



COLEGIO DE POSTGRADUADOS

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN
EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

POSTGRADO DE SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA
ESTADÍSTICA

MODELACIÓN DE LA DISTRIBUCIÓN ESPACIAL DE LA PRECIPITACIÓN EN EL VALLE DE LA CIUDAD DE MÉXICO USANDO TÉCNICAS GEOESTADÍSTICAS

MÉNDEZ VENEGAS JAVIER

TESIS

PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE:

MAESTRO EN CIENCIAS


MONTECILLO, TEXCOCO, ESTADO DE MÉXICO

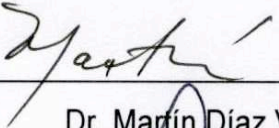
2008


La presente tesis, titulada: **Modelación de la distribución espacial de la precipitación en el valle de la Ciudad de México usando técnicas geoestadísticas**, realizada por el alumno: **Javier Méndez Venegas**, bajo la dirección del Consejo Particular indicado, ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:


**MAESTRO EN CIENCIAS
SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA
ESTADÍSTICA**

CONSEJO PARTICULAR

CONSEJERO 
Dr. Gilberto Rendón Sánchez

DIRECTOR DE TESIS 
Dr. Martín Díaz Viera

ASESOR 
Dr. Vicente González Romero

ASESOR 
Dr. Ricardo Casar González

Montecillo, Texcoco, Estado de México, Mayo de 2008

MODELACIÓN DE LA DISTRIBUCIÓN ESPACIAL DE LA PRECIPITACIÓN EN EL VALLE DE LA CIUDAD DE MÉXICO USANDO TÉCNICAS GEOESTADÍSTICAS

Javier Méndez Venegas, MC.
Colegio de Postgraduados, 2008

La mayor parte de la precipitación en la Ciudad de México es producida por tormentas convectivas, caracterizadas por una alta variabilidad espacial, lo cual implica un alto grado de complejidad en la modelación de su comportamiento. La precipitación es uno de los factores principales del ciclo hidrológico y el conocimiento de su distribución espacial es fundamental para la predicción de otras variables ambientales íntimamente relacionadas como son: el escurrimiento, las inundaciones, la recarga de los acuíferos, etc.

En la presente tesis se propone la aplicación de técnicas geoestadísticas para modelar la distribución espacial de la precipitación en el caso de tormentas convectivas. Con este fin se aplican los procedimientos geoestadísticos de estimación conocidos como Kriging, Cokriging Ordinario y Cokriging Colocado a datos de lluvia. En particular, se usa una modificación del método de estimación conjunta usando pluviómetros y radar, que fuera introducido por Krajewski en 1987. A diferencia del procedimiento usado por Krajewski, se incluye en el método de Cokriging Ordinario la dependencia espacial conjunta de radar-pluviómetros mediante un modelo de correogionalización lineal. Además, se modela la variabilidad espacial de la precipitación usando un método de simulación estocástica con enfoque geoestadístico, el método Secuencial Gaussiano. La metodología propuesta es probada usando datos de pluviómetros y de radar de una tormenta ocurrida en el valle de la Ciudad de México y los resultados son comparados.

Palabras claves: *Geoestadística, distribución espacial de la precipitación, Kriging, Cokriging, modelo de correogionalización lineal, Simulación Secuencial Gaussiana*

RAINFALL SPATIAL DISTRIBUTION MODELING IN THE MEXICO CITY VALLEY USING GEOSTATISTICAL TECHNIQUES

Javier Méndez Venegas, MC.
Colegio de Postgraduados, 2008

Most of Mexico City rainfall is produced by convective storms, characterized by high spatial variability, which implies a high degree of complexity when modeling its behavior. Precipitation is one of the main factors of the hydrologic cycle and the knowledge of its spatial distribution is fundamental for prediction of other environmental variables which are close related, such as: runoff, flooding, aquifer recharge, etc.

In the present thesis, geostatistical techniques for modeling the rainfall spatial distribution in convective storms are proposed. To this end, geostatistical estimation procedures known as: Kriging, Ordinary Cokriging and Colocated Cokriging are applied to rainfall data. In particular, a modification of the geostatistical procedure for spatial rainfall estimation introduced by Krajewski (1987), which applies the ordinary cokriging combining a meteorological radar image with gages data. Here, unlike the Krajewski's procedure, in the ordinary cokriging method the joint radar-raingages spatial dependence is considered by a linear coregionalization model. In addition, the rainfall spatial variability using a geostatistical stochastic simulation method, the Sequential Gaussian Simulation, is modeled.

The proposed methodology is tested using radar and gages data from a storm occurred in the Mexico City Valley and the results are compared.

Keywords: *Geostatistics, rainfall spatial distribution, Kriging, Cokriging, linear coregionalization model, Sequential Gaussian Simulation*

A **Dios** por estar conmigo, guiarme y protegerme.

A mis padres **Valentina Venegas Lázaro y Carmelo Méndez Vásquez**, por que los amo y les doy las gracias por todo el amor y apoyo que siempre me han brindado.

A mis **hermanos, abuelos, tíos**, por el amor, cariño y respeto que les profeso.

A aquellas personas que me han dado su cariño, amistad y apoyo.

AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT), por el otorgamiento de recursos económico para la realización de mis estudios de maestría.

Al Colegio de Postgraduados, por darme la oportunidad de superarme y lograr una las metas que me había propuesto.

A mis profesores del Colegio de Postgraduados, por transmitirme sus conocimientos y experiencia.

A los integrantes de mi consejo particular:

Dr. Martín Alberto Díaz Viera, por su excelente dirección para la realización del presente trabajo.

Dr. Gilberto Rendón Sánchez, Dr. Ricardo Casar González y Dr. Vicente González Romero, por sus valiosas observaciones, sugerencias y por todo el apoyo que me brindaron para concluir de manera satisfactoria este trabajo.

Al personal del Edificio de Estadística del Colegio de Postgraduados, por su extraordinaria amistad, aprecio y apoyo que me brindaron durante esta etapa de mi vida.

A mis amigos, compañeros y a todas las personas que de una u otra forma siempre me han brindado su apoyo incondicional.

CONTENIDO

1. INTRODUCCIÓN	1
2. CONCEPTOS BÁSICOS	5
2.1 Función Aleatoria	5
2.2 Función Aleatoria Gaussiana	5
2.3 Distribución Espacial	6
2.4 Momentos de una Función Aleatoria	6
2.5 Estacionariedad	7
2.5.1 Estacionariedad fuerte o de segundo orden	8
2.5.2 Estacionariedad débil o intrínseca	9
2.5.3 No estacionalidad.....	9
3. METODOLOGÍA PARA LA ESTIMACIÓN GEOESTADÍSTICA	11
3.1 Análisis Exploratorio de Datos	11
3.2 Análisis Estructural	14
3.2.1 Función de Semivarianzas	14
3.2.2 Estimador del Semivariograma	15
3.2.3 Características y Problemas del Estimador	15
3.2.4 Consideraciones Practicas para el Cálculo del Semivariograma Muestral	16
3.2.5 Modelos Teóricos para el Semivariograma	17
3.2.5.1 Parámetros de los Modelos	18
3.2.5.2 Clasificación de los Modelos	20
3.2.6 Combinación de modelos	20
3.2.7 Anisotropía	22
3.2.8 Métodos de ajuste	23

3.2.9	Criterios de selección	24
3.2.10	Validación del modelo.....	26
3.3	Estimación Espacial Univariada: Kriging	27
3.3.1	Kriging puntual	28
3.3.2	Kriging por bloques	28
3.3.3	Kriging en Presencia de Estacionaridad	29
3.3.3.1	Kriging ordinario	29
3.3.4	Kriging en Presencia de no Estacionaridad	31
3.3.4.1	Kriging Universal	31
3.3.4.2	Kriging Residual	32
3.4	Estimación Espacial Bivariada: Cokriging	33
3.4.1	Momentos cruzados de segundo orden	33
3.4.1.1	Covarianza cruzada	33
3.4.1.2	Semivariograma cruzado	33
3.4.2	Modelo de correogionalización lineal	35
3.4.3	Cokriging Ordinario	36
3.4.4	Cokriging Colocado	38
3.4.5	Validación del modelo de semivariograma cruzado	39
4.	METODOLOGÍA DE SIMULACIÓN	40
4.1	Idea básica de la simulación geoestadística.....	41
4.2	Estimación vs. Simulación.....	42
4.3	Métodos de Simulación	43
4.4	Métodos de Tipo Gaussiano	44
4.5	Simulación Secuencial Gaussiana	47
4.6	¿Por qué se eligió el Método Secuencial Gaussiano?	48

5. DESCRIPCIÓN DEL CASO DE ESTUDIO	50
5.1 Información de radar:	50
5.2 Información de pluviómetros:	51
6. ESTIMACIÓN UNIVARIADA USANDO LOS DATOS DE PLUVIÓMETRO.....	52
7. ESTIMACIÓN CONJUNTA USANDO DATOS DE PLUVIÓMETRO E IMAGEN DE RADAR.....	68
8. SIMULACIÓN UNIVARIADA USANDO DATOS DE PLUVIÓMETRO	80
8.1 Análisis Geoestadístico Univariado Usando la Transformación de Anamorfosis	80
8.2 Simulaciones Condicionadas al Kriging Ordinario	83
8.3 Comparación de las Simulaciones con los Valores Originales	86
9. SIMULACIÓN CONJUNTA USANDO DATOS DE PLUVIÓMETRO E IMAGEN DE RADAR.....	89
9.1 Análisis Geoestadístico Bivariado Usando la Transformación de Anamorfosis	89
9.2 Cosimulaciones Condicionadas al Cokriging Ordinario.....	93
9.3 Cosimulaciones Condicionadas al Cokriging Colocado	96
9.4 Comparación de las Cosimulaciones Ordinarias con los Valores Originales	99
9.5 Comparación de las Cosimulaciones Colocadas con los Valores Originales	102

10. DISCUSIÓN Y CONCLUSIONES	105
10.1 Discusión.....	105
10.2 Conclusiones	116
REFERENCIAS	117
ANEXOS.....	120
A1. Datos de pluviómetro	120
A2. Programa en R.....	121

LISTA DE TABLAS

Tabla 3.1: Clasificación de modelos	20
Tabla 4.1: Métodos de Simulación	44
Tabla 6.1: Estadísticas básicas para la variable pluviómetro	52
Tabla 6.2: Análisis de normalidad.....	56
Tabla 6.3: Análisis de puntos aberrantes	56
Tabla 6.4: Análisis de simetría.....	56
Tabla 6.5: Estadísticas básicas para las variables logaritmo de pluviómetro.....	57
Tabla 6.6: Análisis de normalidad de las transformaciones	60
Tabla 6.7: Análisis de puntos aberrantes para las transformaciones	61
Tabla 6.8: Análisis de simetría para las transformaciones	61
Tabla 6.9: Modelos probados para los datos transformados de pluviómetro	63
Tabla 6.10: Modelo elegido	64
Tabla 6.11: Estadísticas del Kriging Ordinario y de la desviación estándar	66
Tabla 7.1: Análisis de normalidad.....	68
Tabla 7.2: Análisis de outliers.....	69
Tabla 7.3: Análisis de simetría.....	69
Tabla 7.4: Análisis de normalidad de las transformaciones.	69
Tabla 7.5: Análisis de outliers para las transformaciones.	70
Tabla 7.6: Coeficiente de Pearson	70
Tabla 7.7: Modelos elegidos para el análisis bivariado	74
Tabla 7.8: Estadísticas básicas del Cokriging Ordinario, Colocado y las desviaciones estándar de los mismos.....	77
Tabla 8.1: Estadísticas para la transformación anamorfosis de pluviómetro.	80
Tabla 8.2: Modelo elegido	81
Tabla 8.3: Estadísticas de las simulaciones secuenciales Gaussianas	83
Tabla 8.4: Modelos	87
Tabla 9.1: Coeficiente de Pearson	89
Tabla 9.2: Modelos elegidos para el análisis bivariado	90
Tabla 9.3: Estadísticas básicas de la imagen de radar	92

Tabla 9.4: Estadísticas de la primera cosimulación condicionada al Cokriging Ordinario	93
Tabla 9.5: Estadísticas de la primera cosimulación condicionada al Cokriging colocado	96
Tabla 9.6: Modelos	100
Tabla 9.7: Modelos	103
Tabla 10.1: Estadísticas de las estimación y simulación univariada	107
Tabla 10.2: Modelos para los semivariogramas	108
Tabla 10.3: Estadísticas de los resultados bivariados usando Cokriging Ordinario	110
Tabla 10.4: Modelos para los resultados bivariados usando Cokriging Ordinario.....	111
Tabla 10.5: Estadísticas de los resultados bivariados usando Cokriging Colocado....	113
Tabla 10.6: Modelos para resultados bivariados usando Cokriging Colocado.....	114

LISTA DE FIGURAS

Figura 2.1: Grafico de grados de estacionaridad.....	10
Figura 3.1: Relación entre las funciones de Semivarianza y Covarianza.....	15
Figura 3.2: Forma del semivariograma sin nugget	19
Figura 3.3: Forma del semivariograma con nugget	19
Figura 3.4: Anisotropía Geométrica	22
Figura 3.5: Anisotropía Zonal	23
Figura 4.1: Esquema general para realizar simulaciones de tipo Gaussianas.....	46
Figura 6.1: Ubicación de puntos, histograma y gráfica de caja de la variable pluviómetro	53
Figura 6.2: Gráfica QQ de la variable pluviómetro	54
Figura 6.3: Gráfico de Estacionaridad de la variable pluviómetro	54
Figura 6.4: Gráficos de valores de pluviómetro vs. coordenadas.....	55
Figura 6.5: Ubicación de los puntos, histograma y gráfica de caja del logaritmo de pluviómetro	58
Figura 6.6: Gráfica QQ del logaritmo de pluviómetro	59
Figura 6.7: Gráfico de Estacionaridad de la variable logaritmo de pluviómetro	59
Figura 6.8: Gráficos de valores de logaritmo de pluviómetro Vs. coordenadas	60
Figura 6.9: Semivariograma adireccional de logaritmo de pluviómetro ($0^\circ, \pm 90^\circ$).....	62
Figura 6.10: Gráfico de anisotropía de logaritmo de pluviómetro.....	62
Figura 6.11: Semivariograma y ajustes de los modelos probados.....	63
Figura 6.12: Semivariograma y modelo ajustado del logaritmo de pluviómetro	64
Figura 6.13: Gráficos de la validación del modelo del logaritmo de pluviómetro.....	65
Figura 6.14: Kriging de los valores estimados para la variable pluviómetro.....	67
Figura 6.15: Desviación estándar del error de estimación para la variable pluviómetro usando Kriging	67
Figura 7.1: Lnradar Vs. Lnpluviómetro	71
Figura 7.2: Semivariogramas adireccionales.....	72
Figura 7.3: Semivariogramas en las diferentes direcciones	73
Figura 7.4: Variogramas adireccionales y modelo ajustado	74

Figura 7.5: Distribución, histograma y diagrama de caja de los errores estandarizados	75
Figura 7.6: Cokriging Ordinario para las predicciones de pluviómetro.	78
Figura 7.7: Desviaciones estándar del error de estimación de pluviómetro usando Cokriging Ordinario	78
Figura 7.8: Cokriging Colocado para las predicciones de pluviómetro.....	79
Figura 7.9: Desviaciones estándar del error de estimación de pluviómetro usando Cokriging Colocado	79
Figura 8.1: Semivariograma y modelo ajustado a la transformación anamorfosis de pluviómetro	81
Figura 8.2: Gráficos de la validación para anamorfosis de pluviómetro	82
Figura 8.3: Imagen de una simulación.....	84
Figura 8.4: Imagen de la media de diez simulaciones.....	84
Figura 8.5: Imagen de la media de cien simulaciones.....	85
Figura 8.6: Histogramas de los datos originales y simulaciones	86
Figura 8.7: Gráficos de caja de los datos originales y simulaciones	87
Figura 8.8: Semivariogramas de los datos originales y simulaciones	88
Figura 9.1: Variogramas adireccionales y modelo ajustado	90
Figura 9.2: Gráficos de la validación del modelo de semivariograma cruzado	91
Figura 9.3: Imagen de radar	92
Figura 9.4: Imagen de una cosimulación usando al Cokriging Ordinario	94
Figura 9.5: Imagen de la media de diez cosimulaciones usando al Cokriging Ordinario	94
Figura 9.6: Imagen de la media de cien cosimulaciones usando al Cokriging Ordinario	95
Figura 9.7: Imagen de una cosimulación usando al Cokriging colocado.....	97
Figura 9.8: Imagen de la media de diez cosimulaciones usando al Cokriging Colocado	97
Figura 9.9: Imagen de la media de cien cosimulaciones usando al Cokriging Colocado	98

Figura 9.10: Histogramas de los datos originales y cosimulaciones ordinarias	99
Figura 9.11: Gráficos de caja de los datos originales y cosimulaciones ordinarias.....	100
Figura 9.12: Semivariogramas de los datos originales y cosimulaciones ordinarias..	101
Figura 9.13: Histogramas de los datos originales y cosimulaciones colocadas	102
Figura 9.14: Gráficos de caja de los datos originales y cosimulaciones colocadas	103
Figura 9.15: Semivariogramas de los datos originales y cosimulaciones colocadas .	104
Figura 10.1: Histogramas de los resultados univariados.....	107
Figura 10.2: Gráficos de caja de los resultados univariados	108
Figura 10.3: Semivariogramas de los resultados univariados	109
Figura 10.4: Histogramas de los resultados bivariados usando Cokriging Ordinario ..	110
Figura 10.5: Gráficos de caja de los resultados bivariados usando Cokriging Ordinario	111
Figura 10.6: Semivariogramas de los resultados bivariados usando Cokriging Ordinario	112
Figura 10.7: Histogramas de los resultados bivariados usando Cokriging Colocado..	113
Figura 10.8: Gráficos de caja de los resultados bivariados usando Cokriging Colocado	114
Figura 10.9: Semivariogramas de los resultados bivariados usando Cokriging Colocado	115

1. Introducción

La precipitación es uno de los factores principales del ciclo hidrológico y su conocimiento es indispensable para los modelos de predicción del escurrimiento, el cálculo de líneas de flujo, la estimación de áreas diferenciadas de precipitación, así como en modelos de cuencas, entre otros. La modelación de los patrones espaciales de la precipitación es relevante tanto en ciencias de la atmósfera como en hidrología.

Debido a que la precipitación con frecuencia se comporta con una gran variabilidad en espacio y tiempo, como es el caso de la mayoría de la lluvia que cae sobre la Ciudad de México, por tanto realizar su estimación de manera que se capture el patrón de su variabilidad espacial no es una tarea fácil.

Se conoce que los pluviómetros suministran mediciones puntuales bastante precisas pero para poder satisfacer los requerimientos de precisión que permitan representar adecuadamente la distribución espacial de la lluvia se necesitaría de una red meteorológica de monitoreo muy densa. Por otro lado existe un instrumento alternativo, el radar meteorológico, el cual puede medir la precipitación de manera indirecta pero continua en espacio y tiempo, abarcando a su vez una gran área alrededor del mismo. Desafortunadamente, los datos de radar contienen errores grandes asociados a limitaciones del equipo y a la variabilidad meteorológica. Los datos de estos dos instrumentos pueden complementarse para obtener mejores estimaciones de la precipitación.

Por las razones anteriormente expuestas, resulta importante contar con metodologías que permitan modelar la variabilidad espacial de la precipitación combinando de una manera apropiada la información proveniente de los pluviómetros con las imágenes de radar.

La estimación de la precipitación se ha realizado tradicionalmente usando datos de pluviómetros. Desafortunadamente, se requiere de un gran número de pluviómetros para obtener mediciones adecuadas de la distribución espacial de la precipitación asociada a diferentes eventos meteorológicos, como son las tormentas convectivas, las cuales se caracterizan por una alta variabilidad espacial y temporal. Hildebrand et al. (1979) encontraron que al menos se necesita un pluviómetro por cada 300 km² para medir la precipitación con precisión. No obstante, los elevados costos de mantenimiento y operación de las redes de monitoreo de lluvia hacen que éstas sean prohibitivas para muchos usuarios, en particular en los países pobres y subdesarrollados.

El radar meteorológico emerge como un instrumento alternativo para estimar la precipitación debido a su capacidad de registrar de manera continua grandes áreas alrededor del mismo. Sin embargo el radar también posee sus desventajas entre las que se destacan el mal funcionamiento del equipo de radar, bloqueo orográfico, ecos falsos, propagación anómala, atenuación, evaporación de la lluvia antes de llegar al suelo, llenado parcial del haz del radar, afectación por la altura sobre el terreno y los diferentes tipos de tormentas. Una buena estrategia consiste en combinar las mediciones de ambas fuentes: pluviómetros y radar, para obtener mejores estimaciones de la distribución espacio-temporal de la precipitación.

En la presente tesis se propone como objetivo principal la aplicación de técnicas geoestadísticas para modelar la distribución espacial de la precipitación en la Ciudad de México en el caso de tormentas convectivas.

La geoestadística es una rama de la estadística aplicada que se especializa en el análisis y la modelación de la variabilidad espacial en Ciencias de la Tierra. Su objeto de estudio es el análisis y la predicción de fenómenos en espacio y/o tiempo, tales como: contenido de minerales (minería), porosidad (petróleo), concentraciones de un contaminante (ingeniería ambiental), etc.

El término de Geoestadística fue acuñado en los años 60 por Matheron quién es reconocido como el padre de esta disciplina. Matheron formalizó y generalizó matemáticamente un conjunto de técnicas desarrolladas por D. G. Krige (1941) que explotaban la correlación espacial para hacer predicciones en la evaluación de reservas de las minas de oro en Sudáfrica. Él definió a la Geoestadística como "la aplicación del formalismo de las **funciones aleatorias** al reconocimiento y estimación de fenómenos naturales"

Aquí se aplican los procedimientos geoestadísticos de estimación conocidos como Kriging, Cokriging Ordinario y Cokriging Colocado a datos de lluvia. En particular, se usa una modificación del método de estimación conjunta con pluviómetros y radar, que fuera introducido por Krajewski en 1987. A diferencia del procedimiento usado por Krajewski, se incluye en el método de Cokriging Ordinario la dependencia espacial conjunta de radar-pluviómetros mediante un modelo de correogionalización lineal. Además se modela la variabilidad espacial de la precipitación usando un método de simulación estocástica con enfoque geoestadístico, el método Secuencial Gaussiano fue introducido por Alabert y Massonat en 1990.

Para la aplicación de estas metodologías se propone el uso de datos de pluviómetros y de radar obtenidos de una tormenta ocurrida en el valle de la Ciudad de México. Esto último resulta de gran interés, puesto que la mayor parte de la precipitación en la Ciudad de México es producida por tormentas convectivas, caracterizadas por una alta variabilidad espacial y temporal.

Es bueno señalar que para poder implementar estas metodologías será necesario el uso de la programación en el software R 2.6 (<http://www.R-project.org>), esto con el fin que nosotros podamos controlar los parámetros tanto de las estimaciones, como de las simulaciones, así también como obtener gráficos mas acordes a estas técnicas.

La tesis está estructurada en 10 capítulos:

Capítulo 2: Se introducen los conceptos básicos de Geoestadística que se requieren para la comprensión de los procedimientos geoestadísticos que serán aplicados.

Capítulos 3 y 4: Se explican las metodologías de los procedimientos univariados y bivariados de estimación y de simulación, respectivamente.

Capítulo 5: Se describe el caso de estudio al cual se le aplicaran las metodologías antes mencionadas.

Capítulos 6 y 7: Se presentan los resultados de la aplicación de los procedimientos de estimación univariado y bivariado, respectivamente.

Capítulos 8 y 9: Se muestran los resultados de la aplicación de las correspondientes simulaciones estocásticas univariadas y bivariadas.

Capítulo 10: Se ofrece una discusión de los resultados así como las conclusiones del trabajo.

Anexos: Se presenta el programa implementado en R 2.6 para la obtención de los resultados.

2. Conceptos Básicos

2.1 Función Aleatoria

Si a cada punto $\underline{x} \in \Omega$ le hacemos corresponder una variable aleatoria $z(\underline{x})$, entonces al conjunto de variables aleatorias $\{z(\underline{x}), \underline{x} \in \Omega\}$ espacialmente distribuidas se le conoce como **función aleatoria** $Z(\underline{x})$.

En la literatura a una función aleatoria también se le llama proceso estocástico cuando \underline{x} se encuentra en un espacio de una dimensión y como un campo aleatorio cuando \underline{x} se localiza en espacio de más de una dimensión.

Al tomar una muestra de la función aleatoria $Z(\underline{x})$, se obtiene una función espacial discreta $Z' = \{Z(x_i), x_i \in \Omega, i = 1, 2, \dots, n\}$, la cual se conoce como **variable regionalizada**. En otras palabras una variable regionalizada Z' es una realización de una función aleatoria $Z(\underline{x})$, en algunos textos geoestadísticos se utiliza indistintamente estos dos términos, pero es bueno aclarar que en el presente trabajo se utilizará el término de **función aleatoria**.

2.2 Función Aleatoria Gaussiana

Se dice que la función aleatoria X es **Gaussiana** si su función de densidad de probabilidades esta dada por:

$$f_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - m_x)^2\right) \quad (2.1)$$

Donde m_x es la media y σ^2 es la varianza de la función aleatoria X . Esta función aleatoria puede ser generalizada al caso multivariado y en este caso la función de densidad de probabilidades esta dada por:

$$f_Y(\mathbf{y}) = \frac{1}{(2\pi)^{N/2} (\det C_Y)^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{Y} - \mathbf{m}_Y)^T C_Y^{-1} (\mathbf{Y} - \mathbf{m}_Y) \right] \quad (2.2)$$

Donde \mathbf{Y} es un vector (y_1, y_2, \dots, y_n) , mientras que \mathbf{m}_Y es el vector de medias y C_Y representa la matriz de varianzas y covarianzas del vector \mathbf{Y} .

2.3 Distribución Espacial

Una función aleatoria $Z(\underline{x})$ definida en \mathbb{R}^d es descrita por su función de distribución, es decir se deben conocer las distribuciones del vector aleatorio $\{Z(\underline{x}_1), Z(\underline{x}_2), \dots, Z(\underline{x}_k)\}$ para todo valor finito k y todos los puntos $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_k$, esto define a una **distribución espacial** de la función aleatoria $Z(\underline{x})$.

2.4 Momentos de una Función Aleatoria

Los Momentos univariados y bivariados de una función aleatoria son:

Esperanza de $Z(\underline{x}_i)$ o momento de primer orden:

$$E(Z(\underline{x}_i)) = m(\underline{x}_i) \quad (2.3)$$

Momentos de segundo orden:

- Varianza de $Z(\underline{x}_i)$.

$$\text{Var}(Z(\underline{x}_i)) = E [Z(\underline{x}_i) - m(\underline{x}_i)]^2 = \sigma_i^2 \quad (2.4)$$

- Covarianza entre $Z(\underline{x}_i)$ y $Z(\underline{x}_j)$ o función de autocovarianza.

$$C(Z(\underline{x}_i), Z(\underline{x}_j)) = E \left\{ [Z(\underline{x}_i) - m(\underline{x}_i)] [Z(\underline{x}_j) - m(\underline{x}_j)] \right\} \quad (2.5)$$

- Semivariograma $\gamma(Z(\underline{x}_i), Z(\underline{x}_j))$ o función de semivarianzas.

$$2\gamma(Z(\underline{x}_i), Z(\underline{x}_j)) = \text{Var} \left\{ [Z(\underline{x}_i) - Z(\underline{x}_j)] \right\}$$

$$\gamma(Z(\underline{x}_i), Z(\underline{x}_j)) = \frac{1}{2} E \left\{ [Z(\underline{x}_i) - Z(\underline{x}_j)]^2 \right\} \quad (2.6)$$

2.5 Estacionariedad

Se dice que una función aleatoria $Z(\underline{x})$ es **estrictamente estacionaria** si su función de distribución de probabilidad conjunta es invariante a cualquier traslación respecto a un vector \underline{h} en otra palabras la función de distribución del vector aleatorio $\{Z(\underline{x}_1), Z(\underline{x}_2), \dots, Z(\underline{x}_n)\}$ es idéntica a la del vector $\{Z(\underline{x}_1 + \underline{h}), Z(\underline{x}_2 + \underline{h}), \dots, Z(\underline{x}_n + \underline{h})\}$ para todo \underline{h} .

En la práctica resulta muy restrictivo este tipo de estacionariedad, por lo que usualmente se trabaja con la estacionariedad en los momentos de primer y segundo orden de la función aleatoria, estos tipos de estacionariedad son descritos a continuación:

2.5.1 Estacionariedad fuerte o de segundo orden

Una función aleatoria $Z(\underline{x})$ se dice que es **estacionaria de segundo orden** si cumple lo siguiente:

- La esperanza existe y no depende de \underline{x} .

$$E(Z(\underline{x})) = m; \quad \forall \underline{x} \quad (2.7)$$

- Para toda pareja $Z(\underline{x})$ y $Z(\underline{x} + \underline{h})$, la covarianza existe y solo depende de \underline{h} .

$$C(\underline{h}) = C(Z(\underline{x}), Z(\underline{x} + \underline{h})) = E[Z(\underline{x})Z(\underline{x} + \underline{h})] - m^2 \quad (2.8)$$

Como se cumple (2.8) entonces la varianza existe, es finita y no depende de \underline{h} , es decir $Var(Z(\underline{x}_i)) = C(\underline{0}) = \sigma^2$. De la misma forma por la estacionariedad de segundo orden se puede establecer la siguiente relación entre las funciones de semivarianza y autocovarianza.

$$\gamma(\underline{h}) = \gamma(Z(\underline{x}), Z(\underline{x} + \underline{h})) = \frac{1}{2} E \left\{ [Z(\underline{x}) - Z(\underline{x} + \underline{h})]^2 \right\} \quad (2.9)$$

También existe una relación directa entre el semivariograma y función de covarianza dada por:

$$\gamma(\underline{h}) = C(\underline{0}) - C(\underline{h}) \quad (2.10)$$

Aunque cabe señalar que con usar una de las dos funciones es suficiente para caracterizar la dependencia espacial.

