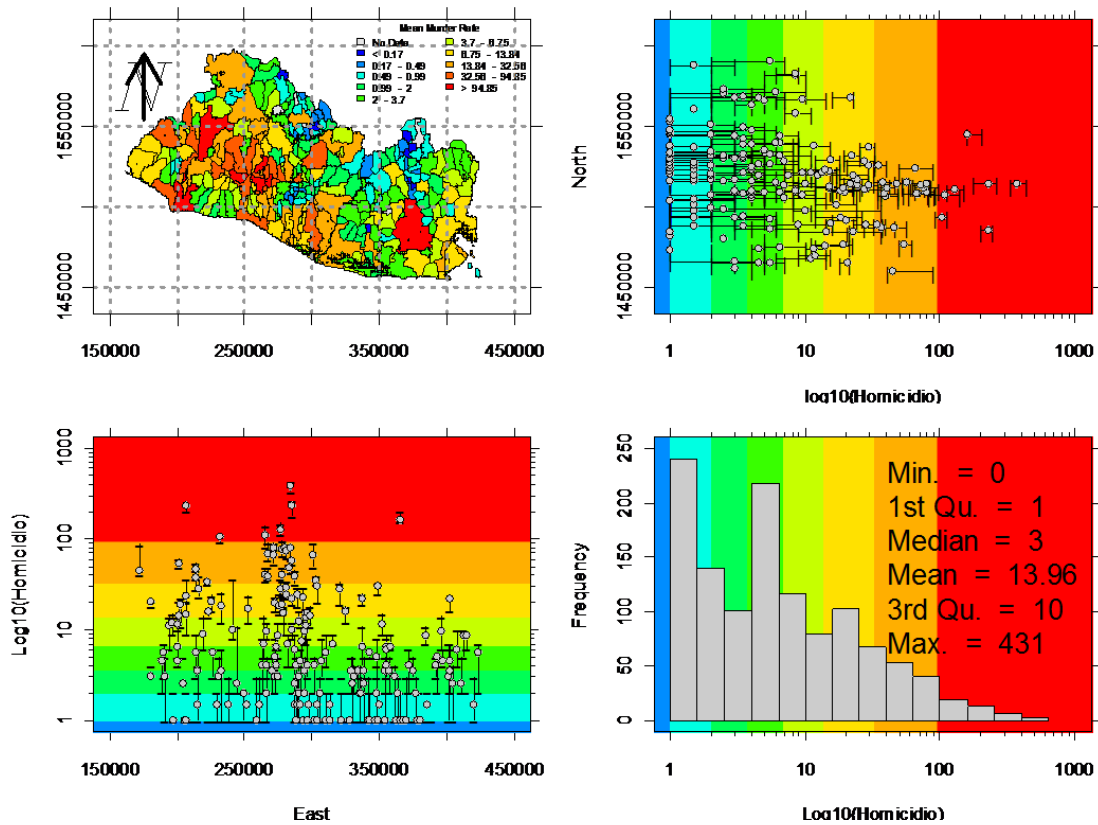


Figura 3.8: Distribución Espacial de las Intensidades de homicidios altas y sus variaciones. Altas intensidades están asociadas con bajas varianzas y vice-versa.

### 3.2.3 Clasificación Espacial de Homicidios en El Salvador.

En este apartado se analiza la clasificación espacial de los homicidios ocurridos en El Salvador. Se estudia el fenómeno mediante el Test I de Moran, visto desde una perspectiva espacio-temporal con el estadístico que mide la autocorrelación o dependencia espacial reflejado en el gráfico “Scatterplot”. También, considerando que los datos originales han sido transformados con el logaritmo, necesario para este tipo de test, ya que la mayoría de intensidades de homicidios se observan a la izquierda de la distribución.

En la figura 3.9 se muestra el indicador local de autocorrelación espacial y la significancia de agrupamiento identificadas por medio del estadístico I de Moran. Se observa que el estadístico es

signficativo, ya que tiene un valor de 0.54. Infiriendo de esta manera que el fenómeno presenta una dependencia espacial y con tendencia al agrupamiento. Este agrupamiento está distribuido por los cuatro cuadrantes del gráfico de Scatterplot, lo cual, el indicador local, se analiza como un coeficiente de regresión de la línea recta sobre la cual caen las intensidades de homicidios. El grupo que presenta una mayor intensidad están distribuidos en el primer cuadrante del gráfico, luego le sigue el grupo donde su intensidad es de bajo a alto, identificado en el segundo cuadrante; en lo que respecta al tercero y cuarto, tienen una peculiaridad, y es el caso en que para los municipios agrupados en el tercero, estos están experimentando una intensidad muy baja. Los del cuarto cuadrante se podría decir que han pasado de una categoría de alta a una categoría menor en lo que respecta a la intensidad de homicidios. Todo esto se puede observar representativamente en el mapa de agrupamiento espacial, donde se observa la significancia existente en aquellos municipios que al compararlos con sus vecinos próximos, estos experimentan una significancia en cuanto a la conformación de los grupos (ver en **negrita** tal observación).

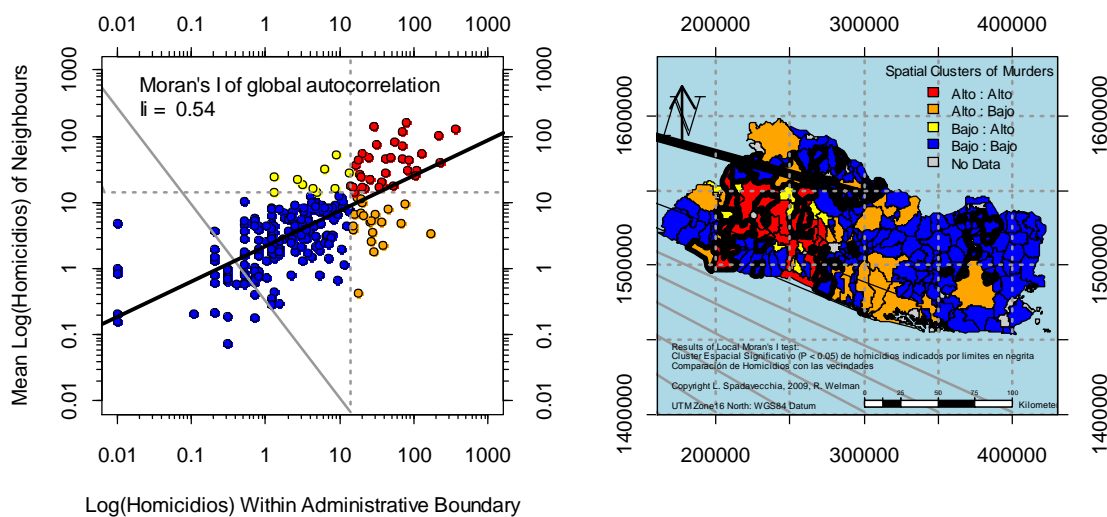


Figura 3.9: Indicador Local de autocorrelación espacial. Grupos significantes están bordeados en color negra, los colores indican el tipo de agrupamiento observado.

