

# **UNIVERSIDAD DE EL SALVADOR**

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

ESCUELA DE MATEMÁTICA



TRABAJO DE GRADUACIÓN

**“METODOLOGÍA PARA EL ANÁLISIS GEOESTADÍSTICO PARA LA  
INTERPOLACIÓN DE LA PRECIPITACIÓN EN EL SALVADOR.”**

**PARA OPTAR A GRADO DE LICENCIADO EN ESTADÍSTICA.**

**PRESENTADO POR:**

BR. GAMALIEL CHÁVEZ HERNÁNDEZ.

**ASESOR:**

MSC. PEDRO ARMANDO RAMOS ALBERTO

**ASESOR ADJUNTO:**

LIC. MARIO GIOVANNI MOLINA MASFERRER

CIUDAD UNIVERSITARIA, FEBRERO DE 2009

## **UNIVERSIDAD DE EL SALVADOR**

**RECTOR:** MSC. RUFINO ANTONIO QUEZADA SÁNCHEZ

**SECRETARIO GENERAL:** LIC. DOUGLAS VLADIMIR ALFARO CHÁVEZ

### **FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA**

**DECANO EN FUNCIONES:** DR. RAFAEL ANTONIO GÓMEZ ESCOTO

**SECRETARIA:** LICDA. MARÍA TRINIDAD TRIGUEROS DE  
CASTRO

### **ESCUELA DE MATEMÁTICA**

**DIRECTOR:** ING. CARLOS MAURICIO CANJURA LINAREZ

**SECRETARIO:** LIC. ERNESTO AMERICO HIDALGO

**TRABAJO DE GRADUACION APROBADO POR:**

**ASESOR:** MSC. PEDRO ARMANDO RAMOS ALBERTO

**ASESOR ADJUNTO:** LIC. MARIO GIOVANNI MOLINA MASFERRER

**DEDICATORIA.**

Toda Gloria y majestad al que en todo es el primero, a Jesucristo mi Salvador, pues Él me dio la sabiduría para el logro de este título. Fue mi refugio en las dificultades y nunca me abandonó y me brindó las fuerzas necesarias que necesité. Gracias Dios.

A mi madre y mi padre, que siempre me apoyaron y me motivaron a seguir adelante.

A mis hermanos, Porfirio Chávez Hernández, José Javier Chávez Hernández y Wilber Giovanni Hernández y a todos los demás miembros de mi Familia, que de una forma u otra me ayudaron para la obtención de este logro.

A mis Asesores, Msc. Pedro Armando Ramos Alberto y Licenciado Mario Giovanni Molina Masferrer, quienes me orientaron para el diseño de este trabajo de graduación.

A la Lic. Yamileth Joya Díaz, quien estuvo a mi lado brindándome apoyo para que concluyera la presente Tesis.

Gamaliel Chávez Hernández.

## ÍNDICE DE CONTENIDOS

	Pág.
Introducción.....	<a href="#">viii</a>
<b>Capítulo I Introducción</b>	
1.1 Proceso de Recolección, tratamiento y análisis de los datos de la precipitación.....	<a href="#">2</a>
1.2 Mapa de Distribución de las estaciones meteorológicas.....	<a href="#">4</a>
<b>Capítulo II: Conceptos, Funciones y Fases para el Análisis Geoestadístico</b>	
2.1 Introducción.....	<a href="#">5</a>
2.2 Conceptos Básicos para el análisis Geoestadísticos y Funciones Básica...	<a href="#">6</a>
2.2.1 Variables Regionalizadas.....	<a href="#">6</a>
2.2.2 Geoestadística.....	<a href="#">6</a>
2.2.3 El Semivariograma. Definición.....	<a href="#">6</a>
2.2.4 Función Aleatoria.....	<a href="#">7</a>
2.2.5 Función de distribución y momentos de una función aleatoria.....	<a href="#">7</a>
2.2.6 Funciones aleatorias estacionarias.....	<a href="#">8</a>
2.2.7 Funciones aleatorias intrínseca.....	<a href="#">10</a>
2.2.8 Funciones aleatorias no estacionarias.....	<a href="#">10</a>
2.3 Análisis exploratorio de datos y supuestos de la geoestadística.....	<a href="#">11</a>
2.3.1 Supuestos de la Geoestadística.....	<a href="#">11</a>
2.4 El análisis estructural.....	<a href="#">14</a>
2.4.1 El semivariograma experimental.....	<a href="#">14</a>
2.4.2 Problemas más comunes encontrados en el cálculo de semivariograma...	<a href="#">16</a>

2.4.3	Condiciones que deben cumplir los modelos del semivariograma.....	<a href="#">19</a>
2.4.4	Modelado de semivariogramas.....	<a href="#">21</a>
2.4.5	Parámetros del semivariograma.....	<a href="#">21</a>
2.4.6	Modelos teóricos de semivariograma.....	<a href="#">22</a>
2.4.6.1	Validación del modelo teórico.....	<a href="#">25</a>
2.4.6.2	Métodos de ajuste.....	<a href="#">26</a>
2.4.6.3	Criterio de Selección y Validación de los Modelos.....	<a href="#">27</a>
2.4.7	Análisis de anisotropía.....	<a href="#">29</a>
2.4.8	Efecto proporcional.....	<a href="#">30</a>
2.4.9	Problemas en el modelaje de semivariogramas.....	<a href="#">31</a>
2.5	Estimación.....	<a href="#">33</a>
2.6	Métodos de Interpolación.....	<a href="#">33</a>
2.6.1	Método de la Triangulación.....	<a href="#">33</a>
2.6.2	Método Inverso de la distancia.....	<a href="#">34</a>
2.6.3	Métodos Krigeaje.....	<a href="#">35</a>
2.6.3.1	El mejor estimador lineal insesgado.....	<a href="#">35</a>
2.6.3.2	Ecuaciones del krigeaje.....	<a href="#">35</a>
2.6.3.3	Método Krigeaje Simple.....	<a href="#">36</a>
2.6.3.4	Método Krigeaje Ordinario.....	<a href="#">36</a>
2.6.3.4.1	Ecuaciones del Kriging ordinario en forma matricial.....	<a href="#">37</a>
2.6.4	El caso no estacionario, Método Krigeaje Universal (KU).....	<a href="#">40</a>
2.6.4.1	Aspectos Prácticos del Kriging Universal.....	<a href="#">41</a>

**Capítulo III: Ajuste del Modelo**

<b>3.1</b>	Introducción.....	<a href="#">42</a>
<b>3.2</b>	Fases de la investigación.....	<a href="#">42</a>
<b>3.3</b>	Aplicación de las fases a la variable precipitación en El Salvador	<a href="#">45</a>
<b>3.3.1</b>	Fase I. Análisis exploratorio de los datos.....	<a href="#">45</a>
<b>3.3.1.1</b>	Histograma de Frecuencias.....	<a href="#">46</a>
<b>3.3.2</b>	Fase II. Análisis Estructural.....	<a href="#">47</a>
<b>3.3.2.1</b>	Selección y búsqueda de Modelos de Semivariogramas.....	<a href="#">48</a>
<b>3.3.4</b>	Fase III. Estimación de los Métodos de Interpolación.....	<a href="#">53</a>
<b>3.3.4.1</b>	Selección del mejor Método de Estimación.....	<a href="#">53</a>
	Conclusiones y Recomendaciones.....	<a href="#">57</a>
	Referencias Bibliográficas.....	<a href="#">58</a>
	Anexos y Tablas.....	<a href="#">59</a>

## **Introducción.**

El presente documento tiene como propósito proporcionar información sobre el análisis geoestadístico que es de utilidad en muchas áreas y todas aquellas áreas que se encargan de recolectar información de fenómenos georreferenciados. La geoestadística actualmente es considerada como una nueva área de investigación ya que proporciona conjunto de técnicas utilizadas para predecir valores de una propiedad distribuida en espacio y tiempo. Es así que en este documento se toma como base para la aplicación, información recolectada, de la base de datos de la variable precipitación del año 2007 en El Salvador. Esta fue proporcionada por el Servicio Nacional de Estudios Territoriales (SNET) de El Salvador, con el objetivo fundamental de desarrollar una guía metodológica para la interpolación de la precipitación y con ella poder realizar análisis, estimaciones del día anterior para el día siguiente.

En el primer capítulo de esta investigación se hace una revisión de información general tal como la forma en que el SNET recolecta y analiza la información de la precipitación y como están distribuidas las estaciones meteorológicas en todo el territorio nacional.

El segundo capítulo se hace referencia a los conceptos, elementos de la geoestadística definiendo conceptos básicos, funciones a utilizar y posteriormente describir las etapas que componen el análisis geoestadístico. En su primera etapa se realiza el análisis exploratorio de datos, esta busca identificar variabilidad, forma y observaciones extremas. En su segunda etapa corresponde a la del análisis estructural el cual consta de la estimación del variograma y la comparación para el ajuste de este a un modelo teórico y posteriormente validar dicho modelo. Finalmente la tercer etapa se realiza la estimación, aquí se describen algunos métodos de estimación utilizados en la interpolación como lo es el Inverso de la Distancia, Triangulación y Kriging.

En el tercer capítulo se describen los procedimientos empleados para identificar de manera experimental (con base en datos muestrales) el análisis exploratorio y la estructura espacial (variograma) de la precipitación. Una vez detectada la estructura espacial, el siguiente paso es

proyectar el mapa de El Salvador con la estimación en sitios de la región de estudio siguiendo criterios, análisis de los resultados

Posteriormente utilizaremos el software **GRASS GIS** [1] y paquete estadístico **R** [2], para completar el análisis geoestadístico, con estos recursos se proyecta y profundiza de los resultados obtenidos durante el análisis, ya que de forma manual puede llevarnos tiempo y pérdida de información.

## CAPITULO 1: INTRODUCCIÓN.

El Salvador, por su ubicación geográfica, tiene las características de poseer un clima de tipo tropical, al igual que el resto de los países de Centroamérica, posee dos tipos de estaciones durante el año; **Época Seca (verano) y Época Lluviosa (invierno)**.

Cuando se trata de la época seca el fenómeno que lo caracteriza es poca lluvia y con abundante sol, cuando se trata de la época lluviosa es todo lo contrario, abundante agua, y como consecuencia de ello se producen en nuestro país una serie de situaciones de riesgo. Es en ese sentido que el comportamiento de las lluvias, se ponen de manifiesto durante los meses de mayo a octubre. Las lluvias son en alguna medida constantes y es así que algunas instituciones registran las lluvias en  $mm^3$ , que les permite cuantificar la lluvia que cae en todo el territorio nacional. Dicha información es recolectada por el SNET y analizada por la Gerencia de Hidrología del Ministerio de Medio Ambiente y Recursos Naturales de El Salvador. Cuenta con información de tal fenómeno, están atentos diariamente a cuantificar para lo que se pretende realizar un estudio con colaboración del SNET, ya que dicha institución es la encargada a nivel nacional de recolectar y analizar la información y esta institución lo hace efectuando estudios muy cortos y con herramientas tradicionales. Actualmente, en la evolución de la estadística se han creado herramientas para el abordaje de información de todos los tipos. Además, contando con el recurso informático y software se facilita el tratamiento, análisis y más aún efectuarse predicciones o estimaciones si son necesarias para estar prevenidos por cualquier eventualidad de la naturaleza. Es así que, se efectuara una aplicación y un análisis utilizando las herramientas de la geoestadística. La motivación de realizar dicho trabajo es también, que en dicha institución no cuenta con una metodología adecuada para el abordaje de dicho fenómeno y es de interés aplicar los conocimientos adquiridos en el área de la estadística.

Como herramienta para el análisis de la información se pretende utilizar técnicas desarrolladas de D. G. Krige y Matheron en la geoestadística, la cual es un conjunto de técnicas usadas para analizar y predecir valores de una propiedad distribuida en espacio y tiempo.

Pretendo analizar el fenómeno de la lluvia tomando como base la teoría de la geoestadística para luego elaborar un documento como una propuesta metodológica que trate de la interpolación de la precipitación en El Salvador y que sea de utilidad para aquellas instituciones ó personas interesadas en el tema de modelaje del fenómenos geoespaciales.

Utilizando diferentes métodos, aplicando criterios y modelos estadísticos propuestos en esta investigación para poder identificar y seleccionar el o los modelos apropiados que mejor se ajusten y describa el comportamiento de la precipitación en El Salvador.

### 1.1 Proceso de Recolección, tratamiento y análisis de los datos de la precipitación

La recolección de los datos de la precipitación por parte del SNET se realiza a través de las diferentes estaciones meteorológicas ubicada en diferentes partes del territorio nacional; cuenta con dos procesos el que se obtiene de manera manual y el otro automático (telemétrica).

El instrumento comúnmente utilizado es el pluviómetro, El diseño básico de un pluviómetro consiste en un recipiente de entrada, llamado balancín, por donde el agua ingresa a través de un embudo hacia un colector donde el agua se recoge y puede medirse visualmente con una regla graduada o mediante el peso del agua depositada, por ello recibe el nombre de manual, como se muestra en la **Figura 1**.



**Figura 1.** Pluviómetro

El primer proceso consiste en que una persona encargada de la estación metereológica verifica cual es la cantidad de lluvia que captada en el pluviómetro y luego la envía a la base de datos que se tiene en el SNET.

El segundo proceso de recolección es el de manera automática (telemétrica). En este caso, es un balancín que registra las lluvias, cada vez que el recipiente se inclina, un sensor detecta una señal que es transmitida en una pantalla visualizando la cantidad de lluvia caída en el lugar. No hay necesidad de vaciarlo pues un pulsador en el contador reajusta a cero después de cada registro de precipitación. Este pluviómetro está conectado a una computadora que envía los valores hacia el operador, en este caso el SNET, en un sistema de telemetría que se realiza mediante comunicación satelital, aunque también se puede realizar por otros medios (teléfono, redes de ordenadores, enlace de fibra óptica, etcétera).

La información de los dos procesos es recolectada continuamente y se va formando lo que es una base de datos, como lo muestra la Tabla A1. Esta es la información de la que dispone el SNET y esta es con la que se contará para la realización del presente trabajo, éstas llenan las condiciones necesarias para el análisis geostadístico, ya que son condiciones que se dan a través del tiempo y espacio, lo más importante es que es información confiable que las que se recolectan de manera manual.

A	B	C	D	E	F	G	H	I	J	K
<b>CENTRO DE INFORMACION Y AGROMETEOROLOGIA- SERVICIO METEOROLÓGICO NACIONAL - SNET</b>										
<b>Estaciones con reporte de lluvia (mm) diaria 2007</b>										
ESTACIÓN/día-mes	1-6	2-6	3-6	4-6	5-6	6-6	7-6	8-6	9-6	10-6
Estacion 1	1.7	0.0	5.5	0.0	23.7	21.9	4.4	11.7	0.1	1.4
Estacion 2		12.8	12.5	0.0	13.4	0.2	0.6	6.2	0.0	54.3
Estacion 3	0.2	0.5	3.0	0.0	31.4	6.1	0.3	4.1	0.0	0.0
Estacion 4	0.4	1.5	7.1		3.0	3.1		5.5	0.0	2.7
Estacion 5		4.8	11.7	0.0	3.0	3.3	5.3	2.0	0.0	31.0
Estacion 6	9.0	7.0	10.0	0.0	13.0	26.0	13.5	12.0	0.0	55.0
Estacion 7	16.2	0.5	2.9	0.0	1.0	0.0	7.5		0.2	17.8
Estacion 8	2.5				17.4	3.6	1.9	55.2		
Estacion 9	2.5	2.4	7.9	0.0	7.8	2.7	2.3	3.3	1.9	23.5
Estacion 10	6.0	2.5	16.4	0.0	53.7	24.2	5.6	45.4	0.0	1.1
Estacion 11	2.8	3.5	6.8	0.0	16.9	4.3	53.7	25.2	6.5	0.0
Estacion 12	2.6	3.5	8.7	0.0	11.3	0.0	1.4	1.5	0.0	0.2
Estacion 13	1.5	2.0	1.7	0.0	3.4	1.6	11.0	1.0	0.0	17.2

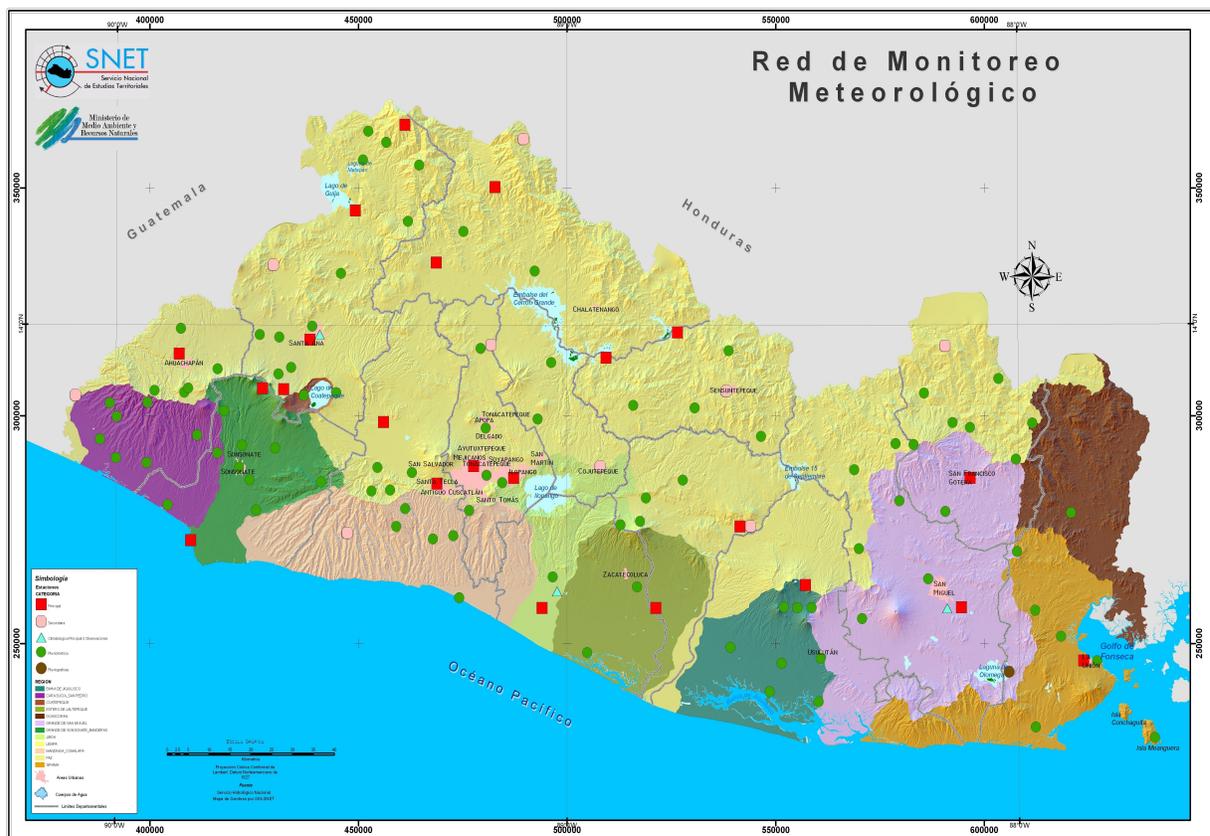
Tabla A1. Base de datos de la Precipitación.

La lectura que nos proporciona esta base de datos es por ejemplo la estación 1 el primero de junio(1-6) se captó una cantidad  $1.7 \text{ mm}^3$  de lluvia este valor representa la cantidad de agua

caída en ese lugar. Los espacios vacíos de esta base representan datos faltantes y los datos en rojo son datos atípicos y de igual forma se leen los demás datos de la tabla.

Las estaciones con que se cuenta para el estudio son 65 como se muestra en el mapa 1

## 1.2 Mapa de Distribución de las estaciones meteorológicas.



Los puntos verdes representan la distribución de las estaciones meteorológicas que miden la precipitación en todo el país.

El objetivo de contar con la base de datos de las precipitaciones obtenidas en el año 2007, es para hacer estimaciones de las precipitaciones del día anterior para el día siguiente.