2.5.2 Estacionariedad débil o intrínseca

Cuando la función aleatoria $Z(\underline{x})$ no es estacionaria de segundo orden pero sus diferencias si, entonces se dice que la función aleatoria presenta **estacionariedad débil**.

En decir $Z(\underline{x})$ cumple con:

- $E[Z(\underline{x} + \underline{h}) - Z(\underline{x})] = 0$ (2.11)

- $Var[Z(\underline{x} + \underline{h}) - Z(\underline{x})] = 2\gamma(\underline{h})$ (2.12)

•

A las condiciones (2.11) y (2.12) se le conocen como **Hipótesis Intrínseca**, es importante recalcar que una función aleatoria estacionaria de segundo orden es siempre intrínseca, pero lo contrario no se cumple.

2.5.3 No estacionalidad

Cuando la función aleatoria $Z(\underline{x})$ no cumple con la hipótesis intrínseca, se dice que la función es **no estacionaria**.

Considerando a la función aleatoria $Z(\underline{x})$ como una suma de componentes, es decir:

$$Z(\underline{x}) = m(\underline{x}) + R(\underline{x}) \quad (2.13)$$

Donde $m(\underline{x})$ es la tendencia y $R(\underline{x})$ es el residuo, el cual puede ser estacionario de segundo orden o intrínseco. En lo que se refiere a la tendencia $m(\underline{x})$, esta es conocida también como función determinística, esta función se define a partir de polinomios obtenidos de las realizaciones de $Z(\underline{x})$.

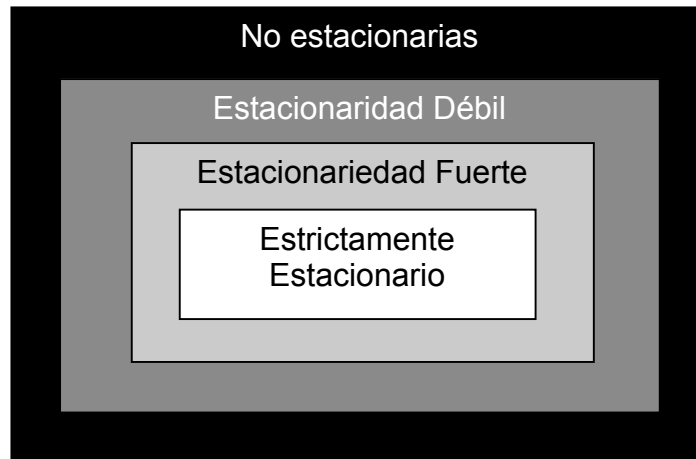


Figura 2.1: Grafico de grados de estacionariedad

En la Figura 2.1 se muestran los grados de estacionariedad, en donde las funciones aleatorias no estacionarias son únicamente las que se muestran en color negro. En lo que se refiere a las funciones aleatorias estacionarias, el grupo más pequeño son las estrictamente estacionarias (color blanco), mientras que la estacionarias fuertes son las mostradas en gris claro y también este grupo contiene a las funciones estrictamente estacionarias, por ultimo la estacionarias débil se muestran en gris oscuro e incluye también a los otros dos tipos de estacionariedad mencionados con anterioridad.

3. Metodología para la Estimación Geoestadística

En este capítulo se describirán a detalle las siguientes etapas básicas:

- a. Análisis Exploratorio de Datos
- b. Análisis Estructural
- c. Estimación Univariada: Kriging
- d. Estimación Bivariada: Cokriging

Las etapas anteriores son necesarias para la realización de la estimación geoestadística tanto univariada como bivariada.

3.1 Análisis Exploratorio de Datos

Este análisis consiste en un conjunto de técnicas estadísticas y gráficas que servirán para tener un entendimiento básico de los datos, identificar datos atípicos, ver si la distribución de la información presenta normalidad o al menos simetría, ver si existe o no tendencia en los datos, estos pasos son necesarios para tener una buena estimación, a continuación se describirán estas técnicas.

Estadísticas básicas

Son medidas de posición, dispersión y forma que nos proporcionan un resumen de la distribución y su variabilidad, algunas de estas medidas son:

- Media aritmética
- Mediana: Es el valor que después de ordenar la información de menor a mayor (estadísticas de orden), deja el mismo número de observaciones a la derecha y a la izquierda.
- Cuartiles: Una vez teniendo las estadísticas de orden, los cuartiles dividen a la distribución en cuatro intervalos, cada uno de ellos con el 25% de las observaciones.

- Varianza y Desviación estándar: Estas medidas miden la variabilidad de los datos.
- Asimetría: Mide la distribución horizontal de la información.

Histograma de frecuencias

Con este gráfico se puede intuir la distribución de probabilidad, existencia de normalidad, simetría de los datos.

Gráfico de caja

Este gráfico es una representación simple de la información que permite estudiar la simetría y detectar valores atípicos. La parte central de la caja indica la mediana, el extremo izquierdo es el primer cuartil Q_1 y el derecho corresponde al tercer cuartil Q_3 , mientras que las extensiones de la caja son $Q_1 - 1.5R_I$ y $Q_3 + 1.5R_I$ e indican el límite inferior y superior respectivamente, donde R_I es el ancho del cuartil. Los puntos que estén por encima o por debajo de los límites se consideran puntos aberrantes.

Gráfico de dispersión

Este gráfico permite ver si existe relación entre dos o más variables. Aquí se grafican las variables en pares y en función de la nube de puntos se intuye la correlación lineal entre las variables.

Normalidad

La mayoría de los métodos estadísticos basan sus supuestos en la existencia de normalidad en la variable en estudio, en este estudio es deseable que nuestra variable presente normalidad, pero si esto no se logra al menos buscaremos que nuestra variable presente simetría, la cual es fundamental para el calculo del semivariograma.

Los métodos que usaremos para ver si existe normalidad son:

- Histograma

- Gráfico Q-Q: Aquí se grafican los cuartiles teóricos vs. cuartiles muestrales, si los datos presentan normalidad estos se adecuan a lo largo de una recta a 45°.
- Pruebas de normalidad: Se realizaran las pruebas de Shapiro-Wilk, Anderson-Darling y Cramer-von Mises.
- Pruebas de simetría: Las pruebas de simetría que se realizaran son Mira test, Cabilio-Masaro test y MGG test.

Si al realizar los gráficos y pruebas se concluye que existe asimetría en los datos, para corregir la asimetría buscaremos aplicar una transformación adecuada (logarítmica, inversa, anamorfosis etc.) que permita considerar a la muestra transformada al menos simétrica.

Análisis de Outliers

Los casos atípicos son observaciones aisladas cuyo comportamiento se diferencia claramente del comportamiento del grueso de la información. Este tipo de valores se pueden clasificar en dos grupos:

- Outliers distribucionales que son los que tienen valores que se alejan significativamente del valor medio de la muestra, para detectar este tipo de valores se utilizaran pruebas como la de chi-squared, Grubbs y outlier.
- Outliers los espaciales son los que tienen un valor muy diferente al de sus vecinos más cercanos, para detectar estos se analizara la distribución espacial de los datos.

Análisis de la Tendencia

Para ver si existe o no tendencia se realizará un gráfico de la información donde se espera ver que los puntos estén distribuidos de manera aleatoria alrededor de la media y otro gráfico de la información vs. las coordenadas para observar si no hay alguna tendencia en alguna dirección preferencial. Otra forma de detectar la tendencia es por medio del semivariograma, ya que en presencia de tendencia el semivariograma tendría un crecimiento más rápido que una función cuadrática.

3.2 Análisis Estructural

El análisis estructural es una de las etapas fundamentales de todo estudio geoestadístico. Tiene como objetivo la caracterización de la estructura espacial del fenómeno estudiado, para cumplir el objetivo anterior se puede utilizar una o varias de las siguientes tres funciones: el semivariograma, la función de covarianzas y el correlograma, los cuales relacionan la información muestral y modelos teóricos de semivarianzas o semivariogramas.

En este trabajo debido a las características de estacionaridad de nuestra información, se utilizará para la determinación de la estructura espacial el semivariograma, a continuación se describirá con más detalle esta función:

3.2.1 Función de Semivarianzas

Dada una función aleatoria $Z(\underline{x})$ que cumple al menos la hipótesis intrínseca, entonces a la función $\gamma(\underline{h})$ se le conoce como función de semivarianzas o semivariograma, en algunos textos geoestadísticos a esta función también suele llamársele simplemente variograma y se define como:

$$\gamma(\underline{h}) = \frac{1}{2} \text{Var}(Z(\underline{x} + \underline{h}) - Z(\underline{x})) = \frac{1}{2} E \left[(Z(\underline{x} + \underline{h}) - Z(\underline{x}))^2 \right] \quad (3.1)$$

La función de semivarianzas es la que se encarga de caracterizar las propiedades de dependencia espacial del fenómeno.

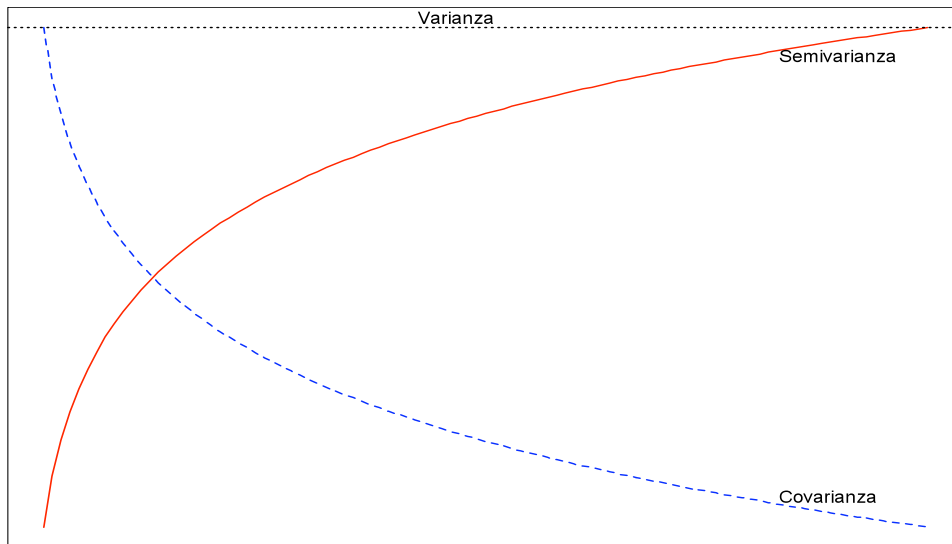


Figura 3.1: Relación entre las funciones de Semivarianza y Covarianza

3.2.2 Estimador del Semivariograma

Dada la información disponible del fenómeno, el estimador más común del semivariograma es $\gamma^*(\underline{h})$, se obtiene por el método de momentos y esta dado por:

$$\gamma^*(\underline{h}) = \frac{1}{2N(\underline{h})} \sum_{i=1}^{N(\underline{h})} [Z(\underline{x}_i + \underline{h}) - Z(\underline{x}_i)]^2 \quad (3.2)$$

Donde $N(\underline{h})$ es el numero de pares y $h = \|\underline{h}\|$ es la distancia o lag de separación entre $Z(\underline{x})$ y $Z(\underline{x} + \underline{h})$.

3.2.3 Características y Problemas del Estimador

Como se observa $\gamma^*(\underline{h})$ es esencialmente una media muestral, por tanto tiene las características asociadas a este tipo de estimador como lo son:

- Es un estimador no robusto.

- Es un estimador no paramétrico.
- Es óptimo si existe distribución normal y una malla regular de muestreo que sea representativa, bajo estas condiciones el sesgo es el mínimo.

A menudo el uso de este estimador produce semivariogramas experimentales erráticos, esto es debido a desviaciones del caso ideal para la aplicación del mismo, algunas de estas desviaciones son:

- La distribución es asimétrica.
- Heterogeneidad de varianzas
- Existencia de puntos aberrantes.

Cabe señalar que si se realiza análisis exploratorio de datos de acuerdo a lo establecido en la sección 3.1, se pueden aproximar estas desviaciones al caso ideal, de modo que ya no afecten a la estimación de $\gamma^*(\underline{h})$.

3.2.4 Consideraciones Prácticas para el Cálculo del Semivariograma Muestral

Para llevar acabo el cálculo del semivariograma se debe tomar en cuenta algunas reglas prácticas para que exista una mejor eficiencia y la calidad de la estimación, esto independientemente del tipo de estimador utilizado, estas reglas son:

- Para la estimación del semivariograma se toman en cuenta el número de parejas que se encuentran en un intervalo $h = \|\underline{h}\|$ con una tolerancia $\pm \frac{\Delta h}{2}$ y dentro de una dirección θ con una tolerancia $\pm \frac{\Delta \theta}{2}$. Al semivariograma así estimado se le conoce como suavizado o regularizado.

- El semivariograma muestral tiene mejor eficiencia y calidad si se calcula para valores de $\|h\|$ menores que la mitad de la distancia máxima $\left(\|h\| < \frac{d_{\max}}{2}\right)$.
- Para determinar con precisión el rango y la meseta del semivariograma se deben tener de 10 a 25 intervalos, esto dependiendo de la cantidad de información con la que se dispone.
- El ancho del intervalo se debe elegir de tal forma que el número de parejas sea relativamente grande, para que la estimación sea estable. Se considera que con más de 30 parejas se obtiene una buena estimación.
- El valor estimado se debe graficar contra el promedio de la distancias de la parejas que se encuentran en el dicho intervalo.

3.2.5 Modelos Teóricos para el Semivariograma

El semivariograma aunque solo es calculado para algunas distancias promedios particulares proporciona gran información del comportamiento espacial de la variable en estudio. Para la predicción del estimador espacial por medio de Kriging, el cual se describirá posteriormente, se requiere del conocimiento de la estructura de autocorrelación para cualquier posible distancia entre sitios dentro del área de estudio. Por lo cual se hace necesario el ajuste de modelos que generalicen lo observado en el semivariograma muestral.

Existen diversos modelos teóricos que se pueden ajustar al semivariograma, estos modelos pueden dividirse en no acotados (lineal, pepita puro, potencia) y acotados (esférico, exponencial, gaussiano).

Los modelos acotados garantizan que la covarianza de los incrementos sea finita, por lo cual son ampliamente usados cuando hay evidencia de que existe buen ajuste.

3.2.5.1 Parámetros de los Modelos

Todos estos modelos tienen tres parámetros comunes que son descritos a continuación:

Efecto pepita

Este efecto también es llamado varianza en el origen o nugget y usualmente es denotado por C_0 , en otras palabras el efecto pepita no es más que una discontinuidad o salto en el origen, y puede ser debido a la escala de medición o a errores en las mismas.

Meseta

La meseta o sill se puede definir como la cota superior del semivariograma, o también como el límite cuando h tiende a infinito. Este parámetro del modelo es denotado por C_1 , cuando no existe efecto pepita y cuando existe efecto pepita la meseta es denotada por $C_0 + C_1$.

Rango

El rango o alcance usualmente denotado por a , es la distancia a partir de la cual dos observaciones se pueden considerar independientes o en otras palabras autocorrelación es nula. Cabe señalar que en algunos modelos de semivariograma no existe una distancia finita para la cual dos observaciones sean independientes; por ello se llama rango efectivo a la distancia para la cual el semivariograma alcanza el 95% de la meseta.

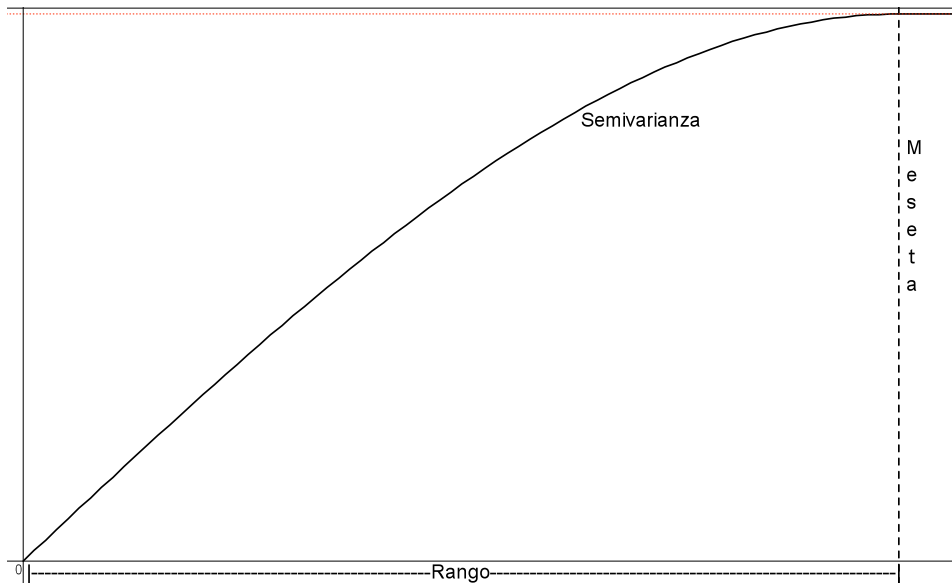


Figura 3.2: Forma del semivariograma sin nugget

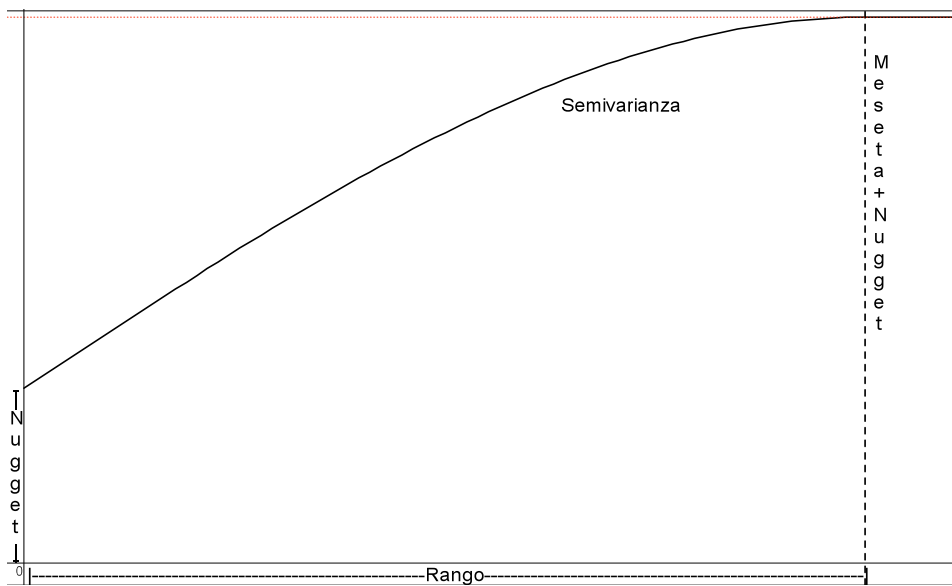


Figura 3.3: Forma del semivariograma con nugget

3.2.5.2 Clasificación de los Modelos

Tabla 3.1: Clasificación de modelos

Modelo	Tipo de Modelo	Expresión Matemática	Rango Efectivo
Esférico	Acotado	$\gamma(h) = \begin{cases} C_1 \left[\frac{3}{2} \left(\frac{h}{a} \right) - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right], & 0 < h \leq a \\ C_1, & h > a \end{cases}$	a
Exponencial	Acotado	$\gamma(h) = C_1 \left[1 - \exp\left(\frac{-3h}{a}\right) \right], \quad h > 0$	$\frac{a}{3}$
Gaussiano	Acotado	$\gamma(h) = C_1 \left[1 - \exp\left(\frac{-h^2}{a^2}\right) \right], \quad h > 0$	$\frac{a}{\sqrt{3}}$
Efecto Pepita	No Acotado	$\gamma(h) = \begin{cases} C_0, & h > 0 \\ 0, & h = 0 \end{cases}$	Sin rango efectivo
Lineal	No Acotado	$\gamma(h) = C_0 + bh, \quad h > 0$	Sin rango efectivo
Potencia	No Acotado	$\gamma(h) = \frac{1}{2} h^\theta, \quad h > 0 \text{ \& } 0 < \theta < 2$	Sin rango efectivo

Donde $\gamma(h)$ es la semivarianza en el intervalo de distancia h , C_0 es el efecto pepita, C_1 es la meseta, a es el rango y b es la pendiente de la recta.

3.2.6 Combinación de modelos

- a) Una combinación lineal de modelos de semivariogramas con coeficientes positivos representa un modelo de semivariograma válido, si consideramos la función aleatoria $Z(\underline{x})$ como una combinación lineal de n funciones aleatorias independientes $Y_i(\underline{x})$:

$$Z(\underline{x}) = \sum_{i=1}^n \lambda_i Y_i(\underline{x}) \quad (3.3)$$

De esta forma el semivariograma queda dado por:

$$\gamma(h) = \sum_{i=1}^n \lambda_i^2 \gamma_i(h) \quad (3.4)$$

Donde $\gamma_i(h)$ es el semivariograma de la función aleatoria $Y_i(\underline{x})$.

b) El producto de funciones de autocovarianza es también otra función de autocovarianza. Es decir considerando a $Z(\underline{x})$ como el producto de n funciones aleatorias independientes $Y_i(\underline{x})$:

$$Z(\underline{x}) = \prod_{i=1}^n Y_i(\underline{x}) \quad (3.5)$$

Entonces la función de autocovarianza queda dada por:

$$C(h) = \prod_{i=1}^n C_i(h) \quad (3.6)$$

Y utilizando (2.8) el semivariograma queda dado por:

$$\gamma(h) = \prod_{i=1}^n C_i(0) - \prod_{i=1}^n C_i(h) \quad (3.7)$$

Usando las dos propiedades anteriores se pueden formar modelos más complejos mediante la combinación de los modelos simples mostrados en la Tabla 3.1.

De hecho algunos de los modelos de la Tabla 3.1 pueden ser escritos como una combinación como el que se presenta a continuación:

$$\gamma(h) = \gamma_0(h) + \gamma_1(h) \quad (3.8)$$

Donde $\gamma_0(h)$ es el modelo efecto pepita y $\gamma_1(h)$ es cualquier otro modelo de semivariograma.

3.2.7 Anisotropía

Los modelos que se han presentado hasta ahora son isotrópicos, esto quiere decir que la variabilidad espacial no depende de la dirección. En muchas ocasiones al cambiar la dirección se obtienen cambios importantes en el rango o en la meseta del semivariograma, a esta variación se le conoce como **anisotropía**. La anisotropía se puede clasificar en dos tipos:

Anisotropía Geométrica

En este caso al obtener los semivariogramas en diferentes direcciones, estos tendrán aproximadamente la misma meseta, pero diferentes rangos (Figura 3.4).

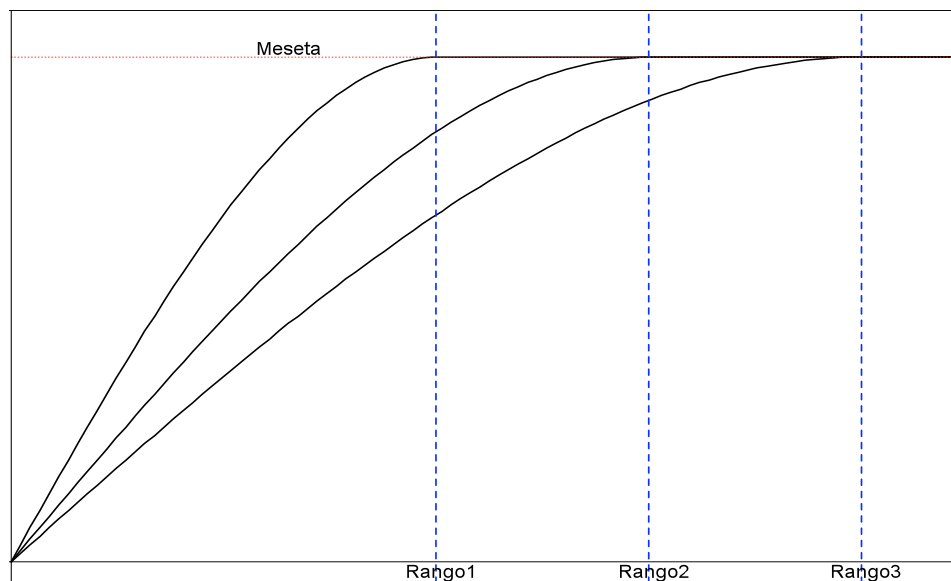


Figura 3.4: Anisotropía Geométrica

Anisotropía Zonal

En este tipo de anisotropía al calcular los semivariogramas en las diferentes direcciones, el valor de la meseta cambia y el valor del rango se mantiene constante (Figura 3.5).

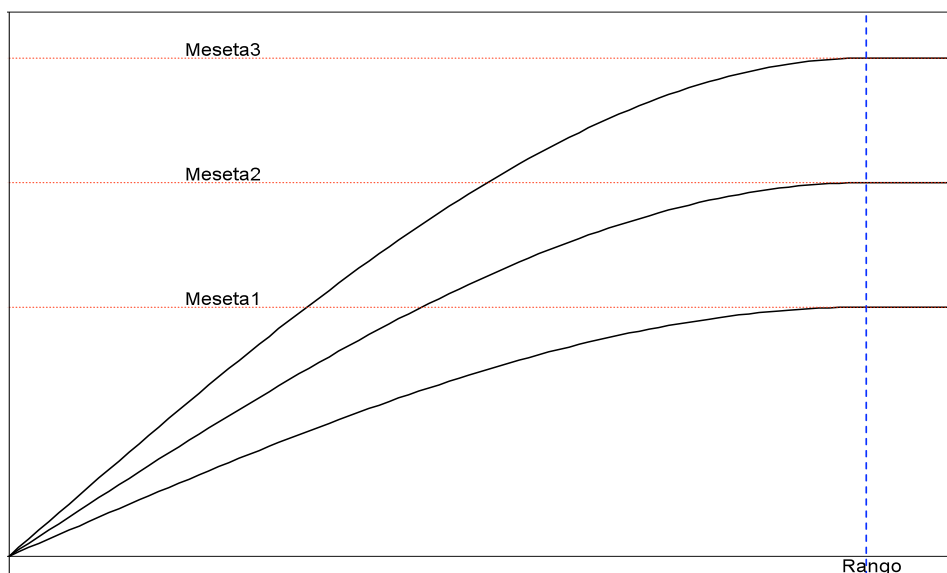


Figura 3.5: Anisotropía Zonal

En la práctica usualmente se estudian las 4 direcciones (0° , 45° , 90° y 135°), estimando los semivariogramas y determinando los rangos y mesetas de los mismos. Si los rangos y las mesetas son aproximadamente iguales no hay necesidad de construir modelos anisotrópicos, por lo que se puede trabajar con los modelos presentados en la Tabla 3.1. Cabe señalar que se pueden presentar los dos tipos de anisotropía al mismo tiempo.

3.2.8 Métodos de ajuste

Existen algunos métodos de ajuste del modelo al semivariograma, lo que hace que nuestro ajuste sea mucho más fiable, a continuación se enlistan los métodos de ajuste más usuales.

- Mínimos Cuadrados Ordinarios (MCO)

En este caso este método realiza supuestos como que los errores son independientes e idénticamente distribuidos con distribución normal, además de que

las semivarianzas estimadas poseen igual varianza, lo cual en ocasiones no se cumple por lo que es necesario utilizar otros métodos para el ajuste.

- Mínimos Cuadrados Generalizados (MCG)

Con este método se soluciona algunos de los problemas de los MCO, aunque para trabajar con este procedimiento se requiere la utilización de matrices grandes, lo que hace muchas veces este método poco práctico.

- Mínimos Cuadrados Ponderados

En este caso lo que se intenta es que la expresión siguiente se minimice.

$$W(\tau, \gamma_k) = \sum_{i=1}^k \frac{[\gamma(h_i) - \gamma(h_i, \tau)]^2}{Var[\gamma(h_i)]} \quad (3.9)$$

Donde τ es el vector de parámetros del modelo y $Var[\gamma(h_i)]$ es la varianza del valor estimado del semivariograma en h_i .

Al igual que los otros métodos este también presenta una dificultad, en este caso es que no siempre se conoce la varianza, pero en este caso lo que hace más factible la utilización de este método es que esta varianza se puede estimar mediante $n(h_i)$, que no es más que el número de parejas de observaciones que se encuentran en h_i , y se considera que este valor es inversamente proporcional a la varianza de los estimados.

3.2.9 Criterios de selección

Los criterios para la selección del modelo son muy variables, pero siempre dependen de los objetivos del trabajo y del conocimiento que se tenga del caso en estudio. Cuando el objetivo primordial es solo encontrar al mejor ajuste para el semivariograma muestral, podemos dejar que determinado software nos ajuste automáticamente el

mejor modelo. Los criterios que usualmente utiliza el software para ajustar el mejor modelo son:

- El modelo elegido tenga mínima la suma de cuadrados del error.
- El modelo que tenga la R^2 más cercana a uno.
- El modelo tenga el menor Criterio de información de Akaike (AIC).

Este último criterio relaciona la bondad de ajuste con el número de parámetros, esta relación esta dada por la siguiente expresión:

$$AIC = \left[n \ln \left(\frac{2\pi}{n} \right) + n + 2 \right] n \ln(R) + 2p \quad (3.10)$$

Se observa claramente en la expresión del AIC , que $\left[n \ln \left(\frac{2\pi}{n} \right) + n + 2 \right]$ es una constante independiente del tipo de modelo usado, por lo tanto para fines prácticos se puede obtener una expresión más simple, la cual es un estimador del AIC y esta dada por:

$$AIC^* = n \ln(R) + 2p \quad (3.11)$$

Donde n es el numero de valores estimados ($\gamma^*(h_i), i = 1, 2, \dots, n$) del semivariograma muestral, R es la suma de los errores de los cuadrados de la diferencia entre valores muestrales $\gamma^*(h_i)$ y ajustados por el modelo $\gamma(h_i)$, es decir $R = \sum_{i=1}^n [\gamma(h_i) - \gamma^*(h_i)]^2$, por ultimo p es el numero de parámetros del modelo de semivariograma ajustado $\gamma(h)$.