### 3.2.4 Estimación de Parámetros de Modelos Espacio-Temporal.

Predecir o generar un escenario de homicidios en El Salvador es necesario realizarlo mediante un modelo estocástico espacio-temporal como el de covarianza no-separable producto-suma. Un tipo de funciones nuevas y con gran precisión en cuanto a la predicción espacio-temporal. Para ello, es importante conocer la naturaleza del fenómeno al cual se le aplicará el modelo estocástico bajo condiciones que necesariamente tienen que considerarse a la hora de generar una predicción. Generalmente en la estadística clásica en lo que respecta a la predicción de series de tiempo, el método a seguir para llegar a fase de predicción incluye: identificación, estimación, chequeo y



predicción. En lo que se refiere al caso de predicción geoestadística espacio-temporal, la situación se complica más cuando agregamos conjuntamente el espacio y el tiempo, pero la filosofía que se utiliza en geoestadística espacio-temporal es equivalente a la que se utiliza en series cronológicas para predecir un fenómeno.

Una metodología para predecir espacio-temporal va desde un análisis estructural hasta una fase de predicción donde incluye una serie de herramientas que hasta la fecha están en desarrollo, pero que se hace el esfuerzo en aplicar unas disponibles y aunque a pesar de ser tediosas brindan resultados de los que uno espera.

La fase de poder implementar estas técnicas geoestadísticas son las siguientes:

Caso 1: Un análisis geoestadístico sin utilizar una transformación anamorfosis a los datos.

Fase 1: Cálculo de la función de covarianza no-separable vía producto-suma.

Fase 2: Simulación Secuencia Gaussiana vía producto-suma.

Caso 2: Un análisis geoestadístico utilizando una transformación anamorfosis a los datos.

Fase 1: Transformación anamorfosis a los datos originales.

Fase 2: Cálculo de la función de covarianza no-separable vía producto-suma.

Fase 3: Simulación Secuencial Gaussiana vía producto-suma.

Fase 4: Transformación Gaussiana Inversa (back-transformation).

### **Caso 1 sin Transformación anamorfosis.**

Para este caso se inicia calculando la función de semivariograma utilizando el modelo de covarianza no-separable producto-suma. Este cálculo es generado a partir de una base de datos espacio-temporal en el programa R. El resultado a partir de este caso donde no se considera una transformación de tipo gaussiana permitirá determinar si en la separación espacial existe un efecto "hole". También, para poder realizar simulaciones de manera que por medio de esto se genere una predicción espacio-temporal del crimen en El Salvador a un tiempo futuro (en este caso, el año 2009).

En la figura 3.10 se presenta el semivariograma para la parte espacial y temporal. Tal como se observa, la separación para ambos componentes muestra la estructura de autocorrelación o dependencia espacio-temporal. Específicamente en la separación espacial presenta un efecto "hole" a una distancia de aproximadamente 100-120 kilómetros. A medida que la distancia se aleja su dependencia espacial predomina, existiendo de esta manera, que el fenómeno de la criminalidad sea explicado mediante un modelo exponencial (ver tabla 3.11). La estimación de dichos parámetros para el modelo exponencial en la parte espacial son: una contribución 1122.73 y un rango de 3969.01; para el modelo con efecto "hole" son: una contribución de 233.95 y un

rango de 42835.95. En lo que respecta al parámetro sill, su valor es de 13.56.68. Por otro lado, para la parte temporal se observa que el fenómeno tiende a ser lineal respecto al tiempo, este tipo de modelos son muy difíciles de poder modelar, por ello, al generar un semivariograma con un modelo exponencial su contribución llega a ser igual a 223.65 con un rango de 22.98, un valor sill igual a 243.54, presentando además, un efecto pepita (nugget) de 19.89. La sill global del modelo de covarianza no-separable es igual a 1384.22 y las estimaciones de los coeficientes para la parte espacial, temporal e interacción es: 0, 0.84, 0.11, respectivamente.

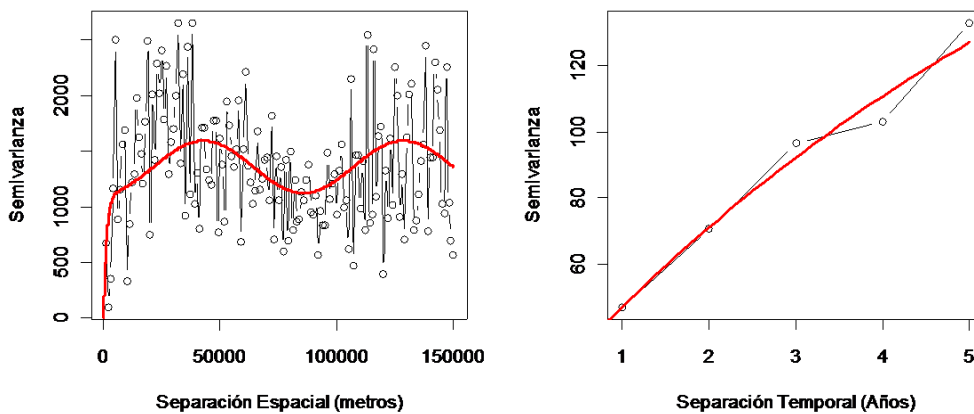


Figura 3.10: Semivariogramas espacio-temporal para los homicidios ocurridos en El Salvador. Sin transformación en los datos.