## CAPITULO 2: CONCEPTOS, FUNCIONES Y FASES PARA EL ANÁLISIS GEOESTADÍSTICO.

### 2.1 Introducción.

En los años 60, Matheron define el término de *Geoestadística*. Reconocido como el padre de esta disciplina, formalizó y generalizó matemáticamente un conjunto de técnicas desarrolladas por D. G. Krige en 1941, que explotaban la *correlación espacial* para hacer predicciones en la evaluación de reservas de las minas de oro en Sudáfrica. Además, formula la definición de Geoestadística como "la aplicación del formalismo de las *funciones aleatorias* al reconocimiento y estimación de fenómenos naturales".

Con las técnicas que propone Matheron se estudian las características más notorias al conjunto de datos, considerándose así, el paso importante para el análisis geoestadístico y para que los modelos sean útiles deben de ser consistentes en las propiedades estocásticas de los datos y ofrecer información relevante. En este sentido los modelos estructurales de la geoestadística constituyen un enfoque adecuado que basado en las propiedades de correlación espacial que permite presentar los rangos más sobresalientes.

Será necesario para el análisis de la información de la que se dispone - las precipitaciones - utilizar herramientas que permitan hacer un análisis considerando toda la información recolectada a nivel nacional, la cual es difícil realizarla de manera manual por que podría llevarnos a la pérdida de información, mayor cantidad de tiempo y estimaciones no muy fiables, es por ello que utilizaremos el **paquete estadístico R** y el programa de análisis geoestadístico **GRASS GIS**, para estimar el modelo más adecuado a nuestra información.

El proceso de estimación y modelación de la función que describe la correlación espacial es conocido en forma general como "análisis estructural". Una vez realizado el análisis estructural, la predicción de valores en puntos no muestrales se aplicará la técnica de interpolación. En resumen, se dice que un análisis geoestadístico está compuesto por tres etapas: (a) el análisis exploratorio de los datos, (b) el análisis estructural y (c) las predicciones

o estimaciones. Estas etapas se describen en el presente capítulo, como también se expondrá los diferentes métodos de interpolación.

Como un paso inicial para hacer el estudio y análisis geoestadístico definimos a continuación los conceptos elementales que sustentaran la investigación.

## 2.2 Conceptos Básicos para el análisis Geoestadísticos y Funciones Básicas

### 2.2.1 Variables Regionalizadas

Se nombra como *variable regionalizada*,  $z(x)$ , a la variable distribuida en el espacio de manera tal que presenta una estructura espacial de correlación, matemáticamente, consistiría en una variable regionalizada de una variable aleatoria  $z$ , definida en un punto del espacio, es decir

$$X = (x_1, x_2, x_3).$$

### 2.2.2 Geoestadística:

Es una metodología para el análisis de datos espacialmente correlacionados para la modelación de la variable espacial. Uno de los elementos principales para el análisis es el uso de variogramas, técnicas relacionadas que se usaran para cuantificar y modelar la correlación espacial de la estructura, la cual incluye diferentes métodos, los utiliza como modelos de correlación espacial. Los métodos geoestadísticos son aplicables en todas las ciencias. Pueden aplicarse para explorar los procesos responsables de la variación espacial. También, pueden aplicarse donde existe una información completa obtenida por percepción remota u otra fuente, para determinar un muestreo eficiente, así como también para estimar el valor de propiedades en localidades no muestreadas.

### 2.2.3 El Semivariograma. Definición.

Es una función que relaciona la semivarianza con el vector  $h$  conocido como "lag", el que denota la separación en distancia y dirección de cualquier par de valores  $Z(x)$  y  $Z(x+h)$  ubicadas en el espacio, entonces la función semivarianza se define como:

$$Var [Z(x+h) - Z(x)] = E \{ [Z(x+h) - Z(x)]^2 \} = 2 \gamma(h), \quad \forall x \in \Omega$$

### 2.2.4 Función Aleatoria

Si a cada punto,  $x \in \Omega$ , donde  $\Omega$  es el dominio en el espacio, le hacemos corresponder una variable aleatoria  $z(x)$ , que en sentido general pueden ser dependientes, entonces el conjunto de variables aleatorias espacialmente distribuidas,  $\{z(x), x \in \Omega\}$  será una función aleatoria  $Z(x)$ .

Al tomar una muestra de una función aleatoria  $Z(x)$ , a la que llamaremos realización, se obtendrá una función espacial discreta  $Z' = \{Z(x_i), x_i \in \Omega, i = 1, 2, \dots, n\}$ , la cual constituye una variable regionalizada. Es decir, una realización de una función aleatoria  $Z(x)$  es una variable regionalizada  $Z'$ .

En lo sucesivo haremos distinción entre la función aleatoria y su realización. La realización es una parte de la función aleatoria.

### 2.2.5 Función de distribución y momentos de una función aleatoria

Sea  $Z(x)$  una función aleatoria definida en  $R^3$ , entonces el vector aleatorio definida por  $\{Z(x_1), Z(x_2), \dots, Z(x_n)\}$ , se caracteriza por su función de distribución de probabilidad  $n$ -variada dada por:

$$F_{Z(x_1), Z(x_2), \dots, Z(x_n)}(z_1, z_2, \dots, z_n) = P[Z(x_1) \leq z_1, Z(x_2) \leq z_2, \dots, Z(x_n) \leq z_n] \quad (1)$$

El conjunto de todas las distribuciones para todo valor de  $n$  y para cualquier selección de puntos en  $R^3$  constituye la ley espacial de probabilidad de la función aleatoria  $Z(x)$ . Esta función en la práctica es imposible de determinar y sólo se puede esperar inferir los primeros momentos de la distribución de  $Z(x)$ . En las aplicaciones en geoestadística lineal resulta suficiente estimar los momentos hasta de segundo orden, no obstante en la mayoría de los casos la información disponible no permite inferir momentos de orden superior.

### Momentos de la distribución de la función $Z(x)$

- El momento de primer orden de la función  $Z(x)$  es la esperanza matemática definida como:

$$m = E(Z(x)) \quad (2)$$

- Los momentos de segundo orden considerados en geoestadística son:

1. La varianza de  $Z(x)$

$$\sigma(x) = Var[Z(x)] = E[\{Z(x) - m\}^2] \quad (3)$$

2. La covarianza de dos variables aleatorias  $Z(x_i)$  y  $Z(x_j)$  definida como:

$$C(x_i, x_j) = E[\{Z(x_i) - m\}\{Z(x_j) - m\}] \quad (4)$$

Esta función es también conocida como función de autocovarianza.

3. El semivariograma  $\gamma(x_i, x_j)$  que se define como:

$$2\gamma(x_i, x_j) = Var[Z(x_i) - Z(x_j)] \quad (5)$$

$$\gamma(x_i, x_j) = \frac{E[\{Z(x_i) - Z(x_j)\}^2]}{2} \quad (6)$$

También conocido como función de semivarianza. Además, el variograma también se define como  $2\gamma(x_i, x_j)$  pero con frecuencia se usa el término indistintamente para designar a  $\gamma(x_i, x_j)$ , esta función nos permitirá estudiar la correlación espacial de la precipitación.

Se debe tomar en cuenta que tanto la varianza como el variograma son siempre positivos, mientras que la covarianza puede tomar valores negativos.

### 2.2.6 Funciones aleatorias estacionarias

Se dice que una función aleatoria es *estrictamente estacionaria* si su función de distribución – Ec. (1)– es invariante a cualquier traslación respecto a un vector  $h$  o lo que es equivalente, la función de distribución del vector aleatorio  $\{Z(x_1), Z(x_2), \dots, Z(x_n)\}$  es idéntica a  $\{Z(x_1 + h), Z(x_2 + h), \dots, Z(x_n + h)\}$  para cualquier  $h$ .

Puesto que, como se planteó, usualmente se trabaja sólo con los momentos hasta de segundo orden, resulta práctico limitar la hipótesis de estacionaridad a estos primeros momentos.

Se dice que la función aleatoria  $Z(x)$  es *estacionaria de segundo orden* si se cumple que:

- i. Su valor esperado existe y no depende de  $x$ ,

$$E(Z(x)) = m \quad (7)$$

para cualquier par de variables aleatorias  $Z(x)$  y  $Z(x+h)$ , su covarianza existe y sólo depende del vector de separación  $h$ .

- ii.  $C(h) \equiv C(x+h, x) = E[Z(x+h)Z(x)] - m^2$  (8)

La estacionaridad de la varianza implica que la varianza existe, es finita y no depende de  $x$ , es decir

$$\sigma^2 = C(0) = \text{Var}[Z(x)] \quad (9)$$

Así mismo bajo esta hipótesis el semivariograma también es estacionario además, se cumple que:

- iii.  $\gamma(h) = \gamma(x+h, x) = \frac{1}{2} E[\{Z(x+h) - Z(x)\}^2]$  (10)

Por lo que, existe una relación directa entre el semivariograma y la función de covarianza

- iv.  $\gamma(h) = C(0) - C(h)$  (11)

En este caso resulta suficiente usar una de las dos funciones para caracterizar la dependencia espacial.

$\gamma(h)$  y  $C(h)$  permiten expresar la correlación entre la variable aleatoria regionalizada  $Z(x)$  y  $Z(x+h)$  separadas por el vector  $h$ .

### 2.2.7 Funciones aleatorias intrínsecas.

Existen funciones aleatorias  $Z(x)$  que representan a fenómenos físicos que muestran una capacidad casi ilimitada de variación, por lo que para estas funciones no están definidas la varianza ni la covarianza. Sin embargo, existen casos en que sus incrementos o diferencias  $Z(x+h) - Z(x)$  tienen una varianza finita. En otras palabras, significa que las diferencias son estacionarias de segundo orden.

Por lo tanto las *funciones aleatorias intrínsecas* son aquellas que cumplen las siguientes condiciones:

- i. El valor esperado de las diferencias es

$$E[Z(x+h) - Z(x)] \quad (12)$$

- ii. La varianza de las diferencias es

$$Var[Z(x+h) - Z(x)] = 2\gamma(h) \quad (13)$$

Estas condiciones se conocen como *Hipótesis Intrínseca*. Es evidente que una función aleatoria estacionaria de segundo orden es siempre intrínseca. Lo contrario no se cumple. A las funciones que cumplen con la hipótesis intrínseca se les considera como *débilmente estacionarias*.

### 2.2.8 Funciones aleatorias no estacionarias

Las *funciones aleatorias no estacionarias* son aquellas cuya esperanza matemática depende de  $x$ :

$$E[Z(x)] = m(x) \quad (14)$$

A  $m(x)$  se le conoce como función de deriva o tendencia.

Si consideramos a la función aleatoria  $Z(x)$  como la suma de una componente determinística  $m(x)$  y de un residuo  $R(x)$  estacionario con media nula, es decir:

$$Z(x) = m(x) + R(x) \quad (15)$$

Se tiene la función que representa la detección de la existencia de tendencia.

Entonces, observamos que el semivariograma  $Z(x)$  depende de  $x$ .

$$\gamma(x+h, x) = \gamma_R(h) + \frac{1}{2}\{m(x+h) - m(x)\}^2 \quad (16)$$

En el caso en que la deriva sea lineal  $m(x) = m_0 + m_1 x$ , el semivariograma no depende de la variable  $x$ , entonces

$$\gamma(x+h, x) = \gamma_R(h) + \frac{1}{2}(m_1 h)^2 \quad (17)$$

pero crece con el cuadrado de  $h$ , lo cual puede utilizarse como un indicador para la detección de la existencia de tendencia.

Existe un enfoque que considera a las funciones aleatorias no estacionarias como intrínsecas de orden  $k$ . Esto quiere decir que si se toman las diferencias de un orden  $k$  apropiado éstas resultan ser estacionarias.

## 2.3 Análisis exploratorio de datos y supuestos de la geoestadística

Antes de realizar un estudio geoestadístico se debe estudiar la estadística descriptiva de los datos disponibles de una variable por medio de un análisis exploratorio de datos. Los estadísticos descriptivos consisten en las medidas de tendencia central y la dispersión.

La aplicación de métodos geoestadísticos de interpolación espacial de datos requiere que la serie de datos de una variable presente una distribución normal o lognormal, lógicamente con cierto sesgo hacia valores más pequeños (sesgo positivo), aunque no existen valores negativos para nuestro caso.

### 2.3.1 Supuestos de la Geoestadística

Como la forma en que se presenta la información es muy diversa, la geoestadística se construye asumiendo condiciones de estacionaridad. Por lo que es necesario aceptar el cumplimiento de ciertas hipótesis sobre el carácter de la función aleatoria o procesos estocásticos estudiados, llamadas Hipótesis de la Geoestadística. Estas son: La Estacionaridad

Estricta, La Estacionaridad de Segundo Orden, La Hipótesis Intrínseca y los Procesos Cuasiestacionarios. A continuación se presenta en que consiste cada una de estas condiciones.

### I- Estacionaridad Estricta.

Se dice que  $Z(x)$  es estrictamente estacionaria si la función de distribución de probabilidades de las variables aleatorias regionalizadas  $Z(x_i)$  -ecuación (1)- son iguales entre sí, independiente de la localización  $x_i$ , lo que requiere que los momentos de distinto orden para cada variable aleatoria regionalizada sean completamente independientes de la localización  $x_i$ .

### II- Estacionaridad de Segundo Orden.

Esta condición es más frecuente en la práctica, la misma exige que:

1)  $E[Z(x_i)] = m$ , existe y no depende de la localización  $x_i$ , donde  $m$  representa la media de las observaciones

2) La función covarianza,  $Cov[Z(x_i) - Z(x_j)]$ , exista y sólo dependa de la longitud del vector  $h = x_i - x_j$  o sea;  $C(h) = Cov[Z(x_i), Z(x_j)] = E[Z(x_i)Z(x_i + h)] - m^2$ , donde  $C(h)$  se conoce como la covarianza de las observaciones

La primera condición requiere la estacionaridad sólo para la media y para la función de covarianza de la variable aleatoria regionalizada. La segunda condición implica, estacionaridad de la varianza y del variograma como se describe a continuación:

$$1^\circ \quad Var[Z(x_i)] = E \{ [Z(x_i) - m]^2 \} = C(0), \quad \forall x$$

$$2^\circ \quad \gamma(h) = E \{ [Z(x_i)]^2 \} - E \{ Z(x_i), Z(x_i + h) \}, \quad \forall x$$

$$\text{Como } E[Z(x_i), Z(x_i + h)] = C(h) + m^2$$

$$y \quad E[Z^2(x_i)] = C(0) + m^2$$

$$\gamma(h) = C(0) + m^2 - [C(h) + m^2]$$

$$\gamma(h) = C(0) - C(h).$$

Como se observa en la última expresión  $\gamma(h)$  y  $C(h)$  se expreso anticipadamente, son dos herramientas que permiten expresar la correlación entre la variable aleatoria regionalizada  $Z(x_i)$  y  $Z(x_i + h)$ .

### III- Hipótesis Intrínseca.

Una función aleatoria  $Z(x)$  se dice intrínseca cuando:

- a) Su esperanza matemática existe y no depende de la localización  $x_i$ .

$$E[Z(x)] = m, \quad \forall x$$

- b) Para todo vector  $h$  el incremento  $[Z(x+h) - Z(x)]$  tiene varianza finita y no depende de la localización  $x_i$ ; es decir

$$Var [Z(x+h) - Z(x)] = E \{ [Z(x+h) - Z(x)]^2 \} = 2 \gamma(h), \quad \forall x$$

donde  $\gamma(h)$ , se denomina variograma (Ec (6)).

Cuando se cumple esta condición se dice que la función aleatoria  $Z(x)$  es homogénea.

### -IV Procesos Cuasiestacionarios.

En la práctica la función estructural, covarianza o semivariograma, es usada para la desigualdad  $|h| \leq b$ , donde  $b$  es el límite que representa la extensión de la región en la que el fenómeno estudiado conserva cierta homogeneidad del comportamiento de  $Z(x_i)$ . En otros casos,  $b$  es la magnitud de una zona homogénea y dos variables  $Z(x)$  y  $Z(x+h)$  no pueden ser consideradas en la misma homogenización, para el caso en el que  $|h| > b$ . En tales casos, podemos, y verdaderamente debemos, estar satisfechos

con una función estructural *covarianza*  $C(x, x+h)$  o *variograma*  $\gamma(x, x+h)$ , lo que no es más que estacionaridad local (para distancias  $h$  menores que límite  $b$ ).

## 2.4 El análisis estructural

El análisis estructural, tópico importante de la geoestadística, se encarga de la caracterización de la estructura espacial de una propiedad o fenómeno regionalizado. Es el proceso en el marco del cual se obtiene un modelo geoestadístico para la función aleatoria que se estudia.

Diremos que el análisis estructural consiste en estimar y modelar una función que refleje la correlación espacial de la variable regionalizada a partir de la adopción razonada de la hipótesis más adecuada acerca de su variabilidad. Esto quiere decir, que en dependencia de las características de estacionaridad del fenómeno se modelará la función de covarianzas o la de semivarianzas. Por su importancia y generalidad estudiaremos el proceso de estimación y modelación de la función de semivarianzas o semivariograma.

El análisis estructural o estudio variográfico está compuesto por:

- El cálculo del semivariograma experimental.
- El ajuste a este de un modelo teórico conocido.

Estos se describen como sigue:

### 2.4.1 El semivariograma experimental

El semivariograma experimental es la que determina las características de variabilidad y correlación espacial del fenómeno estudiado, es decir, permite conocer cómo la variable cambia de una localización a otra.

El variograma se define como la varianza de los incrementos de la variable regionalizada en las localizaciones separadas una distancia  $h$ . Es decir;

$$Var\{Z(x+h) - Z(x)\} = 2\gamma(h)$$

La función  $\gamma(h)$  se denomina **semivariograma**, la cual puede ser obtenida por la expresión.

$$\gamma(h) = \frac{1}{2Np(h)} \sum_{i=1}^{Np(h)} [Z(x_i + h) - Z(x_i)]^2 \quad (18)$$

donde:  $Np(h)$  es el número de pares a la distancia  $h$ .

$h$ : vector de separación la cual determina la distancia entre dos observaciones en una cierta dirección

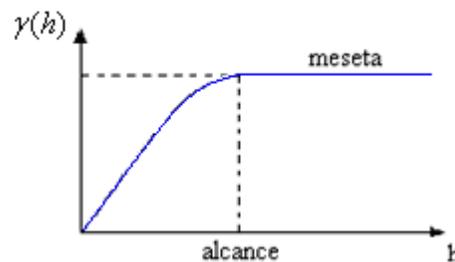
$Z(x_i)$ : son los valores experimentales.

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

Sus características son:

- Es un estimador no paramétrico.
- Es óptimo cuando se dispone de una malla regular de muestreo que sea representativa y la distribución sea normal. En estas condiciones el sesgo es el mínimo posible.

Esta expresión de  $\gamma(h)$  es la herramienta según Arsmtron y Carignan, la que nos proporciona representación gráfica de los valores de  $\gamma(h)$ , en función de  $h$  cuyo grafico es el siguiente.

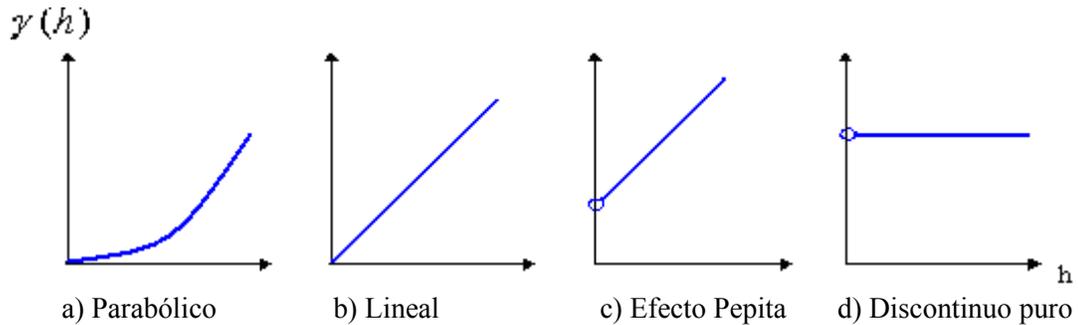


**Figura 2.** Forma típica del semivariograma

El gráfico de  $\gamma(h)$  tiene las siguientes características (figura 2).