Cabe señalar que no siempre los modelos en los que el AIC o la suma de cuadrados del error son mínimos o la R^2 es la mas cercana a uno, son los mejores, ya que muchas veces estos tipos de modelos tienen poco sentido en la realidad o están muy alejados de la misma, muchas veces también se obtiene pequeñas mejoras en alguno o varios

de estos criterios pero los modelos obtenidos son muy complejos, por lo que no se justifica su utilización. Lo mas recomendable es que al mismo tiempo que se busca un buen ajuste, se tenga suficiente conocimiento acerca del fenómeno que se esta intentando modelar, en este trabajo se utilizará como criterio de selección la suma de cuadrados del error debido a que los modelos presentados en la Tabla 3.1 son simples, por tanto los resultados obtenidos con este método es equivalente a utilizar el *AIC*.

3.2.10 Validación del modelo

Existen varios métodos para ver si las decisiones tomadas acerca del ajuste del modelo de semivariograma son correctas, uno de los métodos mas usados, por su sencillez y eficiencia es la validación cruzada (cross validation) mediante el método leave one out, este consiste en de modo secuencial excluir una de las n observaciones muestrales y mediante las $n-1$ observaciones restantes, el modelo de semivariograma escogido y las ecuaciones del Kriging se predice el valor de la variable en estudio que se excluyo. Posteriormente se calcula la diferencia entre el valor muestral $Z(x_i)$ y el valor predicho $Z^*(x_i)$, de esta forma se obtienen los errores de predicción, ahora para ver si el modelo es correcto se analizan los errores de predicción mediante algunos estadísticos y gráficos.

Los principales estadísticos que muestran que el modelo elegido es correcto son:

- $\frac{1}{n} \sum_{i=1}^n [Z(x_i) - Z^*(x_i)] \approx 0$
- $\frac{1}{n} \sum_{i=1}^n \left[\frac{Z(x_i) - Z^*(x_i)}{\sigma_i} \right]^2 \approx 1$

- $Cor [Z(\underline{x}_i), Z^*(\underline{x}_i)] \approx 1$
- $Cor \left[Z(\underline{x}_i), \frac{Z(\underline{x}_i) - Z^*(\underline{x}_i)}{\sigma_i} \right] \approx 0$

Donde $Z(\underline{x}_i)$, $i = 1, 2, \dots, n$ representa el valor muestral de la variable en el punto \underline{x}_i , $Z^*(\underline{x}_i)$, $i = 1, 2, \dots, n$ es el valor estimado de la variable en el punto \underline{x}_i , \approx significa aproximadamente, σ_i es la desviación estándar en el punto \underline{x}_i y Cor es la correlación muestral.

Se debe recalcar que para considerar al modelo como el mejor se deben de satisfacer todos los estadísticos simultáneamente, también es muy recomendable realizar gráficos de los errores estandarizados, como por ejemplo histogramas, ya que en estos se podría ver tanto la forma de la distribución de estos errores, así como también ver si existen valores atípicos.

3.3 Estimación Espacial Univariada: Kriging

Esta palabra es anglosajona y es en honor al ingeniero en minas sudafricano D. G. Krige, quien fue probablemente el primero en hacer uso de la correlación espacial y del mejor estimador lineal insesgado, que es como se designa al Kriging, Krige aplico estos métodos de interpolación espacial en el campo de la minería, mas específicamente para la predicción de reservas de oro.

Esta técnica de interpolación que ofrece el mejor estimador lineal insesgado de la variable en estudio, también posee limitaciones, la principal limitación radica en que para su implementación se requiere de conocer del momento de segundo orden de la función aleatoria (la covarianza o el semivariograma), pero en la práctica es posible inferir a partir de una realización de la misma. Como se ha dicho la parte fundamental

para el Kriging es el semivariograma, ya que de esta función se obtendrán los parámetros a utilizar en el Kriging.

Para el uso de esta técnica también se deben tomar algunas consideraciones, como por ejemplo procurar que el radio de búsqueda no se inferior al rango, ya que en este caso no se estaría utilizando toda la información proporcionada por el semivariograma, aunque al mismo tiempo que toma en cuenta la consideración anterior se debe observar que con el radio de búsqueda que estamos proporcionando están entrando una cantidad de parejas que se pueda considerar suficientes para lograr una estimación estable. El resultado final del Kriging es una imagen o mapa con los valores interpolados del fenómeno que se estudia.

Ahora se describirán los tipos de Kriging de acuerdo al soporte de medición que tiene la información y de acuerdo al tipo de estacionaridad que se presenta:

3.3.1 Kriging puntual

Este tipo de interpolación busca estimar el valor de la variable o fenómeno en estudio en un punto del espacio.

3.3.2 Kriging por bloques

A diferencia de la interpolación anterior, este tipo de Kriging busca estimar el valor promedio dentro de un área determinada que rodea al punto en cuestión, en este caso al tomar el valor promedio los valores que se obtendrán estarán mas suavizados que el Kriging puntual.

Ahora describiremos dos tipos de Kriging, los cuales se podrían considerar de los más importantes, ya que en el primero la función aleatoria al menos cumple con la hipótesis intrínseca y en el segundo la función aleatoria presenta tendencia.

3.3.3 Kriging en Presencia de Estacionaridad

3.3.3.1 Kriging ordinario

Suponga que se tiene una función aleatoria $Z(\underline{x})$ donde se cumple que:

- La esperanza de $Z(\underline{x})$ es constante y generalmente desconocida.

$$m(\underline{x}_i) = E(Z(\underline{x}_i)) = m \quad \text{para } \forall i = 1, 2, \dots, n \quad (3.12)$$

Se requiere conocer al menos uno de los dos momentos de segundo orden de la función aleatoria $Z(\underline{x})$ mencionados a continuación.

- Semivariograma

$$\gamma(\underline{h}) = \frac{1}{2} Var [Z(\underline{x} + \underline{h}) - Z(\underline{x})] \quad (3.13)$$

- Covarianza

$$C(\underline{h}) = E [Z(\underline{x} + \underline{h})Z(\underline{x})] \quad (3.14)$$

Es importante recalcar que el Kriging lo que busca es estimar el valor $Z(\underline{x}_0)$ en el punto \underline{x}_0 donde no hubo medición de la variable en estudio, para este Kriging se utilizará una combinación lineal de las n variables aleatorias en las cuales se tienen mediciones, esto es:

$$Z^*(\underline{x}_0) = \lambda_1 Z(\underline{x}_1) + \lambda_2 Z(\underline{x}_2) + \dots + \lambda_n Z(\underline{x}_n) = \sum_{i=1}^n \lambda_i Z(\underline{x}_i) \quad (3.15)$$

Donde los λ_i representan los pesos o ponderaciones de los valores muestrales. Dichos pesos los pesos son los que minimizan la varianza del error de estimación sujeta a la restricción de que el predictor sea insesgado.

Se obtendrá el sistema de ecuaciones que se utilizarán para este tipo de Kriging:

- Condición de insesgamiento

$$\begin{aligned}
 E[Z^*(x_0)] &= E\left[\sum_{i=1}^n \lambda_i Z(x_i)\right] = E[Z(x_0)] = m \\
 \Rightarrow E\left[\sum_{i=1}^n \lambda_i Z(x_i)\right] &= m \\
 \Rightarrow \sum_{i=1}^n \lambda_i \underbrace{E[Z(x_i)]}_m &= m \\
 \Rightarrow \sum_{i=1}^n \lambda_i &= 1
 \end{aligned} \tag{3.16}$$

- Condición de que la varianza del error de predicción sea mínima

La varianza del error se expresa como:

$$\begin{aligned}
 \sigma_e^2 &= Var[Z(x_0) - Z^*(x_0)] = E\left[\left(Z(x_0) - Z^*(x_0)\right)^2\right] \\
 &= Var[Z(x_0)] + Var[Z^*(x_0)] - 2Cov[Z(x_0), Z^*(x_0)]
 \end{aligned} \tag{3.17}$$

Ahora sustituyendo es estimador de $Z^*(x_0)$ se tiene que:

$$\begin{aligned}
 \sigma_e^2 &= Var[Z(x_0)] + Var\left[\sum_{i=1}^n \lambda_i Z(x_i)\right] - 2Cov\left[Z(x_0), \sum_{i=1}^n \lambda_i Z(x_i)\right] \\
 \Rightarrow \sigma_e^2 &= \sigma_{Z(x_0)}^2 + \sum_{i=1}^n \sum_{j=i}^n \lambda_i \lambda_j \sigma_{Z(x_i)Z(x_j)}^2 - 2 \sum_{i=1}^n \lambda_i \sigma_{Z(x_i)Z(x_0)}
 \end{aligned} \tag{3.18}$$

De este modo para satisfacer la condición de insesgamiento hay que minimizar la siguiente función objetivo:

$$F = \sigma_e^2 - 2\beta \left(\sum_{i=1}^n \lambda_i - 1 \right) \tag{3.19}$$

Donde σ_e^2 es la varianza del error de estimación y β es un multiplicador de Lagrange.

Ahora para obtener el sistema de ecuaciones basta con derivar parcialmente F , con respecto a λ_i y β , de donde se obtiene lo siguiente:

$$\left\{ \begin{array}{l} \sum_{j=1}^n \lambda_j \sigma_{Z(\underline{x}_j)Z(\underline{x}_i)} - \beta = \sigma_{Z(\underline{x}_i)Z(\underline{x}_0)} \text{ para } \forall i = 1, 2, \dots, n \\ \sum_{i=1}^n \lambda_i = 1 \end{array} \right. \quad (3.20)$$

De esta forma la varianza del error de estimación se calcula de la siguiente manera, solo se necesita sustituir el valor de β .

$$\sigma_e^2 = \sigma_{Z(\underline{x}_0)}^2 - \sum_{i=1}^n \lambda_i \sigma_{Z(\underline{x}_i)Z(\underline{x}_0)} + \beta \quad (3.21)$$

3.3.4 Kriging en Presencia de no Estacionaridad

3.3.4.1 Kriging Universal

En el Kriging Ordinario se realizo el supuesto respecto a que variable regionalizada es estacionaria $Z(\underline{x})$ o que al menos cumple con la hipótesis intrínseca, pero en ocasiones $Z(\underline{x})$ no cumple con esta condición, mas bien se caracteriza por exhibir tendencia, por lo que hay que utilizar un tipo de Kriging que tome en cuenta la tendencia, como lo es el Kriging Universal.

Para la aplicación de este tipo de Kriging es necesario hacer ciertos supuestos como lo son:

- Tener conocimiento del comportamiento de la tendencia $m(\underline{x})$ de la variable $Z(\underline{x})$, expresada mediante funciones conocidas $\phi_l(\underline{x})$, usualmente polinomios.
- Conocer la función de covarianzas o semivariogramas de $Z(\underline{x})$ sin tendencia.

En este caso el predictor Kriging universal esta dado por:

$$\tilde{Z}(\underline{x}_0) = \sum_{i=1}^n \lambda_i Z(\underline{x}_i)$$

Y el sistema de ecuaciones en este caso es:

$$\left\{ \begin{array}{l} \sum_{i=1}^n \lambda_i \sigma_{Z(x_i)Z(x_j)} - \sum_{l=1}^L \beta_l \phi_l(x_i) = \sigma_{Z(x_j)Z(x_0)} \quad \text{para } \forall i = 1, 2, \dots, n \\ \sum_{i=1}^L \lambda_i \phi_l(x_i) = \phi_l(x_0) \quad \text{para } \forall l = 1, 2, \dots, L \end{array} \right.$$

Finalmente la varianza del error de estimación es:

$$\sigma_e^2 = \sigma_{Z(x_0)}^2 - \sum_{i=1}^n \lambda_i \sigma_{Z(x_i)Z(x_0)} + \sum_{l=1}^L \beta_l \phi_l(x_0)$$

Nótese que si $L=1$ y $\phi_l(x)=1$ el sistema de ecuaciones del Kriging universal y la varianza de predicción coinciden con las del Kriging ordinario.

3.3.4.2 Kriging Residual

Este tipo de Kriging parte de las mismas circunstancias del Kriging universal, ya que en este también se tiene que $Z(x)$ no es estacionaria aunque la diferencia radica que en este caso se conoce el orden de la tendencia $m(x)$ o se puede estimar mediante el método de mínimos cuadrados $m^*(x)$ y a partir de esta se obtienen los residuos y posteriormente se aplica Kriging ordinario, descrito con anterioridad en este capítulo.

Después de haber descrito las técnicas o métodos necesarios para tener una buena estimación geoestadística de una variable en estudio, continuaremos ahora por describir las técnicas para poder realizar una estimación multivariada, en este caso mas que desarrollar todas las técnicas multivariadas, nos enfocaremos en un caso mas específico, que es el caso bivariado y el cual describiremos a continuación.

3.4 Estimación Espacial Bivariada: Cokriging

Hasta este momento se ha descrito la forma de estimar y modelar la variación de un fenómeno espacial a partir de los valores muestreados de la variable de interés, pero es posible utilizar información de otras variables o mejor conocidas como covariables que estén altamente correlacionadas con nuestra variable primaria o de interés.

Esta técnica de interpolación puede ser aplicada a cualquier campo de estudio donde se tenga información de dos o más variables que están altamente correlacionadas.

3.4.1 Momentos cruzados de segundo orden

Si $Z_i(\underline{x})$ y $Z_j(\underline{x})$ son dos funciones aleatorias tomadas en la región de estudio y las cuales son estacionarias o al menos cumplen la hipótesis intrínseca, entonces se tienen las siguientes definiciones:

3.4.1.1 Covarianza cruzada

$$C_{ij}(\underline{h}) = E \left\{ [Z_i(\underline{x} + \underline{h}) - m_i] [Z_j(\underline{x}) - m_j] \right\} \quad (3.22)$$

3.4.1.2 Semivariograma cruzado

El semivariograma cruzado esta dado por la siguiente expresión:

$$\gamma_{ij}(\underline{h}) = \frac{1}{2} E \left\{ [Z_i(\underline{x} + \underline{h}) - Z_i(\underline{x})] [Z_j(\underline{x} + \underline{h}) - Z_j(\underline{x})] \right\} \quad (3.23)$$

Donde $m_i = E[Z_i(\underline{x})]$ y $m_j = E[Z_j(\underline{x})]$ son las esperanzas de $Z_i(\underline{x})$ y $Z_j(\underline{x})$ respectivamente.

Cuando $i = j$ se tiene que los momentos cruzados se convierten en la covarianza y semivariograma de $Z_i(\underline{x})$.

Para estimar el semivariograma cruzado, se utiliza la siguiente expresión, que corresponde al semivariograma experimental:

$$\gamma_{ij}^*(\underline{h}) = \frac{1}{2N(\underline{h})} \sum_{k=1}^{N(\underline{h})} [Z_i(x_k + \underline{h}) - Z_i(x_k)] [Z_j(x_k + \underline{h}) - Z_j(x_k)] \quad (3.24)$$

Cabe señalar que el semivariograma cruzado puede tomar valores negativos. Estos valores negativos se deben a que el incremento de una variable implica en promedio el decremento en la otra.

Algunas propiedades del semivariogramas son:

- $\gamma_{ij}(\underline{0}) = \underline{0}$
- $\gamma_{ij}(\underline{h}) = \gamma_{ij}(-\underline{h})$
- $\gamma_{ij}(\underline{h}) = \gamma_{ji}(\underline{h})$

La función de covarianzas y el semivariograma se relaciona a través de la siguiente ecuación:

$$\gamma_{ij}(\underline{h}) = C_{ij}(\underline{0}) - \frac{1}{2} [C_{ij}(\underline{h}) + C_{ji}(\underline{h})] \quad (3.25)$$

Esta expresión se debe a que en general la función de covarianzas no es simétrica, es decir:

$$C_{ij}(\underline{h}) \neq C_{ji}(\underline{h}) \quad (3.26)$$

En la práctica comúnmente se asume que la función de covarianza es simétrica. Lo cual simplifica los cálculos asociados a la estimación de la función de covarianza conjunta, entonces la expresión que relaciona la función de covarianzas y el semivariograma se simplifica a:

$$\gamma_{ij}(\underline{h}) = C_{ij}(\underline{0}) - C_{ij}(\underline{h}) \quad (3.27)$$

Es importante considerar que existe relación de dependencia entre el semivariograma cruzado y los semivariogramas simples, la cual esta dada por:

$$|\gamma_{ij}(\underline{h})|^2 \leq \gamma_i(\underline{h})\gamma_j(\underline{h}) \quad (3.28)$$

Conocida como desigualdad de Holder.

De igual forma se tienen una relación entre la meseta del semivariograma cruzado y los semivariogramas simples, y es la siguiente:

$$(\sigma_{ij}^2)^2 \leq \sigma_i^2 \sigma_j^2 \quad (3.29)$$

3.4.2 Modelo de correogionalización lineal

Cuando se describió la parte de estimación de una sola variable, se tenía que al utilizar los modelos de semivariograma la varianza siempre era positiva para todas las combinaciones lineales. En este caso como ya se mencionó esa condición no siempre se cumple, por lo que hay que buscar procedimientos por medio de los cuales esta condición se satisfaga.

Un método que es muy usado para lograr lo propuesto en el párrafo inmediato anterior, es el uso del modelo de correogionalización lineal, el cual establece que todos los semivariogramas individuales y cruzados pueden ser expresados como combinaciones

lineales de modelos de semivariograma básicos (exponencial, esférico, Gaussiano, etc.) idénticos.

Para el caso de dos variables, se tiene que:

$$\begin{aligned}
 \gamma_{Z_1}(\underline{h}) &= \alpha_0 \gamma_0(\underline{h}) + \alpha_1 \gamma_1(\underline{h}) + \dots + \alpha_S \gamma_S(\underline{h}) \\
 \gamma_{Z_2}(\underline{h}) &= \delta_0 \gamma_0(\underline{h}) + \delta_1 \gamma_1(\underline{h}) + \dots + \delta_S \gamma_S(\underline{h}) \\
 \gamma_{Z_1 Z_2}(\underline{h}) &= \omega_0 \gamma_0(\underline{h}) + \omega_1 \gamma_1(\underline{h}) + \dots + \omega_S \gamma_S(\underline{h})
 \end{aligned}
 \tag{3.30}$$

Donde $\gamma_{Z_1}(\underline{h})$ y $\gamma_{Z_2}(\underline{h})$ son los semivariogramas simples y $\gamma_{Z_1 Z_2}(\underline{h})$ es el semivariograma cruzado, mientras que $\gamma_0(\underline{h}), \dots, \gamma_S(\underline{h})$ son los modelos de semivariogramas y α_i, δ_i y $\omega_i, i = 0, 1, \dots, S$, son constantes.

Matricialmente se puede ver de la siguiente forma:

$$\underbrace{\begin{bmatrix} \gamma_{Z_1}(\underline{h}) & \gamma_{Z_1 Z_2}(\underline{h}) \\ \gamma_{Z_1 Z_2}(\underline{h}) & \gamma_{Z_2}(\underline{h}) \end{bmatrix}}_{\text{Matriz de Corregionalización}} = \begin{bmatrix} \alpha_0 & \omega_0 \\ \omega_0 & \delta_0 \end{bmatrix} \begin{bmatrix} \gamma_0(\underline{h}) & 0 \\ 0 & \gamma_0(\underline{h}) \end{bmatrix} + \dots + \begin{bmatrix} \alpha_S & \omega_S \\ \omega_S & \delta_S \end{bmatrix} \begin{bmatrix} \gamma_S(\underline{h}) & 0 \\ 0 & \gamma_S(\underline{h}) \end{bmatrix}
 \tag{3.31}$$

De esta forma para mostrar que el modelo este definido correctamente basta con mostrar que:

$$\alpha_i > 0, \delta_i > 0 \text{ y } \alpha_i \delta_i - \omega_i^2 > 0, \forall i = 0, 1, \dots, S
 \tag{3.32}$$

3.4.3 Cokriging Ordinario

Se tienen dos funciones aleatorias $Z_1(\underline{x})$ y $Z_2(\underline{x})$, las cuales no presentan tendencia, o al menos cumplen con la hipótesis intrínseca.

El semivariograma cruzado se calcula utilizando la ecuación (3.24) y con la información de la variable primaria $Z_1(\underline{x})$, una variable auxiliar o covariable $Z_2(\underline{x})$ y una elevada correlación espacial entre $Z_1(\underline{x})$ y $Z_2(\underline{x})$, se realiza la estimación utilizando el estimador de Cokriging Ordinario dado por:

$$Z_1^*(\underline{x}_0) = \sum_{i=1}^{n_1} \alpha_i Z_1(\underline{x}_i) + \sum_{j=1}^{n_2} \delta_j Z_2(\underline{x}_j) \quad (3.33)$$

La parte izquierda de la expresión (3.33) simboliza la predicción de la variable primaria en el punto \underline{x}_0 , el cual no fue muestreado. Mientras que $Z_1(\underline{x}_i)$, $i=1,2,\dots,n_1$ y $Z_2(\underline{x}_j)$, $j=1,2,\dots,n_2$, representan los puntos muestreados de la variable primaria y covariable respectivamente y α_i , $i=1,2,\dots,n_1$ y δ_j , $j=1,2,\dots,n_2$ son los pesos o ponderaciones de la variable primaria y la covariable. La asignación de los pesos en este método, se asignan de igual forma que en el Kriging, ya que en este caso los pesos son los que minimizan la varianza del error de estimación sujeta a la restricción de que el predictor sea insesgado.

La varianza del error de estimación se obtiene al resolver un sistema de ecuaciones, solo que en este caso el sistema se presentara en forma matricial.

$$\underbrace{\begin{bmatrix} \gamma_{Z_1}(1,1) & \cdots & \gamma_{Z_1}(n_1,1) & \gamma_{Z_1 Z_2}(1,1) & \cdots & \gamma_{Z_1 Z_2}(n_2,1) & 1 & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \gamma_{Z_1}(1,n_1) & \cdots & \gamma_{Z_1}(n_1,n_1) & \gamma_{Z_1 Z_2}(1,n_1) & \cdots & \gamma_{Z_1 Z_2}(n_2,n_1) & 1 & 0 \\ \gamma_{Z_1 Z_2}(1,1) & \cdots & \gamma_{Z_1 Z_2}(n_1,1) & \gamma_{Z_2}(1,1) & \cdots & \gamma_{Z_2}(n_2,1) & 0 & 1 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \gamma_{Z_1 Z_2}(1,n_2) & \cdots & \gamma_{Z_1 Z_2}(n_1,n_2) & \gamma_{Z_2}(1,n_2) & \cdots & \gamma_{Z_2}(n_2,n_2) & 0 & 1 \\ 1 & \cdots & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & 0 \end{bmatrix}}_K \underbrace{\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{n_1} \\ \delta_1 \\ \vdots \\ \delta_{n_2} \\ \beta_1 \\ \beta_2 \end{bmatrix}}_T = \underbrace{\begin{bmatrix} \gamma_{Z_1}(0,1) \\ \vdots \\ \gamma_{Z_1}(0,n_1) \\ \gamma_{Z_1 Z_2}(0,1) \\ \vdots \\ \gamma_{Z_1 Z_2}(0,n_2) \\ 1 \\ 0 \end{bmatrix}}_A \quad (3.34)$$

Donde K contiene los valores de la semivarianza simple y cruzada se todas las distancias entre pares de puntos, las dos ultimas filas de esta matriz contiene los valores correspondientes a la restricción de insesgamiento del predictor. El vector Γ contiene los parámetros $\alpha_i, i = 1, 2, \dots, n_1$ y $\delta_j, j = 1, 2, \dots, n_2$ que se desean estimar, además de β_1 y β_2 , que no son mas que los multiplicadores de Lagrange asignados a la restricción, finalmente el vector Δ contiene los valores de la semivarianza simple y cruzada evaluados para las distancia entre los sitios de muestreo y el sitio donde se desea hacer la predicción, en este caso también las dos ultimas filas corresponden a la condición de insesgamiento.

De esta forma la varianza del error de estimación se calcula mediante expresión (3.35):

$$\sigma_e^2 = Cov[Z_1(\underline{x}_0), Z_1(\underline{x}_0)] + \sum_{i=1}^{n_1} \alpha_i Cov[Z_1(\underline{x}_i), Z_1(\underline{x}_0)] + \sum_{j=1}^{n_2} \delta_j Cov[Z_2(\underline{x}_j), Z_2(\underline{x}_0)] + \beta_1 \quad (3.35)$$

3.4.4 Cokriging Colocado

Se podría decir que esta es una simplificación del Cokriging Ordinario, ya que en este caso se tienen dos funciones aleatorias $Z_1(\underline{x})$ y $Z_2(\underline{x})$, las cuales al menos cumplen la hipótesis intrínseca, la simplicidad de este tipo de estimación con respecto a la anterior es que aquí no se toman n_2 puntos de la covariable, en este caso solo se toma uno, entonces el estimador esta dado por:

$$Z_1^*(\underline{x}_0) = \sum_{i=1}^{n_1} \alpha_i Z_1(\underline{x}_i) + \delta Z_2(\underline{x}_0) \quad (3.36)$$

El procedimiento es análogo al de Cokriging Ordinario, tomando ahora en cuenta el estimador presentado en (3.36).

3.4.5 Validación del modelo de semivariograma cruzado

Para validar este modelo se realizará al igual que en el caso univariado usando validación cruzada, por medio del procedimiento leave one out, el cual ya se explico con anterioridad en que consiste, en este caso se pondrá énfasis en la estimación de la variable primaria $Z_1(\underline{x})$ y su varianza del error de estimación correspondiente σ_i^2 , cabe señalar que se usaran los criterios convencionales (error medio, error cuadrático medio, etc.) para ver la calidad de ajuste.

Para el Cokriging la validación cruzada no es óptima, debido a que la validación del semivariograma cruzado debe estar ligada a la validación de los semivariogramas simples, en otras palabras se debe validar primero $\gamma_1(\underline{h})$ y $\gamma_2(\underline{h})$ por separado y posteriormente validar $\gamma_{12}(\underline{h}) = \gamma_{21}(\underline{h})$ de manera

4. Metodología de Simulación

Las estimaciones espaciales de un fenómeno que se pueda modelar mediante una función aleatoria $Z(\underline{x})$ son con frecuencia insuficientes para describir toda la riqueza y complejidad del mismo debido a múltiples factores, pero sobre todo debido a la carencia de información.

Como desafortunadamente no se dispone de un conocimiento exacto de la realidad "*in situ*" y la información disponible en muchos casos está usualmente muy fragmentada y se limita fundamentalmente al conocimiento de unos pocos puntos muestrales, las estimaciones obtenidas a partir de esta información, incluso empleando el estimador Kriging, son demasiadas imprecisas para obtener una estimación representativa de las dispersiones del fenómeno estudiado, las cuales se requieren conocer en ciertas aplicaciones.

¿Cómo es imposible representar la realidad "*in situ*" correctamente, con suficiente detalle?

Una idea simple es simular esta realidad en base a un modelo, por lo que la realidad y la simulación son variables diferentes de un mismo fenómeno.

Consideremos el siguiente ejemplo. Tenemos mediciones reales del fenómeno, es decir, una realización $\{Z_M(\underline{x}_i), i = 1, \dots, n\}$ de la función $Z(\underline{x})$ en ciertos puntos \underline{x}_i de la región a estudiar. El enfoque geoestadístico consiste en interpretar la distribución espacial de la muestra $Z_M(\underline{x})$ como una realización particular de la función aleatoria $Z(\underline{x})$. Esta función aleatoria está caracterizada por su función de distribución de probabilidad espacial, o en su defecto por sus dos primeros momentos, los cuales son estimados a partir de los datos experimentales.

Este modelo es entonces adecuado para el problema práctico de la determinación de medidas de la dispersión de los valores observados $Z_M(\underline{x}_i)$, ya que las varianzas de la dispersión de $Z(\underline{x})$ pueden ser expresadas como una función del momento de segundo orden solamente (covarianza o variograma). Una simulación entonces consiste en obtener otra realización $Z_S(\underline{x})$ de esta función aleatoria $Z(\underline{x})$. Las dos realizaciones la real y la simulada difiere una de la otra en determinadas localizaciones pero ambas pertenecen a la misma función aleatoria $Z(\underline{x})$, es decir tienen la misma función de distribución y los mismos momentos de primer y segundo órdenes por lo se dice que son **estadísticamente equivalentes**.

El fenómeno simulado tiene la ventaja de ser conocido en todos los puntos y no solamente en los puntos experimentales $\{Z_M(\underline{x}_i), i = 1, \dots, n\}$. Con frecuencia al fenómeno simulado se le llama **modelo numérico** del fenómeno real.

4.1 Idea básica de la simulación geoestadística

La idea básica detrás de la simulación estocástica con un enfoque geoestadístico consiste en obtener nuevas realizaciones “artificiales” $Z_S(\underline{x})$ de la función aleatoria $Z(\underline{x})$, de manera tal que éstas reflejen las mismas propiedades estadísticas que se esperan que posea la función aleatoria $Z(\underline{x})$.

Por lo general no conocemos con precisión las propiedades estadísticas de $Z(\underline{x})$ y cuando más lo que podemos hacer es inferirlas a través de una sola realización o muestra de la función aleatoria, entonces lo que se hace es intentar obtener realizaciones simuladas $Z_S(\underline{x})$ que sean estadísticamente equivalentes a la muestra que se posee de la función aleatoria.

La equivalencia estadística en un sentido estricto significa que todas las realizaciones $Z_s(\underline{x})$ sean isomorfas a $Z(\underline{x})$, es decir que tengan la misma distribución espacial de la función aleatoria $Z(\underline{x})$ que se simula, pero en la mayoría de los casos nos tenemos que conformar con que al menos posean los mismos momentos de primer y segundo orden que inferimos a partir de una muestra de $Z(\underline{x})$.

Resulta deseable en muchas aplicaciones quedarse solamente con aquellas realizaciones $Z_s(\underline{x})$ que en los puntos muestrales $\{\underline{x}_i, i = 1, \dots, n\}$ coinciden los valores simulados $Z_s(\underline{x}_i)$ con los valores reales o experimentales $Z_M(\underline{x}_i)$. A estas realizaciones $Z_s(\underline{x})$ de la función aleatoria $Z(\underline{x})$ se les conoce como **simulaciones condicionales** $Z_{SC}(\underline{x})$ del fenómeno $Z(\underline{x})$. Formalmente se puede representar como:

$$Z_{SC}(\underline{x}_i) = Z_M(\underline{x}_i) \text{ para } i = 1, 2, \dots, n \quad (4.1)$$

4.2 Estimación vs. Simulación

La estimación por Kriging permite estimar a la función aleatoria por medio de una combinación lineal ponderada de los datos. Algunas propiedades de la estimación por Kriging son:

- El error de estimación es insesgado.

$$E[Z^*(\underline{x}) - Z(\underline{x})] = 0 \quad (4.2)$$

- El error de estimación tiene varianza mínima.

$$\min_{\underline{x}} E \left\{ [Z^*(\underline{x}) - Z(\underline{x})]^2 \right\} = 0 \quad (4.3)$$

- La estimación es suavizada lo cual significa que la variabilidad de los valores estimados es menor que la variabilidad de los valores reales.