Con los parámetros y coeficientes estimados es posible generar simulaciones mediante un tipo de simulación geoestadística como la secuencial gaussiana. Considerando aquí que debido al efecto pepita identificado en el modelo y la no contribución espacio-temporal (esto por tener un valor de 0 para la interacción espacio-temporal del modelo) las estimaciones generadas mediante tal simulación serán muy sensibles a este tipo de comportamiento. Advirtiéndolo de esta manera, que la estimación de los homicidios en los diferentes municipios se distribuirá considerando tal “efecto hole o periodicidad espacial” y la no interacción.

Las simulaciones son realizadas con un paquete que considera estos parámetros y un condicionamiento de los datos para tal estimación. Los parámetros que son incluidos son los que conforman los modelos identificados y el acondicionamiento depende en gran medida de la capacidad del procesador en el ordenador para la simulación. En nuestro caso, el número de condicionamiento de datos es igual a el número de vecinos + (vecinos \* ventana) (LukeSpadavecchia, 2007); por ejemplo, si se considera 62 vecinos, y 2 para Windows, se generarían un total de 186 condicionamiento de datos. Un valor suficiente para la generación de las estimaciones debido que son 262 municipios en total.

```

*** Covariance model is permissible. ***

Spatial Model:
-----
Exponential: Contribution = 1122.73 Range = 3969.01
Hole Effect:Contribution = 233.95 Range = 42835.95

Temporal Model:
-----
Nugget Effect:Contribution = 19.89
Exponential:Contribution = 223.65 Range = 22.98

Sills:
-----
Spatial Partial Sill = 1356.68
Temporal Partial Sill = 243.54
Global Sill = 1384.22

Scaling constants (k):
-----
k[1] = 0
k[2] = 0.84
k[3] = 0.11

```

Figura 3.11: Modelo de covarianza no-separable producto-suma para realizar las simulaciones.

En la figura 3.12 se presentan cuatro realizaciones de las intensidades de homicidios ocurridos en El Salvador. Mencionando en este caso que el número de simulaciones para tal cálculo fue de 100, lo cual, para efectos de ejemplo se han representado únicamente cuatro simulaciones. La clasificación utilizada para la representación en el mapa fue a partir de la función `classIntervals` del paquete `classInt` del programa R, en dicha función existe una variedad de algoritmos para la clasificación geográfica de la información como: el Fisher, quantiles, k-media, entre otros. El que se ha considerado para tal representación ha sido mediante el algoritmo Fisher. Se mira que la distribución muestra resultados que no son muy representativos en cuanto a la realidad con la que ocurren los homicidios en los municipios, esto, esencialmente en aquellos que son más sensibles o en los que no experimentan una intensidad mayor en la ocurrencia del homicidio.

El problema aquí radica en que se está trabajando con un método de simulación que el requisito es que los datos tienen que ser necesariamente de tipo gaussiano, cosa que en la realidad es difícil que fenómenos como la criminalidad sean de este tipo. Lo sensible que son los datos por existir unos muy extremos origina que en el momento de la clasificación busquen y se agrupen en los que mayormente ocurren los homicidios, es decir, aquellos que experimentan frecuentemente mayor intensidad; por ejemplo, en la zona identificada como más violenta del país, el mayor porcentaje de municipios experimentan homicidios por arriba de los 40, ocasionando así, que aquellos donde experimentan una intensidad menor, sean clasificados sobre este conjunto de municipios violentos.

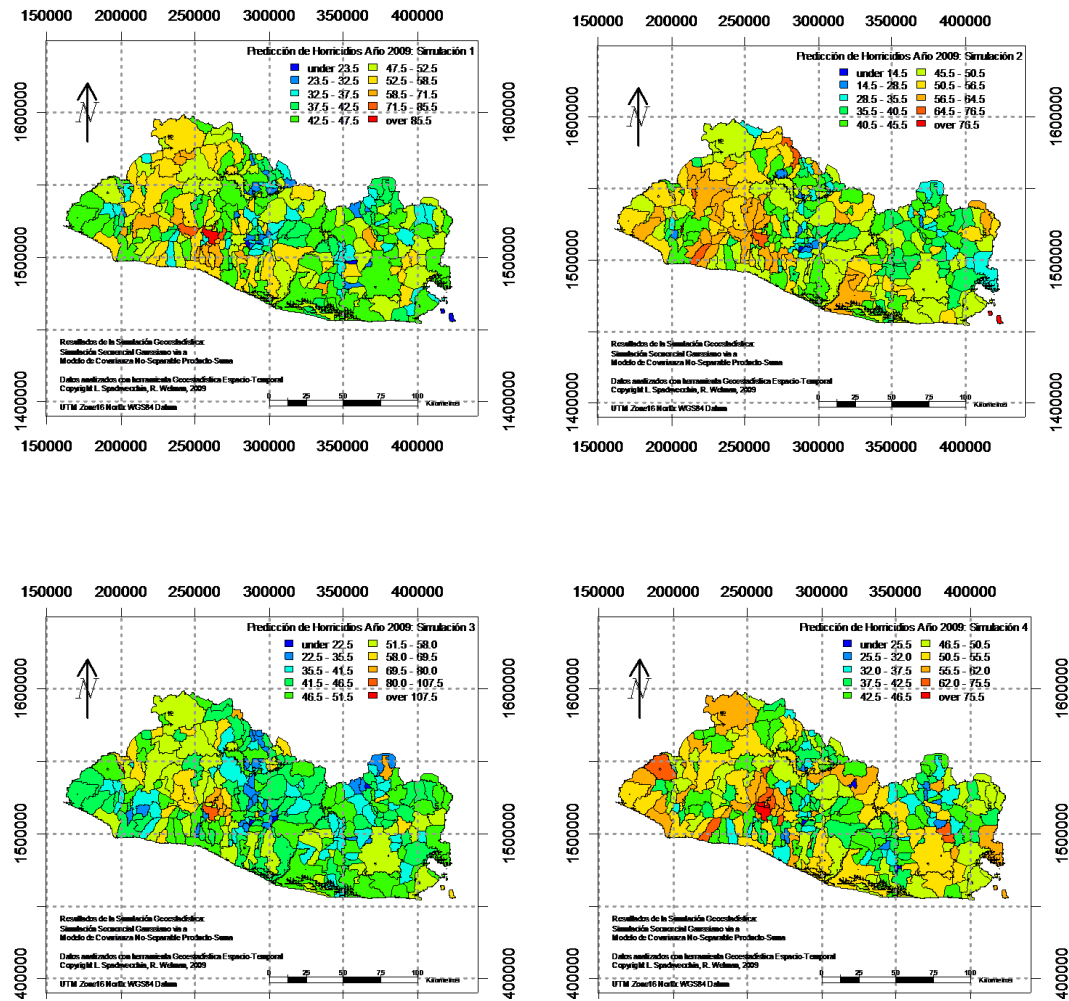


Figura 3.12: Simulaciones para el año 2009 sin realizarle una Transformación Gaussiana.