- Pasa por el origen (para  $h = 0$ ,  $\gamma(h) = 0$ )
- Es en general una función creciente de  $h$ .

En la mayor parte de los casos  $\gamma(h)$  crece hasta cierto límite llamado meseta, en otros casos puede crecer indefinidamente. El comportamiento en el origen puede tener diferentes formas, se detallan en la Figura 3.



**Figura 3.** Comportamiento del variograma en el origen

- a) **Parabólico:** Caracteriza a una variable muy regular, siendo continua y diferenciable.
- b) **Lineal:** Caracteriza a una variable continua.
- c) **Discontinuidad en el origen:** “Efecto de pepita”, es el caso en que  $\gamma(h)$  no tiende a cero cuando  $h$  tiende a cero. Representa a una variable muy irregular.
- d) **Discontinuo puro:** Llamado también ruido blanco, representa el caso de mayor discontinuidad, siendo el caso límite de ausencia de estructura, donde los valores de dos puntos cualesquiera no tienen correlación alguna.

#### 2.4.2 Problemas más comunes encontrados en el cálculo de semivariograma

La fuente de problemas que se pueden presentar en la realización de un análisis estructural es muy variada, lo que está en correspondencia con la variedad de casos que se presentan en la naturaleza. Algunos de los problemas más comunes son:

➤ **El valor idóneo del incremento h:**

Una inadecuada selección de h puede proporcionar un semivariograma errático, aunque no se puede dar un criterio exacto o aproximado sobre cual es el mejor valor de  $h$ , es recomendable recalcular  $\gamma(h)$  para distintos valores de h, hasta encontrar una forma suavizada del mismo.

➤ **Distribuciones con valores extremos.**

La existencia de valores extremos, altos o bajos, en una distribución, puede conducir a la obtención de un variograma fuertemente errático. En este caso la solución puede ser, eliminar los datos extremos porque pueden ser ocasionados por errores, en otros casos pueden encontrarse en zonas geográficamente distintas y pueden ser tratados de manera separada. Una herramienta útil para la detección de valores extremos y encontrar el incremento adecuado puede ser, calculado a través de lo que se conoce como la “**Nube de Variogramas**”, el cual consiste en representar los valores de  $[Z(x_i + h) - Z(x_i)]^2 / 2$  contra  $h$ , para cada par posible de la información inicial.

➤ **La existencia de poblaciones mixtas.**

Existen datos que pueden mostrar diferentes poblaciones. En muchos casos las poblaciones están geográficamente diferenciadas, donde se recomienda tratar las zonas por separado. En otros casos las poblaciones se presentan mezcladas geográficamente, en este caso una solución puede ser un cambio de escala, con lo que se logra reducir la diferencia de los valores extremos.

➤ **Heterocedasticidad.**

Se dice que los datos son heterocedásticos, si el diagrama del valor medio contra la desviación estándar  $\sigma$ , calculadas en vecindades móviles, muestra que la dispersión de los valores (medidos por  $\sigma$ ) está relacionada con su magnitud (medida por  $m$ ). Por la tanto, por ser el semivariograma una medida de la dispersión, su magnitud está ligada a la magnitud de los valores de los datos. Cuando se da este tipo de fenómeno, se dice que existe un efecto proporcional.

➤ **Desviaciones en el muestreo.**

En este caso, las observaciones  $\{Z(x_i), i = 1, \dots, n\}$  de la función aleatoria  $Z(x)$  aparece que no están distribuidas espacialmente en forma aleatoria. Consecuentemente se produce un sesgo en el muestreo, es decir, en la elección de las localizaciones de las observaciones, tendientes a estar agrupadas en áreas de altos o bajos valores. Estas desviaciones pueden ser detectadas mediante un gráfico del número de muestras contra el valor medio para vecindades de un tamaño determinado.

Otras razones por lo que los semivariogramas son erráticos:

1. No hay suficientes muestras.
2. Las muestras no son representativas del fenómeno.
3. Las clasificaciones de las muestras no son válidas.
4. El área estudiada es no homogénea.
5. Pequeños o largos conjuntos de datos son necesarios.
6. Pequeñas o largas distancias deben ser calculadas.
7. Más o menos distancias deben ser calculadas.
8. Pequeñas tolerancias son necesarias.
9. Las muestras pueden tener localizaciones incorrectas.
10. Los valores muestreados pueden ser erróneos.

El problema fundamental en la obtención de un semivariograma correcto es, la elección adecuada de los intervalos de distancias para los cuales será calculado el semivariograma, de

modo que en éstos la cantidad de pares encontrados sea suficiente, desde el punto de vista estadístico.

### 2.4.3 Condiciones que deben cumplir los modelos del semivariograma.

En la selección de una función adecuada para ser ajustada a un semivariograma muestral, se debe tener en cuenta tres elementos: un intercepto con la ordenada, una sección monótonamente creciente y una meseta. Sin embargo, no servirá cualquier modelo que aparente ajustarse a los valores empíricos debido a la siguiente razón.

Supongamos que  $Z(x)$  es una función aleatoria estacionaria de segundo orden de la cual obtenemos la variable regionalizada  $\{Z(x_i), i = 1, 2, \dots, n\}$ , donde su semivariograma y su función de covarianza son  $\gamma(h)$  y  $C(h)$  respectivamente.

Sea  $Y$  una combinación lineal de  $Z(x_i)$  tal que

$$Y = \sum_{i=1}^n \lambda_i Z(x_i) \quad (19)$$

donde  $\lambda_i, i = 1, 2, \dots, n$ , son pesos arbitrarios.

La magnitud  $Y$  es también una variable regionalizada con varianza:

$$Var[y] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(x_i, x_j) \quad (20)$$

La varianza de  $Y$  puede ser positiva o cero, pero no negativa y la función de covarianza en la parte derecha de la expresión anterior debe asegurar que esta condición sea satisfecha.

La matriz de covarianzas.

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$$

Donde  $\sigma_{ij} = C(x_i, x_j)$  debe ser definida positiva, es decir, el determinante y todos sus menores principales son positivos o cero. Aquí las covarianzas están definidas como  $\sigma_{ij} = \text{cov}[Z(x_i), Z(x_j)]$ ,  $i, j = 1, 2, \dots, n$ .

La función de covarianzas  $C(h)$  si existe, debe ser positiva por lo que serán permisibles funciones que cumplan este criterio.

Las propiedades que no tengan covarianzas *a priori* finitas no tendrán definida la función de covarianzas. En el caso de que cumplan la hipótesis intrínseca, entonces estará definido el semivariograma y se cumple que:

$$\text{Var}[Y] = C(0) \sum_{i=1}^n \lambda_i \sum_{j=1}^n \lambda_j - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(x_i, x_j) \quad (21)$$

Si la  $\sum_{i=1}^n \lambda_i = 0$  la expresión se reduce a:

$$\text{Var}[Y] = - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(x_i, x_j) \quad (22)$$

Esto implica que,  $-\gamma(h)$  debe ser condicionalmente positiva semidefinida (incluye el cero y los positivos), con la condición de que la  $\sum_{i=1}^n \lambda_i = 0$ . Esto, es equivalente a decir, que debe ser *condicionalmente negativa semidefinida*.

Como consecuencia de esta propiedad, se puede demostrar que el semivariograma debe tener un ritmo de crecimiento inferior a  $h^2$ , es decir, se debe cumplir que:

$$\lim_{h \rightarrow \infty} \frac{\gamma(h)}{h^2} = 0$$

Cuando no se satisface esta condición puede ser un indicador de que la función aleatoria no es estacionaria. Por otro lado, de la definición de semivariograma, se infiere que éste debe anularse en el origen, es decir,  $\gamma(h) = 0$ .

#### 2.4.4 Modelado de semivariogramas

El modelado de semivariogramas incluye dos etapas fundamentales, la primera es el semivariograma experimental o empírico y la segunda etapa es necesario ajustar a éste a un modelo teórico, con el objetivo de determinar los parámetros descriptivos del semivariograma, que posteriormente serán usados en la estimación.

#### 2.4.5 Parámetros del semivariograma

Los parámetros del semivariograma caracterizan tres elementos importantes en la variabilidad de un atributo que son: 1) la discontinuidad en el origen  $C_0$ , efecto de pepita 2) el valor máximo de variabilidad  $C$  (meseta), y 3) el área de influencia de la correlación  $a$ , alcance que se presenta en la figura 4.

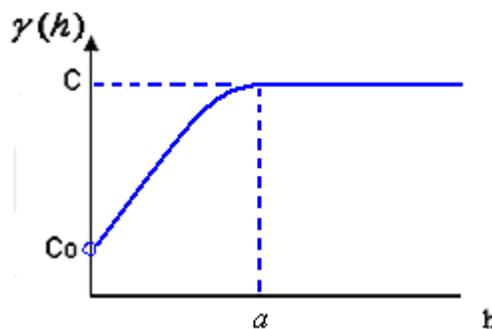


Figura 4. Parámetros del semivariograma.

Los parámetros del semivariograma consisten en:

1) **El Efecto Pepita (Nugget):**

Cuando el semivariograma por definición es nulo en el origen, pero en la práctica las funciones obtenidas pueden presentar discontinuidad en el origen, a esta discontinuidad se le llama **efecto de pepita**, (en inglés Nugget effect) y se denota por  $C_0$ . Si esta por debajo de cero, el valor asumido por este efecto es cero, para valores negativos de  $\gamma(0)$  no tienen significado y no es común.

2) **La Meseta (Sill):**

Es cuando el valor de  $\gamma(h)$  para el cual, con el aumento de  $h$ , su valor permanece constante. Se representa como  $C$  y se denomina meseta. Puede obtenerse trazando una línea paralela a la abscisa y que se ajuste a los puntos de mayor valor del semivariograma y su valor se lee en la intersección de esta línea con la ordenada.

3) **El Alcance (Range):**

La distancia  $h$  para la cual las variables  $Z(x)$  y  $Z(x+h)$  son independientes, se denomina alcance y se representa por  $a$ , es decir, las distancias para la cual los valores de la variable dejan de estar correlacionados, o lo que es lo mismo, la distancia para la cual el semivariograma alcanza su meseta.

#### 2.4.6 Modelos teóricos de semivariogramas

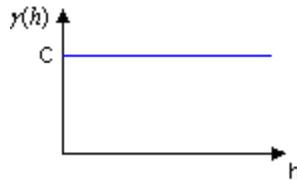
Los modelos teóricos de semivariogramas admisibles o más utilizados en la práctica se describen atendiendo a las dos características en el modelado de semivariogramas:

- 1) Su comportamiento en el origen, el cual puede ser lineal, parabólico y con Efecto de Pepita
- 2) La presencia o ausencia de meseta.

A continuación se presentan estos modelos con su respectiva descripción:

- a) **Efecto de Pepita:** Corresponde a un fenómeno puramente aleatorio (ruido blanco), sin correlación entre las muestras, cualquiera sea la distancia que las separe, se presenta en la figura 5, donde C representa el valor de la meseta.

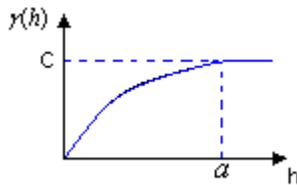
$$\begin{aligned} \gamma(h) &= 0 & h &= 0 \\ \gamma(h) &= C & |h| &> 0 \end{aligned}$$



**Figura 5.** Modelo Efecto Pepita

- b) **Modelo Esférico:** Este modelo es probablemente el más utilizado, es una expresión polinomial simple (figura 6), se puede observar un crecimiento casi lineal y después a cierta distancia finita del origen se alcanza una estabilización, la meseta. La tangente en el origen encuentra a la meseta en el punto de abscisa  $(2/3)a$ , donde  $a$  representa el valor del alcance.

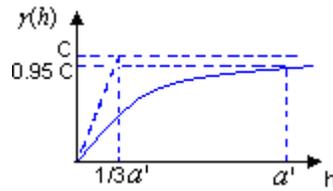
$$\begin{aligned} \gamma(h) &= C \left[ \left( \frac{3}{2} \right) \left( \frac{h}{a} \right) - \frac{1}{2} \left( \frac{h}{a} \right)^3 \right] & h &\leq a \\ \gamma(h) &= C & h &> a \end{aligned} \quad (23)$$



**Figura 6.** Modelo Esférico

- c) **Modelo Exponencial:** Este modelo, a diferencia del esférico, crece inicialmente más rápido y después se estabiliza de forma asintótica (figura 7). Como la meseta no se alcanza a una distancia finita, se usa con fines prácticos el “alcance efectivo” o “alcance práctico”  $a'$ , cuyo valor se obtiene en el punto de la abscisa para el cual el modelo obtiene el 95% de la meseta, con un valor  $a' = 3a$ , donde  $a$  es el parámetro de escala. La tangente en el origen encuentra a la meseta en el punto  $a = (1/3)a'$ .

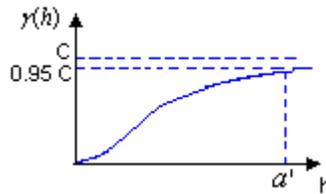
$$\gamma(h) = C [1 - \text{Exp}(-|h|/a)], \quad |h| > 0 \quad (24)$$



**Figura 7.** Modelo Exponencial

- d) Modelo Gaussiano:** Es un modelo extremadamente continuo (figura 8), inicialmente presenta un comportamiento parabólico en el origen, después al igual que en el modelo Exponencial se alcanza la meseta de forma asintótica. El alcance práctico tiene un valor de  $a' = 1.73a$ , que es el valor de la abscisa donde se alcanza el 95% de la meseta.

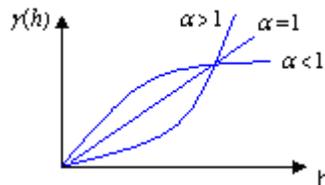
$$\gamma(h) = C [1 - \text{Exp}(-|h|^2/a^2)], \quad |h| > 0 \quad (25)$$



**Figura 8.** Modelo Gaussiano.

- e) Modelo con Función Potencia:** Es un modelo que se caracteriza porque no tiene meseta (figura 9), se obtiene para valores de  $\alpha$  correspondientes a 0.5, 1.0 y 1.5.

$$\gamma(h) = |h|^\alpha \quad \text{con } \alpha \in ]0, 2[ \quad (26)$$



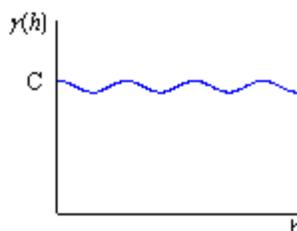
**Figura 9.** Modelo con Función Potencia.

Obsérvese que para el valor de  $\alpha=1$  se obtiene el modelo Lineal, el cual no tiene ni meseta ni alcance.

### f) Modelo de Efecto agujero (Periódico)

Este modelo por su comportamiento es indicativo de fenómenos con componentes periódicas es decir que tiene una forma senoidal (figura 10). La ecuación de este modelo viene dada por:

$$\gamma(h) = C \left( 1 - \frac{\text{sen}(h)}{h} \right) \quad \text{para } h > 0 \quad (27)$$



**Figura 10.** Modelo efecto agujero

Este modelo puede ser usado para representar procesos regularmente continuos y con un comportamiento periódico, el cual es frecuentemente encontrado, donde existe una sucesión de zonas ricas y pobres.

Otra alternativa que se utiliza para representar el modelo efecto agujero es:

$$\gamma(h) = C(1 - \cos(h)) \quad \text{para } h \geq 0 \quad (28)$$

#### 2.4.6.1 Validación del modelo teórico

En esta etapa se procede a realizar la evaluación y adecuación de los modelos a través del ajuste a los modelos teóricos utilizando el semivariograma experimental. Se realiza por medio de una curva aproximada, variando los valores  $C_0$  (efecto de pepita),  $C$  (meseta) y  $a$  (alcance). Tratando de seleccionar los posibles modelos candidatos y discriminando algunos de ellos, comprobando la calidad de la estimación, el cumplimiento de las condiciones de los parámetros hasta seleccionar aquel con los parámetros que mejor se ajusten al semivariograma teórico utilizando métodos de ajuste que nos proporciona el análisis geostatístico para la validación del modelo.

### 2.4.6.2 Métodos de ajuste

Algunos analistas de la geoestadística ajustan los modelos de forma visual, pero esta práctica no es fiable y es preferible usar algún procedimiento estadístico para tales fines.

Dentro de estos métodos se tiene:

1. Mínimos Cuadrados Ordinarios (MCO).

Este método es utilizado con frecuencia y como un primer momento para el ajuste. En este método se utilizan los residuos en la cual se estudian y se analizan para observar si los residuos están normalmente distribuidos, que son independientes y también que las semivarianzas estimadas poseen igual varianza.

2. Mínimos Cuadrados Generalizados (MCG).

Este método es utilizado con frecuencia y es utilizado también para el ajuste, pero este método la asume la relación entre las variables se considera que no es lineal y que la varianza no es constante en el tiempo, todo lo contrario al MCO, este método requiere de la inversión de grandes matrices tales como la matriz de varianza-covarianza, de información y procedimientos, lo que hace muy complicado su cálculo. Para resolverlo hay que usar el método de MCG que intenta estimar el cambiante valor de la varianza y utilizarlo para corregir los valores de la muestra.

3. Mínimos Cuadrados Ponderados.

Para ajustar por este método se requiere de minimizar la expresión siguiente:

$$W(\tau, \gamma_k) = \sum_{j=1}^k \frac{[\gamma(h_j) - \gamma(h_j, \tau)]^2}{Var(h_j)} \quad (29)$$

donde

$\tau$  es el vector de los parámetros del modelo,

$Var[\gamma(h_j)]$  es la varianza del valor estimado del semivariograma en  $h_j$ .

Desafortunadamente esta varianza es desconocida casi siempre. Con frecuencia se escoge en su lugar el número de pares de observaciones  $m(h_j)$  con que se estima la misma, suponiendo que es inversamente proporcional a la varianza de los estimados.

Cressie (1985), ha demostrado teóricamente que un mejor procedimiento es reemplazar  $Var[\gamma(h_j)]$  por  $[\gamma(h_j, \tau)]^2 / m(h_j)$ , dando un peso considerablemente mayor a las semivarianzas esperadas menores. La diferencia en los resultados de la aplicación de estos dos tipos de pesos es pequeña y se ha confirmado que cuando se utiliza  $m(h_j)$  se le asigna pesos muy pequeños a los estimados de los intervalos  $h_j$  menores (Martín A. Viera 2002 [3]).

Laslett ha sugerido un mejoramiento. Sus pesos están dados por:

$$\frac{m(h_j) \gamma(h_j)}{[\gamma(h_j, \tau)]^3} \quad (30)$$

y estos son actualizados en cada iteración del proceso cuando  $\gamma(h_j, \tau)$  es reestimada.

Teóricamente esta forma de ponderación en la práctica se ha encontrado que se obtienen resultados de mejor convergencia.

### 2.4.6.3 Criterio de Selección y Validación de los Modelos

Sea  $Z(x)$  una función aleatoria estacionaria con semivariograma  $\gamma(h)$ , su función de covarianza  $C(h)$  viene dada por  $C(h) = \sigma^2 - \gamma(h)$ , donde  $\sigma^2$  es la varianza de  $Z(x)$ . Sea  $Z_{x_1}, Z_{x_2}, \dots, Z_{x_n}$  los valores de  $Z(x)$  en  $n$  puntos medidos. La validación cruzada consiste en suprimir el  $i$ -ésimo valor medido  $Z(x_i)$  y estimarlo a partir del resto de los datos. El valor estimado  $Z^*(x_i)$  se calcula por krigeaje, procedimiento explicado más adelante, si se repite este proceso para los  $N$  puntos, se pueden calcular  $n$  errores de validación:

$$E(x_i) = Z^*(x_i) - Z(x_i) \quad i = 1, 2, \dots, N.$$

Así se van probando diferentes valores de los parámetros del semivariograma hasta que los errores de validación cumplen los siguientes criterios estadísticos:

1. El error medio, dado por  $T_1 = (1/n) \sum_{i=1}^n [Z(x_i) - Z^*(x_i)]$ , debe ser aproximadamente igual a cero.
2. El error medio cuadrado, dado por  $T_2 = (1/n) \sum_{i=1}^n [Z(x_i) - Z^*(x_i)]^2$ , debe ser pequeño.
3. La medida,  $T_3 = (1/n) \sum_{i=1}^n \{ [Z(x_i) - Z^*(x_i)] / \sigma \}^2$ , debe ser igual a uno.
4. La medida,  $T_4 = \text{Corr}\{ [Z(x_i) - Z^*(x_i)] / \sigma, Z^*(x_i) \}$ , debe ser cero.
5. La medida,  $T_5 = |\text{Corr} [Z(x_i), Z^*(x_i)]|$ , debe ser uno.