La simulación busca que las realizaciones artificiales tengan las mismas propiedades estadísticas y de variabilidad espacial que la función aleatoria en estudio. Dicho explícitamente, la simulación $Z_s(\underline{x})$ debe poseer los mismos momentos experimentales (media, varianza, covarianza o variograma, así como el histograma) que caracterizan a $Z(\underline{x})$. Algunas de las propiedades de las simulaciones son:

- La interpolación es exacta, es decir reproduce el mismo valor medido en cada sitio con dato, esto siempre y cuando se use simulación condicional.
- Los valores no están suavizados, ya que los valores simulados tienen la misma variabilidad que los valores reales.
- El error entre el valor real y el simulado no tiene varianza mínima. Por el contrario, se puede mostrar que la varianza del error de la estimación de $Z(\underline{x})$ mediante simulación condicional $Z_{sc}(\underline{x})$ es exactamente el doble de la varianza del error usando Kriging.

La principal diferencia entre la simulación y la estimación radica en que mientras la primera intenta reproducir las propiedades estadísticas y espaciales de la función aleatoria $Z(\underline{x})$, la segunda busca que el valor estimado $Z^*(\underline{x})$ en el punto \underline{x} sea lo más cercano posible al valor real desconocido $Z(\underline{x})$.

4.3 Métodos de Simulación

En los años recientes la simulación geoestadística de funciones aleatorias ha adquirido más importancia, esto sobre todo en campos como la industria petrolera, la geohidrología y la industria ambiental entre otras. A continuación se presenta una clasificación de los métodos más usuales de simulación geoestadística y sus particularidades.

Tabla 4.1: Métodos de Simulación

Método de Simulación	Condicionado	Gaussiano	Malla Regular
Matricial	Si	Si	No
Secuencial Gaussiano	Si	Si	No
Recocido Simulado (Simulated Annealing)	Si	No	Si
Espectral	No	Si	No
Bandas Rotantes	No	Si	No

En este caso se presentara la teoría de los métodos de tipo Gaussiano y de tipo secuenciales, poniendo más énfasis en el método de simulación secuencial Gaussiano, que es el que será utilizado en el presente trabajo.

4.4 Métodos de Tipo Gaussiano

En este caso lo que se pide es que la función aleatoria $Z(\underline{x})$ que se esta estudiando y de la cual se realizará la simulación, presente función de densidad de probabilidades normal multivariada, lo cual es difícil de cumplir, más aun cuando se trabaja con fenómenos naturales.

Ahora para poder tener normalidad multivariada, una condición necesaria es tener normalidad univariada, por lo cual se busca transformar a la función $Z(\underline{x})$, de manera que esta presente normalidad.

La forma sencilla de transformar una función aleatoria buscando que presente normalidad es la siguiente:

Si se tiene una función aleatoria $Y(\underline{x})$ con función de densidad de probabilidades univariada $F_Y(y) = G(y) \sim N(0,1)$.

$$\begin{aligned} \Rightarrow F_Y(y_p) &= F_Z(z_p) = p \text{ para } \forall p \in [0,1] \\ \Rightarrow y &= F_Y(F_Z(z)) \end{aligned} \quad (4.4)$$

En la práctica si contamos con n datos de $Z(\underline{x})$, estos se ordenan de acuerdo a su estadística de orden.

$$z^{(1)} \leq z^{(2)} \leq \dots \leq z^{(n)} \quad (4.5)$$

Entonces la función de densidad de probabilidades acumulada de $Z(\underline{x})$ es:

$$\begin{aligned} F_Z(z^{(k)}) &= \frac{k}{n} \\ \Rightarrow y^{(k)} &= G^{-1}\left(\frac{k}{n}\right) \end{aligned} \quad (4.6)$$

Este tipo de transformación se conoce como **anamorfosis**.

Por último es importante señalar que después de realizar la simulación utilizando los datos transformados, se debe de realizar la transformación inversa para obtener los valores resultantes en la misma escala de medición del fenómeno. A continuación se mostrará un diagrama donde se describe el procedimiento general de las simulaciones Gaussianas.

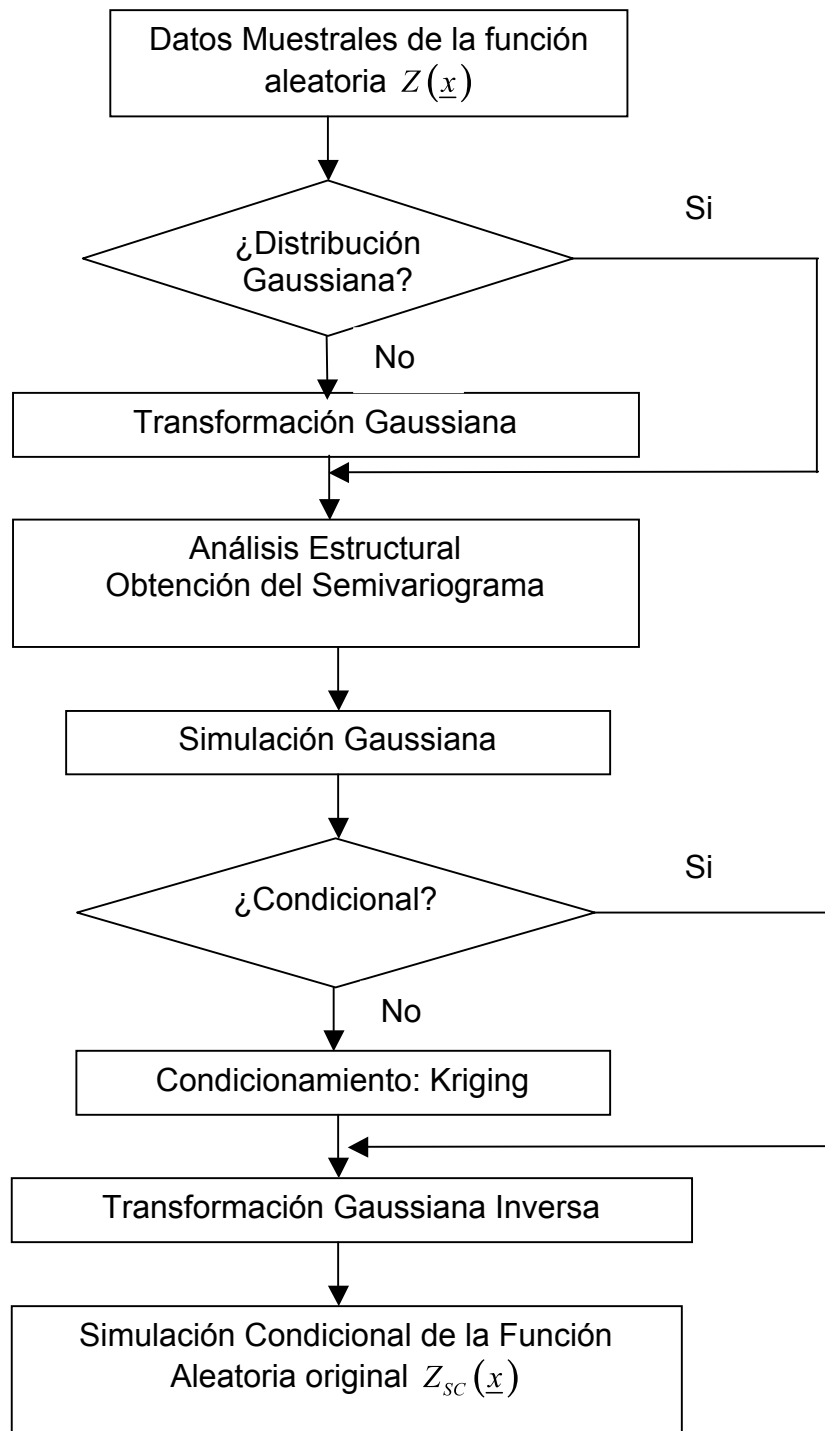


Figura 4.1: Esquema general para realizar simulaciones de tipo Gaussianas

4.5 Simulación Secuencial Gaussiana

Este método fue introducido en la literatura geoestadística por Alabert y Massonat (1990) para simular variaciones de log-permeabilidad a pequeña escala. Si consideramos un vector aleatorio $\mathbf{Z} = (Z_1, Z_2, \dots, Z_N)'$ del cual se tiene un subvector conocido de realizaciones $(Z_1 = z_1, Z_2 = z_2, \dots, Z_M = z_m)'$ donde $0 \leq M < N$. La distribución del vector \mathbf{Z} condicionado en $Z_i = z_i$ para $i = 1, 2, \dots, M$, puede ser dado como el producto de $N - M$ distribuciones condicionales univariadas, es decir:

$$\begin{aligned} \Pr(Z_{M+1} = z_{M+1}, \dots, Z_N = z_N | Z_1 = z_1, \dots, Z_M = z_M) &= \Pr(Z_{M+1} = z_{M+1} | Z_1 = z_1, \dots, Z_M = z_M) \\ &\quad * \Pr(Z_{M+2} = z_{M+2} | Z_1 = z_1, \dots, Z_M = z_M, Z_{M+1} = z_{M+1}) \\ &\quad \vdots \\ &\quad * \Pr(Z_N = z_N | Z_1 = z_1, \dots, Z_M = z_M, Z_{M+1} = z_{M+1}, \dots, Z_{N-1} = z_{N-1}) \end{aligned} \quad (4.7)$$

Por tanto podemos obtener la **simulación secuencial** de \mathbf{Z} de la siguiente forma:

- Se selecciona el primer punto a ser generado Z_{M+1} .
- Obtener la densidad condicional dada la información muestral, es decir se requiere encontrar $\Pr(Z_{M+1} = z_{M+1} | Z_1 = z_1, \dots, Z_M = z_M)$.
- Seleccionar el segundo punto a ser generado Z_{M+2} .
- Obtener la densidad condicional dada la información muestral y el punto Z_{M+1} generado en la etapa anterior es decir se requiere encontrar $\Pr(Z_{M+2} = z_{M+2} | Z_1 = z_1, \dots, Z_M = z_M, Z_{M+1} = z_{M+1})$.
- El procedimiento anterior se repite hasta encontrar los puntos.

La simulación secuencial es un proceso muy general, pero se puede utilizar en particular para Z_i de forma que $Z_i = Z(x_i)$, donde $Z(x)$ es una función aleatoria y

$\underline{x}_i, i = 1, 2, \dots, M$ son los puntos muestrales, mientras que los puntos donde se requiere realizar la simulación de $Z(\underline{x})$ están dados por $\underline{x}_i, i = M + 1, M + 2, \dots, N$. Es importante señalar que cuando $M = 0$ a la simulación se le conoce secuencial no condicional y se conoce como simulación secuencial condicional cuando $M > 0$.

En la práctica la principal dificultad es el cálculo de la función de probabilidades mostrada en (4.7), esto exceptuando el caso ideal en donde el vector aleatorio tiene distribución normal. Cuando \mathbf{Z} tiene distribución normal, entonces la distribución condicional de Z_i es normal, con media Z_i^* y varianza $\sigma_{KQ_i}^2$, donde Z_i^* es el estimador de Kriging Ordinario de Z_i y $\sigma_{KQ_i}^2$ esta asociada a la varianza del Kriging Ordinario.

Este método de simulación pueden ser extendido al ámbito multivariado, mas especifico en este trabajo se tratara el caso bivariado, para este tipo de simulación se requiere que las funciones aleatorias estén correlacionadas. Cabe señalar que al igual que en el caso univariado aquí también el supuesto primordial es que ambas funciones aleatorias tienen distribución normal, este tipo de simulación es conocida como simulación conjunta o cosimulación.

4.6 ¿Por qué se eligió el Método Secuencial Gaussiano?

En la Tabla 4.1 se enlistaron las características de algunos de los algoritmos de simulación alternativos al método Secuencial Gaussiano, utilizado en este trabajo.

La primera consideración que se tomó para elegir el método de simulación fue que el método fuera condicional, debido a que se buscaba que las realizaciones artificiales de la función aleatoria tuvieran la media y varianza muy cercana a la media y varianza teóricas, lo cual también es conocido como ergodicidad. Esta condición nos restringe a los métodos Matricial, Secuencial Gaussiano y de Recocido Simulado.

Para diferenciar entre los estos métodos condicionales se comparó principalmente que tan complejo computacionalmente era su implementación.

El **método matricial** genera realizaciones artificiales de la función aleatoria en las cuales la matriz de covarianza coincide con la obtenida mediante Kriging. Este método es sencillo de implementar pero resulta muy costoso computacionalmente, ya que al incrementar la cantidad de puntos a simular el tiempo de procesamiento es superior al cuadrado del número de puntos debido a que requiere de la inversión de la matriz de covarianzas. Este método clásico fue introducido en aplicaciones geoestadísticas por Davis (1987), y es aplicable hasta límite que sea factible la descomposición de Cholesky, es decir, hasta unos cuantos cientos de puntos simulados.

El **método de recocido simulado** fue introducido por Kirkpatrick (1983) como un método para la solución de problemas combinatorios (discretos) de optimización. Este inicia con una imagen aleatoria y requiere de los datos observados y del modelo de semivariograma, posteriormente el algoritmo modifica la imagen inicial, hasta que esta reproduzca el semivariograma y al mismo tiempo mantenga la distribución de los datos muestrales. Desde el punto de vista computacional este método es muy demandante ya que requiere de millones de iteraciones para que la simulación refleje correctamente la función de semivarianzas.

Finalmente, en el **método secuencial Gaussiano** un nuevo valor simulado $Z_s(\underline{x})$ se obtiene a partir de la función de distribución de probabilidad estimada usando los valores muestrales (reales) y los valores previamente simulados en etapas anteriores en una vecindad del punto \underline{x} . Es un algoritmo simple y eficiente. Sus desventajas estriban en que el tamaño de la región a simular debe ser mucho más grande que el alcance de los variogramas. Su mayor dificultad es la elección del radio de búsqueda. Cuando son pequeños conducen a un pobre condicionamiento y una mala réplica del variograma.

5. Descripción del Caso de Estudio

En 1995 el Servicio Meteorológico Nacional, compro siete nuevos radares y modernizó los cinco existentes, los cuales conforman actualmente la red nacional de radares Doppler, los cuales se encargan de seguimiento de los ciclones tropicales y la estimación de lluvia.

El radar que a nosotros nos interesa se encuentra en el Cerro de la Catedral, cerca de la Ciudad de México y es manejado por la Comisión Nacional del Agua, del cual proviene la información para la realización del trabajo.

El Sistema de Aguas del Departamento del Distrito Federal tiene bajo su control la operación de una red de 61 pluviómetros que están ubicados dentro de la zona de cobertura del radar y tiene una densidad del orden de un pluviómetro por cada 30 km².

5.1 Información de radar:

Como ya se mencionó utilizará la información provista por el radar ubicado en el Cerro de la Catedral. Estas imágenes son de 8 bits de 240 X 240 Km. con una resolución de 1 km², es decir, el píxel abarca un área de por 1 Km. por 1 Km., en la presentación pseudo-CAPPI a 2 Km. de altura sobre el sitio de radar y se obtienen cada quince minutos.

Debido a que el radar registra cada 15 minutos una imagen con valores de la energía devuelta por las gotas de lluvia, mejor conocida como reflectividad (Z), es necesario sumar cuatro imágenes de radar consecutivas para obtener la precipitación acumulada en una hora. Previamente, cada una de estas imágenes debe ser convertida a valores de intensidad de lluvia (R) usando una relación Z-R. En nuestro caso se usó la relación $Z=300R$ recomendada por el fabricante (Valdés-Manzanilla y Herrera Zamarrón, 2000).

5.2 Información de pluviómetros:

Los datos de pluviómetros fueron tomados de una red instalada de 61 pluviómetros en el área metropolitana de la Ciudad de México que reportan vía radio, cada minuto la lluvia acumulada en ese intervalo a una computadora central.

Para los datos de pluviómetros se consideraron los archivos de computadora correspondientes a la fecha y hora de la tormenta a modelar. Considerando que cada pluviómetro cuenta con un contador que se incrementa en uno cada vez que registra una lluvia de un 1/4 de mm (Rosengaus, 2000) se procedió al cálculo de la lluvia acumulada en una hora.

Para el análisis geoestadístico conjunto resulta necesario tener valores de ambas mediciones en ubicaciones comunes. Con este fin, a las coordenadas correspondientes a la ubicación de un pluviómetro dado se le asocia un valor de lluvia acumulada de radar de la siguiente manera: se localiza el píxel que contenía a dichas coordenadas y se toma su valor de lluvia acumulada de radar junto con los valores de los ocho píxeles circundantes, finalmente se escoge el que tenga menor diferencia en valor absoluto respecto al valor de lluvia acumulada medido por el pluviómetro. Esto se hace debido a la incertidumbre que presenta la posición del lugar donde va a caer la lluvia medida por el radar producto de la acción del viento. Como resultado tenemos las mediciones de lluvia acumulada de pluviómetro y radar ubicadas en las mismas coordenadas.

Para poner en práctica la metodología en el trabajo propuesta se eligió solo una tormenta, esta fue la del **día 15 de julio de 1997** debido a que para este evento se tenía buena información en cantidad y calidad tanto de radar como de pluviómetros. Los registros confiables de lluvia para esta tormenta se obtuvieron para 50 pluviómetros y las imágenes de radar correspondientes a la duración de la misma, esta lluvia fue expresada en milímetros aplicando los procedimientos explicados en párrafos anteriores.

6. Estimación Univariada Usando los Datos de Pluviómetro

Para llevar a cabo la estimación univariada seguiremos los pasos ya establecidos en la metodología para la estimación, como ya se mencionó los resultados y gráficos aquí presentados se obtuvieron mediante la utilización del software R 2.6, al software le integramos paquetes complementarios como: sp, geoR, nortest, sgeostat, gstat, outliers, entre otros (<http://cran.r-project.org/web/packages/>).

En la Tabla 6.1 se muestran las estadísticas básicas de los datos de pluviómetro de la información con la que se dispone, al realizar un análisis de estas primeras estadísticas se podría decir que aunque la media y la mediana están bastante cercanas y desviación estándar se podría considerar pequeña lo cual es deseable, el problema en este caso es que aproximadamente el 75% de la información se encuentra entre $[0.25, 2]$ y solo alrededor de 25% de la misma se ubica entre $(2, 7.75]$.

Tabla 6.1: Estadísticas básicas para la variable pluviómetro

Estadístico	Precipitación Pluviómetro (mm)
Numero de Observaciones	50
Distancia Mínima (Km.)	1.5258
Distancia Máxima (Km.)	45.6961
Mínimo	0.2500
Media	1.4700
Máximo	7.7500
Varianza	2.5756
Desviación estándar	1.6048
1° Cuartil	0.3125
Mediana	1.0000
3° Cuartil	1.9380

La Figura 6.1 confirman lo que ya se había podido apreciar en la Tabla 6.1, esto es que la distribución de pluviómetro no es simétrica, por tanto tampoco presenta normalidad, de igual manera en la gráfica de caja se puede distinguir que en los datos analizados existen puntos aberrantes. De hecho al observar la Figura 6.2 se puede concluir que esta información presenta una simetría positiva muy marcada, lo cual se tiene que buscar corregir, ya que para el análisis estructural que llevaremos a cabo más adelante, se requiere que por lo menos la distribución sea simétrica.

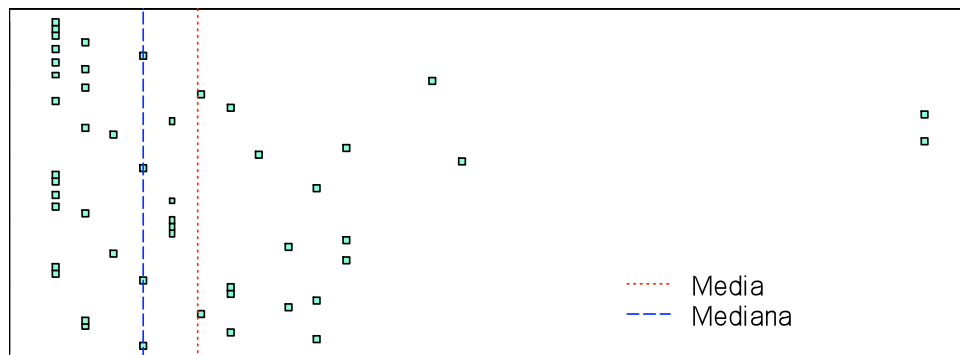


Figura 6.1: Ubicación de puntos, histograma y gráfica de caja de la variable pluviómetro

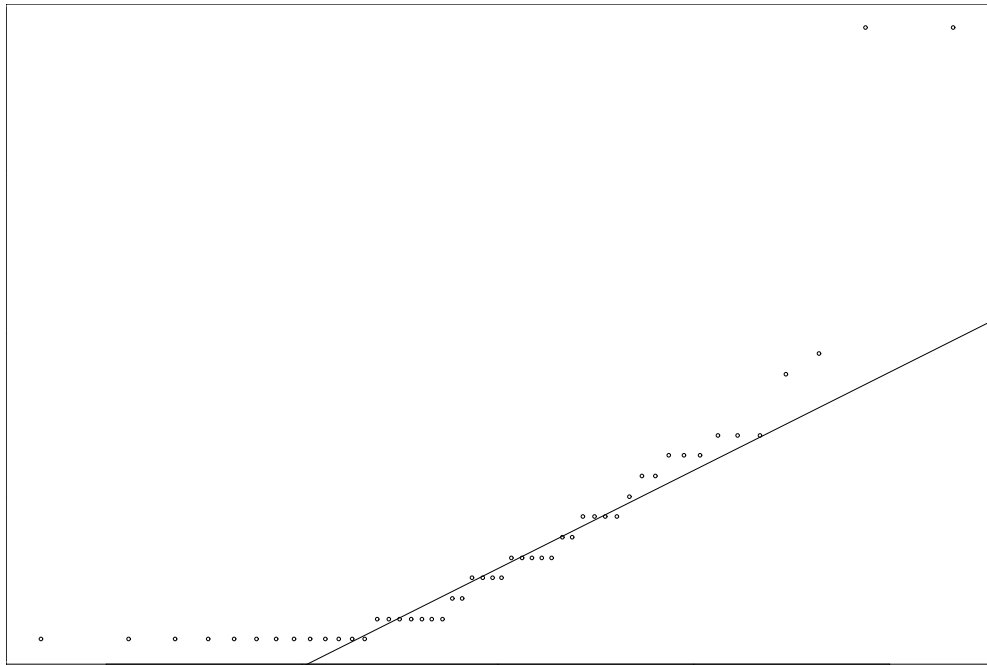


Figura 6.2: Gráfica QQ de la variable pluviómetro

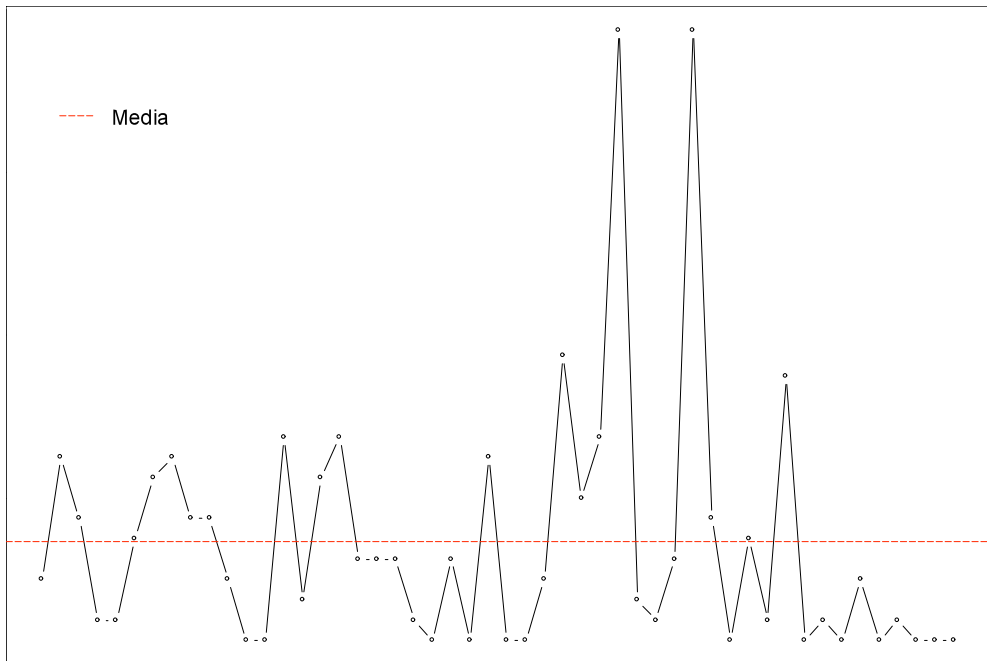


Figura 6.3: Gráfico de Estacionaridad de la variable pluviómetro

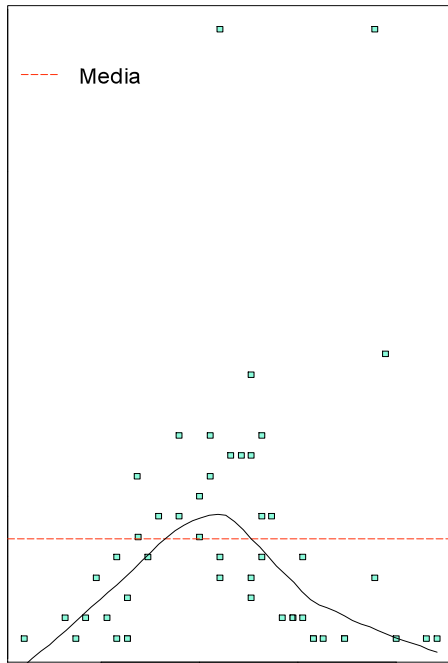


Figura 6.4: Gráficos de valores de pluviómetro vs. coordenadas

En la Figura 6.3 muestra un reflejo muy claro de la existencia de puntos aberrantes, ya que en la gráfica son estos puntos los que están mas alejados de la media, así mismo en la Figura 6.4 en los cuales se presentan gráficos de datos vs. coordenadas se puede ver que mientras en el sentido de las UTM Y los valores se ven distribuidos aleatoriamente alrededor de la media y sin ninguna tendencia, en el sentido UTM X la información podría presentar tendencia.

Se realizaron las pruebas de normalidad, puntos aberrantes y simetría las cuales son pruebas más formales para decidir si existe normalidad, puntos aberrantes y simetría en la información.

Tabla 6.2: Análisis de normalidad

Prueba	p-value
Shapiro-Wilk normality test	1.228e-08
Anderson-Darling normality test	3.433e-09
Cramer-von Mises normality test	7.404e-07

Como lo esperaríamos las tres pruebas de la Tabla 6.2 rechazan H_0 : Existe normalidad, con el nivel de significancia usual ($\alpha = 0.05$) por lo tanto se puede concluir que los datos de pluviómetro no presentan normalidad.

Tabla 6.3: Análisis de puntos aberrantes

Prueba	p-value	conclusión
chi-squared test for outlier	9.112e-05	El valor 7.75 es un punto aberrante
Grubbs test for one outlier	0.000485	El valor 7.75 es un punto aberrante
Outlier		El valor 7.75 es un punto aberrante

Tabla 6.4: Análisis de simetría

Prueba	p-value
MGG Test	0.0005796
Cabilio-Masaro Test	0.006126
Mira Test	0.008688

En la Tabla 6.3 se muestran tres pruebas para ver si existen puntos aberrantes, y se concluye que en la información se puede considerar como punto aberrante distribucional a 7.75. Mientras que en la Tabla de las pruebas de simetría rechazan H_0 : Existe asimetría, con nivel de significancia $\alpha = 0.05$, por tanto los datos no presentan simetría.

Para corregir el problema de asimetría y buscar la normalidad en los datos, se propone realizar una transformación logarítmica, esto con el fin de reducir las escalas de la información. Además de ver si con esta reducción de escalas se puede corregir alguno de los otros problemas como lo es la presencia de puntos aberrantes.

En la Tabla 6.5 se exponen las estadísticas básicas de los datos de pluviómetro transformados, en estas estadísticas se ve que la información tiene una mejor distribución, ya que el 50% de la información esta entre $[-1.386,0]$ y el otro 50% de esta se ubica entre $(0,2.048]$.

Tabla 6.5: Estadísticas básicas para las variables logaritmo de pluviómetro

Estadístico	Precipitación lnPluviometro (mm)
Numero de Observaciones	50
Distancia Mínima (Km.)	1.5258
Distancia Máxima (Km.)	45.6961
Mínimo	-1.3860
Media	-0.0903
Máximo	2.0480
Varianza	1.0004
Desviación estándar	1.0002
1° Cuartil	-1.2130
Mediana	0.0000
3° Cuartil	0.6598

La Figura 6.5 muestra la distribución de la variable ya transformada, donde a simple vista podría existir simetría, además de que en la gráfica de caja se aprecia que ya no existen puntos aberrantes, aunque la Figura 6.6 se observa que la información sigue sin presentar normalidad.

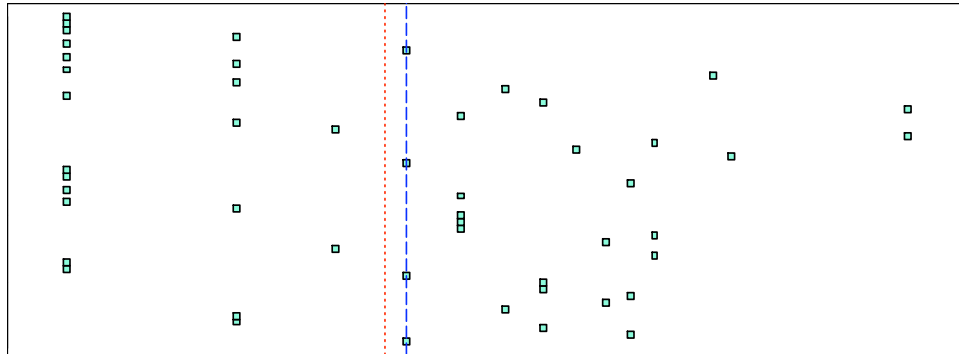


Figura 6.5: Ubicación de los puntos, histograma y gráfica de caja del logaritmo de pluviómetro

En las Figuras 6.7 y 6.8, muestran la estacionaridad de los datos, en la primera se aprecia que los datos aparentemente ya son estacionarios, mientras que en la segunda se ve que los datos en el sentido de la UTM X podrían presentar cierta tendencia, por lo que mas adelante se analizaran el semivariograma, por medio del cual se vera si la tendencia de los datos es grave o solo es una tendencia leve, o no significativa, por lo que se podría seguir adelante con el análisis.