Es de mencionar además que el condicionamiento de los datos en lo que respecta a la estimación se consideró el kriging ordinario y no el simple, ya que por lo general el tipo de simulación secuencial gaussiano utiliza este tipo de estimadores.

Ahora bien, la pregunta es ¿Cuál simulación es la más idónea? La respuesta es el promedio de tales simulaciones. Se elige el promedio y varianza debido al hecho de ser equivalentes a una estimación mediante un estimador kriging.

En la figura 3.13 se muestra el escenario delictual de la intensidad de homicidios en El Salvador para el año 2009 con un promedio de 100 simulaciones vía secuencial gaussiano y utilizando el modelo de covarianza no-separable producto-suma. Tal como se observa, la distribución de los homicidios acorde al algoritmo de clasificación Fisher, su distribución es más representativa hasta cierto punto para municipios donde por lo general ocurren frecuentemente y con mayor

intensidad los homicidios. Municipios como Santa Ana, Colón, San Salvador, Ilopango, Tecoluca y San Miguel experimentan una mayor intensidad de homicidios para el año 2009.

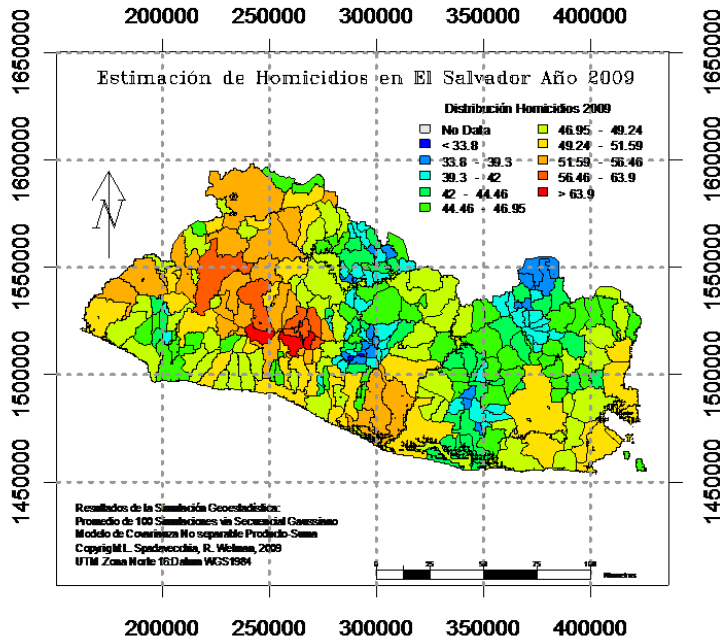


Figura 3.13: Escenario de homicidios año 2009 sin realizarle una Transformación Gaussiana.

### Simulación Geoestadística Espacio-Temporal: Caso 2 con Transformación anamorfosis.

La simulación secuencial gaussiana como se mencionó anteriormente es una técnica geoestadística robusta, aplicable a la generación de realizaciones de variables continuas, y por ende, requiere que al momento de abordar este tipo de técnicas se trabaje con la transformación gaussiana o normal de los datos.

Algunas propiedades de la simulación geoestadística son:

- 1- La variable simulada posee:
  - Idénticos valores en las localizaciones de los datos experimentales (datos reales).
  - Igual valor de la media, covarianza o variograma e histograma que los datos experimentales (datos reales).
- 2- Se reproducen las fluctuaciones que ocurren en la realidad.

- 3- El fenómeno simulado tiene la ventaja de ser conocido en todos los puntos y no solamente en puntos experimentales. Con frecuencia al fenómeno simulado se le llama “modelo numérico” del fenómeno real.

Los pasos para llevar a cabo una simulación secuencial gaussiana son los siguientes:

- 1- Transformar los datos en un Normal Score (conteo normal) antes del cálculo del semivariograma.
- 2- Realizar el análisis variográfico de los valores transformados (normalizados).
- 3- Simular los valores en un “espacio normal o normalizado”, construyendo mediante kriging ordinario la simulación secuencial.
- 4- Transformar los datos condicionalmente simulados (back-transform).
- 5- Chequear los resultados: histograma.

Mediante estos pasos descritos anteriormente se realiza la simulación secuencial gaussiana a los datos de homicidios en El Salvador considerando el caso de transformación anamorfosis. En tal sentido, seguiremos paso a paso para lograr los resultados esperados.

En la figura 3.16 se muestra la transformación gaussiana y su inversa, y tal como se observa en el histograma para los datos de homicidios durante los seis años, su distribución es asimétrica. Procediendo a transformar los datos mediante el programa NSCORE de GSLIB, se calcula el conteo normal (Normal Score) para la serie de datos y se obtiene un fichero de datos el cual contiene una columna adicional con el conteo. Tal como se muestra en la tabla 3.5.

East	North	Tiempo	Homicidio	Nscore.homicidio
405188.291	315421.868	1	87	1.81621
405191.611	315420.388	2	87	1.82457
405193.101	315421.128	3	47	1.45040
...	...	...	...	...
608412.411	268572.108	5	7	.52825
608412.381	268569.938	6	2	-.27365

Tabla 3.5: Forma del fichero generado por nscore de gslib donde se muestra la variable transformada por conteo normal.