Otros autores sólo plantean que los criterios estadísticos fundamentales son los indicados por  $T_1$  y  $T_3$ .

Idealmente, todos los criterios estadísticos anteriores deben satisfacerse simultáneamente, pero en la práctica una mejoría en uno de ellos puede degradar a otro. Por lo que es recomendable hacer un análisis integral de los estadísticos de las diferencias cuando se tiene mucha degradación entre un estadístico y otro. También, estos criterios estadísticos pueden utilizarse en los métodos de interpolación para la validación y selección de uno de ellos.

Otro método importante es el análisis de los errores por medio del histograma de los errores normalizados se compila una lista con los puntos que poseen grandes errores o sea aquellas rangos en los cuales su valor sobre pasa la curva de la normalidad. Esto es útil para la identificación de valores atípicos (outliers), datos sospechosos o anomalías de otra naturaleza.

La eliminación de las localizaciones con elevados errores normalizados (datos atípicos) y el recálculo del semivariograma y su modelación puede producir mejorías significativas.

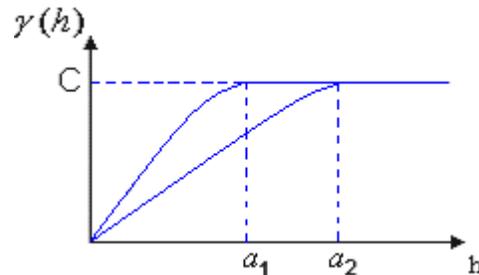
El procedimiento de *leave one out* creado por D. G. Kriging es un caso particular del método conocido como Jackknifing, ya que de disponer de suficientes datos espacialmente distribuidos de forma homogénea de la función aleatoria  $Z(x)$ , se podrían dividir en dos muestras:  $Z_1(x)$  y  $Z_2(x)$ . La primera se usaría para estimar el variograma, mientras que la segunda serviría para validarlo. Es decir, se estimarían los valores empleando el método de kriging Ordinario con el modelo de variograma obtenido usando la primera muestra  $Z_1(x)$  en los puntos correspondientes a las observaciones que pertenecen a la segunda muestra  $Z_2(x)$  y se evaluarían los estadígrafos de las diferencias  $Z_2(x_i) - Z_2^*(x_i)$  de manera análoga, esto significa que se van calculados los diferentes errores entre el valor real y el estimado hasta obtener  $n$ -errores de validación.

#### 2.4.7 Análisis de anisotropía.

Para efectuarlo es necesario realizar un análisis sobre el comportamiento de la variabilidad del atributo en estudio. Se conoce que el semivariograma describe las características de continuidad espacial de la variable regionalizada en una dirección, pero este comportamiento puede variar según la dirección que se analice, Journel y Huijbregts (1978). Se exige por este motivo un análisis del comportamiento de la continuidad en distintas direcciones, a este tipo de procedimiento se le conoce como Análisis de Anisotropía.

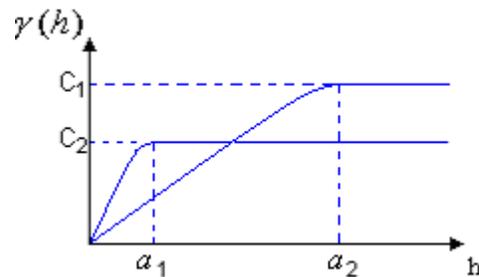
También se dice que el fenómeno es **Isotrópico** cuando el semivariograma calculado en diferentes direcciones (norte-sur, este-oeste, y en direcciones intermedias de  $45^\circ$  o de  $22.5^\circ$ ) muestra similar comportamiento y cuando el semivariograma muestra diferentes comportamientos es llamado **Anisotrópico** (propiedad que presentan ciertos cuerpos consistente en la dependencia de sus propiedades de la dirección que en ellos se considere). Los tipos de anisotropías más comunes son la Geométrica y la Zonal.

**Anisotropía Geométrica:** Está presente cuando los semivariogramas en diferentes direcciones tiene la misma meseta pero distinto alcance.



**Figura 11.** Semivariograma que representa la Anisotropía Geométrica

**Anisotropía Zonal:** Está presente cuando los semivariogramas en diferentes direcciones tiene diferentes mesetas y alcances.



**Figura 12.** Semivariograma que representa la Anisotropía Zonal.

Sin embargo, hay otro tipo de anisotropía que requiere el uso de transformaciones de coordenadas, con el objetivo de obtener modelos Isotrópicos. A continuación se detallan.

#### 2.4.8 Efecto proporcional

Cuando en el cálculo del semivariograma se detecta que existe una relación lineal entre el valor medio de las muestras usadas en el cálculo de cada  $\gamma(h)$  y la desviación estándar correspondiente, se dice que existe un efecto proporcional (heterocedasticidad). Este efecto se puede observar y detectar presentando los valores de  $\bar{x}$  contra  $\sigma$ , es decir, que el coeficiente de variación ( $\sigma / \bar{x}$ ) sea aproximadamente constante, esto ocurre cuando los datos presentan

una distribución lognormal (figura 13). “La solución a este problema consiste en dividir cada valor del semivariograma local por el cuadrado de la media local, y obtener lo que se conoce como semivariograma relativo” (David en 1977), es decir

$$F(h) = \gamma(h) / \bar{x}(h)^2$$

$$F(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \frac{[Z(x_i) - Z(x_i + h)]^2}{[(Z(x_i) - Z(x_i + h))/2]^2} \quad (33)$$

Gráficamente tenemos

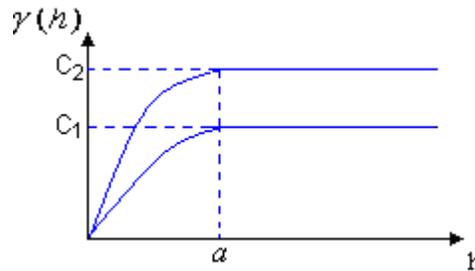


Figura 13. Efecto Proporcional

#### 2.4.9 Problemas en el modelaje de semivariogramas

Los problemas más comunes al modelar semivariogramas que complican este proceso se resumen los siguientes casos (Krajewski y Gibbs 1993).

- 1) **La anisotropía geométrica está presente:** Indica que los semivariogramas direccionales tienen la misma meseta pero diferentes alcances, ésta puede ser corregida a través de una transformación lineal de coordenadas que permita reducir una elipse a un círculo.
- 2) **La anisotropía zonal está presente:** indica que tanto las mesetas como los alcances son diferentes para los semivariogramas direccionales, puede ser corregido separando el semivariograma en sus dos componentes; isotrópicos horizontal y anisotrópico vertical.

- 3) **La tendencia de los datos está presente:** indica que los valores medidos aumentan o disminuyen dramáticamente en la zona estudiada con el aumento de la distancia. Esto puede ser resuelto aplicando polinomios a la ecuación del semivariograma, es decir, un análisis de tendencia.
- 4) **El efecto proporcional está presente:** Indica que la desviación estándar local es proporcional al cuadrado de la media local y que los datos presentan una distribución lognormal, la cual puede ser resuelto dividiendo cada valor del semivariograma local por el cuadrado de la media local, es decir, usando semivariogramas relativos.
- 5) **Existencia de estructuras anidadas:** indica que diferentes procesos operan a diferentes escalas, como por ejemplo a) a muy pequeñas distancias la variabilidad puede estar presente debido a cambios de una composición a otra. b) a pequeñas distancias la variabilidad puede estar presente debido a errores. c) a grandes distancias la variabilidad puede estar presente debido a casos transitorios de desgaste. Este puede ser resuelto aplicando varios modelos simultáneamente.
- 6) **Existencia de efecto hueco:** indica que muy pocos pares están disponible para la comparación a una distancia específica. Y puede ser resuelto recuperando más casos para la distancia definida.
- 7) **La periodicidad está presente:** indica que el comportamiento del semivariograma repite por sí mismo periodicidades, por ejemplo: El valor de la meseta puede aumentar o disminuir sistemáticamente, o bien el caso en que los valores son tomados alternativamente a través de diferentes estratos, como piedras areniscas, esquistos, etc. Esto puede ser resuelto si es un problema real y no un antifaz del análisis, la periodicidad puede ser también un fenómeno real mostrado por zona ricas y pobres repetidas a espacios similares.

## 2.5 Estimación

Lo expresado hasta aquí tiene un único objetivo, conocer la información disponible para realizar estimaciones es decir, estimar valores desconocidos a partir, no sólo de los conocidos, sino también de su estructura de continuidad espacial. Lo que establece la diferencia con la variedad de métodos de interpolación es que no utilizan estas características y que se emplean actualmente con diferentes fines. Algunos métodos de estimación que se utilizaran se comentan a continuación.

## 2.6 Métodos de Interpolación.

Estos métodos consisten en procedimientos matemáticos utilizados para predecir el valor de un atributo en una localización precisa a partir de valores del atributo obtenidos de localizaciones vecinas ubicadas al interior de la misma región se utiliza la interpolación cuando:

- a) En la transformación de un numero finito de observaciones a un espacio continuo.
- b) Obtener un patrón espacial comparable compatrones espaciales ya conocidos.

A continuación se hace referencia a algunos métodos.

### 2.6.1 Método de la Triangulación

El método consiste en ajustar un plano que pase por las tres muestras más cercanas y adyacentes a la localización que se desea estimar.

La ecuación del plano es:

$$Z = ax + by + c \quad (34)$$

Cada punto medido tiene coordenadas (x, y) y z representa el valor muestreado.

Con el objetivo de obtener la ecuación del plano que pase por los tres puntos se construye el siguiente sistema de ecuaciones lineales, la cual se resuelve utilizando:

$$\begin{aligned} ax_1 + by_1 + c &= z_1 \\ ax_2 + by_2 + c &= z_2 \\ ax_3 + by_3 + c &= z_3 \end{aligned}$$

y así obtenemos los coeficientes a, b y c, entonces el valor de  $z$  se puede obtener en cualquier localización dentro del triángulo correspondiente.

### 2.6.2 Método Inverso de la distancia

Este método se basa en una combinación lineal dada por:  $Z^*(x) = \sum_{i=1}^n \lambda_i Z(x_i)$ , donde  $\lambda_i$  representa los pesos proporcionales al inverso de la distancia euclidiana entre las localizaciones muestreadas y la que se desea estimar. Estos pesos se obtienen mediante la ecuación siguiente:

$$d(x_i, x_j) = \sqrt{(x_i - x_j)^T (x_i - x_j)} \quad (35)$$

Entonces;

$$\lambda_i = (1/d_{oi}) / \sum_{j=1}^n 1/d_{oj} \quad (36)$$

De donde:  $d_{oi}$  es la distancia entre la localización a estimar y la localización de la muestra  $i$  y los  $d_{oj}$  es la distancia entre la localización a estimar y la localización de la muestra  $j$ .

Generalizando obtenemos:

$$Z^*(x) = \left[ \sum_{i=1}^n (1/d_{oi})^w Z(x_i) \right] / \sum_{j=1}^n (1/d_{oj})^w \quad (37)$$

Donde  $w$  es el exponente que genera diferentes tipos de ecuaciones, el comúnmente utilizado en la mayoría de los casos es para  $w = 1$  se tiene:

$$Z^*(x) = \left[ \sum_{i=1}^n (1/d_{oi}) Z(x_i) \right] / \sum_{j=1}^n 1/d_{oj} \quad (38)$$

Estas dos técnicas de estimación utilizan directamente los valores muestreados en el proceso de estimación y refieren pesos de acuerdo a las distancias entre los datos, sin tener en cuenta la continuidad espacial de la información disponible. Veamos ahora el **krigeaje**, interpolador de la geoestadística, que sí utiliza los resultados discutidos en el análisis estructural.

### 2.6.3 Métodos de Krigeaje

#### 2.6.3.1 El mejor estimador lineal insesgado.

El kriging es un término que ha sido acuñado para designar al "*mejor estimador lineal insesgado*" de un punto. Este nombre apareció alrededor de 1960 para nombrar una técnica creada en Francia por Matheron, a partir de los trabajos de D. G. Krige quién fue probablemente el primero que hizo uso de la correlación espacial y del mejor estimador lineal insesgado en el campo de la evaluación de yacimientos minerales.

El kriging es una técnica de estimación local que proporciona el mejor estimador lineal insesgado de una característica desconocida que se encuentre en estudio, además proporciona un error de estimación conocido como varianza del krigeaje, que depende del modelo de variograma. Con este tipo de análisis se obtiene una mejor calidad de estimaciones. La limitación a la clase de estimadores lineales es bastante natural ya que esto significa que solamente se requiere el conocimiento del momento de segundo orden de la función aleatoria (la covarianza o el variograma) y que en general en la práctica es posible inferir a partir de una realización de la misma.

#### 2.6.3.2 Ecuaciones del Estimador de krigeaje

Supóngase que se dispone de valores muestreados  $Z(x_i)$ ,  $i=1, \dots, n$  y deseamos estimar un valor de la característica observada en el panel  $Z(v)$ , la cual esta dada por una combinación lineal de  $Z(x_i)$ , es decir

$$Z^*(v) = \lambda_i Z(x_i) \quad (39)$$

$$\sum_{i=1}^n \lambda_i = 1 \quad (40)$$

donde  $Z^*(v)$  es el valor estimado y  $\lambda_i$  son los pesos de krigeaje, de modo que los  $\lambda_i$  sean obtenidos de tal forma que proporcione un estimador insesgado  $E[Z^*(v) - Z(v)] = 0$  y de varianza mínima  $Var[Z^*(v) - Z(v)]$ .

La geoestadística exige en su segunda etapa el conocimiento del comportamiento estructural de la información, es decir, debe contar con el modelo de semivariograma teórico que refleje las características de correlación espacial de la información disponible se presenta el sistema krigéaje para este caso:

### 2.6.3.3 Método Krigéaje Simple

Supóngase que hay una variable regionalizada estacionaria con media  $m$ , y covarianza conocidas y sea  $Z(x)$  la variable de interés medida en el sitio  $x$ , donde  $m$  está dado por  $m = E[Z(x)]$ .

El estimador y el sistema a resolver para encontrar los pesos  $\lambda_i$  y los valores estimados para  $Z^*(x_j)$  está dada por

$$\text{Estimador: } Z^*(v) = \sum_{i=1}^n \lambda_i Z(x_i) + m(1 - \sum_{i=1}^n \lambda_i). \quad (41)$$

$$\text{Sistema: } \sum_{i=1}^n \lambda_i C(x_i, x_j) = C(x_j, v) \quad j = 1, \dots, n \quad (42)$$

$$\text{Varianza de krigéaje: } \sigma^2 = C(v, v) - \sum_{i=1}^n \lambda_i C(x_i, x_j) \quad (43)$$

### 2.6.3.4 Método Krigéaje Ordinario

Suponga que se hacen mediciones de la variable de interés  $Z$  en los puntos  $x_i$ ,  $i = 1, 2, \dots, n$  de la región de estudio, es decir se tienen realizaciones de las variables  $Z(x_1), \dots, Z(x_n)$  y se desea predecir  $Z(v)$ , en el punto  $v$  donde no hubo medición. En esta circunstancia, el método kriging ordinario propone que el valor de la variable puede predecirse como una combinación lineal de las  $n$  variables aleatorias así:

$$Z^*(v) = \lambda_1 Z(x_1) + \lambda_2 Z(x_2) + \lambda_3 Z(x_3) + \dots + \lambda_n Z(x_n) \text{ es decir}$$

$$Z^*(v) = \sum_{i=1}^n \lambda_i Z(x_i). \text{ A continuación se muestra el estimador y el sistema a resolver para}$$

encontrar los pesos  $\lambda_i$  y  $Z^*(v)$

**En términos de la covarianza**

$$\text{Estimador:} \quad Z^*(v) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (44)$$

$$\text{Sistema:} \quad \begin{cases} \sum_{i=1}^n \lambda_i C(x_i, x_j) - \mu = C(x_j, v) & i, j = 1, \dots, n \\ \sum_{i=1}^n \lambda_i = 1 \end{cases} \quad (45)$$

$$\text{Varianza de krigeaje:} \quad \sigma^2 = C(v, v) - \sum_{i=1}^n \lambda_i C(x_i, v) + \mu \quad (46)$$

Donde  $\mu$  es un multiplicador de Lagrange

### En términos del semivariograma

$$\text{Estimador:} \quad Z^*(v) = \sum_{i=1}^n \lambda_i Z(x_i)$$

$$\text{Sistema:} \quad \begin{cases} \sum_{i=1}^n \lambda_i \gamma(x_i, x_j) + \mu = \gamma(x_j, v) & i, j = 1, \dots, n \\ \sum_{i=1}^n \lambda_i = 1 \end{cases} \quad (47)$$

$$\sum_{i=1}^n \lambda_i = 1$$

$$\text{Varianza de krigeaje:} \quad \sigma^2 = \sum_{i=1}^n \lambda_i (\gamma(x_i, v) - \gamma(v, v)) + \mu \quad (48)$$

#### 2.6.3.4.1 Ecuaciones del Kriging ordinario en forma matricial

De forma análoga al sistema de ecuación de Kriging Ordinario podemos representar el sistema de ecuaciones del Kriging en forma matricial como sigue:

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} & 1 \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} & 1 \\ \dots & \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \lambda \\ \dots \\ \lambda \\ -\mu \end{bmatrix} = \begin{bmatrix} \sigma_{k1} \\ \sigma_{k1} \\ \dots \\ \sigma_{kn} \\ 1 \end{bmatrix}$$

donde  $\sigma_{ki} = \sigma_{x_i v} = C(x_i, v)$

Este sistema de ecuaciones tiene solución única sí y solo sí la matriz es no singular, lo cual se cumple cuando la matriz de covarianzas es estrictamente positiva definida. Esto se puede

garantizar si se usan modelos de función de covarianzas estrictamente positivo definidos y no se incluyen puntos duplicados. Es decir, no se puede usar una muestra donde la función aleatoria tome dos valores distintos en un mismo punto.

No obstante la matriz resultante en la práctica resulta no positiva definida, pero este problema se resuelve de manera simple, eliminando datos erróneos y los datos faltantes.

Nótese que bajo la hipótesis intrínseca las covarianzas pueden ser reemplazados por las semivarianzas, usando la expresión:

$$\sigma_{ij} = \sigma^2 - \gamma_{ij}$$

Donde  $\sigma^2 = C(0)$ , es la varianza total o *a priori* de la muestra y representa la covarianza cuando  $h=0$  que es igual a la varianza.

Entonces sustituyendo la expresión de la covarianza en función del semivariograma obtenemos el sistema del Kriging en función del semivariograma:

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1n} & 1 \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2n} & 1 \\ \dots & \dots & \dots & \dots & \dots \\ \gamma_{n1} & \gamma_{n2} & \dots & \gamma_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \lambda \\ \dots \\ \lambda \\ -\mu \end{bmatrix} = \begin{bmatrix} \gamma_{k1} \\ \gamma_{k1} \\ \dots \\ \gamma_{kn} \\ 1 \end{bmatrix}$$

De manera análoga, este sistema de ecuaciones tiene solución única sí y solo sí la matriz es no singular, lo cual se cumple cuando la matriz de covarianzas es condicionalmente negativa definida de manera estricta. Esto se puede garantizar usando un modelo válido de variograma. Sin embargo, a diferencia del caso anterior no existe una simplificación evidente para la solución numérica del sistema dado que sería demasiado complicado al desarrollar cada ecuación.