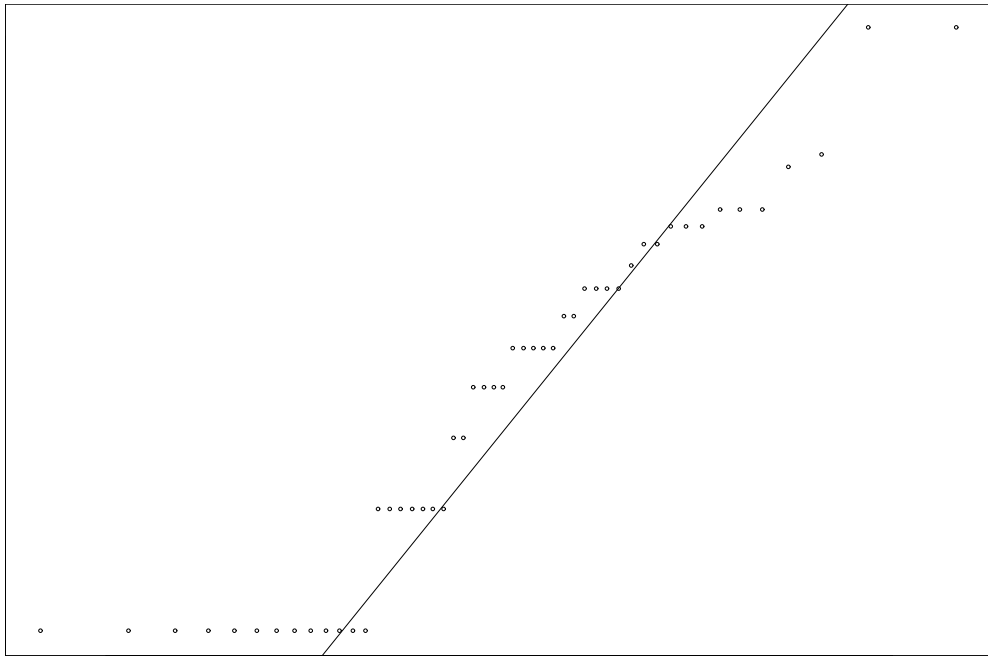


Figura 6.6: Gráfica QQ del logaritmo de pluviómetro

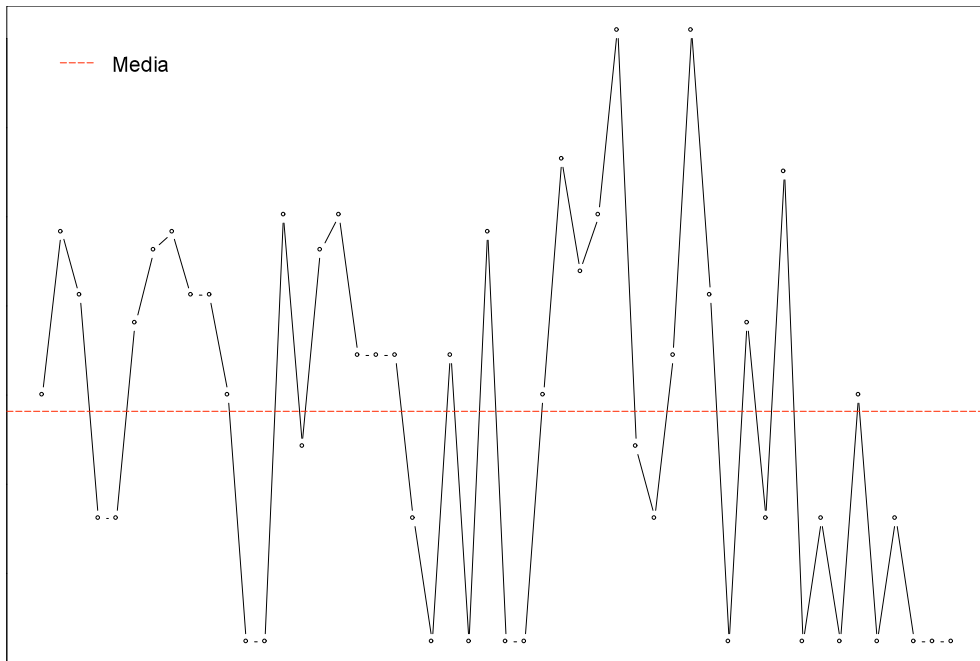


Figura 6.7: Gráfico de Estacionaridad de la variable logaritmo de pluviómetro

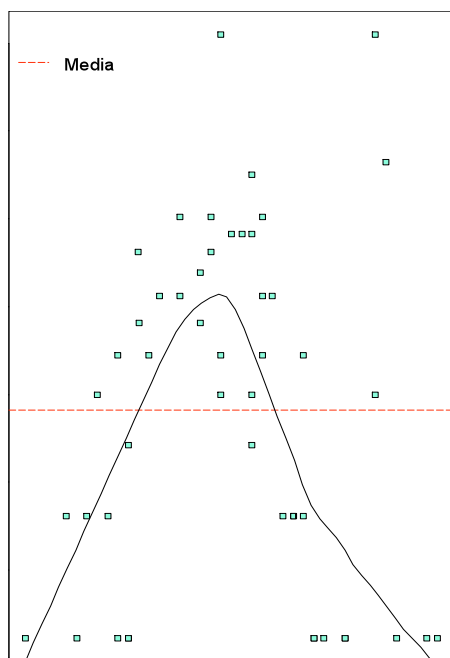


Figura 6.8: Gráficos de valores de logaritmo de pluviómetro Vs. coordenadas

Para los datos con la transformación logarítmica se llevaran acabo las mismas pruebas de normalidad, puntos aberrantes y simetría, que se utilizaron en el caso de los datos originales.

Tabla 6.6: Análisis de normalidad de las transformaciones

Prueba	p-value
Shapiro-Wilk normality test	0.001841
Anderson-Darling normality test	0.001554
Cramer-von Mises normality test	0.009668

Como se aprecia en la Tabla 6.6, se puede decir que datos ya con la transformación sigue rechazando H_0 : Existe normalidad, con un nivel de significancia $\alpha = 0.05$ y se concluye que los datos siguen sin presentar normalidad.

Tabla 6.7: Análisis de puntos aberrantes para las transformaciones

Prueba	p-value	Conclusión
chi-squared test for outlier	0.7325	El valor 2.0476 es un punto aberrante
Grubbs test for one outlier	0.7325	El valor 2.0476 es un punto aberrante
Outlier		El valor 2.0476 es un punto aberrante

Tabla 6.8: Análisis de simetría para las transformaciones

Prueba	p-value
MGG Test	0.4263
Cabilio-Masaro Test	0.3981
Mira Test	0.3953

Las pruebas de la Tabla 6.7 indican que sigue existiendo un punto aberrante, este se presenta en la parte superior de los datos, la decisión será no eliminar este dato ya que al observar los otros valores en los datos se puede ver que este valor es poco diferente, por lo que no sería un punto aberrante espacial, entonces se espera que no afecte nuestras estimaciones, además que si elimináramos se perdería mucha información sobre los lugares mas lluviosos, finalmente de la Tabla 6.8 se concluye que no se rechaza H_0 : Existe simetría, por tanto la distribución de los datos transformados es simétrica, lo cual es suficiente para tener un buen calculo del semivariograma.

En la Figura 6.9, que es donde se presenta el semivariograma adireccional de los datos transformados, se puede llegar a la conclusión de que no se presenta un crecimiento cuadrático, lo cual indica que no existe tendencia o si existe es muy leve, por lo cual se puede seguir adelante con el análisis. Referente a la Figura 6.10 se puede decir que los rangos y mesetas de los semivariogramas en las diferentes direcciones son prácticamente iguales, por tanto se podría concluir que no existe evidencia de anisotropía, con lo cual no se ve la necesidad de la construcción de modelos anisotrópicos.

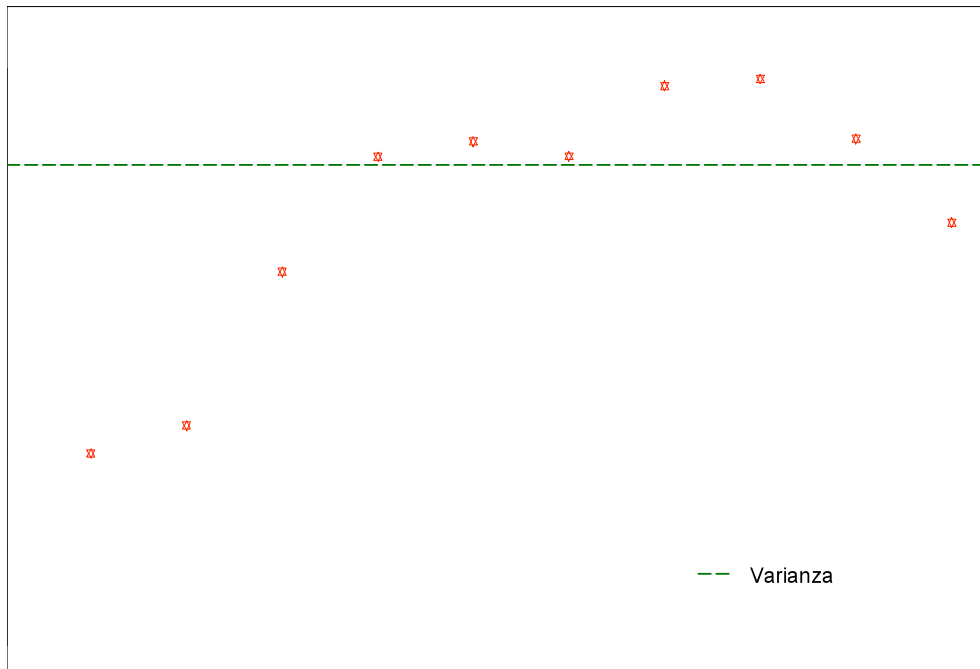


Figura 6.9: Semivariograma adireccional de logaritmo de pluviómetro ($0^\circ, \pm 90^\circ$)

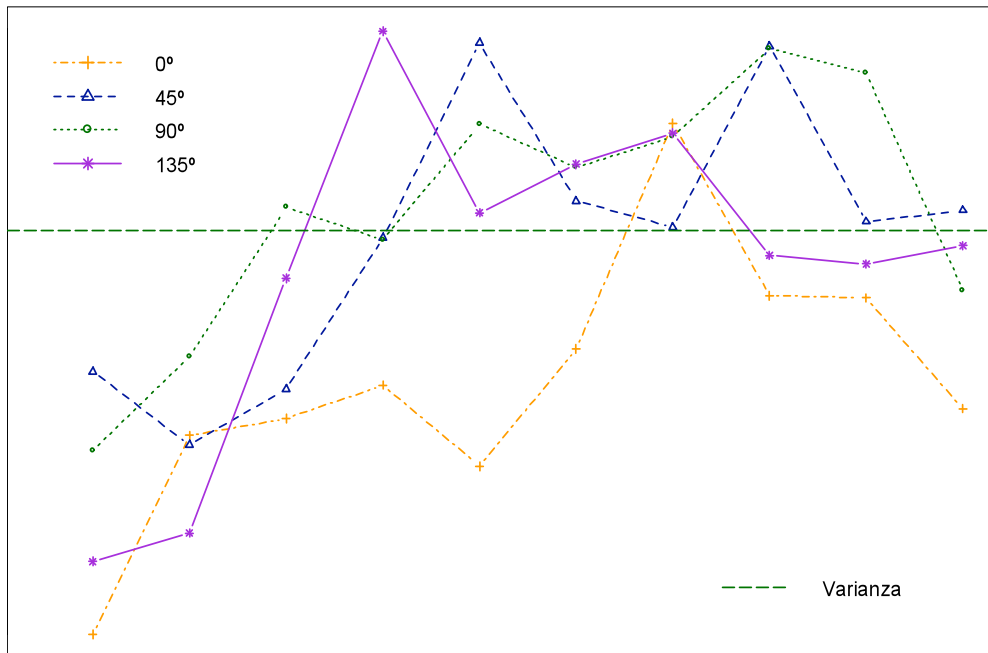


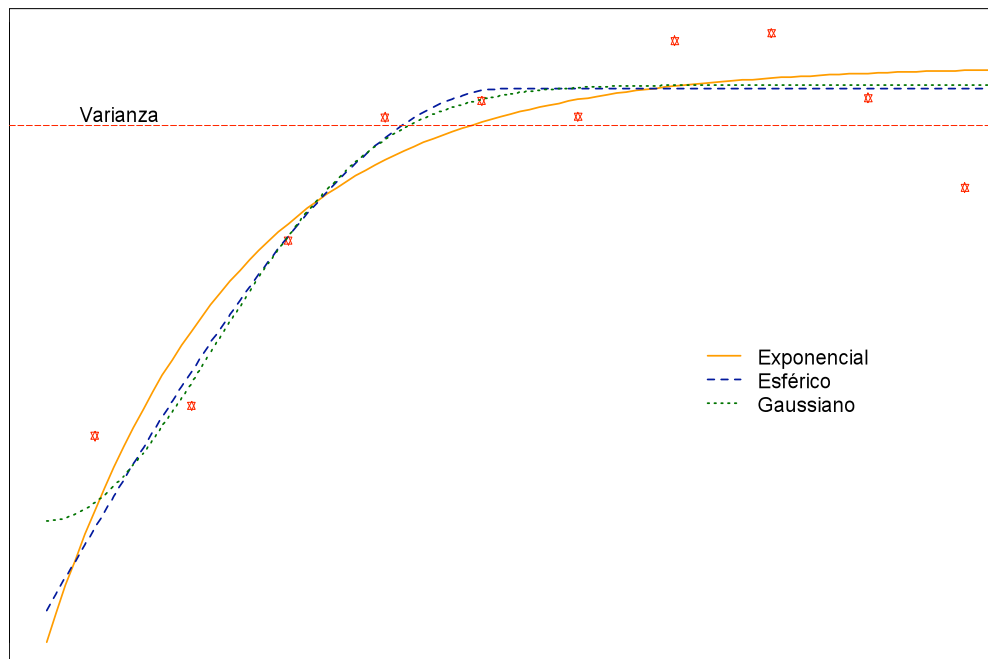
Figura 6.10: Gráfico de anisotropía de logaritmo de pluviómetro

Se realizó el ajuste de los modelos exponencial, Gaussiano y esférico, el método utilizado fue el de mínimos cuadrados ponderados, donde los pesos o ponderaciones fueron tomados de acuerdo al número de parejas en el intervalo correspondiente, para escoger el mejor modelo en este caso se toma en cuenta el modelo que minimiza la suma de cuadrados, a continuación se presentan los modelos.

Tabla 6.9: Modelos probados para los datos transformados de pluviómetro

Modelo	lnPluviometro (mm)			
	nugget	meseta	Alcance	Suma de cuadrados
Exponencial	0	1.1151	5.7860	10.01
Gaussiano	0.2349	0.8438	7.2827	6.8104
Esférico	0.0616	1.0104	14.2595	7.3504

Figura 6.11: Semivariograma y ajustes de los modelos probados.



El modelo elegido fue el esférico, ya que es uno de los modelos con menor suma de cuadrados, además de un buen rango y uno de los menores nugget, lo cual es importante ya que este parámetro lo que hace es aumentar el error del modelo.

Finalmente el modelo elegido y sus parámetros son presentados en la Tabla 6.10, a este modelo se le realizara la validación para ver el ajuste que tiene y posteriormente sus parámetros se usaran para realizar la estimación por medio de Kriging.

Tabla 6.10: Modelo elegido

Modelo	lnPluviometro (mm)			
	nugget	meseta	Alcance	Suma de cuadrados
Esférico	0.06	1.0120	14.2468	7.3505

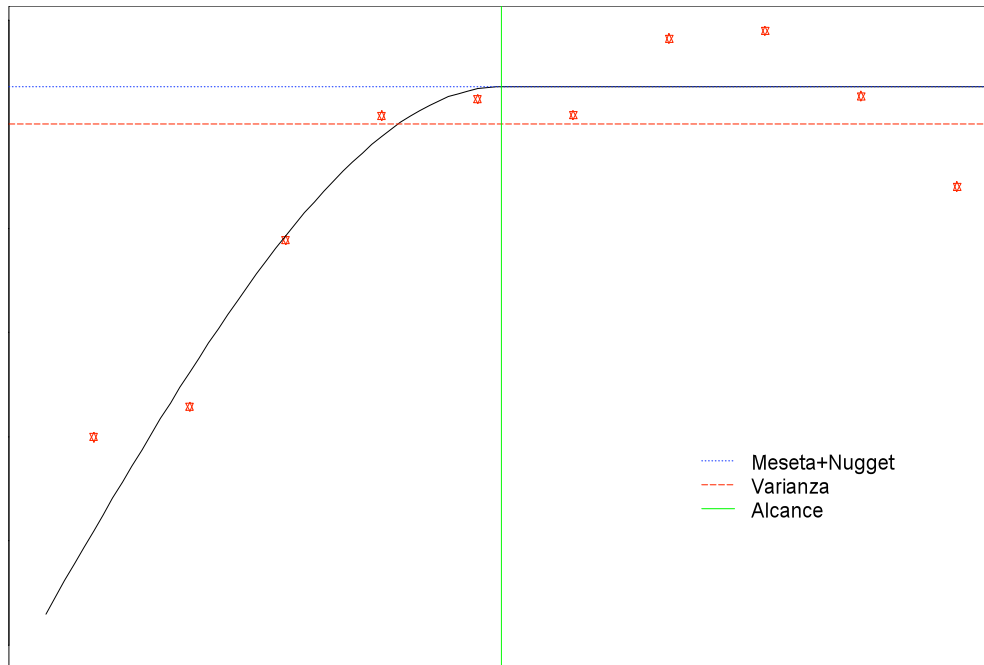


Figura 6.12: Semivariograma y modelo ajustado del logaritmo de pluviómetro

La validación de modelo se realizó utilizando el procedimiento validación cruzada (leave one out), de la validación se utilizaron los errores estandarizados, de los cuales se calcularon la media y varianza.

Media de los errores estandarizados = -0.03239552

Varianza de los errores estandarizados = 0.9148395

Como se puede observar el modelo tiene un buen ajuste, ya que la media de los errores estandarizados es cercana a cero y la varianza de este valor se puede considerar muy cercana a uno, los cuales se consideran los valores ideales, en cuanto a los gráficos, primeramente el de Reales Vs. Estimados, y de los errores (Figura 6.13) se puede ver que aproximadamente es una recta de 45° y que los errores se distribuyen aleatoriamente alrededor de cero.

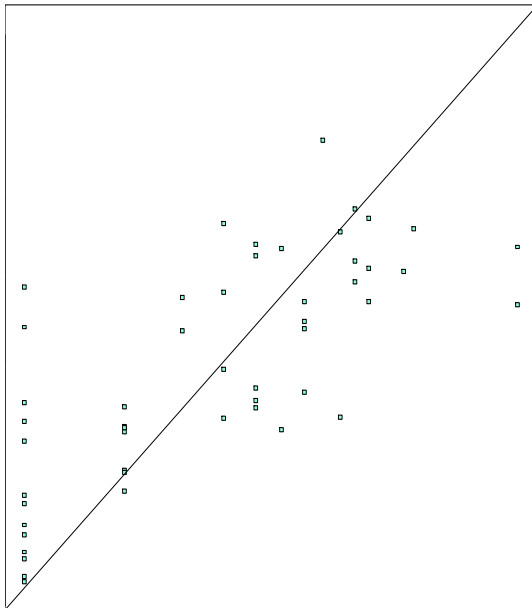


Figura 6.13: Gráficos de la validación del modelo del logaritmo de pluviómetro

Con el modelo ajustado y que como ya se mostró es un modelo bueno se realizara la estimación univariada, se usara el Kriging ordinario puntual, ya que este es el que mejor se adapta a las condiciones de nuestra información, como lo es que las variables en nuestro estudio son estacionarias, es importante mencionar que este procedimiento se aplicara sobre una malla regular de 1 km. por 1 km.

Primeramente en la Tabla 6.11 se muestran las estadísticas del Kriging y de la desviación estándar de su error de estimación, mientras que en las Figuras 6.14 y 6.15 se exhiben la imágenes de Kriging, la Figura 6.14 muestra la información de los valores estimados o predicciones de logaritmo de pluviómetro, mientras que en el Figura 6.15 se observa la información de la desviación estándar del error de la estimaciones que se comete al utilizar Kriging Ordinario Puntual en estos datos, dentro de esta Figura también se puede observar como el error aumenta a medida que se aleja de los puntos, los cuales corresponden a la información que se tiene.

Tabla 6.11: Estadísticas del Kriging Ordinario y de la desviación estándar

Estadístico	Kriging Ordinario	Desviación Estándar del Error
Numero de Observaciones	1892	1892
Distancia Mínima (Km.)	1.0000	1.0000
Distancia Máxima (Km.)	60.1082	60.1082
Mínimo	0.2126	0.3503
Media	1.0250	0.7407
Máximo	6.2120	1.1260
Varianza	0.6809	0.0384
Desviación estándar	0.8252	0.1960
1° Cuartil	0.4565	0.5850
Mediana	0.7458	0.6981
3° Cuartil	1.3490	0.9015

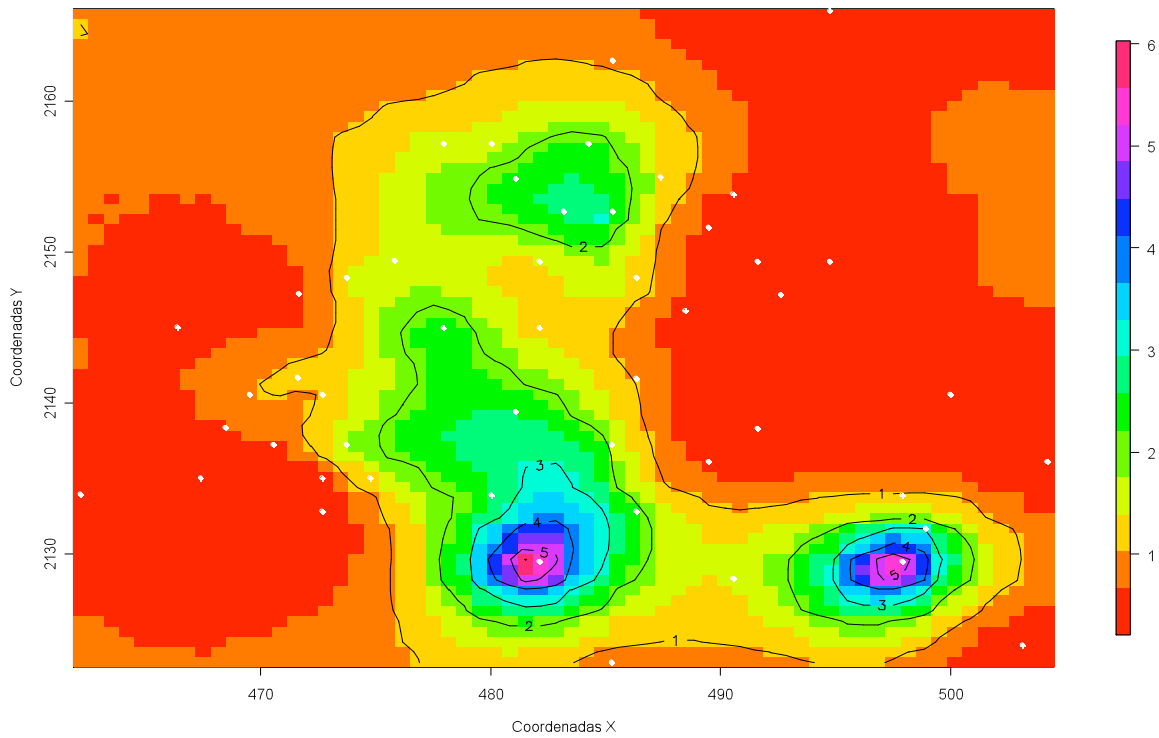


Figura 6.14: Kriging de los valores estimados para la variable pluviómetro

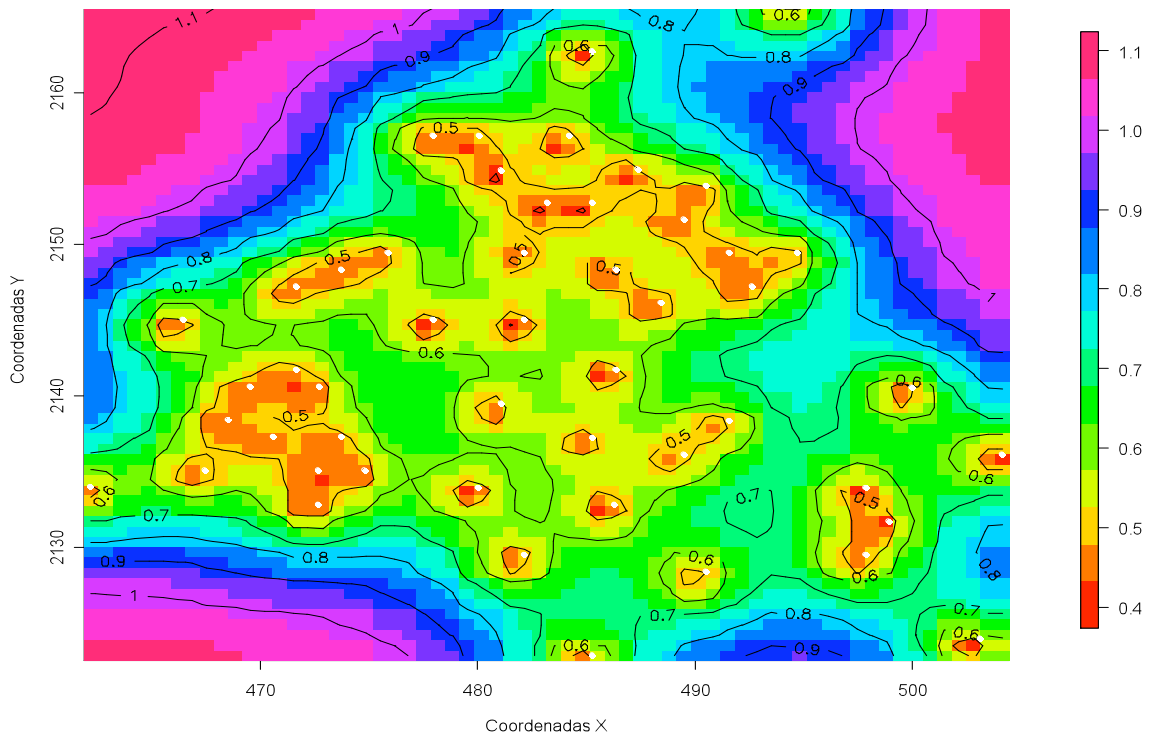


Figura 6.15: Desviación estándar del error de estimación para la variable pluviómetro usando Kriging

7. Estimación Conjunta Usando Datos de Pluviómetro e Imagen de Radar

La realización de este análisis se basa primordialmente en la existencia de correlación entre las variables en estudio, en este caso particular el logaritmo de pluviómetro y logaritmo de radar, puesto que si las variables son no correlacionadas, no tendría caso llevar a cabo el presente estudio, ya que los resultados que se obtendrían variarían poco con los que se obtuvieron al realizar el análisis univariado de la variable pluviómetro.

Cabe señalar que en este estudio la variable pluviómetro se le conocerá como variable primaria o variable que se predecirá, esto siempre y cuando exista una correlación elevada. En este caso la variable radar será la covariable o conocida también como variable auxiliar, esta covariable usualmente es la que esta altamente muestreada.

Para utilizar como covariable a la variable radar es necesario realizar el análisis exploratorio de datos descrito en la Sección 3.1, esto debido a que en este capítulo se requerirá estimar el semivariograma de esta variable y como ya se ha mencionado en anteriores capítulos para esto se requiere entre otras cosas que exista normalidad o al menos simetría.

Se presentaran las pruebas más importantes para esta variable:

Tabla 7.1: Análisis de normalidad

Prueba	p-value
Shapiro-Wilk normality test	1.097e-06
Anderson-Darling normality test	7.615e-06
Cramer-von Mises normality test	9.079e-05

Las tres pruebas de la Tabla 7.1 rechazan H_0 : Existe normalidad, con el nivel de significancia usual ($\alpha = 0.05$).

Tabla 7.2: Análisis de outliers

Prueba	p-value	conclusión
chi-squared test for outlier	0.0001529	El valor 7.79 es un punto aberrante
Grubbs test for one outlier	0.001011	El valor 7.79 es un punto aberrante
Outlier		El valor 7.79 es un punto aberrante

Tabla 7.3: Análisis de simetría

Prueba	p-value
MGG Test	0.003175
Cabilio-Masaro Test	0.01469
Mira Test	0.01898

En la Tabla 7.2 muestra que en la información de radar existen puntos aberrantes y la Tabla de las pruebas de simetría se rechaza H_0 : Existe simetría, con nivel de significancia $\alpha = 0.05$, por tanto para poder utilizar a la variable radar como covariable necesitamos realizar una transformación sobre la misma, en este caso se realizó la transformación logaritmo natural, con lo cual al realizar las mismas pruebas anteriormente expuestas se obtienen los siguientes resultados.

Tabla 7.4: Análisis de normalidad de las transformaciones.

Prueba	p-value
Shapiro-Wilk normality test	0.1318
Anderson-Darling normality test	0.09455
Cramer-von Mises normality test	0.09562

Como se aprecia en la Tabla 7.4, al realizar esta transformación logarítmica se obtiene normalidad en los datos, por consiguiente no se rechaza H_0 : Existe normalidad, con un nivel de significancia $\alpha = 0.05$, por tanto ya no es necesario llevar a cabo las pruebas de simetría, ya que al haber normalidad en los datos, implica que también existe simetría.