Con la variable transformada se procede a verificar su histograma para conocer que efectivamente presenta una forma de tipo gaussiana o normal. Al ver el histograma en la figura 3.16 se observa la

clara normalidad en los datos de homicidios. Mediante este resultado, ahora se procede a realizar el cálculo de la función de covarianza utilizando el modelo producto-suma. Este procedimiento se realiza con el paquete R considerando la variable transformada y la cuadrícula o archivo grid (donde representa todos los puntos generados aleatoriamente en el territorio de El Salvador) para estimar los parámetros del modelo. En la figura 3.15 se muestra el modelo de covarianza no-separable vía producto-suma válido para los datos de homicidios transformados por un conteo normal. Tal como se observa, las constantes de escala que representan en el modelo para la parte espacial, temporal e interacción fueron generados.

```

*** Modelo de Covarianza es Válido ***

Modelo Espacial:
-----
Exponencial:   Contribución = 0.92   Rango = 22745.89
Efecto Hole:   Contribución = 0.07   Rango = 61125.19

Modelo Temporal:
-----
Efecto Nugget: Contribución = 0
Exponencial:   Contribución = 0.22   Rango = 1.24

Sills:
-----
Sill Parcial Espacial = 1
Sill Parcial Temporal = 0.22
Sill Global           = 1.01

Constantes de Escala (k):
-----
k[1] = 0.94
k[2] = 0.79
k[3] = 0.07

```

Figura 3.14: Modelo de covarianza no-separable producto-suma con transformación anamorfofis.

Este resultado es importante para la realización de las simulaciones a partir de un modelo teórico existente en la literatura geoestadística.

En la figura 3. 15 se muestra el gráfico del semivariograma para las dos componentes, espacial y temporal. Se observa que para la separación espacial la autocorrelación es evidente en cuanto comparamos las intensidades a medida la distancia se incrementa, mostrando de esta manera una estacionariedad espacial. El modelo teórico que ajusta a los datos es el exponencial. Mientras tanto, para la parte temporal, el comportamiento es similar que el espacial. Infiriendo de esta manera, que los modelos para este conjunto de datos son el exponencial. Por otro lado, se observa también la evidencia de un efecto pepita o nugeet para la parte temporal y un efecto hole para la parte espacial.

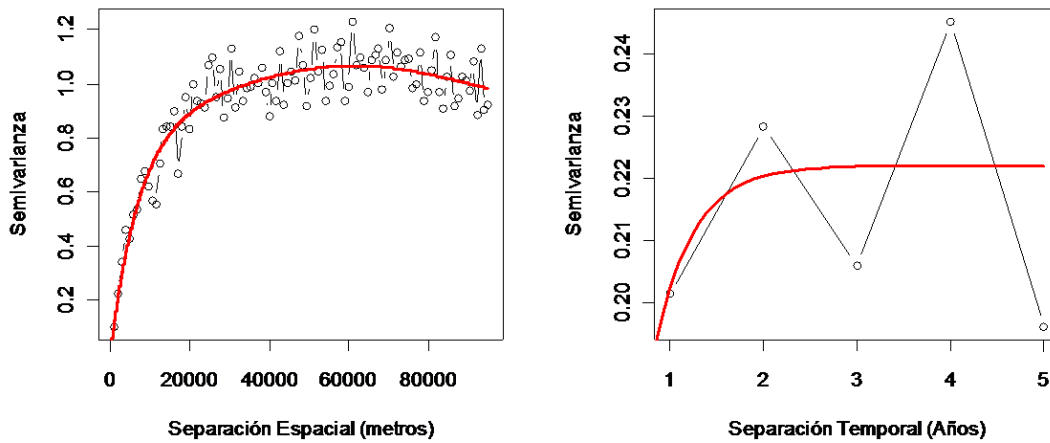


Figura 3.15: Semivariograma espacio-temporal para los homicidios ocurridos en El Salvador. Con transformación en los datos.

Con estos resultados se procede a la simulación geoestadística utilizando el paquete gaussim desarrollado por LukeSpadavecchia (School of GeoSciences, Institute of Atmospheric and Environmental Science, University of Edinburgh, UK). El paquete contiene el algoritmo de simulación secuencial gaussiana vía producto-suma, y que genera por medio de él realizaciones artificiales de la realidad con la que han ocurrido los homicidios (considerando su transformación nscore). Estas simulaciones generadas con un máximo de 100 simulaciones son entonces nuevamente transformados (“backtransformed”) en el programa backtr de gslib empleando una tabla (lookuptable) que es creada cuando inicialmente se transforman los datos en el programa nscore. En la tabla 3.6 se muestra la forma del fichero que genera tal programa en gslib, donde en la columna uno aparece los valores reales de las intensidades de homicidios y en la columna dos aparece el valor de la transformación correspondiente para cada valor real.

Homicidios	Nscore.Homicidios
0	-3.41572
1	-.71612
...	...
382	2.94972
431	3.41572

Tabla 3.5: Tabla lookup generado por el programa backtr de gslib.



Con los dos archivos generados por nscore y backtr se procede a transformar simulación por simulación para representar tal estimación en un mapa el cual se generará con el paquete R. En la figura 3.16 se muestra la verificación de que las simulaciones realizadas mediante el modelo producto-suma con la variable transformada genera un histograma parecido al histograma de la variable transformada de las intensidades de homicidios. Al realizar la transformación gaussiana inversa a las estimaciones producto de la simulación, se observa el histograma a modo de ejemplo para dos simulaciones, su forma equivalente a la escala original de los datos de homicidios.

Una vez se ha realizado la verificación de las propiedades de las realizaciones producto de la simulación secuencia gaussiana. Se procede a generar el escenario de las intensidades de homicidios estimada. El paquete gausim genera dependiendo del número de simulaciones ficheros con formato texto que incluye campos importantes como las coordenadas, el tiempo (año 2009), la media local, la estimación nscore, la estimación backtr, entre otras más. A modo de presentación se muestra una estructura de salida generada por tal programa gausim, el cual se muestra en la tabla 3.6.

East	North	Tiempo	Nscore.estimación	Backtr.estimación
193508	1532837	2009	0.1528	4
382330	1540099	2009	0.0062	3
338756	1521217	2009	0.1124	4
...	...	...	...	...
254512	1541552	2009	1.0777	23
269037	1486358	2009	2.3004	195

Tabla 3.6: Simulación 1 transformada a su escala original luego de la simulación secuencial.