Y la varianza de la estimación se calcula por medio de la ecuación:

$$\sigma_e^2 = \sum_{i=1}^n \lambda_i \gamma_{ki} + \mu \quad (52)$$

Estas últimas expresiones resultan ser la formulación del Kriging más común puesto que se aplica en el caso cuando la variable regionalizada cumple la hipótesis intrínseca la cual es una condición menos exigente que la estacionaridad de segundo orden.

Al sistema krigeaje es necesario hacer algunas observaciones:

- a) El sistema krigeaje tiene solución única si y solo si la matriz es definida estrictamente

positiva, es decir:  $\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(x_i, x_j) \geq 0$  o en términos de variograma

$$- \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(x_i, x_j) \geq 0$$

La primera condición fue demostrada anteriormente en la ecuación (21)

- b) El krigeaje, el cual es un estimador imparcial, es también un interpolador, es decir, para iguales soportes de observación  $v_\alpha$  ( $\alpha=1, \dots, n$ ) y de estimación  $V$ , Los valores real  $Z_\alpha$  y estimado  $Z^*$  son iguales, además de que la varianza de krigeaje  $\sigma_k^2$  es cero.
- c) Las expresiones del sistema krigeaje y de la varianza de krigeaje son completamente generales, es decir, son aplicables cualquiera sean los soportes de observación y estimación y el modelo estructural empleado.
- d) El sistema krigeaje y la varianza de krigeaje dependen sólo: del modelo estructural  $C(h)$  o  $\gamma(h)$  obtenido y de la geometría del soporte de observación. Esta característica da la posibilidad de que la varianza de krigeaje sea usada cuidadosa y convenientemente para el estudio de redes y la clasificación de recursos.

### 2.6.4 El caso no estacionario, Método Krigeaje Universal (KU)

Este es el caso de un fenómeno no estacionario, es decir, lo que hace imposible la aplicación del krigeaje presentado hasta aquí, dado de que los métodos anteriores solo son utilizados para el caso en que la serie es estacionaria. Con el objetivo de solucionar este problema Matheron (1970) propuso dos aproximaciones, primero el Krigeaje Universal (KU), que consiste en extraer de la variable original  $Z(x)$ , la parte no estacionaria, por medio de una componente determinística  $m(x)$  que representa la deriva, hasta encontrar la parte estacionaria del fenómeno, obteniéndose un componente estocástico  $R(x)$  relacionados por la siguiente expresión:  $Z(x) = m(x) + R(x)$ .

Para el componente determinístico se sugiere utilizar una función polinomial de las coordenadas para modelar la tendencia, es decir:

$$m(x) = \sum_{l=0}^K a_l f^l(x) \quad (53)$$

donde  $a_l$  son coeficientes que representan la tendencia de la función y  $f^l(x)$  es la función que describe la tendencia. Así pueden obtenerse derivas simples, lineales, cuadráticas, etc. Para una deriva simple el KU se reduce al Krigeaje Ordinario.

Obteniéndose finalmente el sistema Krigeaje Universal.

$$\sum_{\beta=1}^N \lambda_{\beta} \gamma(x_{\alpha}, x_{\beta}) + \sum_{l=0}^K a_l f^l(x_{\beta}) = \gamma(x_{\alpha}, x_o) \quad (54)$$

$$\sum_{\alpha=1}^N \lambda_{\alpha} f^l(x_{\alpha}) = f^l(x_o)$$

Con varianza de estimación.  $\sigma_{KU}^2 = \sum_{\alpha=1}^N \lambda_{\alpha} \gamma(x_{\alpha}, x_{\beta}) + \sum_{l=0}^K a_l f^l(x_{\beta}) \quad (55)$

Una variante de krigeaje que tiene en cuenta esta situación, fue desarrollada por Goldberger, A, S. en 1962 y descrita por Matheron en 1969, para tratamiento de datos débilmente estacionarios y con tendencia. La aplicación de KU puede resultar difícil por la indeterminación de la tendencia y del semivariograma.

Una aproximación más general, es el estudio del modelo de **Funciones Aleatorias Intrínsecas de orden  $K$** , la cual consiste en construir incrementos de orden creciente hasta alcanzar un orden  $K$  para el cual dichos incrementos son estacionarios. La desventaja de todos estos métodos de interpolación es que nunca van hacer igual a los valores.

#### 2.6.4.1 Aspectos Prácticos del Kriging Universal

En este método deben ser conocidos a priori el orden  $k$  del polinomio que mejor describe o explica la tendencia y la función de semivarianzas o variograma  $\gamma(x_i, x_j)$  de la función aleatoria  $Z(x)$  sin tendencia.

Los supuestos anteriores nos conducen en la práctica a dos problemas:

- a) El orden  $k$  del polinomio nunca es conocido, hay que adivinarlo.
- b) El variograma tampoco es conocido y hay que estimarlo a partir de los residuos, es decir  $R(x) = Z(x) - m(x)$ .

Matheron señaló (1971) que tales estimadores son sesgados y que a menos que el variograma de los residuos sea conocido, digamos que a partir de su estimación en una dirección donde no se perciba la deriva, el kriging universal tiene serias dificultades operacionales.

Como ya se ha visto el kriging universal puede ser expresado de manera muy simple. El problema real es que cuando existe una sola realización de la función aleatoria no estacionaria  $Z(x)$ , resulta imposible estimar el variograma. Lo que uno puede hacer es intentar eliminar la deriva  $m(x)$  y trabajar con los residuos  $R(x) = Z(x) - m(x)$  los cuales son estacionarios o al menos intrínsecos. Esto significa que la deriva tiene que ser estimada a partir de los valores muestrales.

## **CAPITULO 3: AJUSTE DEL MODELO.**

### **3.1 Introducción.**

Preocupa a los investigadores la búsqueda del o los modelos apropiados que describan el fenómeno en estudio y para ello contamos con la teoría clásica y los elementos básicos para el análisis geoestadístico de la información. Es así que a continuación se propone la metodología para la interpolación de la precipitación diaria en El Salvador.

Se muestran aspectos como el tipo de investigación, los procedimientos que se utilizarán para llevar a cabo dicha investigación, además se estudiarán tres modelos para tres variables, una es el promedio de milímetros (mm) de lluvia caída menor que 10 mm, la otra variable será calculado por el promedio que este entre 10 mm y 20 mm, y la última variable se utilizará en las precipitaciones que el promedio sea mayores de 20 mm de lluvia caída por día. La razón por la cual se hace esta diferencia de rangos es por que al calcular un solo modelo de manera global no se podrá ajustar los datos para los días en los cuales la precipitación sea menor que 10 o en aquellos días en que la precipitación media sobrepase los 20 mm de lluvia es por ello que se plantearán y analizarán tres modelos para la automatización de los mismos.

### **3.2 Fases de la Investigación**

Dentro de estas fases se consideran las siguientes:

- (a) El análisis exploratorio de los datos,**
- (b) El análisis estructural**
- (c) Estimaciones**

**La Fase I. Análisis Exploratorio,** Esta basado en el análisis exploratorio de datos, se utilizarán técnicas estadísticas convencionales, tales como el calculo de las medidas de la tendencia central y la dispersión en la que se pretende observar el comportamiento de las variables  $X_1$ ,  $X_2$ ,  $X_3$  estas variables se describen más adelante. Además se requiera que la serie de datos de la variable presente una distribución normal o lognormal con sesgo positivo

en nuestro caso. Los datos atípicos en esta etapa no se modificaran debido a que son valores reales, con respecto a los datos faltantes, el proceso a seguir para continuar el análisis será eliminarlos de la base de datos de la precipitación dado que no existe un mecanismo por parte del SNET para relleno de datos faltantes.

### **Fase 2. Análisis estructural.**

Se trata en esta fase el análisis para realizar el cálculo del variograma experimental con los datos en estudio de la precipitación, luego representamos el gráfico y se le ajustara por medio de la comparación con un variograma teórico, posteriormente se proseguirá a calcular los respectivos parámetros del variograma ajustado y finalmente, la validación del mismo. Si el modelo es adecuado se pasará a la fase siguiente en caso de no serlo se ajustara otro modelo de variograma teórico y así sucesivamente hasta lograr el mejor ajuste.

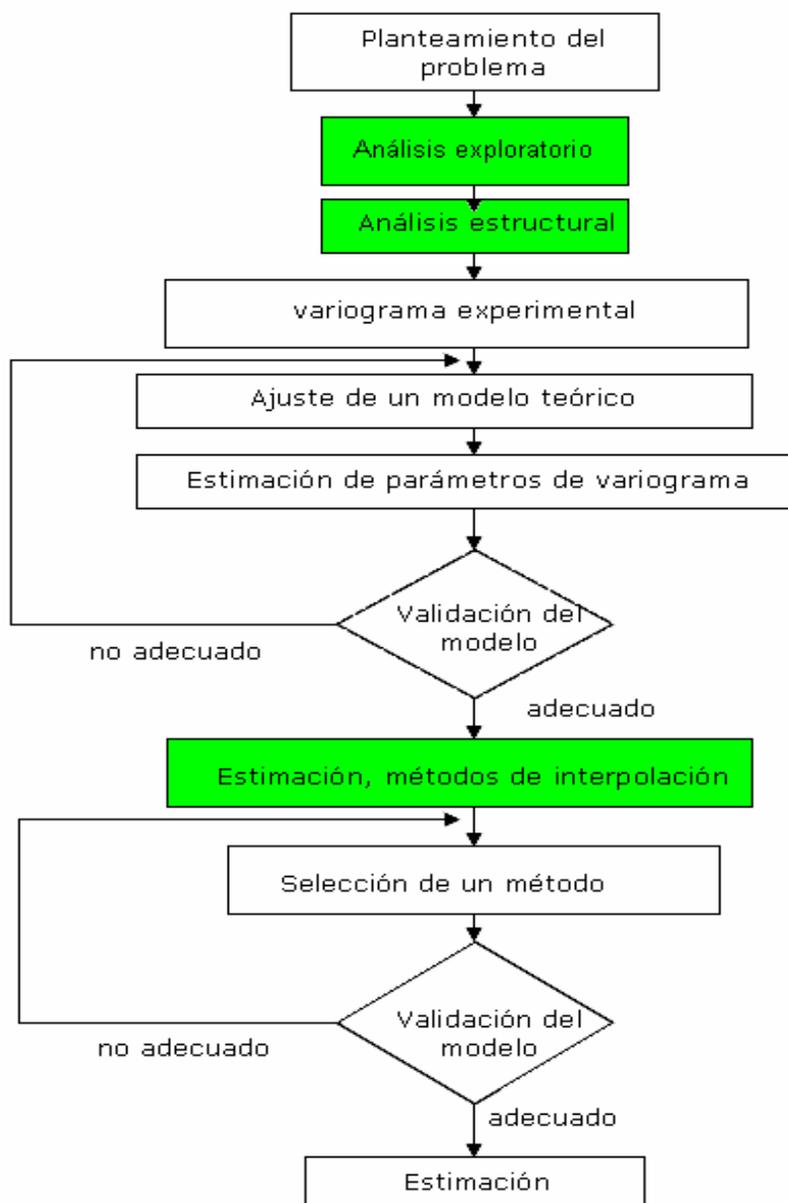
Como última fase tenemos la estimación de la precipitación utilizando los diferentes métodos de interpolación descritos en el capítulo anterior, de los métodos se seleccionará uno de ellos, para cada uno de los rangos o variables descritas X1, X2, X3, después se continuara con la validación del mismo, si es adecuado realizaremos la estimación sino volveremos a seleccionar otro método de interpolación.

Después de describir la metodología a utilizar en la aplicación de análisis geoestadístico para la interpolación de la precipitación se presenta la ejecución de dicha metodología.

Para la realización de estas etapas se utilizaran los paquetes siguientes: el paquete estadístico R y GRASS GIS, estos paquetes se utilizan por dos razones la primera es que no tienen costo algunos para poder obtenerlos y utilizarlos, la segunda razón es que proporcionan buenos resultados y no son muy difíciles de utilizar estos paquetes.

A continuación se presenta el esquema de la metodología a seguir para en la interpolación de la precipitación en El Salvador.

## Metodología para el Análisis Geoestadístico para la Precipitación.



**Figura 14.** Esquema metodológico

### 3.3 Aplicación de las fases a la variable precipitación en El Salvador.

#### 3.3.1. Fase I. Análisis exploratorio de los datos.

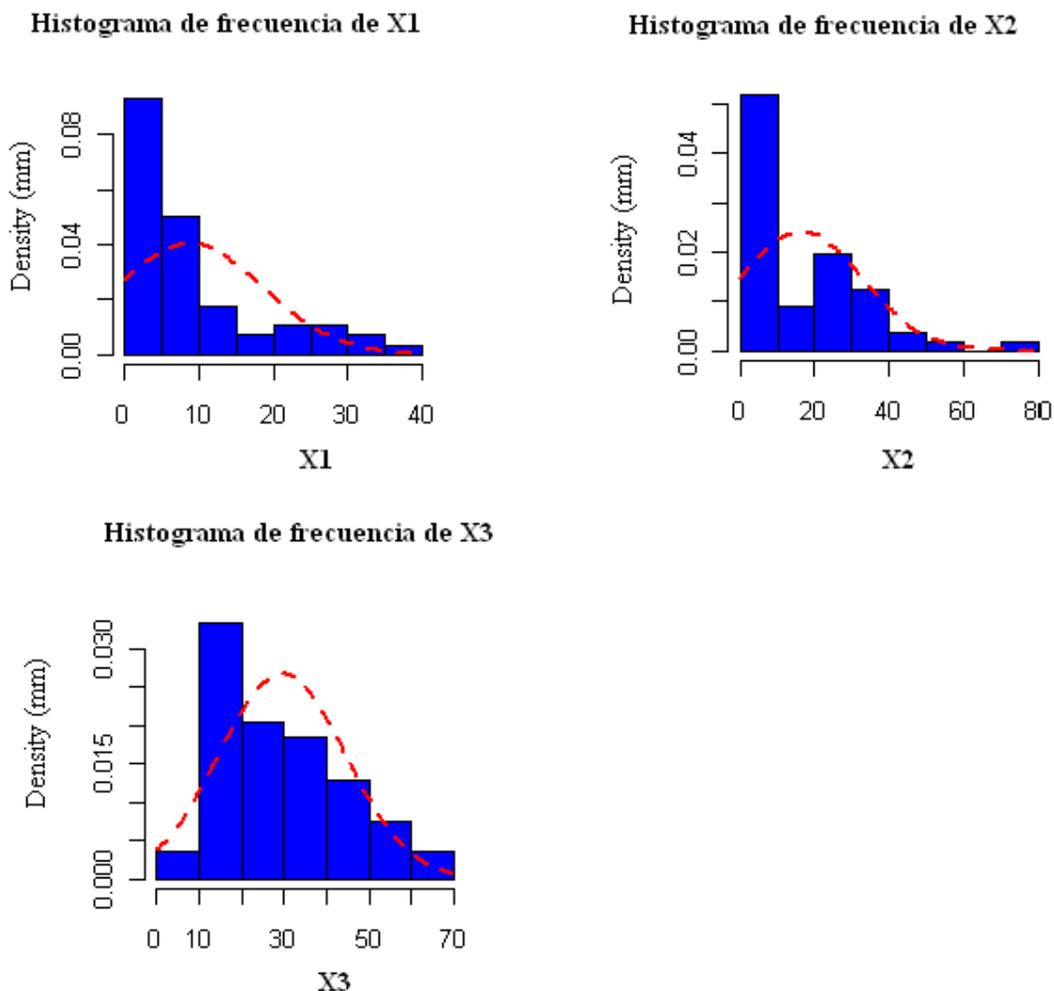
Para iniciar el análisis exploratorio se reviso la base de datos de la precipitación Tabla A1, esto dio lugar a la construcción de tres variables; X1=promedio lluvia caída sea menor que 10 *mm* , X2=Promedio de lluvia que este entre 10 *mm* y 20 *mm* , X3= precipitaciones que el promedio sea mayores de 20 *mm* , que cada una de ellas representa la cantidad de lluvia caída. y darle tratamiento a cada una obteniendo los estadísticos como se muestra en la Tabla A2. También se muestra el respectivo histograma de cada una de las variables descritas anteriormente, todo el análisis exploratorio se realizara utilizando el paquete estadístico **R**.

Tabla A2. Estadísticos Descriptivos.

	X1	X2	X3
<b>Min.</b>	0	0	7
<b>Max.</b>	38	72.40	65
<b>Media</b>	8.716	16.36	29.35
<b>Mediana</b>	5.55	9.45	29.58
<b>Moda</b>	0	31.4	23.6
<b>Varianza</b>	98.19774	272.2464	222.6
<b>Desviación estándar</b>	9.909477	16.49989	14.91978
<b>1st Cuantil</b>	0.775	2.60	17.57
<b>3rd Cuantil</b>	11.275	28.10	39.88

Como se puede observar en la tabla A2 para la variable X1 el valor máximo de lluvia caída para esta es de 38 mm mientras que para la variable X3 es 65 mm, lo cual muestra que hay una gran diferencia en relación los demás estadísticos de las otras variables. También se puede observar que la varianza de X2 es grande con respecto a las demás variables, esto significa que los valores observados con respecto a su media están muy separados.

### 3.3.1.1 Histograma de frecuencia.



**Figura 15.** Histograma de frecuencias

Como se observa en los histogramas anteriores, las variables presentan cierta normalidad es decir, que las precipitaciones presentan un comportamiento normal por que la mayoría de la información de las precipitaciones se encuentra dentro de la curva de normalidad. Además se puede observar que existe la presencia de barras dentro de la curva de normalidad la cual representa el rango de lluvia caída que se registro en las diferentes estaciones meteorológicas y también existen barras altas fuera de la curva de normalidad esto se debe a que existen valores atípicos, los cuales en nuestro caso no se pueden eliminar dado que son reales y que no se deben a errores de medición, la mayoría de datos atípicos se presentan en los últimos rangos dado de que es ahí es donde se encuentran los valores más alejados con respecto a su a media y

esto ocasiona a que el histograma se alargue con respecto al eje horizontal (número clases o rangos).

### 3.3.2. Fase II. Análisis Estructural.

El análisis estructural se realizara en dos etapas estas son:

- El cálculo del semivariograma experimental.
- El ajuste a este de un modelo teórico conocido.

El variograma experimental se define como la varianza de los incrementos de la variable regionalizada en las localizaciones separadas una distancia  $h$ .

$$\text{Var}\{Z(x+h) - Z(x)\} = 2\gamma(h)$$

La función  $\gamma(h)$  se denomina **semivariograma**, la cual puede ser estimada por la función

siguiente.

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

donde:  $Np(h)$  es el número de pares a la distancia  $h$ .

$h$ : determina la distancia entre dos observaciones en una cierta dirección;  $Z(x_i)$ : valores experimentales;  $x_i$ : localizaciones donde son medidos los valores  $Z(x_i)$ .

Una vez se ha estimado la función procedemos a la grafica de la nube de puntos resultantes y luego lo compararemos y le asignaremos un modelo teórico que mejor se ajuste a los datos; estos modelos pueden ser: **Efecto de Pepita, Modelo Esférico, Modelo Exponencial, Modelo Gaussiano, Modelo con función potencia, Modelo de Efecto agujero**, luego dentro del proceso metodológico seleccionamos el mejor modelo utilizando los estadísticos de prueba, por medio de los parámetros de la validación cruzada los cuales son:

$$T_1 = (1/n) \sum_{i=1}^n [Z(x_i) - Z^*(x_i)] \approx 0.$$

$$T_2 = (1/n) \sum_{i=1}^n [Z(x_i) - Z^*(x_i)]^2, \text{ debe ser pequeño.}$$

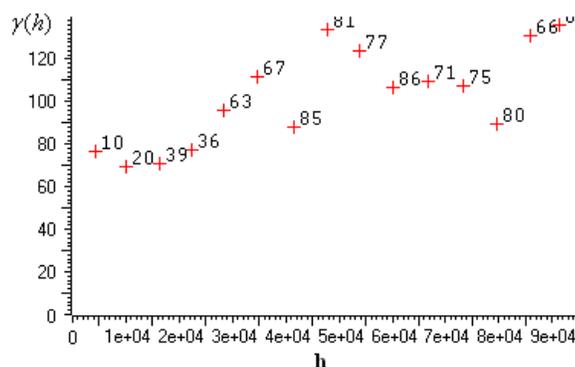
$$T_3 = (1/n) \sum_{i=1}^n \{ [Z(x_i) - Z^*(x_i)] / \sigma \}^2 = 1.$$

Cada uno de ellos representa la diferencia que existe entre los datos reales y las estimaciones, para poder seleccionar nuestro modelo.