Tabla 7.5: Análisis de outliers para las transformaciones.

Prueba	p-value	Conclusión
chi-squared test for outlier	0.0266	El valor -1.7147 es un punto aberrante
Grubbs test for one outlier	0.5873	El valor -1.7147 es un punto aberrante
Outlier		El valor -1.7147 es un punto aberrante

En la Tabla 7.5 las pruebas nos indican que existe un punto aberrante el cual se ubica en la parte superior de los datos, la decisión será no eliminar este dato ya que al observar los otros valores en la información se puede ver que este valor es poco diferente, por lo que no sería un punto aberrante espacial. Por tanto la covariable que se utilizara para realizar el presente trabajo será logaritmo de radar (Lnradar).

Ahora para ver el grado de correlación numéricamente, se calculo el coeficiente de Pearson, el cual mide la correlación lineal entre dos variables.

Tabla 7.6: Coeficiente de Pearson

Variables	Lnradar	Lnpluviometro
Lnradar	1.0000	0.8256
Lnpluviometro	0.8256	1.0000

En la Tabla 7.6 claramente se aprecia con el coeficiente de Pearson que existe una correlación lineal elevada entre las variables del estudio, por lo cual si tiene sentido

llevar acabo el análisis geoestadístico conjunto, ya que se obtendrán resultados mejores que los obtenidos en el caso univariado.

En la Figura 7.1 del análisis se observara que tanto los valores inferiores, como superiores están poco correlacionados, aunque en el centro, donde se concentra el grueso de la información, se ve que existe una correlación elevada.

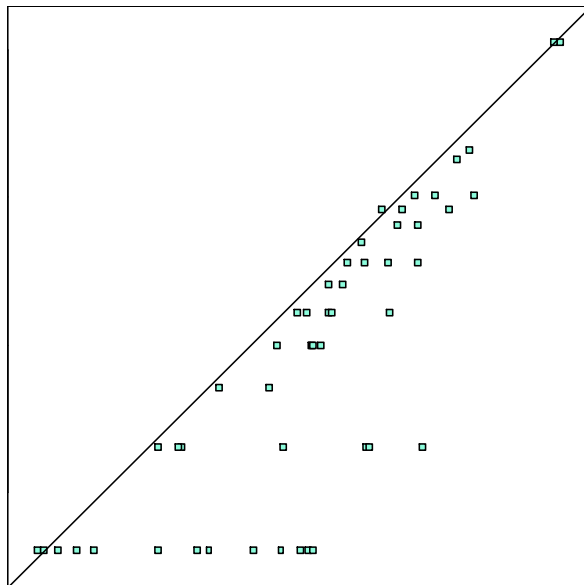


Figura 7.1: Lnradar Vs. Lnpluviometro

Primeramente se calcularan los semivariogramas simples y el semivariograma cruzado, estos se muestran en la Figura 7.2.

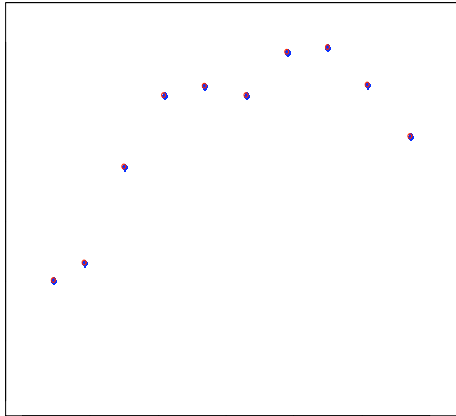


Figura 7.2: Semivariogramas adireccionales

En la Figura 7.3 se muestran los semivariogramas en cuatro diferentes direcciones, donde se puede apreciar que los variogramas aproximadamente tienen el mismo alcance y la misma meseta, por lo cual no es necesario el uso de modelos anisotrópicos.

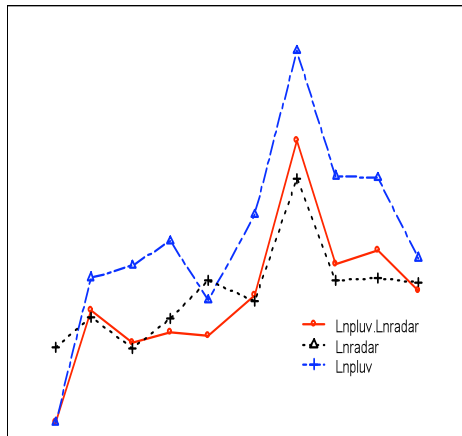


Figura 7.3: Semivariogramas en las diferentes direcciones

Se realizó el ajuste del mejor modelo a cada semivariograma Tabla 7.7, esto se realizó por medio de mínimos cuadrados ponderados, a continuación se presenta el mejor ajuste, que como se verá resultó ser el modelo esférico, en la Figura 7.4 se presenta tanto el semivariograma como su modelo ajustado, así como también los determinantes, los cuales son positivos, con lo cual de acuerdo al modelo de correlograma lineal, el modelo aquí propuesto es válido.

Tabla 7.7: Modelos elegidos para el análisis bivariado

Variable	Modelo	nugget	meseta	Rango
Lnpluviometro	Esférico	0.14	1	19
Lnradar	Esférico	0.03	0.86	19
Inradar.Inpluviometro	Esférico	0.01	0.83	19

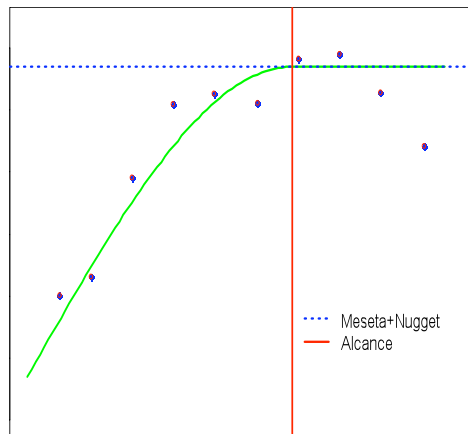


Figura 7.4: Variogramas adireccionales y modelo ajustado

Ahora con el mejor modelo para el semivariograma cruzado, se realizara la validación de este por medio del procedimiento validación cruzada (leave one out), esto se realizara por medio del análisis de los errores o residuales estandarizados así como también por medio de Figura 7.5.

Media de los errores estandarizados= -0.003689175

Varianza de los errores estandarizados = 0.8805832

Se aprecia que los valores tanto de la media como de la varianza son muy aproximados a los ideales, los cuales serian cero y uno respectivamente. Por lo que se puede concluir que con el modelo que se esta trabajando para el análisis conjunto es aceptable, esto mismo se ve en los errores, los cuales se distribuyen de manera aleatoria alrededor de cero.

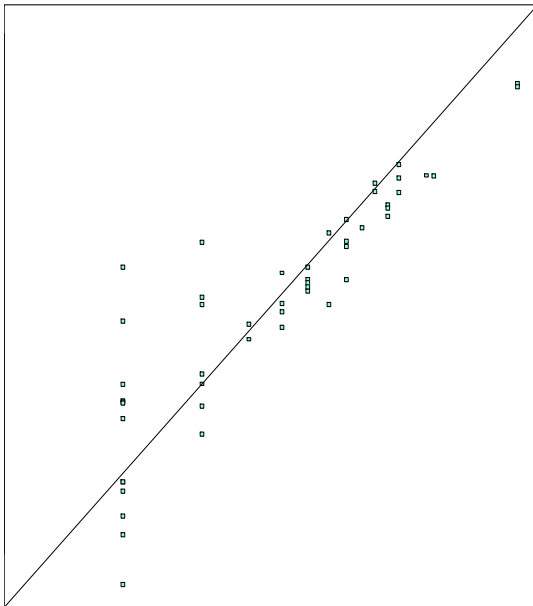


Figura 7.5: Distribución, histograma y diagrama de caja de los errores estandarizados

Finalmente para terminar con este análisis bivariado, se llevara acabo la estimación de la precipitación utilizando Cokriging Ordinario y Cokriging Colocado, ya que se cumple con la condición de estacionaridad en la información.

Al igual que en el caso univariado, para estos procedimientos se utilizara una malla regular de 1 km. Por 1 km. En las Figuras posteriores se mostraran las estimaciones y desviaciones estándar del Cokriging Ordinario y Colocado respectivamente.

Primeramente en la Tabla 7.8 se muestran las estadísticas del Cokriging y sus desviaciones estándar, ahora analizando un poco las imágenes obtenidas por Cokriging Ordinario, se podría decir en lo que se refiere a la Figura 7.6, la cual corresponde a las estimaciones, la imagen obtenida es menos suavizada que la que se obtuvo con Kriging (Figura 6.14), en otras palabras en esta imagen se aprecia más detalle en la estimación.

En la Figura 7.7 en la que se muestran las desviaciones estándar del error de estimación, se puede concluir que estas disminuyeron con respecto a las obtenidas con Kriging (Figura 6.15), además de haber menores errores de estimación en las esquinas, esto debido a la información que nos apporto la covariable.

En lo referente a las imágenes obtenidas por Cokriging Colocado, estas son muy parecidas a las obtenidas con Kriging, la mejora que se obtuvo con esta estimación fue la disminución de los errores de estimación en las esquinas, como sucedió en el Cokriging Ordinario.

Tabla 7.8: Estadísticas básicas del Cokriging Ordinario, Colocado y las desviaciones estándar de los mismos

Estadístico	Cokriging Ordinario pluviómetro (mm)	Desviación Estándar del Error	Cokriging Colocado Pluviómetro (mm)	Desviación Estándar del Error
Numero de Observaciones	1892	1892	1892	1892
Distancia Mínima (Km.)	1.0000	1.0000	1.0000	1.0000
Distancia Máxima (Km.)	60.1082	60.1082	60.1082	60.1082
Mínimo	0.000071	0.4548	0.1855	0.4856
Media	0.9979	0.5576	0.9473	0.7463
Máximo	9.8540	0.8918	4.6450	1.1200
Varianza	1.1759	0.00544	0.5602	0.0266
Desviación estándar	1.0844	0.0737	0.7485	0.1633
1º Cuartil	0.2547	0.5066	0.3784	0.6179
Mediana	0.6057	0.5335	0.6259	0.6963
3º Cuartil	1.4410	0.5893	1.3790	0.8617

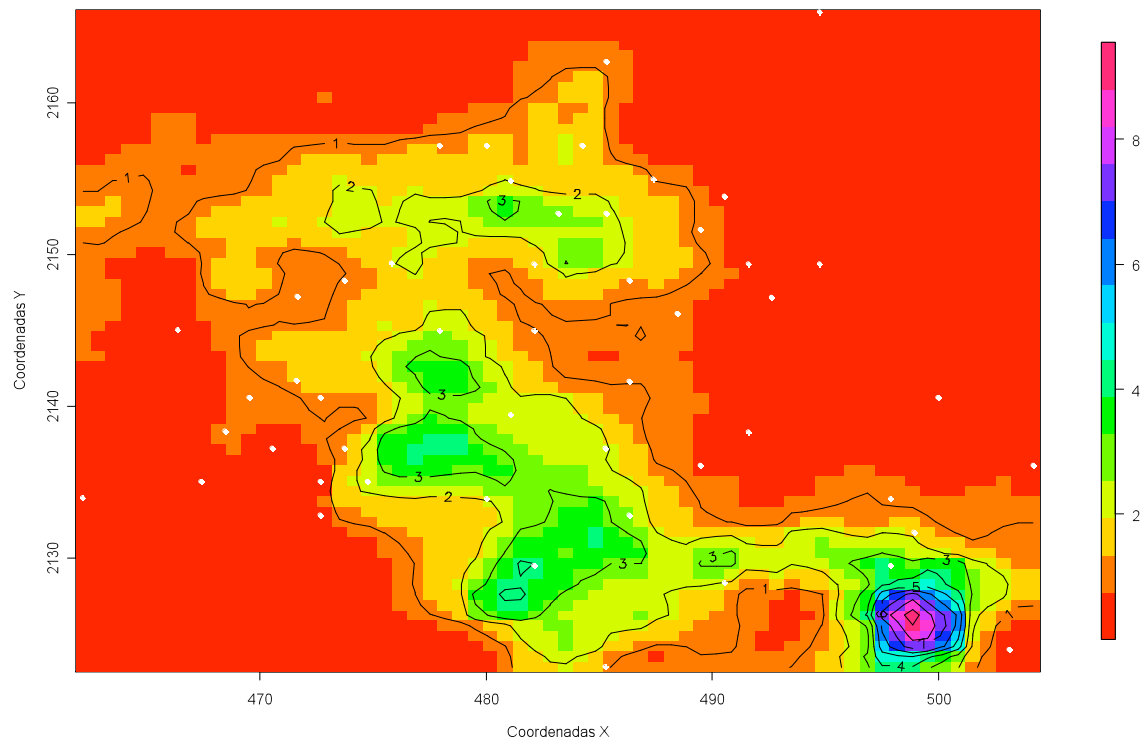


Figura 7.6: Cokriging Ordinario para las predicciones de pluviómetro.

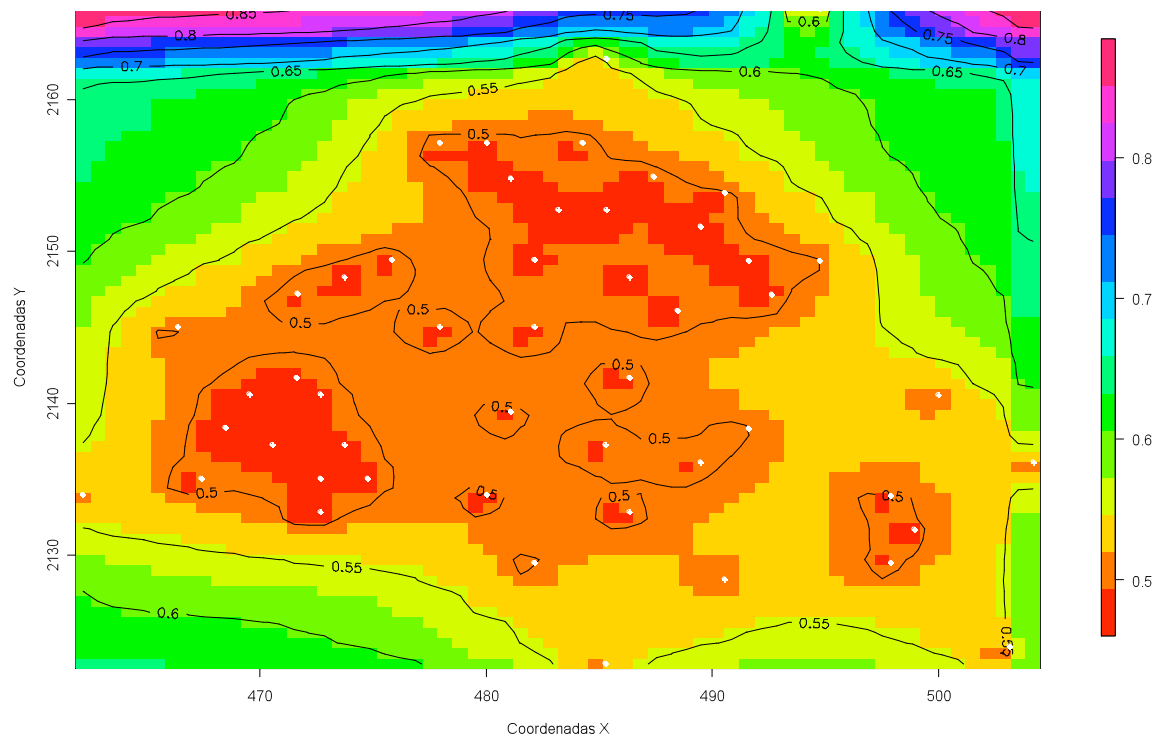


Figura 7.7: Desviaciones estándar del error de estimación de pluviómetro usando Cokriging Ordinario

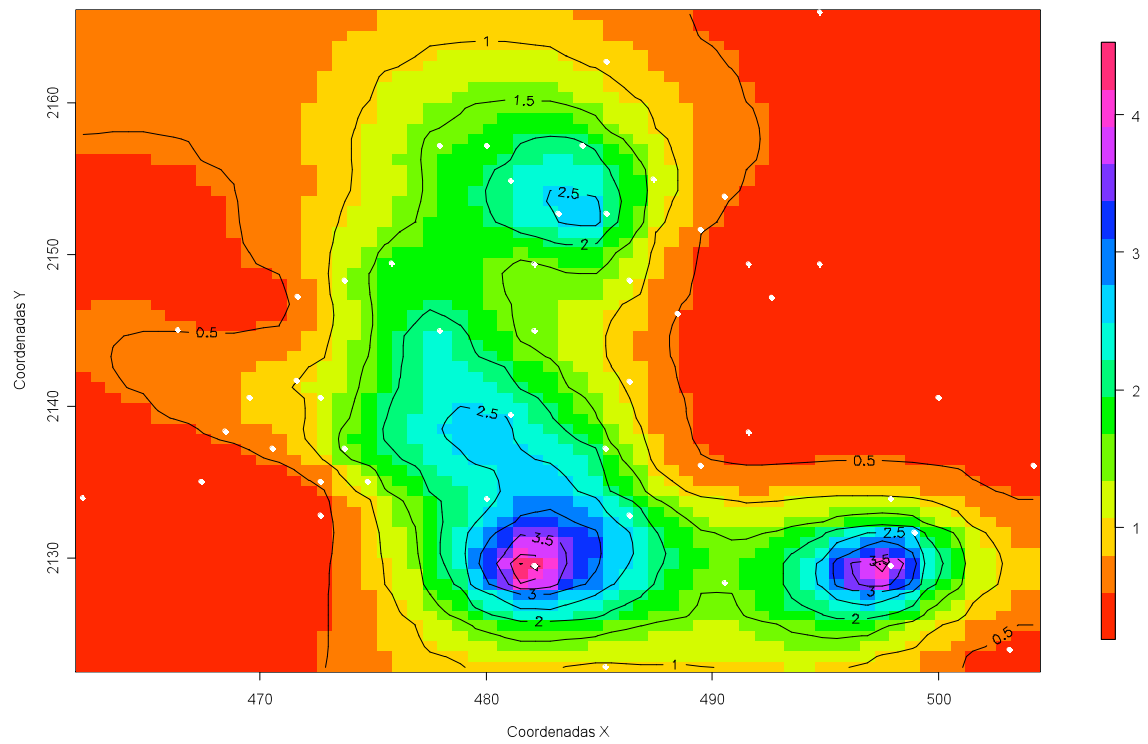


Figura 7.8: Cokriging Colocado para las predicciones de pluviómetro

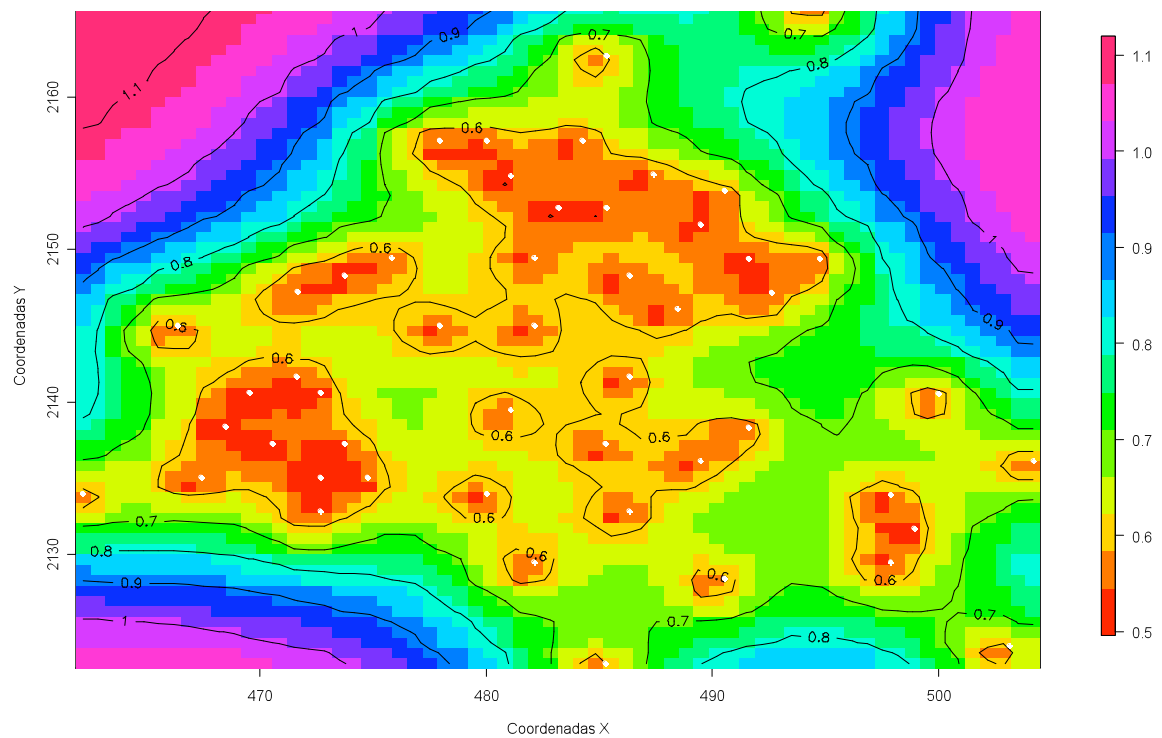


Figura 7.9: Desviaciones estándar del error de estimación de pluviómetro usando Cokriging Colocado

8. Simulación Univariada Usando Datos de Pluviómetro

Para llevar a cabo simulación secuencial Gaussiana usando Kriging Ordinario, como ya se mencionó en el Capítulo 4 se requiere que la información presente normalidad, lo cual como se observó en el Capítulo 6 no se cumple, por tanto para realizar la simulación primero se llevara a cabo la transformación de anamorfosis, la cual nos asegura que la distribución sea normal.

8.1 Análisis Geoestadístico Univariado Usando la Transformación de Anamorfosis

Después de realizar la transformación se obtuvieron todos los cálculos mostrados en el Capítulo 6 para la transformación logarítmica, por tanto en este Capítulo ya no se presentara todo ese desarrollo, solo se presentara los resultados mas relevantes.

Tabla 8.1: Estadísticas para la transformación anamorfosis de pluviómetro.

Estadístico	Precipitación Anamorfosis pluviómetro (mm)
Numero de Observaciones	50
Distancia Mínima (Km.)	1.5258
Distancia Máxima (Km.)	45.6961
Mínimo	-2.0970
Media	0.0484
Máximo	2.8780
Varianza	1.0355
Desviación estándar	1.0176
1° Cuartil	-0.6343
Mediana	0.0200
3° Cuartil	0.6842

Tabla 8.2: Modelo elegido

Modelo	Anamorfosis pluviómetro (mm)			
	nugget	meseta	Alcance	Suma de cuadrados
Esférico	0	1.1114	13.8189	7.7138

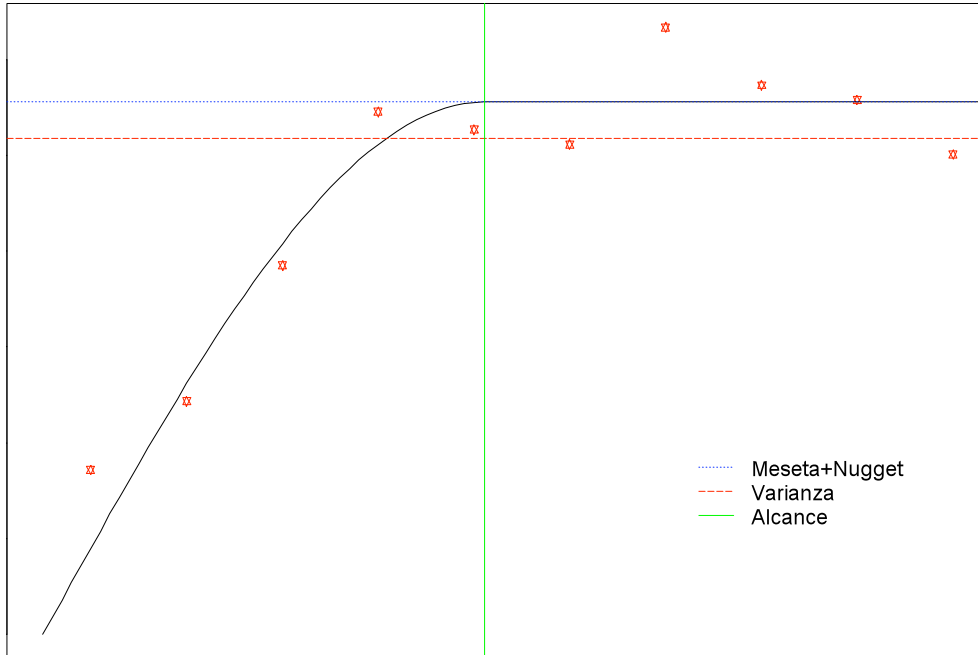


Figura 8.1: Semivariograma y modelo ajustado a la transformación anamorfosis de pluviómetro

Como en los anteriores Capítulos la validación de modelo se realizó utilizando el procedimiento validación cruzada (leave one out), de la validación se utilizaron errores estandarizados, de los cuales se calcularon la media y el varianza.

Media de los errores estandarizados = -0.03477282

Varianza de los errores estandarizados= 1.020995

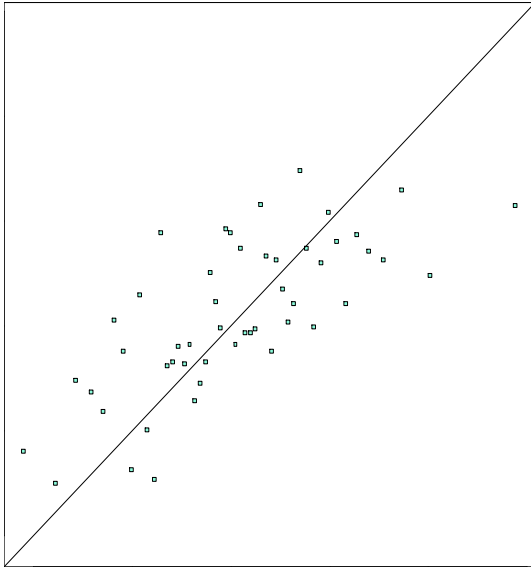


Figura 8.2: Gráficos de la validación para anamorfosis de pluviómetro

Primeramente cabe recalcar que la principal diferencia del análisis presentado en el Capítulo 6 con el presentado en este Capítulo, radica en que en este caso la distribución de los datos transformados si es normal, en lo que se refiere a los modelos obtenidos con la transformación logarítmica y con la transformación de anamorfosis Tabla 6.10 y 8.2 respectivamente, se puede decir que el cambio más significativo fue que en el modelo con la transformación de anamorfosis se elimino completamente el efecto nugget, lo que trascendió positivamente, esto se aprecia en la validación ya que los valores de la media y varianza de los errores estandarizados obtenidos en este Capítulo son mas cercanos a los ideales (media cero y varianza uno) a diferencia del Capítulo 6 en donde están mas alejados.

8.2 Simulaciones Condicionadas al Kriging Ordinario

Se obtuvieron cien simulaciones secuenciales Gaussianas condicionadas al Kriging Ordinario, a continuación se presentan las estadísticas básicas y la imagen de una simulación, de la media de 10 simulaciones y de la media de cien simulaciones.