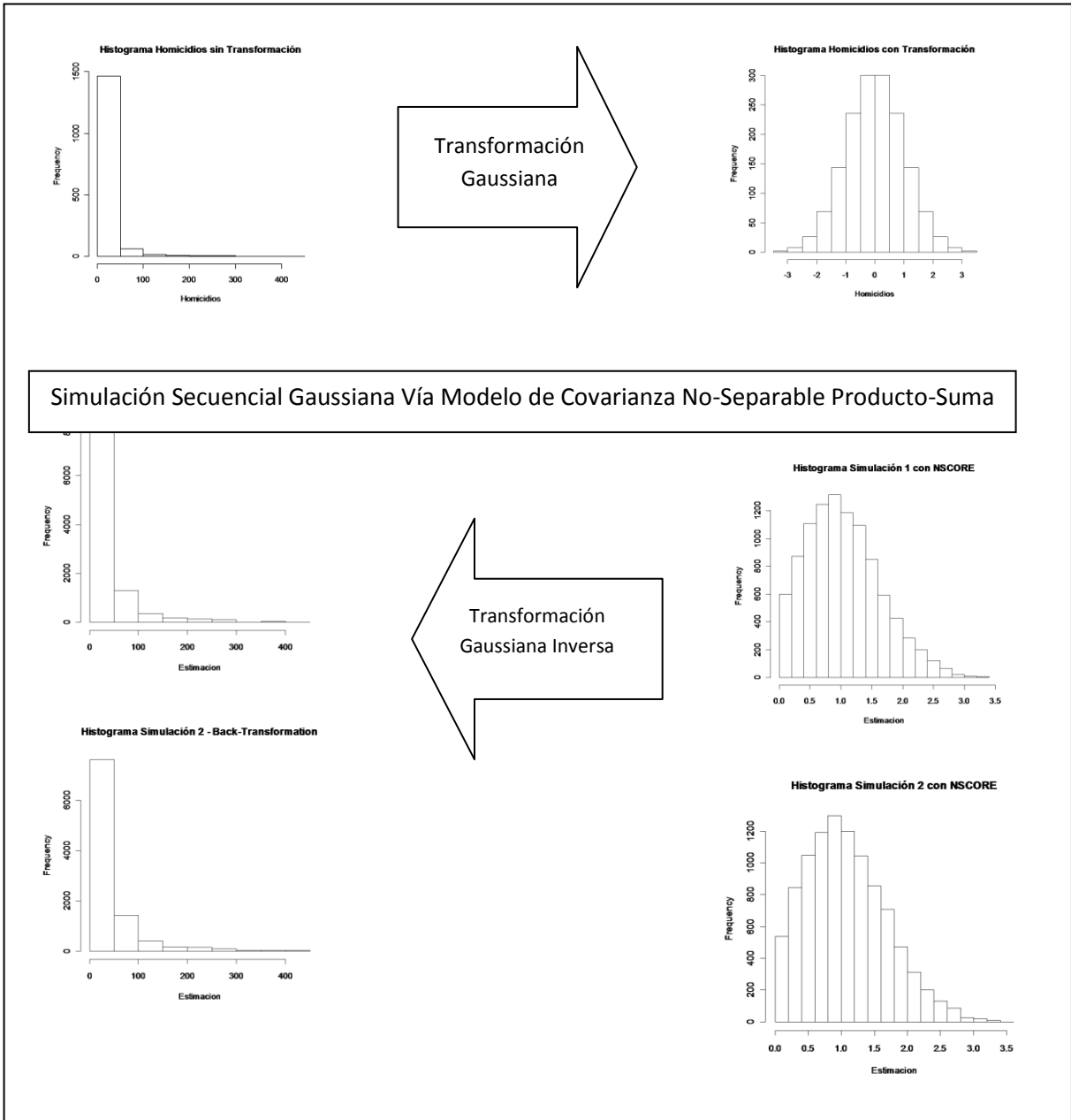


Figura 3.16: Transformación Gaussiana para la realización de las simulaciones de homicidios.

El paso siguiente es la representación de las simulaciones, lo cual, a modo de ejemplo se presentan en la figura 3. 17 cuatro simulaciones producto de la simulación realizada con gaussim. Tal como se observa, en la simulación 1 su distribución de homicidios en El Salvador muestra un comportamiento equivalente en cuanto a la intensidad observada en la realidad. Esta distribución es generada a partir de una clasificación con algoritmo Fisher en el programa R. La mayor intensidad está concentrada en zonas ya antes identificadas (ver figura 3.6, mapa de promedio de homicidios durante 2003 a 2008). De igual forma, se observa en el resto de simulaciones, no así para el caso de la simulación 4, donde por lo general a la hora de clasificar las estimaciones tienen

que ver del tipo de algoritmo de clasificación que se esté utilizando, habría que verificarlo con otro algoritmo de clasificación.