La construcción del semivariograma experimental para cada una de las variables X1, X2, X3, se hizo con un intervalo de distancia de 10,000 m (*lag size*) y una tolerancia en la dirección total ( $45^\circ \pm 90^\circ$ ). Se comprobó que la precipitación tiene un comportamiento isotrópico para cada una de las variables, es decir, que la nube de puntos del semivariograma no varía demasiado según la dirección elegida esto se puede ver en el Anexo 2. la continuidad espacial de la precipitación es igual en dirección noreste-oeste que en dirección noreste-suroeste donde la variable X1, X2, X3 no cambia demasiado, entonces el modelo para cada una de las variables es Isotrópico, Luego, se presentaran a continuación el modelo teórico seleccionado para cada una de las variables y los demás modelos se presentan en el anexo , con la estimación de sus parámetros, estos son: la discontinuidad en el origen  $C_0$  efecto de pepita, el valor máximo de variabilidad  $C$  (meseta), y el área de influencia de la correlación  $a$  alcance con el cálculo.

### 3.3.2.1 Selección y búsqueda de Modelos de Semivariogramas.

A continuación se presenta la grafica de la nube de puntos para la variable X1



**Figura 16.** Nube de puntos para la variable X1

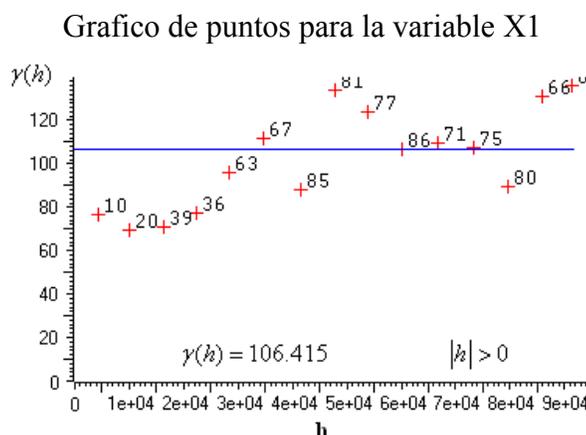
Para la selección del modelo teórico, fue necesario calcular  $T_1$ ,  $T_2$ ,  $T_3$  y por medio de ello fue posible analizar cual modelo teórico se ajustaba mejor.

Tabla. A3. Errores de medición de la Variable X1.

<b>Estadístico \ Modelo</b>	$T_1$	$T_2$	$T_3$
Efecto Pepita	<b>0.00000125</b>	<b>99.983107</b>	<b>1.018118141</b>
Modelo Esférico	-0.11940502	67.533308	0.68772776
Modelo Exponencial	-1.10738786	67.9554264	0.69202642
Modelo Gaussiano	-0.0966915	67.5534096	0.68793247
Efecto Agujero	-4.30989286	42216.9134	429.917387
Modelo Lineal	-0.13491416	65.7233662	0.66929962

La tabla anterior se generó a partir de la estimación de los parámetros de la validación cruzada para los distintos modelos de semivariograma con sus respectivos parámetros, esta validación nos sirvió para poder seleccionar el modelo que se ajuste a la nube de puntos.

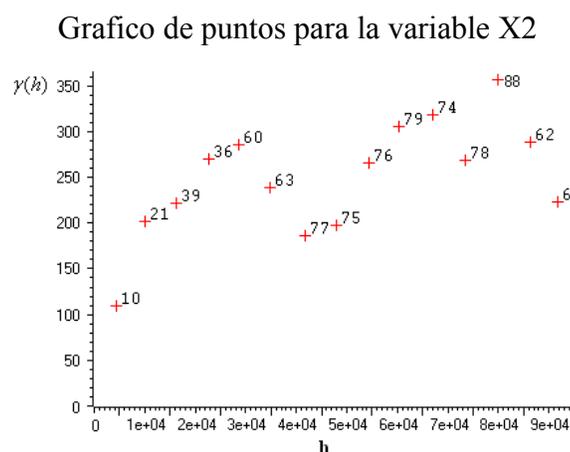
El Modelo teórico seleccionado para la variable X1 según la tabla A3 y de acuerdo a los parámetros  $T_1$  y  $T_3$  que son más pequeños que los demás parámetros de los otros modelos, ajusta al modelo Efecto Pepita dado que los estadísticos  $T_1$  y  $T_3$  cumplen con los criterios de la validación cruzada que se describieron con anticipación y se puede observar en la figura 17 con los parámetros  $C_0 = C = 106.415$ ,  $a = 0$ .



**Figura 17.** Variograma seleccionado para la variable X1

Como vemos en el grafico anterior el comportamiento de las estimaciones del semivariograma están distribuidos de manera tal que el efecto pepita tiene tanto en la parte bajo la línea como sobre la línea la mitad de la observación y explica mejor esta variabilidad.

De la misma forma se elabora el grafico del semivariograma y se estimo el modelo que mejor se ajusta a la variable X2, por medio del cálculo de  $T_1$ ,  $T_2$  y  $T_3$  para la variable X2



**Figura 18.** Nube de puntos para la variable X2

Para la selección del modelo teórico se sigue el mismo análisis del modelo anterior, la tabla A4 muestra los estadísticos de la validaron cruzada  $T_1$ ,  $T_2$  y  $T_3$  para la variable X2.

Para la selección del modelo teórico, fue necesario calcular  $T_1$ ,  $T_2$ ,  $T_3$  y por medio de ello fue posible analizar cual modelo teórico se ajustaba mejor.

Tabla. A4. Errores de medición de la Variable X2.

<b>Estadístico \ Modelo</b>	$T_1$	$T_2$	$T_3$
Efecto Pepita	<b>3.81E-16</b>	<b>277.1965349</b>	<b>1.018182666</b>
Modelo Esférico	-0.64623946	209.568689	0.76977588
Modelo Exponencial	-0.23789786	234.778998	0.86237696
Modelo Gaussiano	-0.62259393	218.355132	0.80204974
Efecto Agujero	-0.21514018	220.764548	0.81089988
Modelo Lineal	-0.54931929	207.256444	0.76128267

EL Modelo teórico seleccionado para la variable X2 según la tabla 3. Se ajusta al modelo Efecto Pepita dado que los estadísticos  $T_1$  y  $T_3$  cumplen con los criterios de la validación además este modelo representa como se distribuyen los datos tanto bajo la línea como sobre ella y se puede observar figura 19 con los parámetros  $C_0=C=261.715$ :

Grafico de puntos para la variable X2

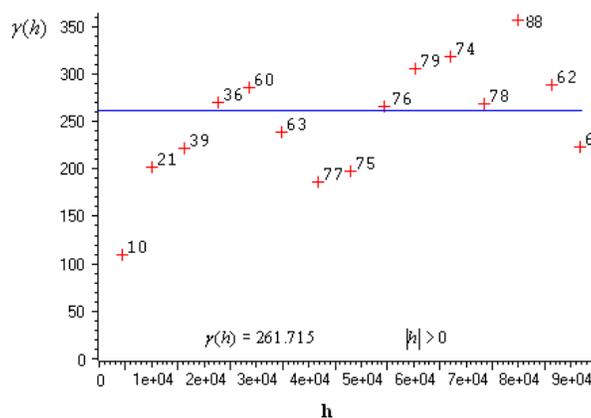


Figura 19. Variograma seleccionado de la variable X2

Como vemos en el grafico anterior el comportamiento de las estimaciones del semivariograma están distribuidos de manera tal que el efecto pepita tiene tanto en la parte bajo la línea como sobre la línea la mitad de la observación y explica mejor esta variabilidad.

De la misma forma se elabora el grafico del semivariograma y se estimo el modelo que mejor se ajusta a la variable X3.

Grafico de puntos para la variable X3

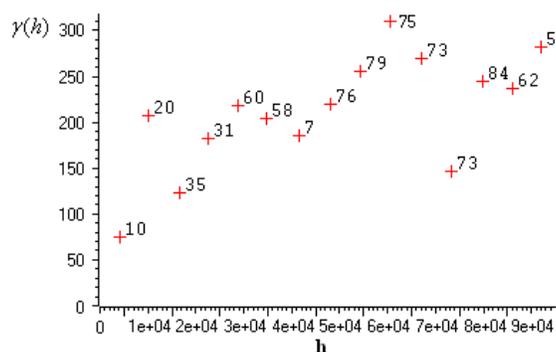


Figura 20. Nube de puntos para la variable X3

Para la selección del modelo teórico se sigue el mismo análisis que se realizó para las variables X1 y X2, la tabla A5 muestra los estadísticos de la validación cruzada  $T_1$ ,  $T_2$  y  $T_3$  para la variable X3.

Tabla. A5. Errores de medición de la Variable X3

<b>Estadístico \ Modelo</b>	$T_1$	$T_2$	$T_3$
Efecto Pepita	<b>0</b>	<b>226.799768</b>	<b>1.01886763</b>
Modelo Esférico	0.0116556	179.951958	0.80841011
Modelo Exponencial	0.02310926	180.847446	0.8124333
Modelo Gaussiano	-0.00719815	179.968507	0.80848446
Efecto Agujero	-218.968944	2197504.21	9871.9939
Modelo Lineal	-0.01623704	174.367407	0.78332227

El Modelo teórico seleccionado para la variable X3 según la tabla A5, sería el Efecto Pepita dado que los estadísticos  $T_1$  y  $T_3$  cumplen con los criterios de la validación y además que es el único modelo que explica mejor la correlación espacial de la precipitación los demás modelos están muy lejos de explicar esta correlación y se puede observar en la figura 21 con los parámetros  $C_0 = C = 223.824$ :

Gráfico de puntos para la variable X3

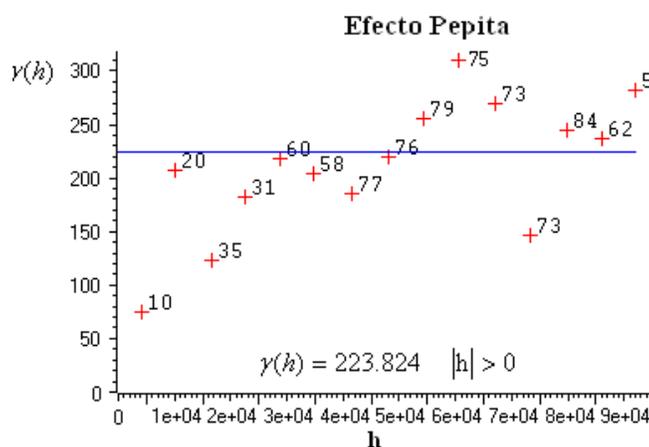


Figura 21. Variograma seleccionado para la variable X3

Como se observa en el gráfico el comportamiento de las estimaciones del semivariograma están distribuidos de manera tal que el efecto pepita tiene tanto en la parte bajo la línea como sobre la línea la mitad de la observación y explica mejor esta variabilidad.

### 3.3.4. Fase III. Estimación de los Métodos de Interpolación.

En esta etapa se generan todos los modelos de interpolación descritos en el capítulo anterior: método de la Triangulación, Inverso de la Distancia, Kriging Simple, Kriging Ordinario y Universal, con el objetivo de seleccionar el modelo que se ajusta a los datos de la precipitación para las diferentes variables, todos estos modelos que no son seleccionados son presentados en el anexo 6 en adelante, y estos modelos no se consideraran adecuados por que no explican de manera adecuada la interpolación de la precipitación.

#### 3.3.4.1 Selección del mejor Método de Estimación.

Para la selección del mejor modelo se va a utilizar el estadístico  $T_5 = |Corr [Z(x_i), Z^*(x_i)]| \simeq 1$  este, se utilizará para compararlo, es decir, para poder seleccionar el mejor método de interpolación y hay que tener en cuenta que es el único estadístico que muestra la librería gstat del paquete GRASS GIS para la selección del mejor método de estimación, donde  $Z(x_i)$  son los valores reales de las precipitaciones y  $Z^*(x_i)$  son los estimaciones según el método de interpolación utilizado.

Para la selección, se calculo cada método de interpolación para cada una de las variables X1, X2, X3 y estos se muestran en la Tabla A6.

Tabla A6. Correlación de los métodos de Estimación

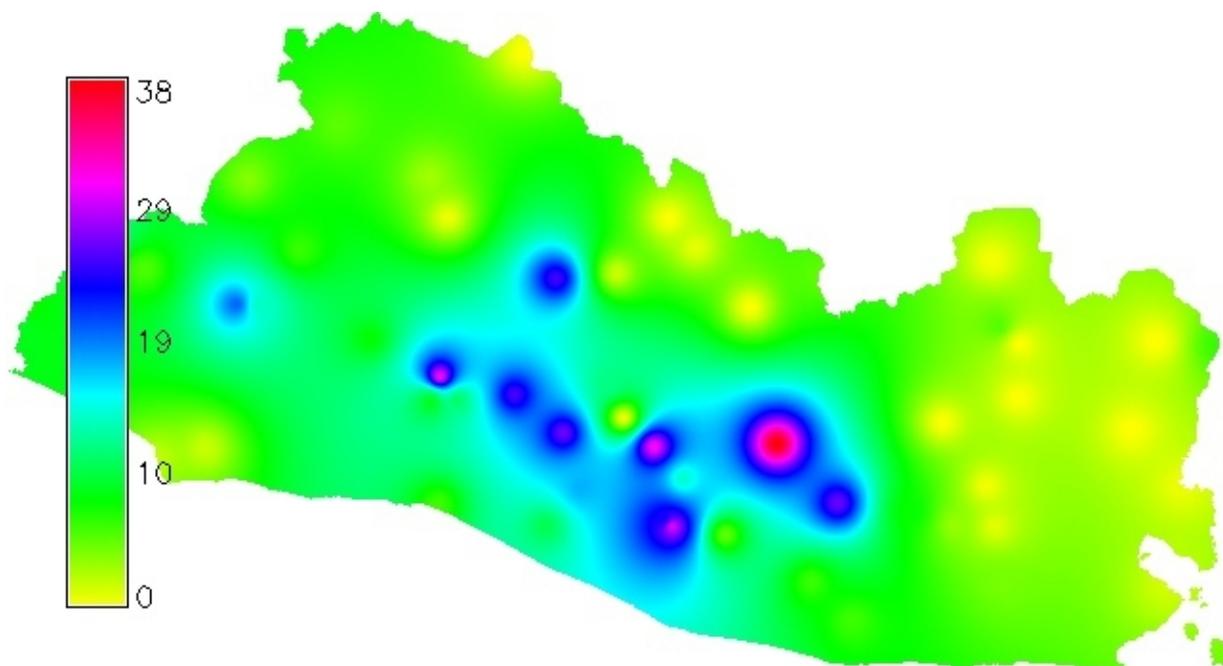
<b>Método</b>	X1	X2	X3
Triangulación	----	----	----
Inverso de la Distancia	<b>0.5463</b>	<b>0.5567</b>	<b>0.5463</b>
Kriging Simple	1.884e-016	5.251e-016	1.958e-016
Kriging Ordinario	0.4102	0.2439	0.3649
Kriging Universal	0.4675	0.1439	0.2577

Como se observa en la tabla A6, una vez aplicados los métodos, se observa que el método de la triangulación no tiene valores dado de que la versión 6.3 de GRASS GIS no lo tiene

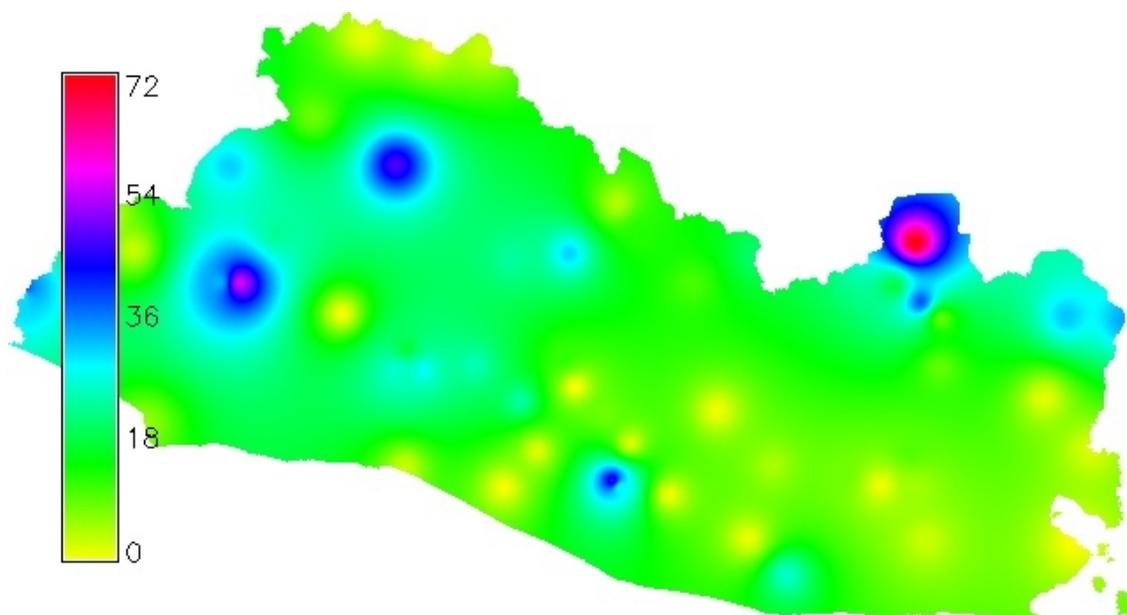
implementado y como se puede observar el mejor modelo obtenido por la aplicación de los diferentes métodos de interpolación según el criterio de la validación cruzada y el estadístico  $T_5$ , se observó que el Inverso de la Distancia para la variable X1, X2, X3, presenta la mejor correlación entre los valores reales y los valores estimados de las precipitación, los mapas de los método seleccionado para las variables anteriores se presentan a continuación.

Los tres mapas que se presentan a continuación reflejan grandes rasgos de la distribución espacial de la precipitación que en algunos lugares es semejante y otros no es semejante. Las concentraciones de precipitación en *mm* más bajas aparecen en las estaciones de color amarillo y las concentraciones medias aparecen en color celeste-azul estas representan los lugares donde la abundancia de la vegetación permite mejor abundancia de humedad y las precipitaciones atípicas (extremas) están en color rojo es en estos lugares donde las lluvia son más abundante y como se puede observar estas ultimas están en el centro de El Salvador esto se debe a que por esta zona pasa lo que es la cadena montañosa de nuestro país y por consiguiente se da más precipitaciones que en otros lugares del territorio nacional, para poder observarlo mejor podemos ver la paleta de colores de cada uno de los mapas que se encuentra al lado izquierdo de cada uno de ellos.