Tabla 8.3: Estadísticas de las simulaciones secuenciales Gaussianas

Estadístico	Una Simulación	Media de diez Simulaciones	Media de cien Simulaciones
Numero de Observaciones	1892	1892	1892
Distancia Mínima (Km.)	1.0000	1.0000	1.0000
Distancia Máxima (Km.)	60.1082	60.1082	60.1082
Mínimo	0.250	0.250	0.250
Media	1.454	1.138	1.042
Máximo	7.750	7.750	7.750
Varianza	2.6507	0.9910	0.8132
Desviación estándar	1.6281	0.9955	0.9017
1° Cuartil	0.250	0.500	0.500
Mediana	1.000	1.000	0.750
3° Cuartil	2.000	1.500	1.250

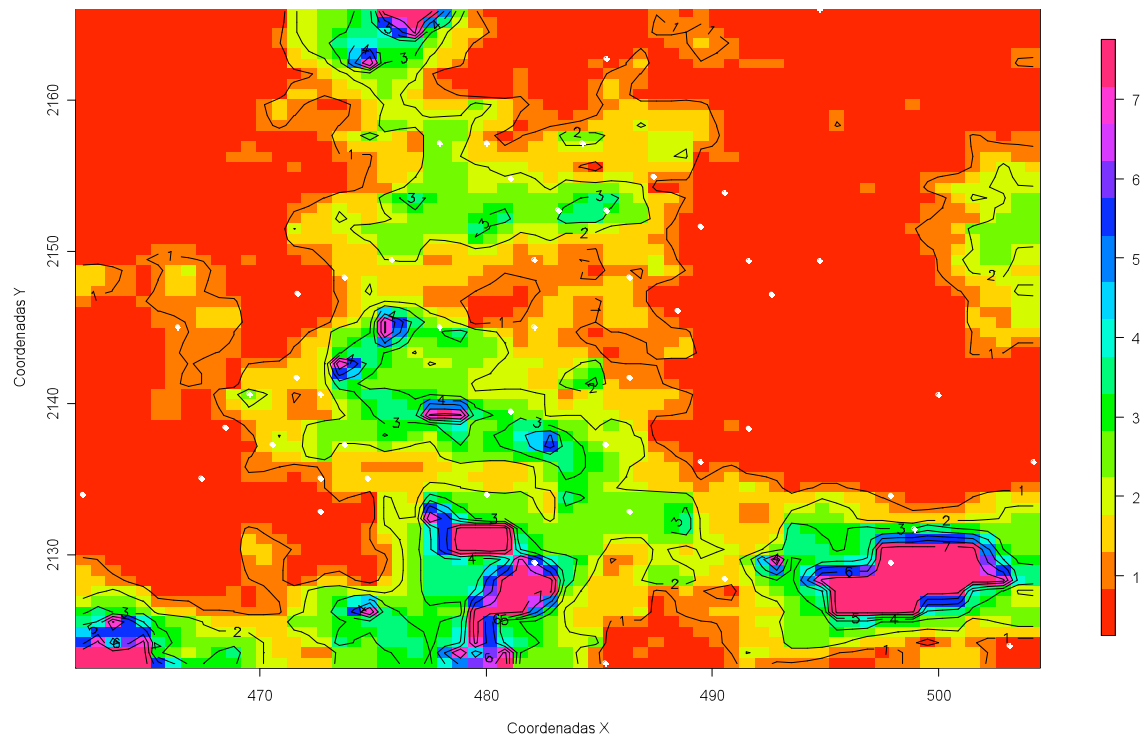


Figura 8.3: Imagen de una simulación

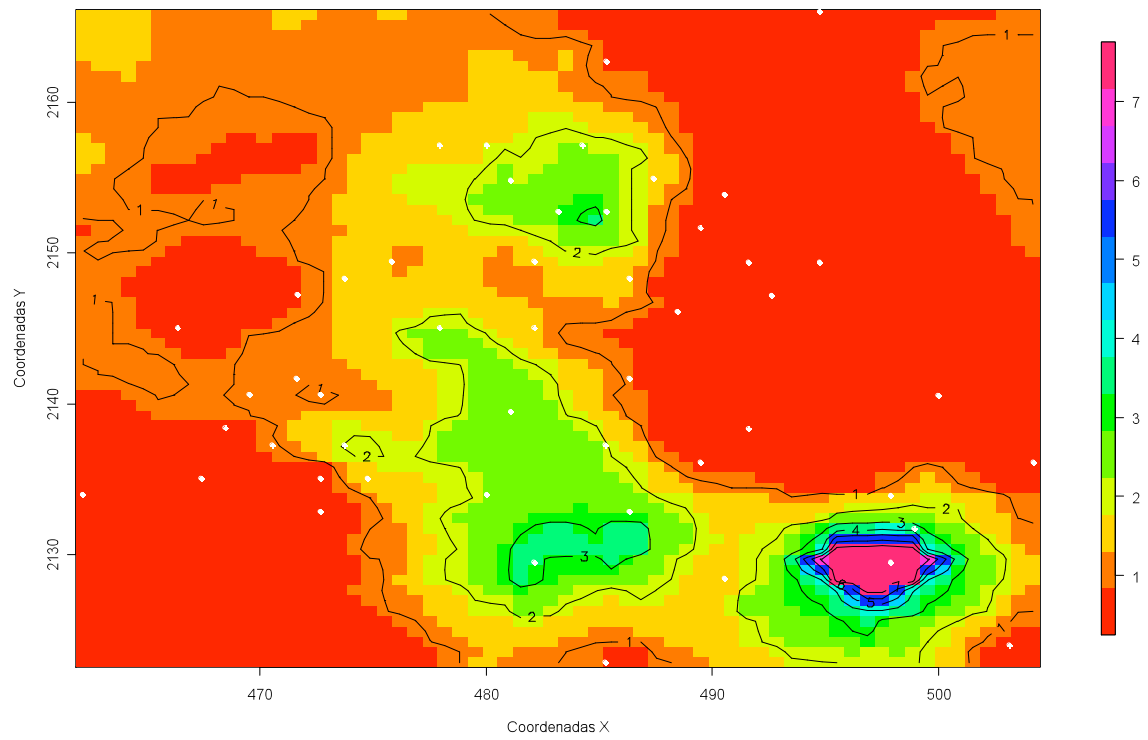


Figura 8.4: Imagen de la media de diez simulaciones

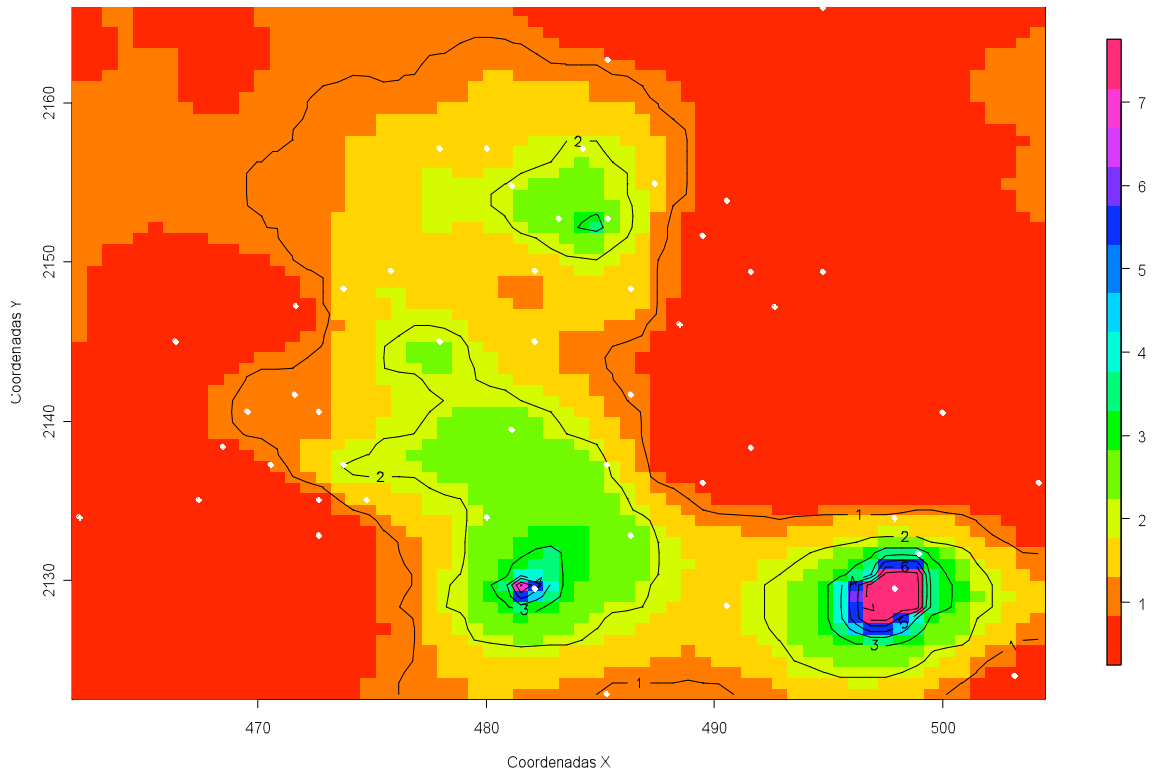


Figura 8.5: Imagen de la media de cien simulaciones

8.3 Comparación de las Simulaciones con los Valores Originales

Para observar si las simulaciones están cumpliendo con su objetivo, se compararan las simulaciones con respecto a los datos originales de pluviómetros (Datos Originales), esto se realizara mediante la observación del histograma, grafico de caja y semivariograma.

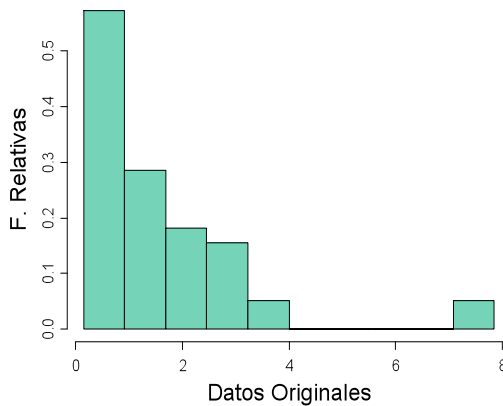


Figura 8.6: Histogramas de los datos originales y simulaciones

Es la Figura 8.6, 8.7, 8.8 y Tabla 8.4 la comparación importante es entre los datos originales, con lo de una simulación, que como se observa en el grafico inmediato anterior el histograma de una simulación esta reproduciendo en buena manera el histograma de los datos originales.

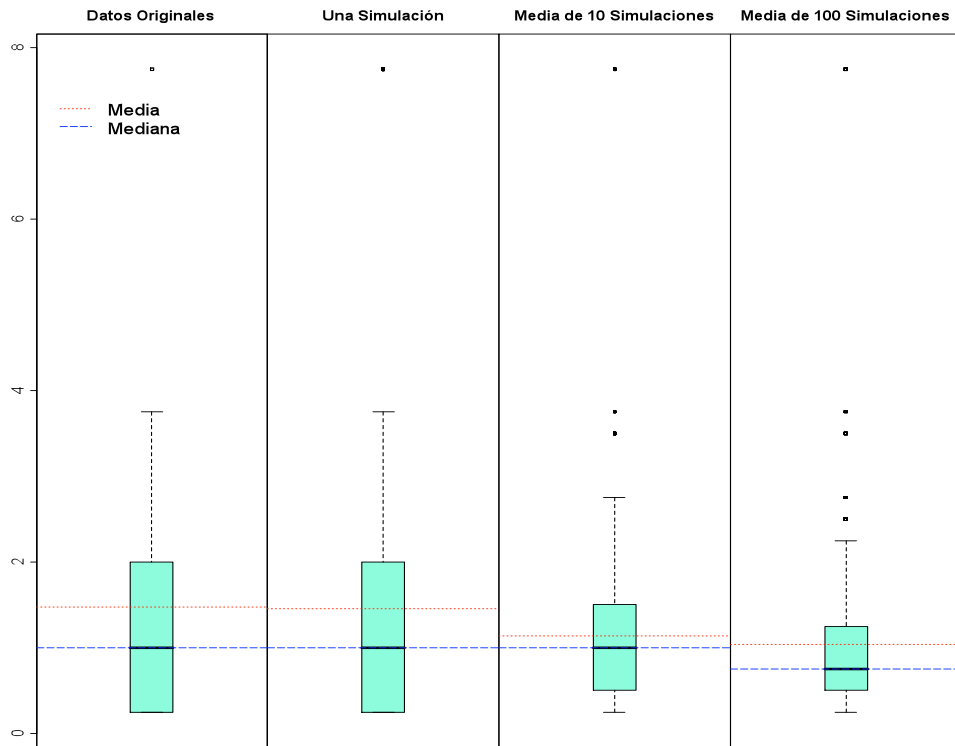


Figura 8.7: Gráficos de caja de los datos originales y simulaciones

Es la grafica de caja se aprecia que una simulación reproduce casi en su totalidad la grafica de caja de los datos de pluviómetro, así mismo se observa en la Tabla 8.4 y Figura 8.8, que una simulación también esta reproduciendo el histograma de la información original de pluviómetros.

Tabla 8.4: Modelos

Modelos	Variable	nugget	meseta	Alcance
Esférico	Datos Originales	0	1.1114	13.8189
Esférico	Una Simulación	0	1.2571	14.1037
Esférico	Media de 10 Simulaciones	0	0.7856	19.2791
Esférico	Media de 100 Simulaciones	0	0.7430	18.2406

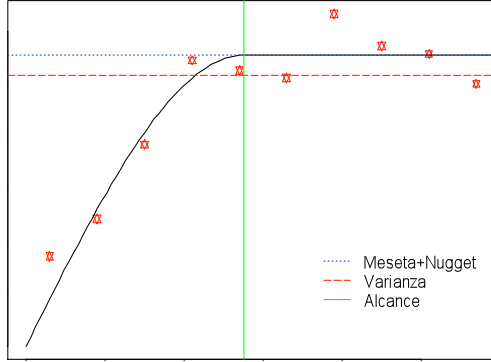


Figura 8.8: Semivariogramas de los datos originales y simulaciones

9. Simulación Conjunta Usando Datos de Pluviómetro e Imagen de Radar

Para realizar las simulaciones secuenciales Gaussianas conjuntas usando Cokriging Ordinario y la otra usando Cokriging Colocado, como ya se mencionó en el Capítulo 4 se requiere que las variables a utilizar tengan distribución normal y que además estén correlacionadas.

9.1 Análisis Geoestadístico Bivariado Usando la Transformación de Anamorfosis

Para realizar la cosimulación se utilizaran las variables transformación anamorfosis de pluviómetro y transformación logarítmica de radar, ya que como se mostró en Capítulo 7 y 8 respectivamente, estas variables presentan normalidad, ahora se calculara el coeficiente de Pearson, para medir el grado de correlación.

Tabla 9.1: Coeficiente de Pearson

Variables	Logaritmo de radar (Lnradar)	Anamorfosis de pluviómetro
Lnradar	1.0000	0.8067
Anamorfosis de pluviómetro	0.8067	1.0000

Claramente se aprecia con el coeficiente de Pearson que existe una correlación lineal elevada entre las variables. Por lo que se realizara un análisis similar al presentado en el Capítulo 7, pero al igual que en Capitulo 8, aquí solo se presentaran los resultados mas relevantes.

Tabla 9.2: Modelos elegidos para el análisis bivariado

Variable	Modelo	nugget	meseta	Rango
Anamorfosis de pluviómetro	Esférico	0.16	0.98	19
Lnradar	Esférico	0.08	0.81	19
Anamorfosis de pluviómetro.Lnradar	Esférico	0.05	0.79	19

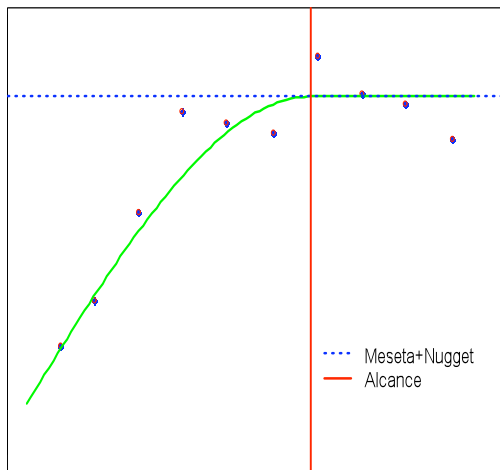


Figura 9.1: Variogramas adireccionales y modelo ajustado

Con el modelo de semivariograma cruzado, se realizó la validación por medio del procedimiento validación cruzada (leave one out), utilizando los errores estandarizados.

Media de los errores estandarizados= -0.003836478

Varianza de los errores estandarizados = 0.9520466

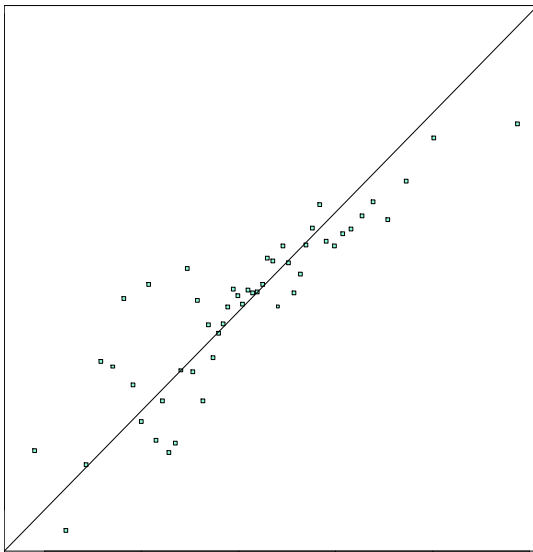


Figura 9.2: Gráficos de la validación del modelo de semivariograma cruzado

Los modelos obtenidos en el presente Capítulo, presentan un efecto nugget mas elevado que los obtenidos en el Capítulo 7, pero al igual que en el Capítulo 8, esto tiene un efecto positivo, ya que al realizar la validación de los modelos para la transformación de anamorfosis, los valores para los errores estandarizados tienen mas cercanía a los valores ideales, que los obtenidos en el Capítulo 7.

La imagen de radar que a continuación se presentara es la correspondiente a la tormenta en estudio, en el Capítulo 5 se describió mas ampliamente como fueron obtenidos los datos para la realización de la imagen.

Tabla 9.3: Estadísticas básicas de la imagen de radar

Estadístico	Radar (mm)
Numero de Observaciones	1638
Distancia Mínima (Km.)	1.0000
Distancia Máxima (Km.)	55.9017
Mínimo	0.0001
Media	2.0510
Máximo	15.5000
Varianza	4.1547
Desviación estándar	2.03832
1° Cuartil	0.5000
Mediana	1.5000
3° Cuartil	2.9000

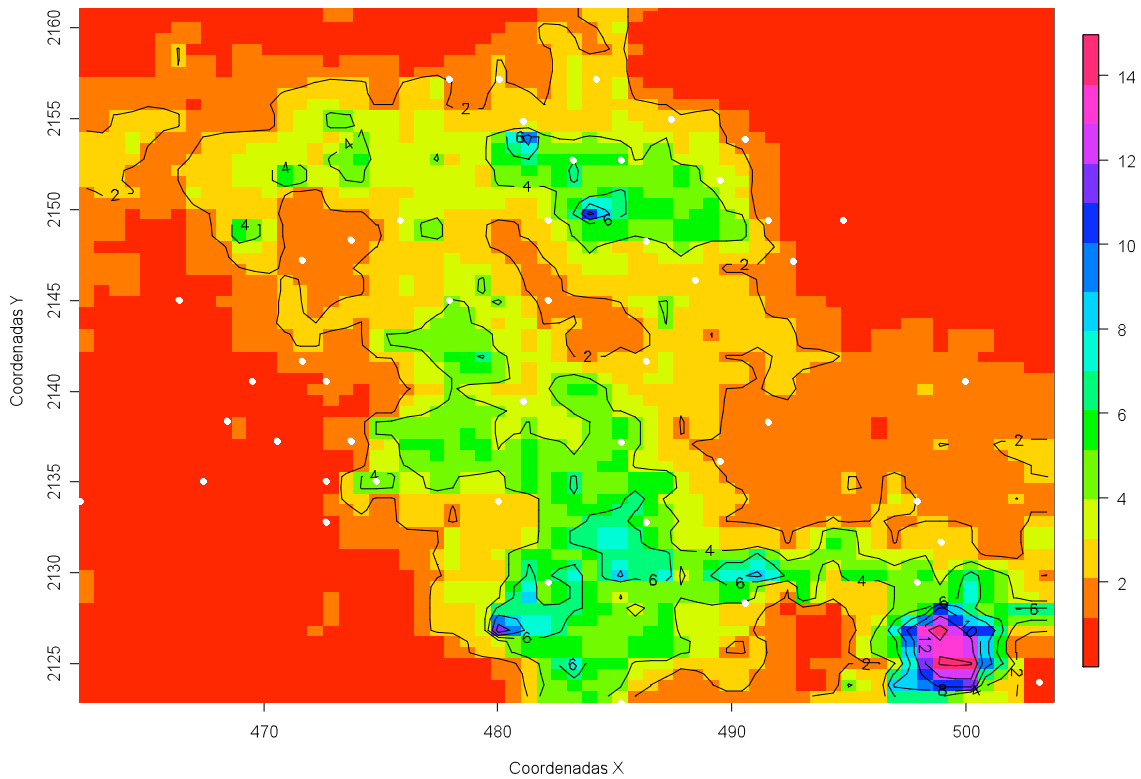


Figura 9.3: Imagen de radar

9.2 Cosimulaciones Condicionadas al Cokriging Ordinario

Se obtuvieron cien simulaciones secuenciales Gaussianas conjuntas usando Cokriging Ordinario, a continuación se presentan las estadísticas básicas y la imagen de una cosimulación, de la media de diez cosimulaciones y de la media de cien cosimulaciones.

Tabla 9.4: Estadísticas de la primera cosimulación condicionada al Cokriging Ordinario

Estadístico	Una Cosimulación	Media de diez Cosimulaciones	Media de cien Cosimulaciones
Numero de Observaciones	1892	1892	1892
Distancia Mínima (Km.)	1.0000	1.0000	1.0000
Distancia Máxima (Km.)	60.1082	60.1082	60.1082
Mínimo	0.250	0.250	0.250
Media	1.351	1.043	1.0.16
Máximo	7.750	7.750	7.750
Varianza	2.3703	1.1011	0.9592
Desviación estándar	1.5396	1.0493	0.9793
1° Cuartil	0.250	0.250	0.250
Mediana	0.750	0.500	0.500
3° Cuartil	1.750	1.750	1.500

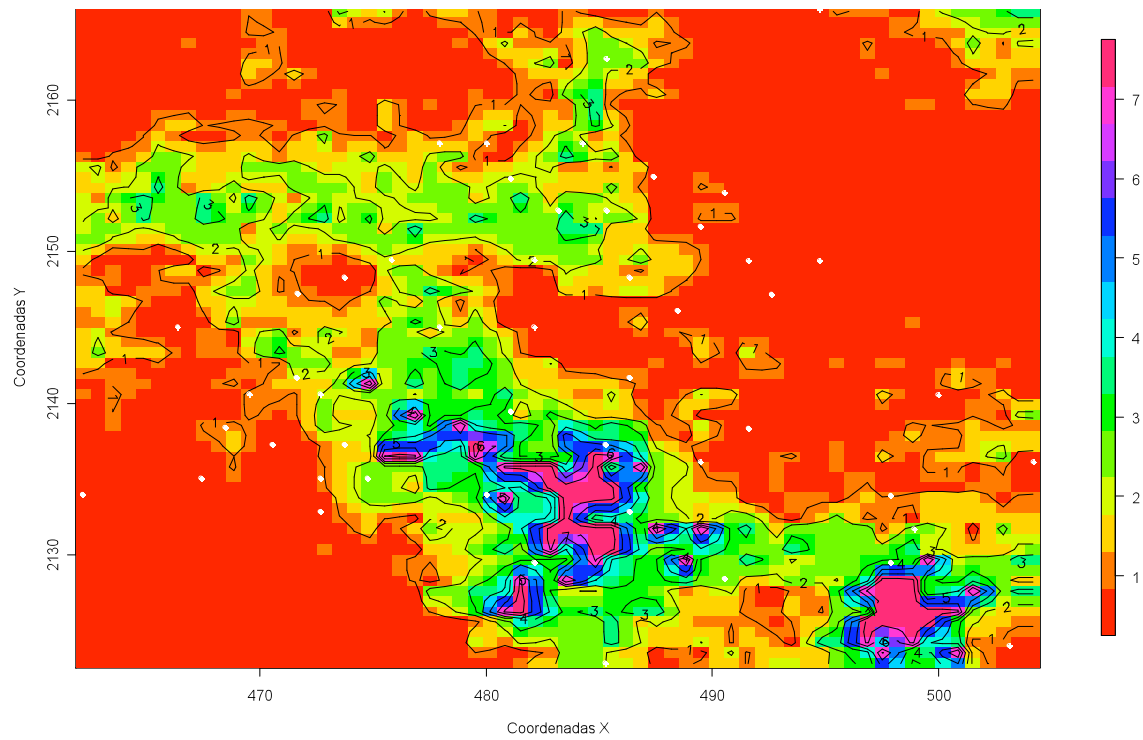


Figura 9.4: Imagen de una cosimulación usando al Cokriging Ordinario

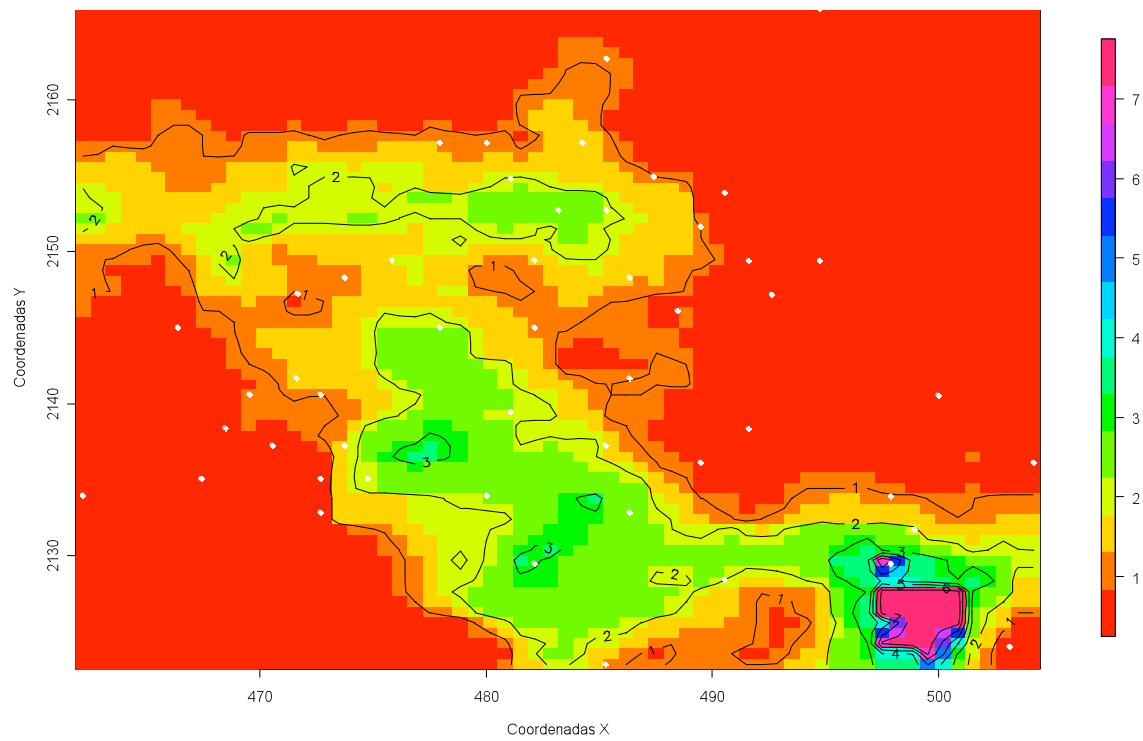


Figura 9.5: Imagen de la media de diez cosimulaciones usando al Cokriging Ordinario

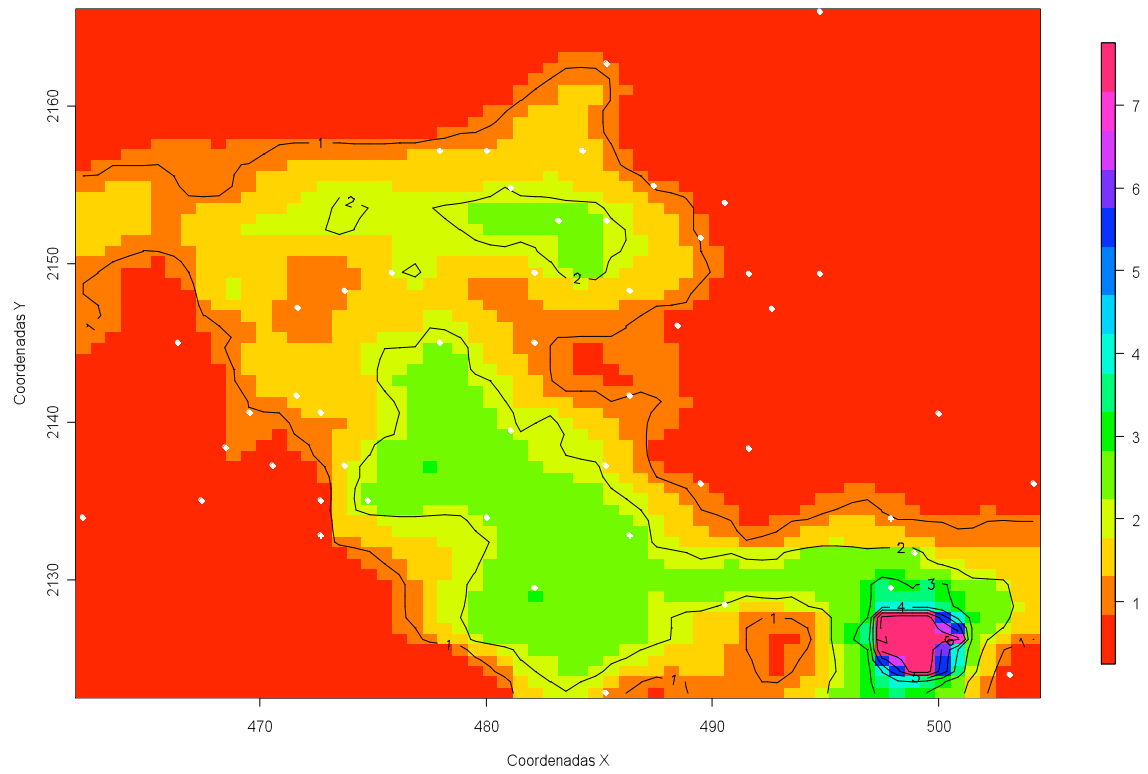


Figura 9.6: Imagen de la media de cien cosimulaciones usando al Cokriging Ordinario

9.3 Cosimulaciones Condicionadas al Cokriging Colocado

Al igual que para el caso anterior también se obtuvieron cien simulaciones secuenciales Gaussianas conjuntas solo que en este caso usando Cokriging Colocado, en este caso también se presentan las estadísticas básicas y la imagen de una cosimulación, de la media de diez cosimulaciones y de la media de cien cosimulaciones.