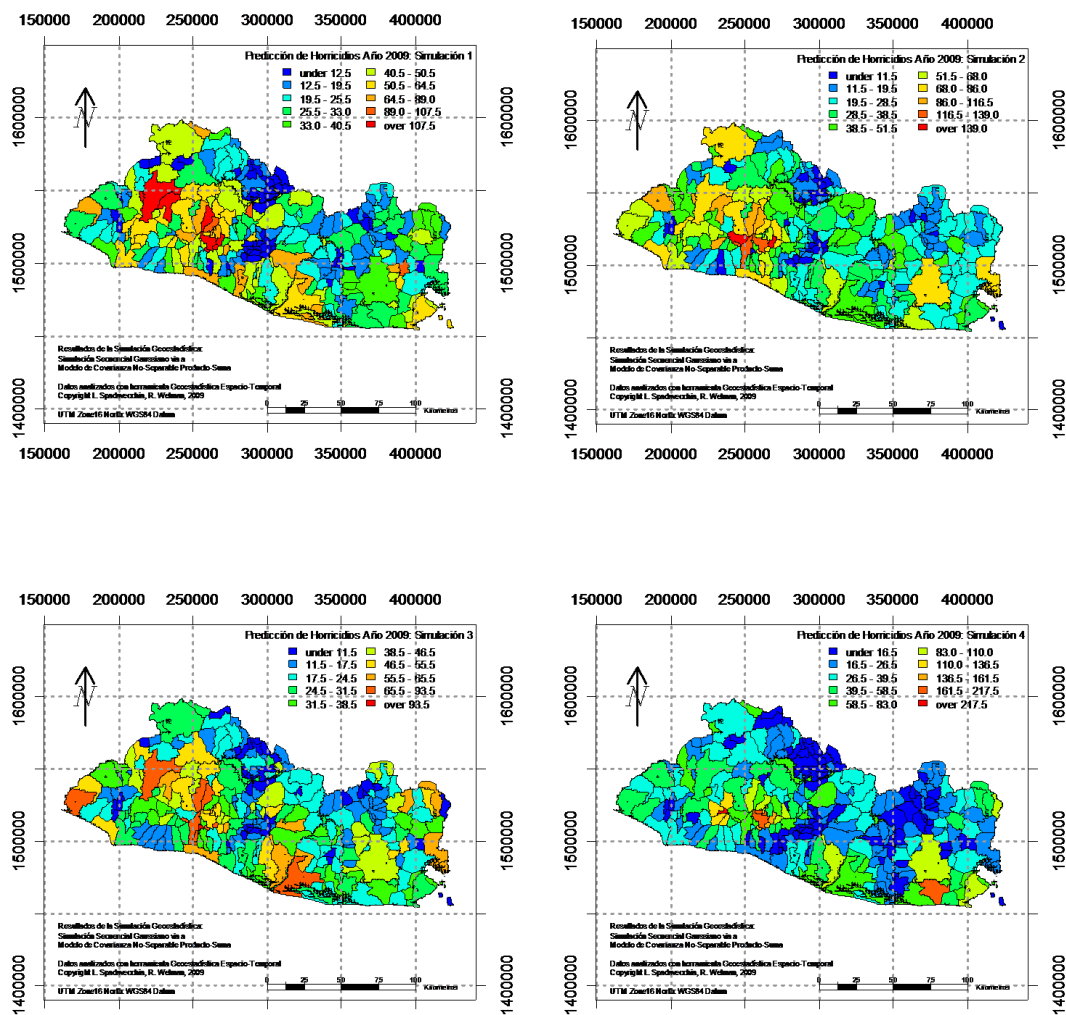


Figura 3.17: Simulaciones para el año 2009 con Transformación Gaussiana Inversa (backtr).

Finalmente se puede observar en la figura 3.18 el escenario delictual de las intensidades de homicidios para el año 2009 como el promedio de 50 simulaciones generadas mediante la simulación secuencial gaussiana, y tal como se muestra, el escenario muestra las mismas concentraciones de homicidios en las zonas que se habían identificado en la clasificación espacial contrastado con el estadístico de I de Moran (gráfico Scatterplot). Se decide la media de tales simulaciones por el hecho que la generación de las realizaciones de intensidades es estimada partiendo del condicionamiento de estimación por medio del estimador de kriging ordinario. El mapa presenta evidentemente aquellos municipios donde experimentarán una intensidad mayor del delito de homicidio para el año 2009.

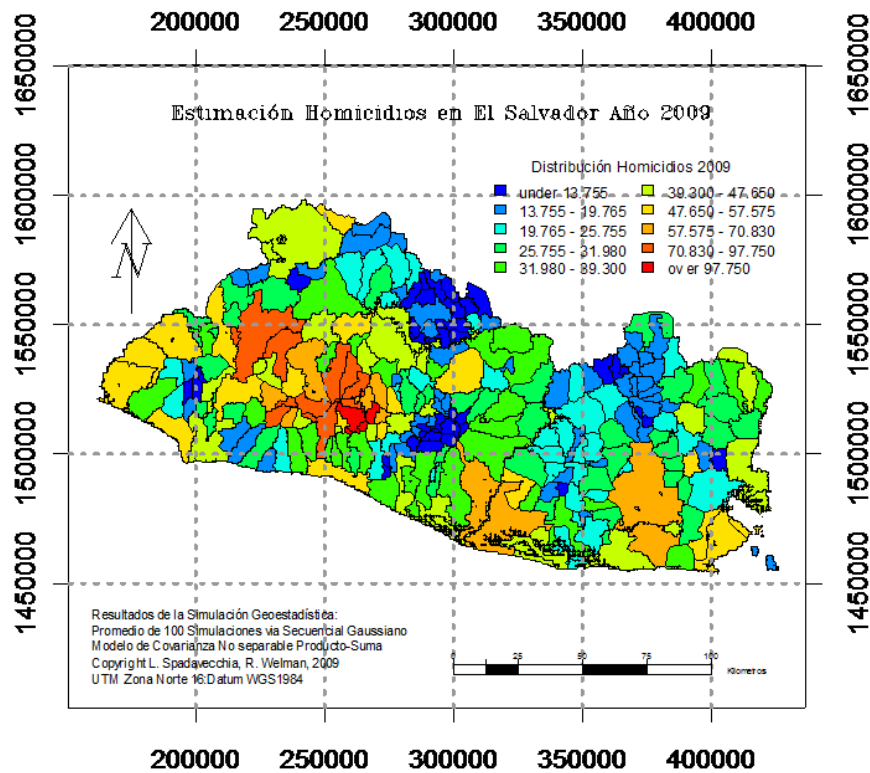


Figura 3.17: Escenario de homicidios año 2009 con Transformación Gaussiana Inversa (backtr).

En resumen se puede decir que los resultados obtenidos mediante una simulación secuencial gaussiana con transformación de conteo normal y basado en la identificación y estimación de un modelo de covarianza producto-suma, la predicción espacio-temporal realizada ha pasado la prueba de fuego, en el sentido de haber comprobado los supuestos que tienen este tipo de modelo geoestadístico espacio-temporal para la realización del escenario delictual del crimen en El Salvador. Considerando además, que a la fecha es un método geoestadístico que está en desarrollo por la comunidad científica, especialmente en algunos países europeos como Inglaterra.