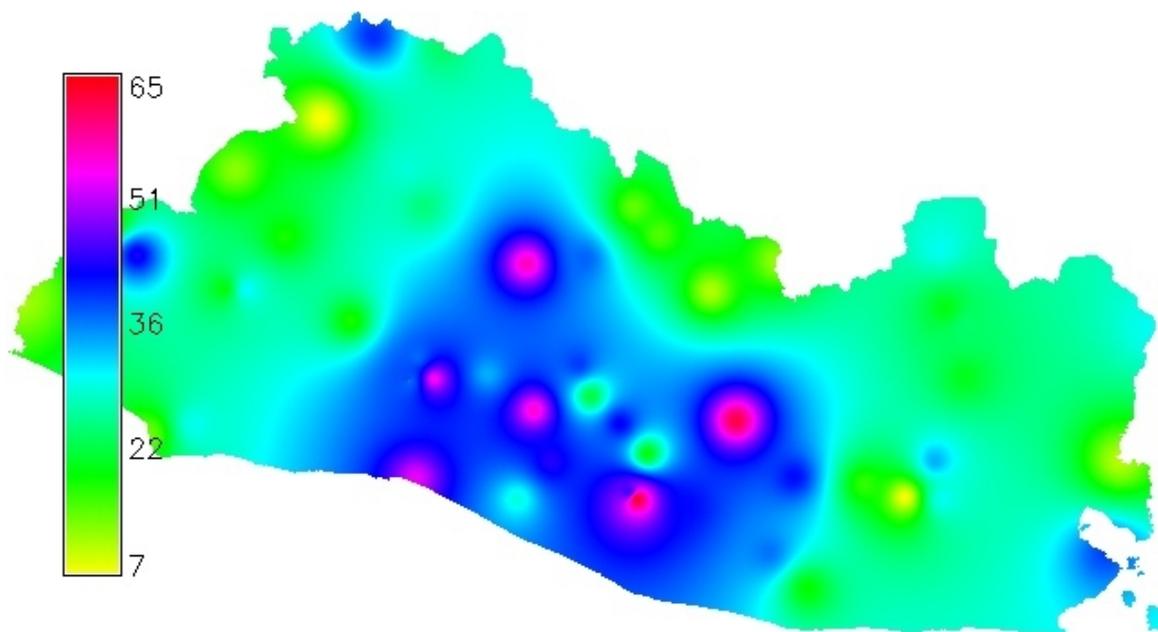
**Figura 22.** Mapa del Inverso de la Distancia la variable X1



**Figura 23.** Mapa del Inverso de la Distancia de la variable X2



**Figura 24.** Mapa del Inverso de la Distancia de la variable X3



Al realizar una comparación de estos mapas con los se presentan en el anexo 6 en adelante, estos mapas no presentan un tonalidad de color uniforme como lo que presenta el Inverso de la Distancia y además algunos están muy separados de la realidad que se refleja las estaciones ubicadas en todo el territorio nacional.

Por lo tanto estos mapas reflejan como se estima las precipitaciones del día anterior para al día siguiente.

Los datos analizados y tratados en el estudio al observar uno de los datos por ejemplo en el mapa del Inverso de la distancia de la variable X2, en la zona Oriental por el departamento de Morazán se puede notar que el punto rojo que se observa, representa la estimación de la estación meteorológica ubicada en ese punto de dicho departamento la cual refleja precipitación abundante y este se puede observar de acuerdo a la base de datos proporcionada por el SNET que se observa en la tabla A1.

Se observa en cada mapa el ajuste que representa los lugares donde más aconteció el fenómeno lluvioso, aunque hay que considerar en algunos lugares la ausencia de estaciones meteorológicas, que con su complementación se puede contar con mayor información que permita ampliar más las regiones en los que las lluvias fueron más frecuentes ó no fueron más frecuentes.

## CONCLUSIONES Y RECOMENDACIONES.

### 1. CONCLUSIONES

1. El análisis de la base de datos se realizó con el fin de estudiar y crear una aplicación de la geoestadística y diseñar una metodología para el tratamiento de fenómenos georreferenciados tales como la precipitación, temperatura, humedad etc.
2. Con el análisis exploratorio realizado a la base de datos de la tabla A1 puede observarse que los niveles de precipitación en el territorio nacional hay ausencia de datos y esto se debe a que hay errores en el envío de datos a esta base, además hay valores altos y esto se debe a que en algunos lugares como las zonas montañosas llueve más que en aquellas que no son montañosas.
3. Con el estudio de la precipitación en El Salvador durante la fase del análisis estructural se observó que al comparar los semivariogramas teórico que mejor se ajusta a los diferentes promedios de lluvia caída arroja como resultado del análisis de los criterios de selección es el modelo llamado el Efecto Pepita.
4. El estudio permitió que para muestras pequeñas el Kriging no proporciona buenas estimaciones porque todos estos métodos fueron erráticos en nuestro análisis, también puede observarse que al aumentar el tamaño de la muestra por medio de las iteraciones en búsqueda del mejor modelo nos proporciona un mejor ajuste.
5. Que el análisis de la información de las precipitaciones estudiado en los diferentes rangos permitió el análisis, descripción y selección del mejor método y se concluye que el mejor método de estimación que nos proporcionó el análisis de selección es el método del Inverso de la Distancia según los criterios de selección.

6. El paquete estadístico R y GRASS GIS fueron dos herramientas importantes en nuestros análisis.
  
7. El trabajo de investigación demuestra el carácter interdisciplinario que en la actualidad es necesario para los trabajos y que hacer científico.

## 2. RECOMENDACIONES

- Diseñar mecanismos para la Depuración de los datos de la precipitación para minimizar errores.
  
- Tratar de distribuir más estaciones meteorológicas para medir la precipitación en diferentes partes del territorio nacional y así tener mejores estimaciones en futuras investigaciones.
  
- Continuar realizando el análisis geoestadístico en nuestro país, para poder realizar diferentes análisis no solo del SNET sino en cualquier estudio donde pueda utilizarse la Geoestadística para representar variables regionalizadas

## REFERENCIAS BIBLIOGRÁFICAS

1. Paquete para el análisis de Información Geográfica, Grass Gis, [www.grass.itc.com](http://www.grass.itc.com), 2008
2. Paquete para el análisis estadístico R. <http://www.r-project.org>, 2008
3. **Martín A. Díaz Viera (2002)**, Geoestadística Aplicada, Instituto de Geofísica, UNAM México, Instituto de Geofísica y Astronomía, CITMA, Cuba, 144 Pág.
4. **Dr. C. José Quintín Cuador Gil**, “Elementos de la Geoestadística”, Departamento de Informática, Universidad de Pinar del Río, Cuba 2003, 30 Pág. [www.monografias.com/trabajos14/geoestadistica/geoestadistica.shtml](http://www.monografias.com/trabajos14/geoestadistica/geoestadistica.shtml)
5. **D. G. Rossiter**, geostatistic computing (Lecture 1), International for Geo-information Science & Earth Observation (ITC), 26–January–2007, USA, 84 Pág. <http://www.itc.nl/personal/rossiter>
6. **D. G. Rossiter**, Exploring and visualizing spatial data (Lecture 2), International for Geo-information Science & Earth Observation (ITC), 26–January–2007, USA, 107 Pág.
7. **D. G. Rossiter**, Modelling spatial structure from point samples (Lecture 3), International for Geo-information Science & Earth Observation (ITC), 26–January–2007, USA, 91 Pág.
8. **D. G. Rossiter**, Spatial prediction from point samples (Lecture 4, 5, 6), International for Geo-information Science & Earth Observation (ITC), 26–January–2007, USA, 65 Pág, 87 Pág., 36 Pág.
9. **Edzer J. Pebesma**, gstat user's manual <http://www.geog.uu.nl/gstat/manual>, Dept of Physical Geography, Utrecht University, the Netherlands 2001, 100 Pág.
10. **Markus Neteler**, Manual Grass GIS <http://mapserver.gis.umn.edu/mum/mtg2005.html>, University of Minnesota, USA 2005, 37 Pág.
11. **Ramón Giraldo Henao**, INTRODUCCION A LA GEOESTADISTICA, UNIVERSIDAD NACIONAL DE COLOMBIA, Facultad de Ciencias, Departamento de Estadística. 1997, 94 Pág.
12. **Georges Matheron (1969)**, Curso de Geoestadística fascículo II, Centro de Geoestadística de la Escuela de Minas de París, 78 Pág.

# Anexos

## Anexo 1. Script utilizado en el software R del Análisis Exploratorio.

```

#Activar el directorio de trabajo
getwd()
setwd("C:/Program Files/R/R-2.7.2rc/Gamaliel")
#limpiar el area de trabajo
ls()
rm(list=ls(all TRUE))
ls()
x1<-c(4.9,13.4,3.7,5.9,0.7,... 2); x1
x2<-c(8.4,56,31.4,1.4,23.3,12.6,10...); x2
x3<-c(7,30.5,12.3,40.2...); x3

summary(x1)
summary(x2)
summary(x3)

V1<-var(x1); V1
V2<-var(x2); V2
V3<-var(x3); V3

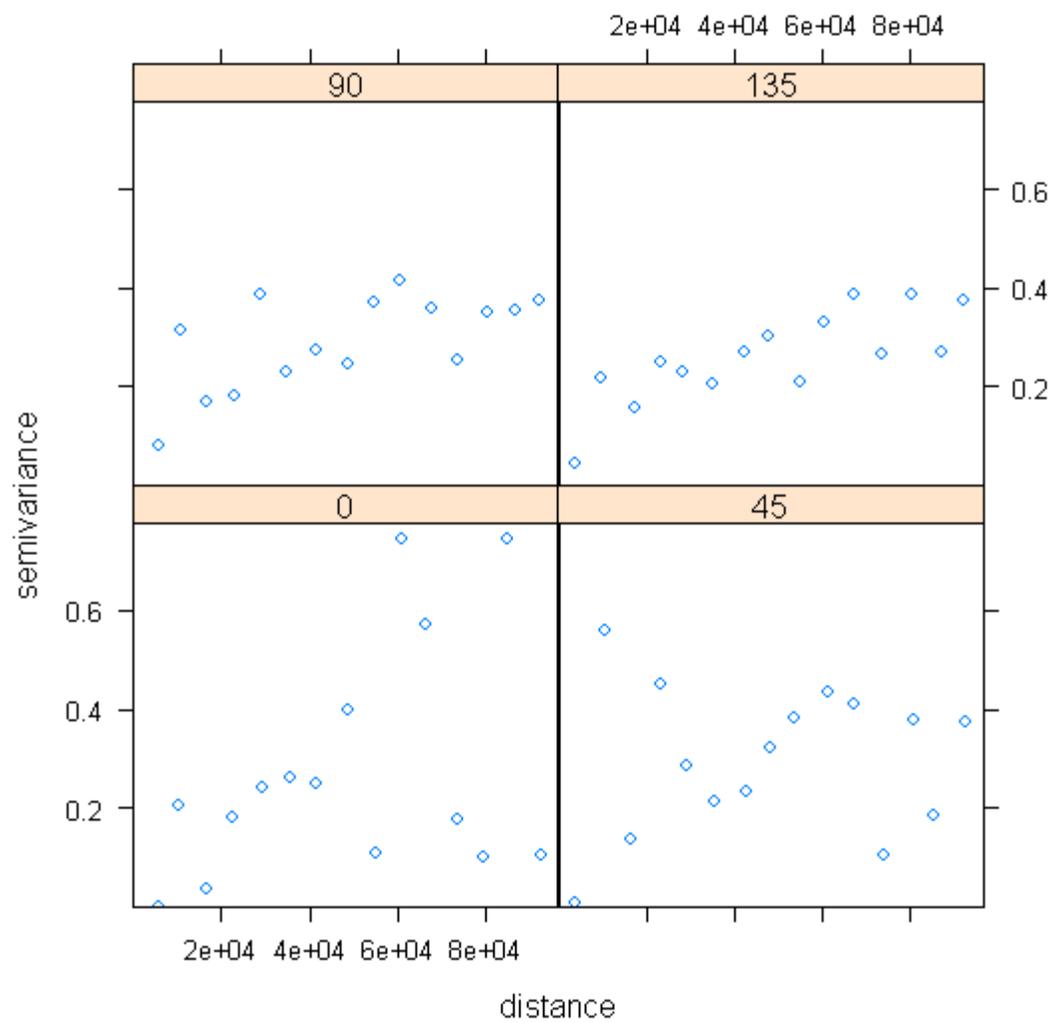
des1<-sd(x1); des1
des2<-sd(x2); des2
des3<-sd(x3); des3
#Moda
n<-length(x1); n
k=1+3.322*logb(n,10)
k<-round(k); k
rango<-max(x1)-min(x1); rango
a<-rango/k; a
frec<-table(cut(x1, breaks=k, labels=NULL, include.lower=F, right=T,
dig.lab=3)); frec
for(i in 1:length(x1)) if (frec[i]==max(frec)) break()
moda<-names(frec[i]); moda

# Grafico de los Histogramas
par(mfrow = c(2, 2))
hist(x1, main= "Histograma de Frecuencias", col="blue", xlab="Lluvia",
axes=T, labels=F, prob=TRUE)
curve(dnorm(x, 8.716,9.909477), col="red", lty=2, lwd=2, add=T)

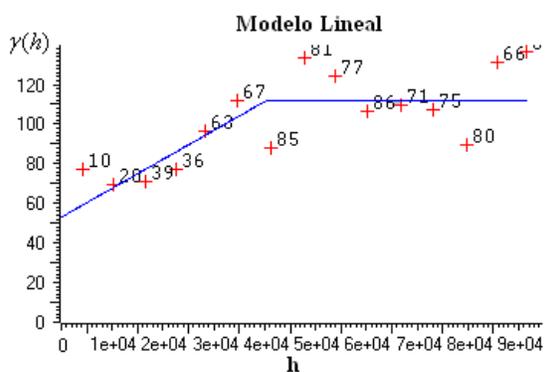
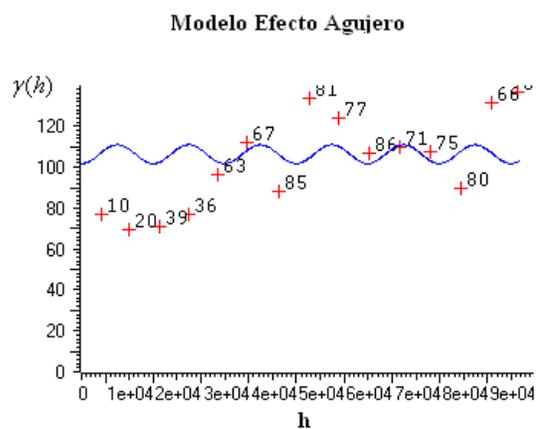
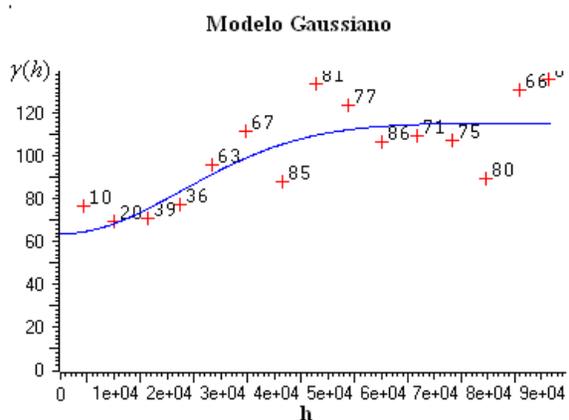
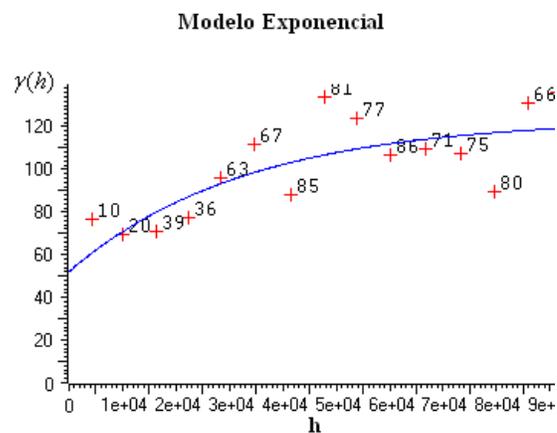
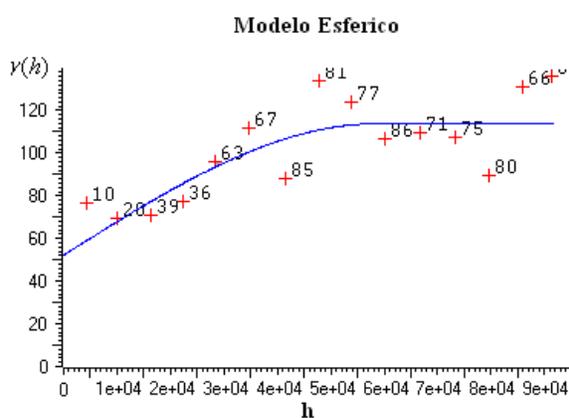
hist(x2, main= "Histograma de Frecuencias", col="blue", xlab="Lluvia",
axes=T, labels=F, prob=TRUE)
curve(dnorm(x, 16.36, 16.49989), col="red", lty=2, lwd=2, add=T)

hist(x3, main= "Histograma de Frecuencias", col="blue", xlab="Lluvia",
axes=T, labels=F, prob=TRUE)
curve(dnorm(x, 29.58, 14.91978), col="red", lty=2, lwd=2, add=T)

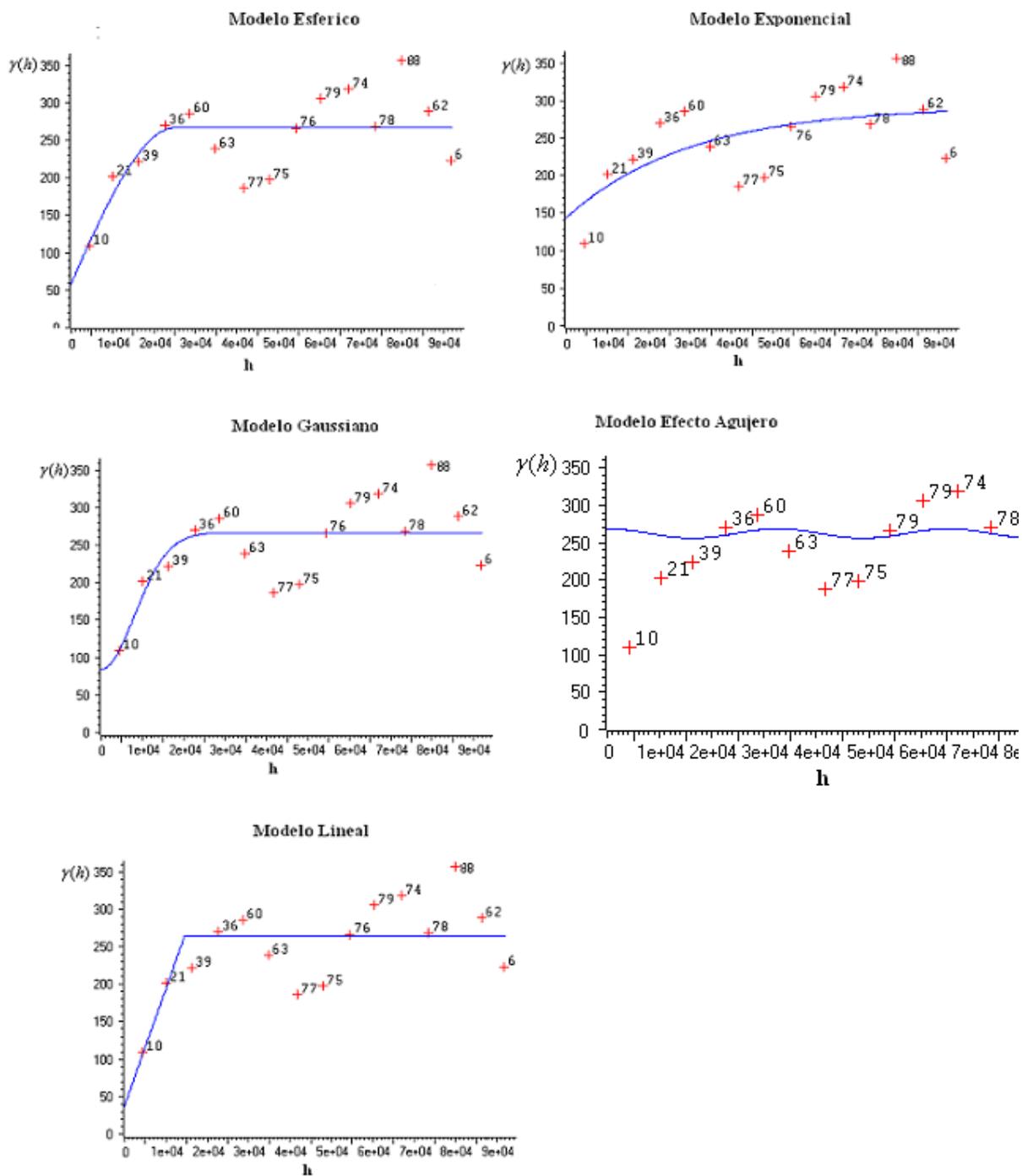
```