## CONCLUSIONES

Son muchas las conclusiones que se llegan acerca de esta investigación realizada sobre la criminalidad en El Salvador pero se han resumido en cuatro, las cuales se detallan a continuación:

- Se concluye que el fenómeno de la criminalidad en El Salvador en cuanto al análisis estructural espacial presenta un patrón no aleatorio con tendencia al agrupamiento, esto, al analizarlo mediante el análisis de distancia de primer y segundo orden considerando dos tipos de datos como el coordenado o puntual y campo escalar. En cuanto a la interpolación espacial, esto fue posible realizarlo, por el hecho de existir una variabilidad espacial fuerte en el fenómeno del crimen, haciendo de esta manera que la estimación o predicción espacial del crimen identificara por medio del mapa predictivo las zonas más vulnerables en cuanto a la ocurrencia del homicidio. Además, en lo que se refiere a la parte geoestadística espacio-temporal, específicamente en el análisis estructural fue posible identificar aquellos municipios más violentos, esto, visto desde una clasificación espacio-temporal mediante un contraste del estadístico de I de Moran.
- Al realizar una predicción espacial para los años de 2003 hasta 2008, se constató que las concentraciones de intensidades de homicidios en los diferentes municipios de El Salvador son las mismas para todos los años. Aunque la migración del homicidio tiene que ver hasta cierto punto cuando existen zonas de influencia que hacen que pueda desplazarse el fenómeno a municipios donde nunca han experimentado una intensidad mayor o menor de homicidios. Son seis concentraciones más representativas en cuanto a las incidencia del homicidio, distribuyéndose geográficamente tres concentraciones en el área centro-occidente, dos en el área paracentral y una el área oriental del país. La mayor concentración y que abarca territorialmente tres departamentos está situada en el centro del país, específicamente en los departamentos de San Salvador, La Libertad y zona limítrofe con Santa Ana. Mientras tanto, en lo que respecta a zonas de influencia que hacen que municipios menos violentos experimenten una mayor intensidad en la ocurrencia del homicidio son: Sonsonate, Acajutla, Ilobasco, Tecoluca, San Miguel, La Unión y otros municipios situados en la zona norte del departamento de la Unión.
- Partiendo del enfoque Cokriging para la generación de un escenario delictual del homicidio en El Salvador, se observó que la interpolación generó concentraciones pequeñas comparándolo mediante las interpolaciones que se realizaron para cada año con el estimador de kriging disyuntivo. Estas concentraciones mostraron un escenario que al relacionarlo con años anteriores como 2004 y 2006, la intensidad es mayor, concluyendo que el escenario para tal año 2009 mediante este enfoque es semejante al experimentado en el año 2004 y 2005.
- Finalmente, en lo que respecta a la aplicación del enfoque mediante el modelo de lineal de correogionalización producto-suma para generar un escenario delictual del homicidio vía simulación secuencial gaussiano se pudo concluir lo siguiente: – En la simulación geoestadística espacio-temporal vía secuencial gaussiano, fue importante encontrar

herramientas que permitieran llevar una variable regionalizada como las intensidades de homicidios a una variable regionalizada de tipo gaussiano, por el hecho de que el modelo de simulación a aplicar necesariamente la variable de interés tiene que ser de tipo gaussiano. En ese sentido, se hizo el esfuerzo para lograr transformar dichas intensidades de homicidios para la simulación y luego volver a llevarla a su escala original mediante una transformación gaussiana inversa. - El modelo lineal de correogionalización producto-suma a pesar de ser más óptimo en la estimación o predicción cuando la variable de interés es correlacionada con otra secundaria, en el caso nuestro, donde el fenómeno de la criminalidad fue enfocado a tal método geoestadístico y debido a lo escaso de herramientas para su aplicación, se llegaron a resultados como el promedio de 100 simulaciones que muestran claramente la identificación de las mismas concentraciones de intensidades de homicidios observadas en los enfoques anteriores como el Kriging Disyuntivo y Cokriging.

## REFERENCIAS BIBLIOGRAFICAS.

1. Anselin, L. 1995. Local indicators of spatial association - LISA. *Geographical Analysis* 27: 93–115.
2. Castillejo, Maximiano. (2007). *Teoría Geoestadística aplicada al análisis de la Variabilidad Espacial Arqueológica Intra-Site*.
3. Cressie, Noel & Christopher K. Wile. (1999). Space-Time Kalman Filter. Volume 4, pp 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. Páginas 19-23.
5. Díaz, Martín. (2002). *Geoestadística Aplicada*. Página 5.
6. E. Porcu, J. Mateu y F. Saura. (2000). Nuevos Modelos Espacio-Temporales de Covarianza y Densidad Espectral en Epidemiología. Department of Mathematics. Universitat Jaume I.
7. G. Gneiting. (1999). *Geostatistical Space-Time Models, Stationarity, Separability and Full Symmetry*. Department of Statistics, University of Washington Seattle, Washington, USA.
8. H. –C. Huang, F. Martinez, J. Mateu, F. Montes, Model Comparison and Selection for stationary space-time models. *Computational Statistics & Data Analysis*.
9. H. Giraldo. *Introducción a la Geoestadística. Teoría y Práctica*. Universidad Nacional de Colombia.
10. Ibáñez, M. Victoria. (2003). Modelos estadísticos espacio-temporales en perimetría. Página 51.
11. J. Berlanga and J. Obregon. (1970). *Geoestadística*.
12. M. Georges. (1969). *Curso de Geoestadística*.
13. M. Victoria Ibáñez. *Modelos Estadísticos espacio-temporales en perimétrica*. Escuela Superior de Tecnología y Ciencias Experimentales, Universitat Jaume I. Julio 2003.
14. Phaedon C. Kyriakidis and André G. Journel. *Geostatistical Space-Time Models: A Review*. *Mathematical Geology*, Vol. 31, No. 6, 1999.
15. S. De Iaco, D.E. Myers, y D. Posa. *The Linear Coregionalization Model y The Product-Sum Space-Time Variogram*. 2003.
16. Sanz, Felipe. (1999): *La escuela cartografica de criminologibritanica. Antecedentes de la Geografia del crimen*. Páginas 11-22.
17. Shekhar Srinivasan, B. E. (2006). *Direct Spatiotemporal Interpolation of Reservoir Flow Responses*. University of Texas at Austin.