**Anexo 2.** Variogramas direccional para las tres variables.

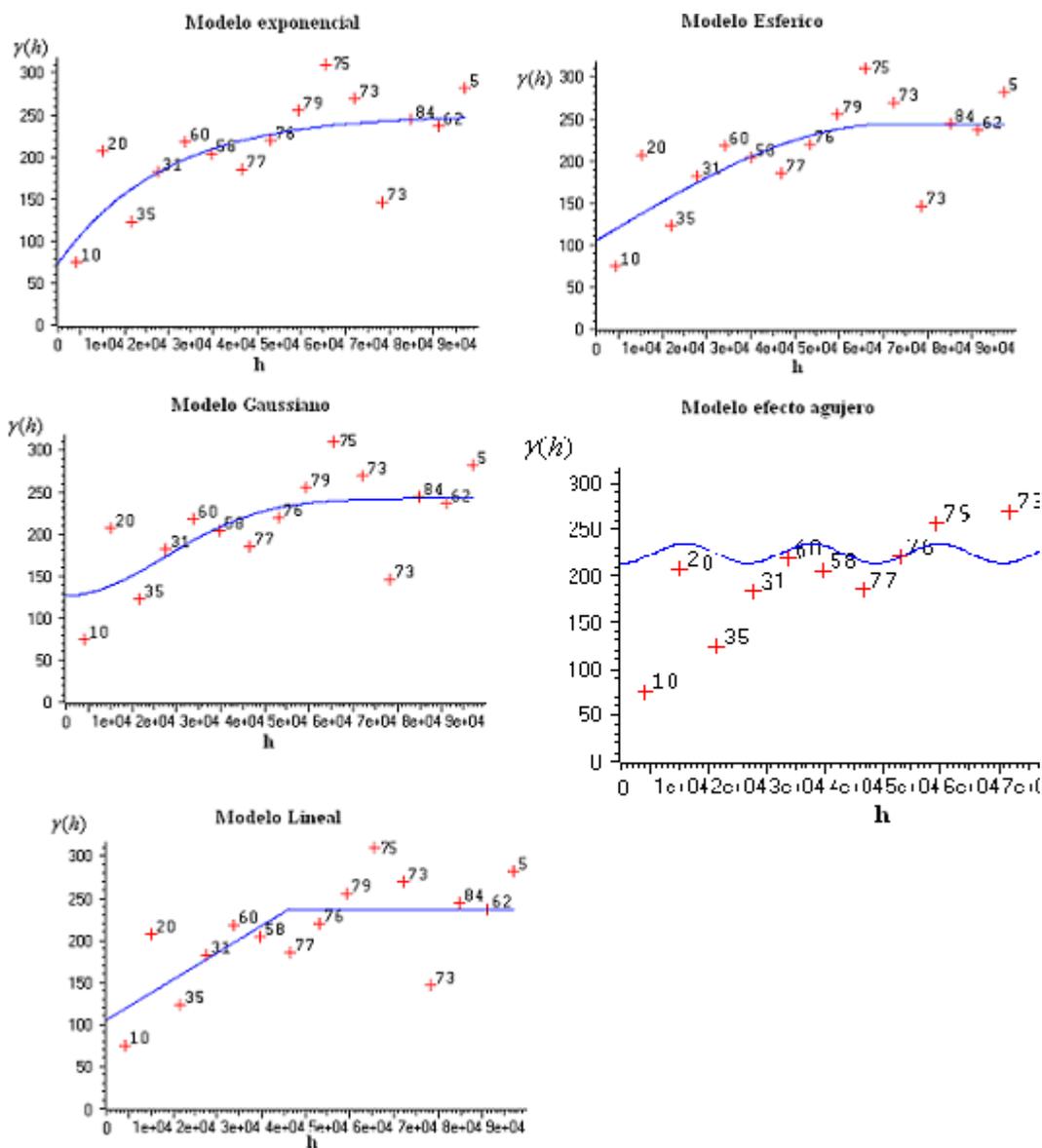
### Anexo 3. Variogramas de la variable X1.



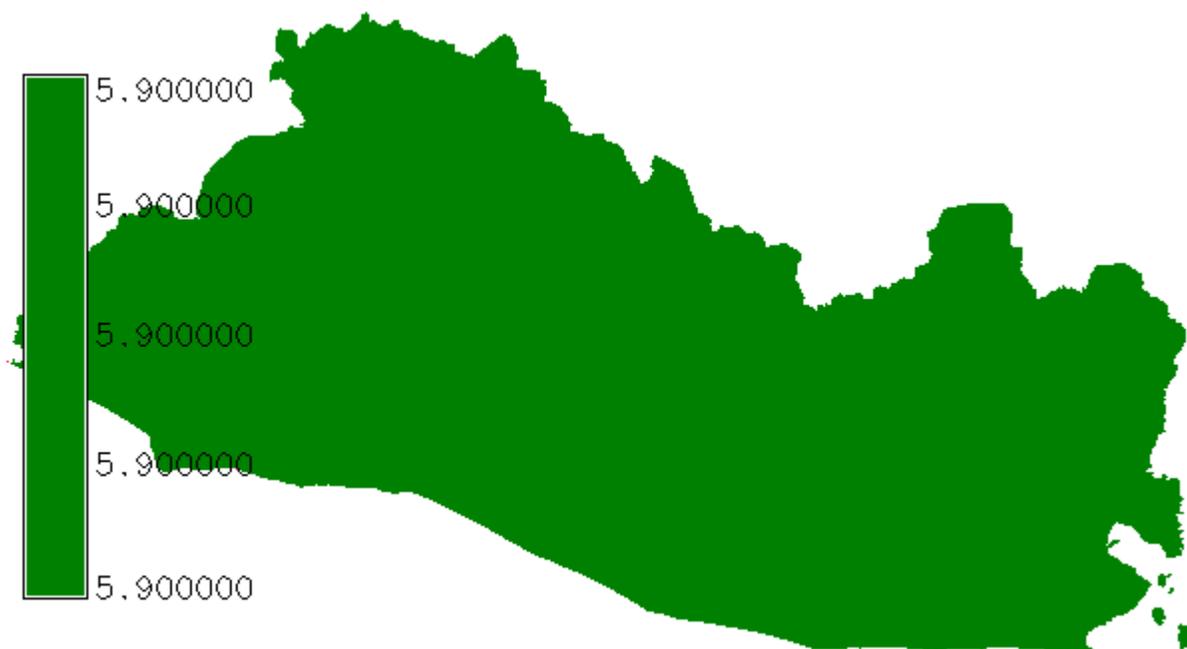
#### Anexo 4. . Variogramas de la Variable X2.



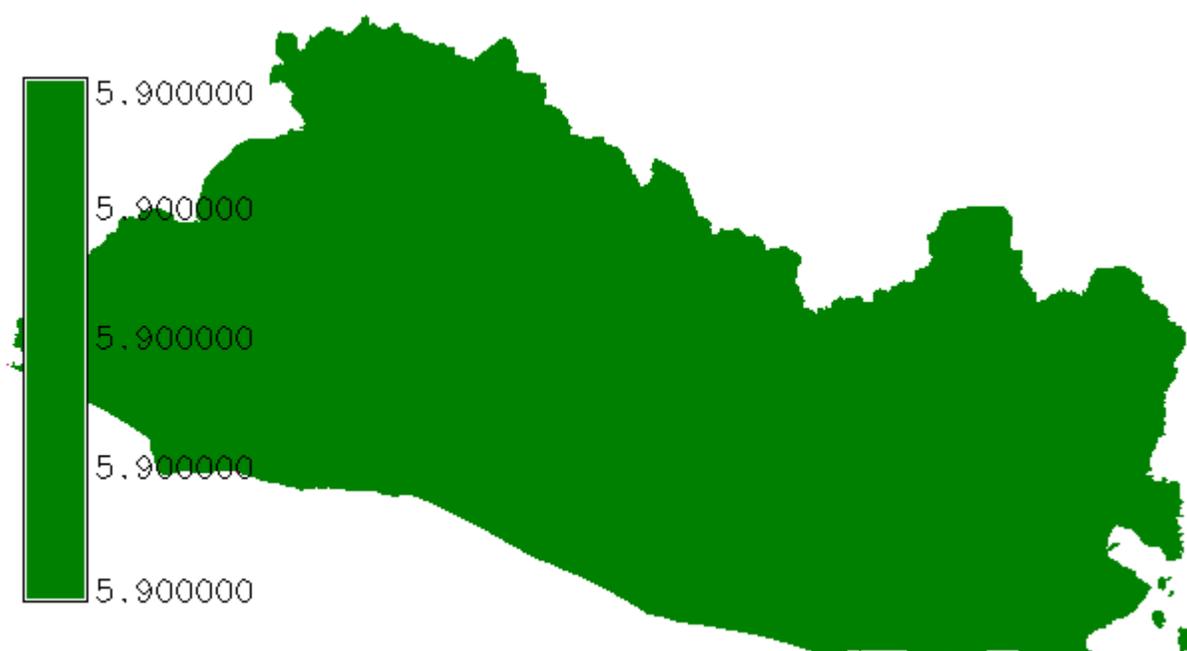
### Anexo 5. Variogramas de la Variable X3.



**Anexo 6.** Mapa de Kriging Simple para la variable X1



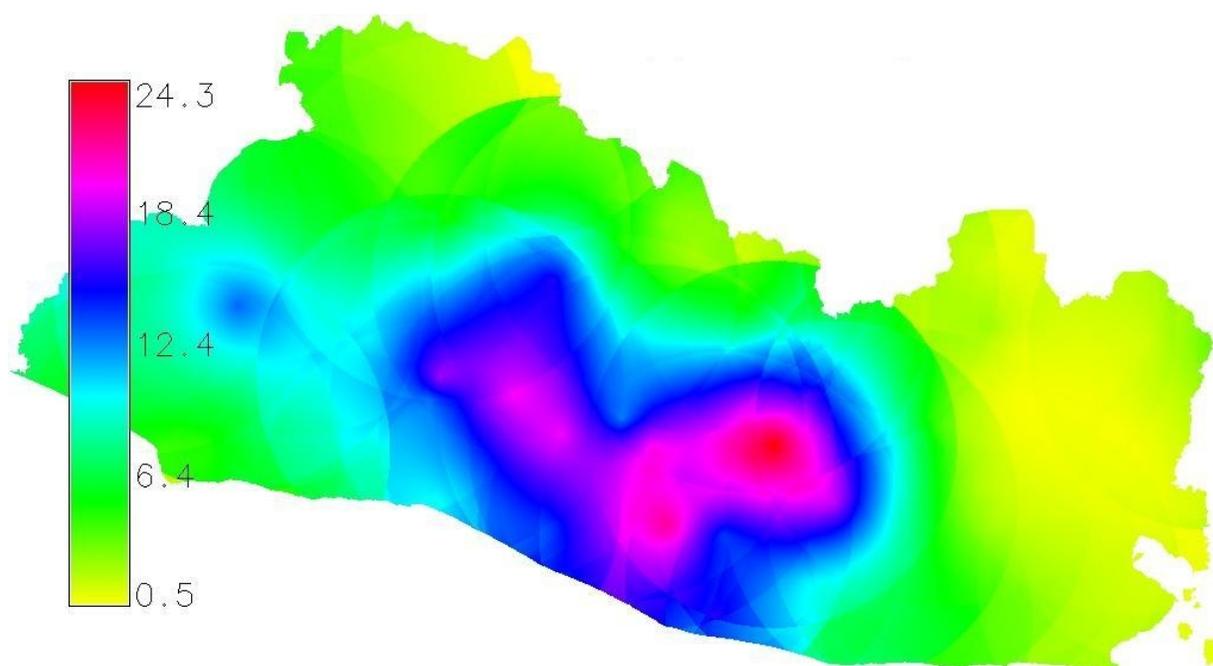
**Anexo 7.** Mapa de Kriging Simple para la variable X2



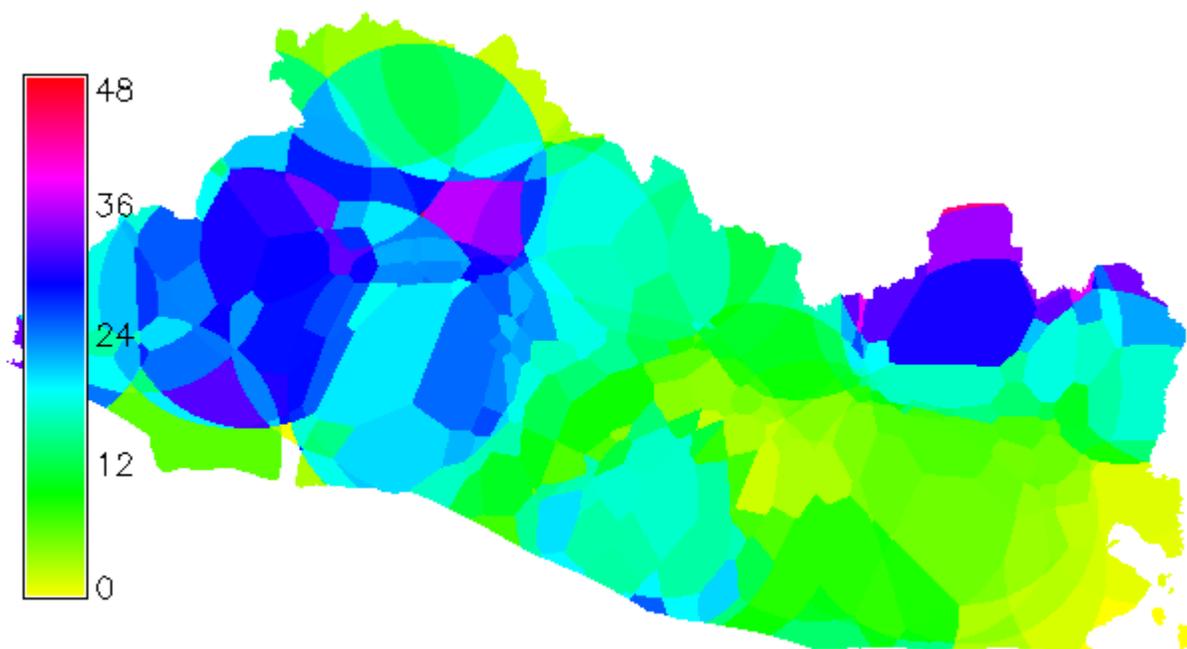
**Anexo 8.** Mapa de Kriging Simple para la variable X3



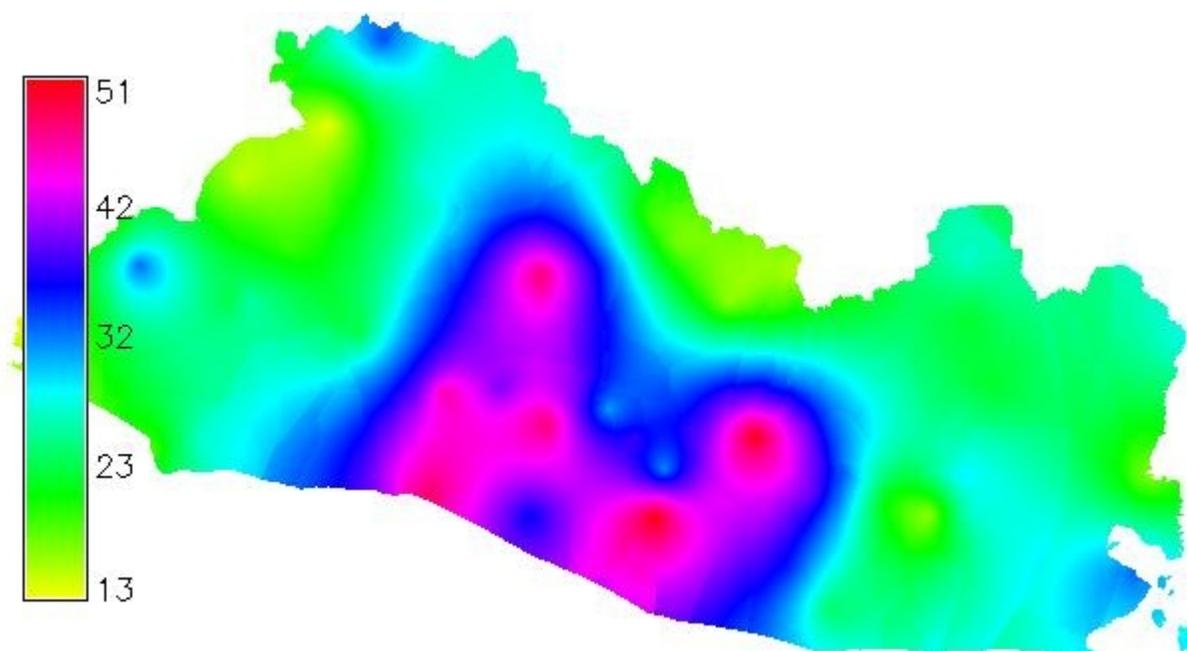
**Anexo 9.** Mapa de Kriging Ordinario para la variable X1.



**Anexo 10.** Mapa de Kriging Ordinario para la variable X2



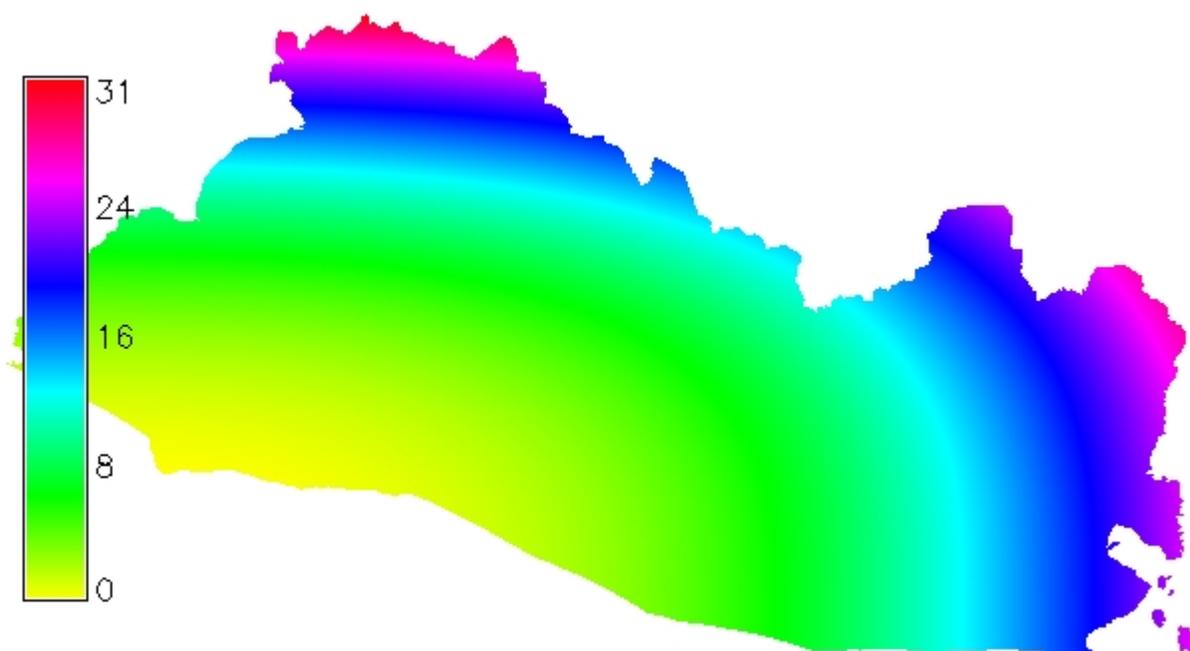
**Anexo 11.** Mapa de Kriging Ordinario para la variable X3



**Anexo 12.** Mapa de Kriging Universal para la variable X1



**Anexo 13.** Mapa de Kriging Universal para la variable X2



Anexo 14. Mapa de Kriging Universal para la variable X3