Tabla 9.5: Estadísticas de la primera cosimulación condicionada al Cokriging colocado

Estadístico	Una Cosimulación	Media de diez Cosimulaciones	Media de cien Cosimulaciones
Numero de Observaciones	1892	1892	1892
Distancia Mínima (Km.)	1.0000	1.0000	1.0000
Distancia Máxima (Km.)	60.1082	60.1082	60.1082
Mínimo	0.250	0.250	0.250
Media	1.419	1.111	1.037
Máximo	7.750	7.750	7.750
Varianza	2.4688	0.9500	0.7640
Desviación estándar	1.5712	0.9747	0.8741
1° Cuartil	0.250	0.250	0.250
Mediana	1.000	1.000	0.750
3° Cuartil	2.250	1.750	1.750

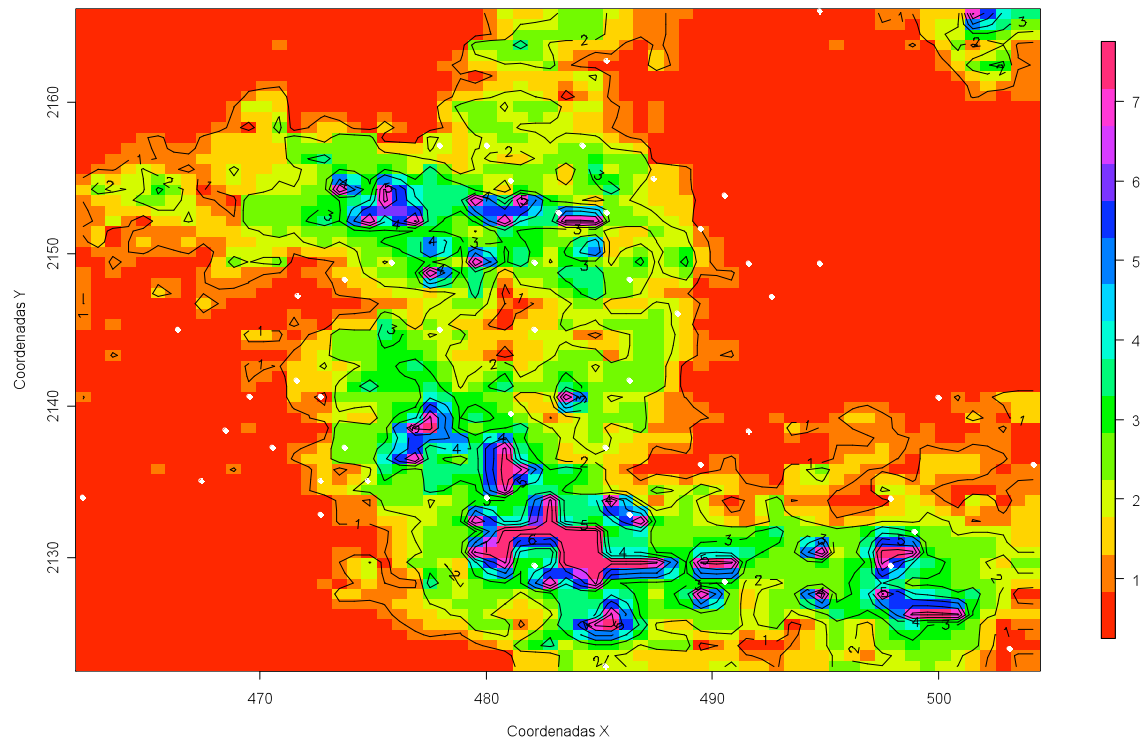


Figura 9.7: Imagen de una cosimulación usando al Cokriging colocado

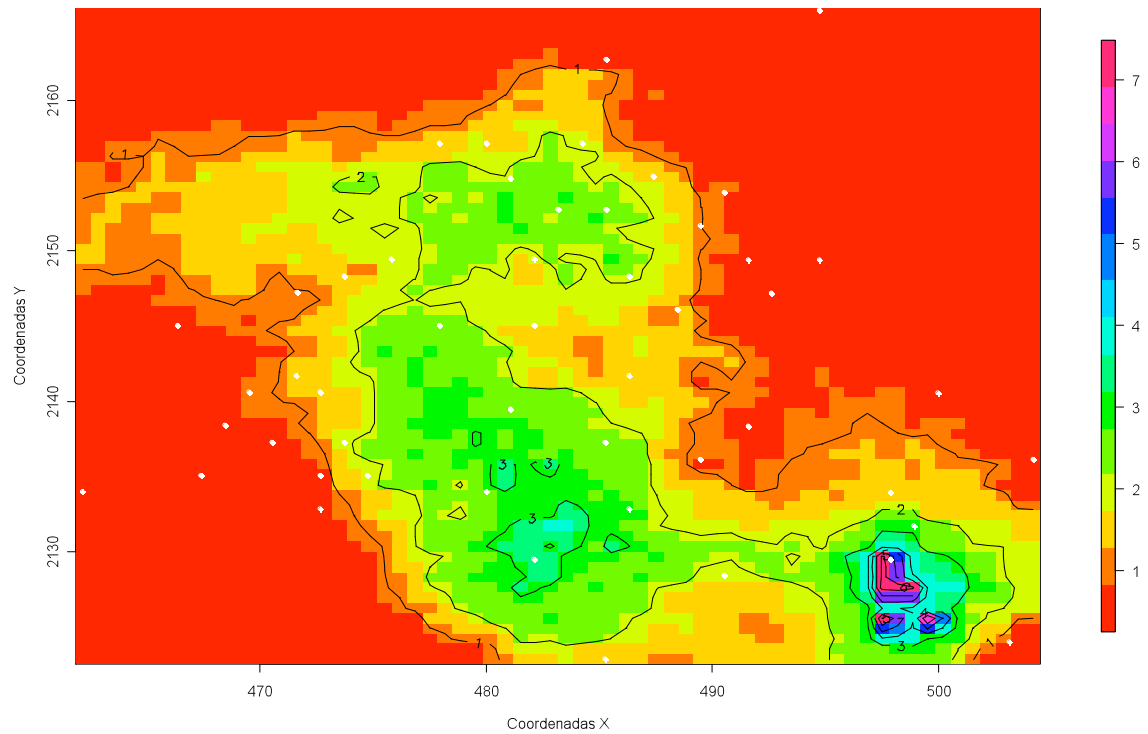


Figura 9.8: Imagen de la media de diez cosimulaciones usando al Cokriging Colocado

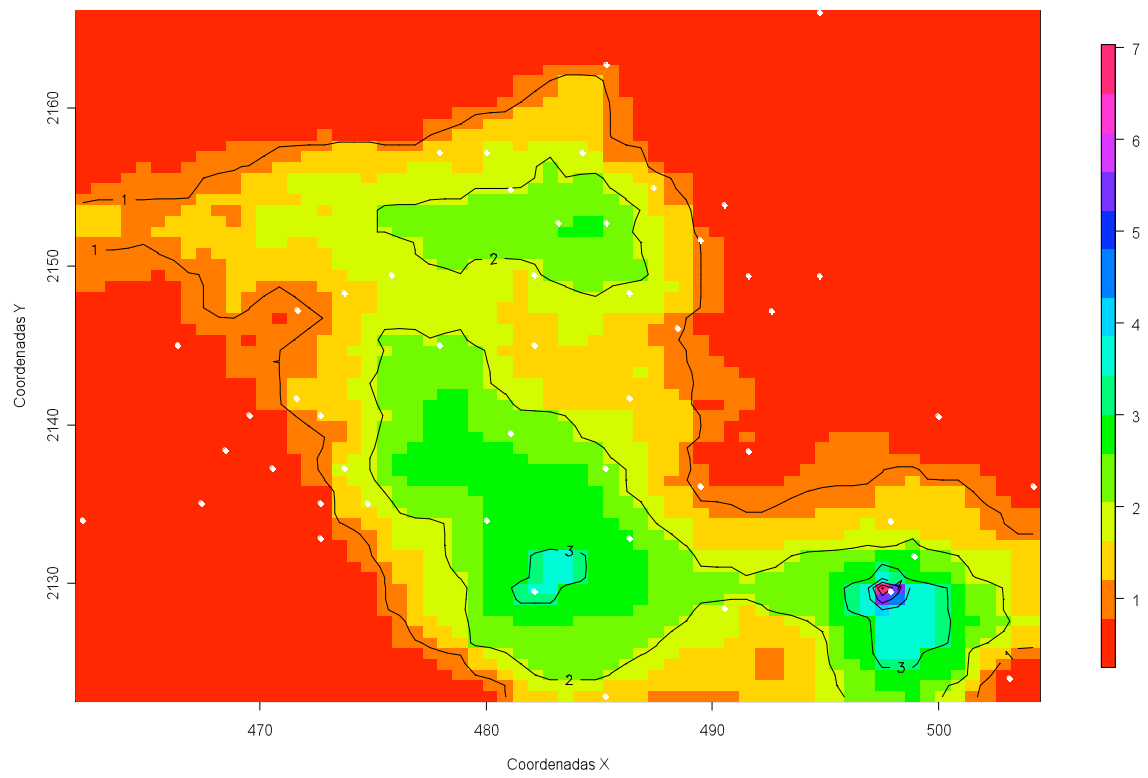


Figura 9.9: Imagen de la media de cien cosimulaciones usando al Cokriging Colocado

9.4 Comparación de las Cosimulaciones Ordinarias con los Valores Originales

Al igual que en el Capítulo 8, aquí también se presentaran los histogramas, gráficos de caja, modelos y semivariogramas con los modelos ajustados.

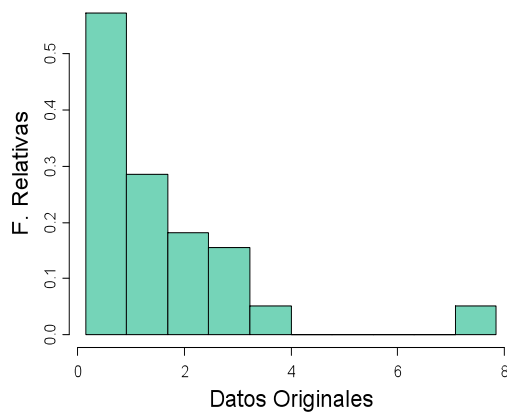


Figura 9.10: Histogramas de los datos originales y cosimulaciones ordinarias

Como se aprecia en los histogramas, una simulación reproduce en buena medida el histograma de los datos originales, mientras que al promediar las simulaciones tanto en el Capítulo anterior, como en este Capítulo, los valores se suavizan y se pierde información de los lugares con mayor precipitación.

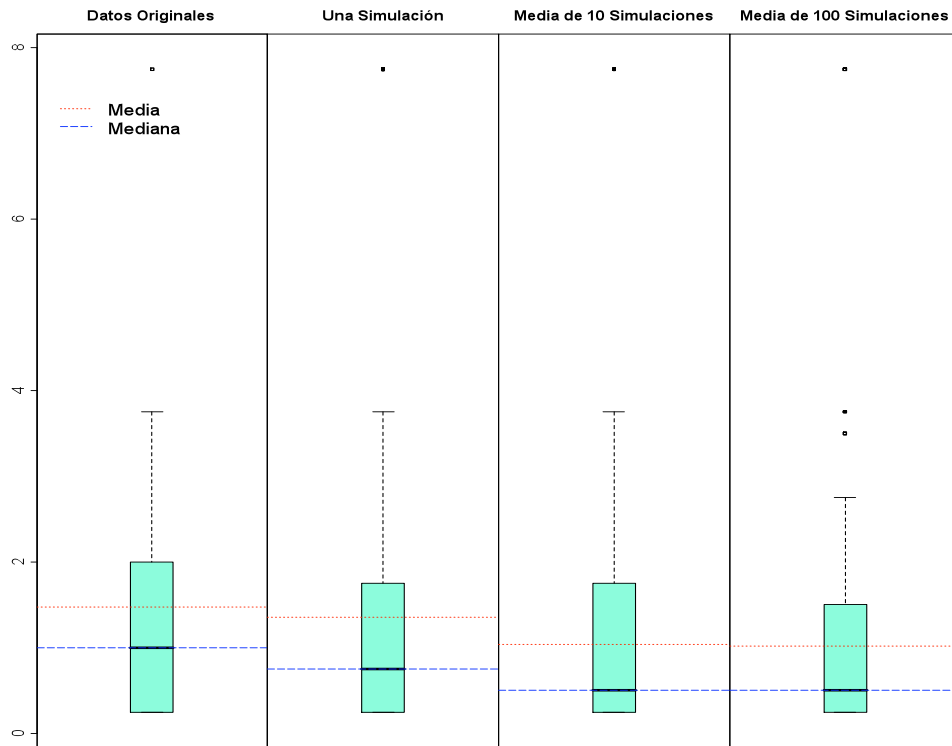


Figura 9.11: Gráficos de caja de los datos originales y cosimulaciones ordinarias

En este caso también una simulación reproduce el gráfico de caja de la información de pluviómetros, lo que se logra al incluir la información de la imagen de radar, es que una simulación reproduzca de una mejor manera los parámetros el variograma de los datos originales de pluviómetro.

Tabla 9.6: Modelos

Modelos	Variable	nugget	meseta	Alcance
Esférico	Datos Originales	0.16	0.98	19
Esférico	Una Simulación	0.05	1.1081	20.2828
Esférico	Media de 10 Simulaciones	0	1.1715	22.7875
Esférico	Media de 100 Simulaciones	0	1.1817	22.8491

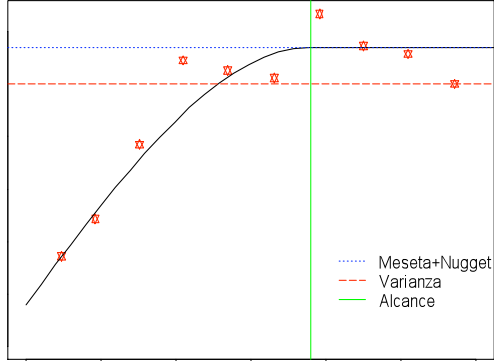


Figura 9.12: Semivariogramas de los datos originales y cosimulaciones ordinarias

9.5 Comparación de las Cosimulaciones Colocadas con los Valores Originales

Al igual que en las comparaciones anteriores, para este caso se mostraran los histogramas, gráficos de caja, modelos y semivariogramas con los modelos ajustados.

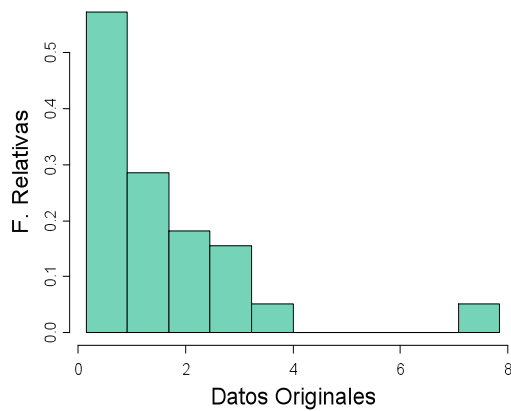


Figura 9.13: Histogramas de los datos originales y cosimulaciones colocadas

Para este caso los histogramas de los promedios de las simulaciones reproducen mejor el histograma de los datos originales, aunque estos pierden información de los lugares con mayor precipitación.

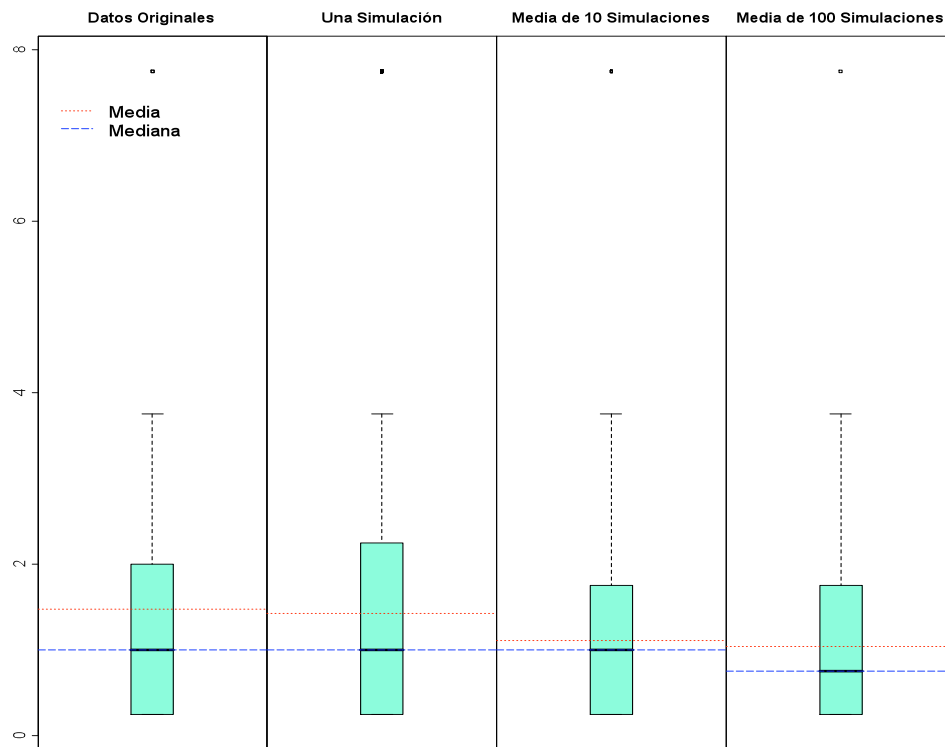


Figura 9.14: Gráficos de caja de los datos originales y cosimulaciones colocadas

La grafica de caja es bien reproducida por una simulación, de igual manera una simulación reproduce los parámetros del variograma de la información original.

Tabla 9.7: Modelos

Modelos	Variable	nugget	meseta	Alcance
Esférico	Datos Originales	0.16	0.98	19
Esférico	Una Simulación	0.05	1.1274	20.8477
Esférico	Media de 10 Simulaciones	0	1.2167	23.8754
Esférico	Media de 100 Simulaciones	0	1.2203	23.8325

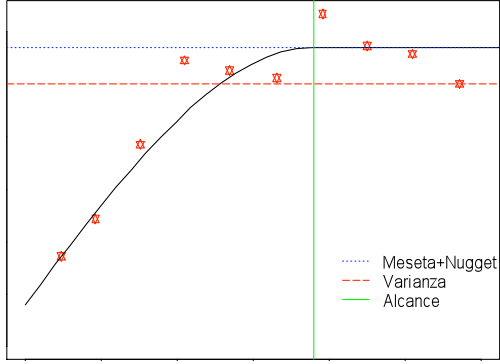


Figura 9.15: Semivariogramas de los datos originales y cosimulaciones colocadas

10. Discusión y Conclusiones

10.1 Discusión

El propósito del presente trabajo es la estimación de la distribución de la precipitación acumulada en una hora a nivel del suelo para la tormenta ocurrida el día 15 de julio de 1997 en el valle de la Ciudad de México usando las contribuciones de las mediciones en una red de pluviómetros y una imagen de radar meteorológico.

Al aplicar el método Kriging en el Capítulo 6 usando exclusivamente los datos de los pluviómetros se puede observar en el mapa de las distribuciones espaciales del error de la estimación (Fig. 6.15) que los mayores errores se concentran en las zonas pobremente muestreadas, es decir, donde existen pocos puntos de medición. Si se compara la imagen de la distribución de la precipitación obtenida usando Kriging (Fig. 6.14) con la de radar (Fig. 9.3) se puede constatar que la primera está mucho más suave respecto a la segunda, esto se debe a la cantidad de pluviómetros (50) con los que se hizo la estimación. Además, se observa que las ubicaciones de los valores más altos en la imagen del Kriging están desplazadas en la dirección noroeste con respecto a la imagen de radar. Esto se debe fundamentalmente a la influencia del viento, puesto que la imagen de radar se registra a una altura de alrededor de 2000 m por encima del nivel del terreno, e implica que el viento soplabá predominantemente del sureste hacia el noroeste durante la tormenta que se analiza.

Si aplicamos de manera conjunta a los datos de pluviómetros y a la imagen de radar el método de Cokriging Ordinario usando un modelo de correogionalización lineal (Figura 7.6 y Tabla 7.8) se producen estimaciones que reflejan con un mayor detalle la variabilidad espacial existente en la imagen de radar y al mismo tiempo los valores estimados se comportan dentro del rango de las mediciones de los pluviómetros.

Si comparamos los valores de la desviación estándar del error de la estimación utilizando el Cokriging Ordinario (Figura 7.6) con los obtenidos por Kriging (Figura 6.14), vemos que dentro de la región con valores de pluviómetros el primero es menor (0.55) respecto al segundo (0.7), y los valores medios de la desviación estándar del error de la estimación son 0.5576 y 0.7407, respectivamente (ver Tablas 7.8 y 6.11). Por lo que se puede afirmar que la estimación usando Cokriging Ordinario es más precisa que la obtenida con Kriging.

El Cokriging Ordinario y el Cokriging Colocado a diferencia del Kriging Ordinario ofrecen estimaciones más precisas en las regiones donde los valores de pluviómetro son pobres, lo anterior se ve reflejado claramente en la varianza, ya que la varianza de los métodos de Cokriging es menor que la que presenta el Kriging (Tablas 6.11 y 7.8).

El lo referente a las Simulaciones Secuenciales Gaussianas tanto usando Kriging como Cokriging, se puede decir que estas reproducen adecuadamente los primeros momentos de la función aleatoria, los valores simulados respetan las zonas con los valores altos de las mediciones de los pluviómetros además de comportarse dentro del rango de los mismos, por consiguiente también reproducen de buena manera el histograma y el gráfico de caja de los datos originales de pluviómetro, cabe señalar que la principal diferencia entre las simulaciones y las cosimulaciones, es que las segundas reproducen de mejor manera el variograma de los datos de pluviómetro.

Resulta conveniente además comparar las estimaciones y las simulaciones respecto a los valores de precipitación en pluviómetros ya que son nuestros datos duros por ser ésta la variable primaria. Para esto se presentarán a continuación un resumen comparativo en forma gráfica y de tablas tanto para el caso univariado así como para el bivariado.

Tabla 10.1: Estadísticas de las estimación y simulación univariada

Estadístico	Datos Pluviómetros	Kriging Ordinario	Una Simulación Secuencial Gaussiana
Numero de Observaciones	50	1892	1892
Distancia Mínima (Km.)	1.5258	1.0000	1.0000
Distancia Máxima (Km.)	45.6961	60.1082	60.1082
Mínimo	0.2500	0.2126	0.250
Media	1.4700	1.0250	1.454
Máximo	7.7500	6.2120	7.750
Varianza	2.5756	0.6809	2.6507
Desviación estándar	1.6048	0.8252	1.6281
1° Cuartil	0.3125	0.4565	0.250
Mediana	1.0000	0.7458	1.000
3° Cuartil	1.9380	1.3490	2.000

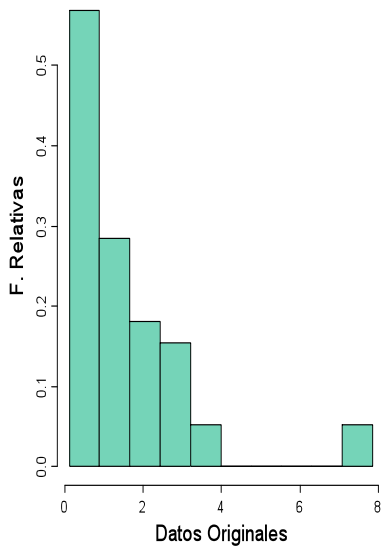


Figura 10.1: Histogramas de los resultados univariados

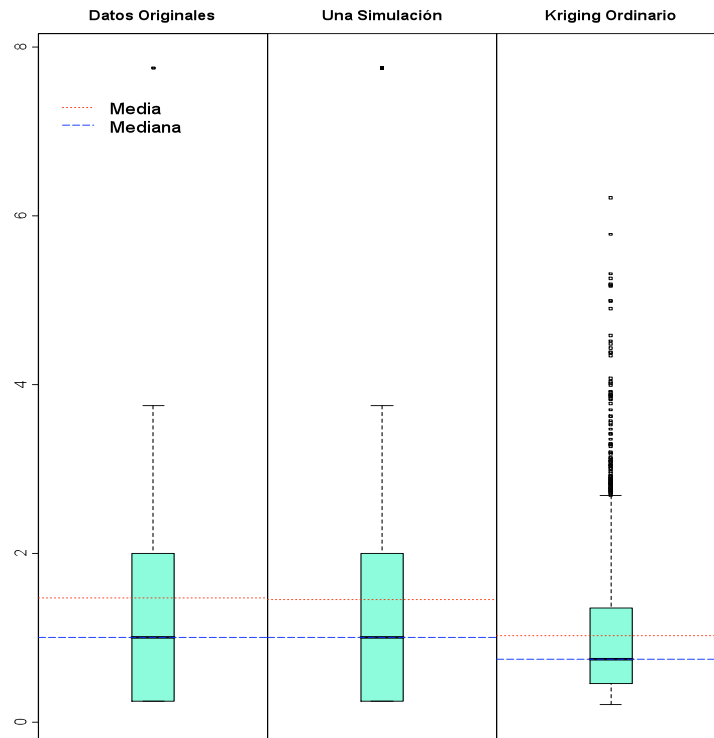


Figura 10.2: Gráficos de caja de los resultados univariados

Tabla 10.2: Modelos para los semivariogramas

Modelos	Variable	nugget	meseta	Alcance
Esférico	Datos Originales	0	1.1114	13.8189
Esférico	Una Simulación	0	1.2571	14.1037
Esférico	Kriging Ordinario	0	1.2197	17.2645

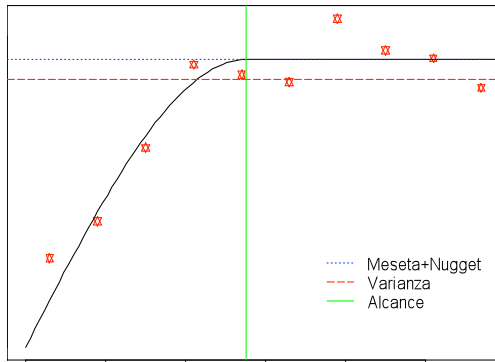


Figura 10.3: Semivariogramas de los resultados univariados

Tabla 10.3: Estadísticas de los resultados bivariados usando Cokriging Ordinario

Estadístico	Datos Pluviómetros	Cokriging Ordinario	Una Cosimulación
Numero de Observaciones	50	1892	1892
Distancia Mínima (Km.)	1.5258	1.0000	1.0000
Distancia Máxima (Km.)	45.6961	60.1082	60.1082
Mínimo	0.2500	0.000071	0.250
Media	1.4700	0.9979	1.351
Máximo	7.7500	9.8540	7.750
Varianza	2.5756	1.1759	2.3703
Desviación estándar	1.6048	1.0844	1.5396
1° Cuartil	0.3125	0.2547	0.250
Mediana	1.0000	0.6057	0.750
3° Cuartil	1.9380	1.4410	1.750

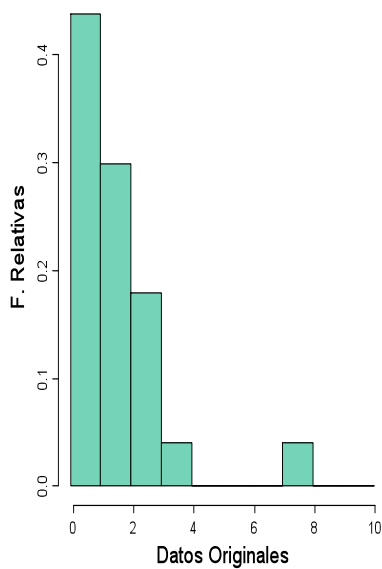


Figura 10.4: Histogramas de los resultados bivariados usando Cokriging Ordinario

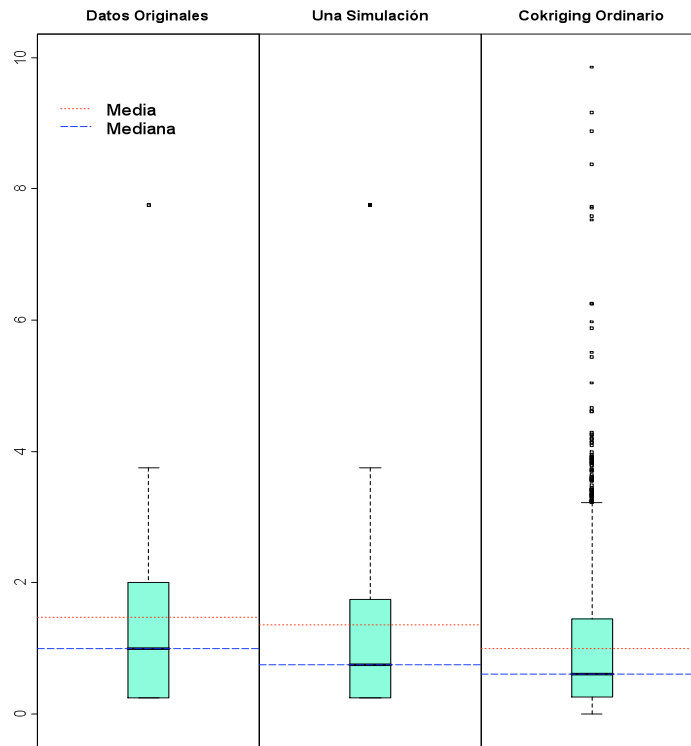


Figura 10.5: Gráficos de caja de los resultados bivariados usando Cokriging Ordinario

Tabla 10.4: Modelos para los resultados bivariados usando Cokriging Ordinario

Modelos	Variable	nugget	meseta	Alcance
Esférico	Datos Originales	0.16	0.98	19
Esférico	Una Simulación	0.05	1.1081	20.2828
Esférico	Cokriging Ordinario	0	1.2408	26.0889

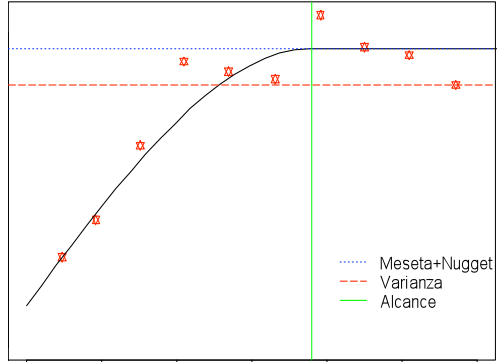


Figura 10.6: Semivariogramas de los resultados bivariados usando Cokriging Ordinario

Tabla 10.5: Estadísticas de los resultados bivariados usando Cokriging Colocado

Estadístico	Datos Pluviómetros	Cokriging Colocado	Una Cosimulación
Numero de Observaciones	50	1892	1892
Distancia Mínima (Km.)	1.5258	1.0000	1.0000
Distancia Máxima (Km.)	45.6961	60.1082	60.1082
Mínimo	0.2500	0.1855	0.250
Media	1.4700	0.9473	1.419
Máximo	7.7500	4.6450	7.750
Varianza	2.5756	0.5602	2.4688
Desviación estándar	1.6048	0.7485	1.5712
1° Cuartil	0.3125	0.3784	0.250
Mediana	1.0000	0.6259	1.000
3° Cuartil	1.9380	1.3790	2.250

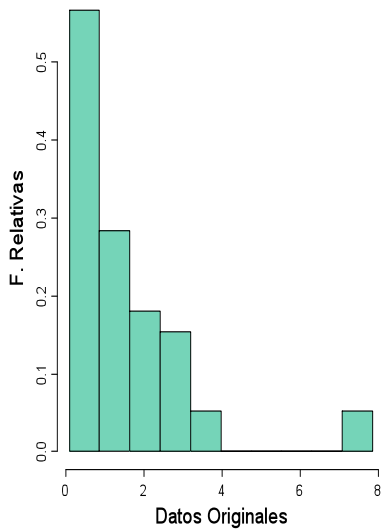


Figura 10.7: Histogramas de los resultados bivariados usando Cokriging Colocado

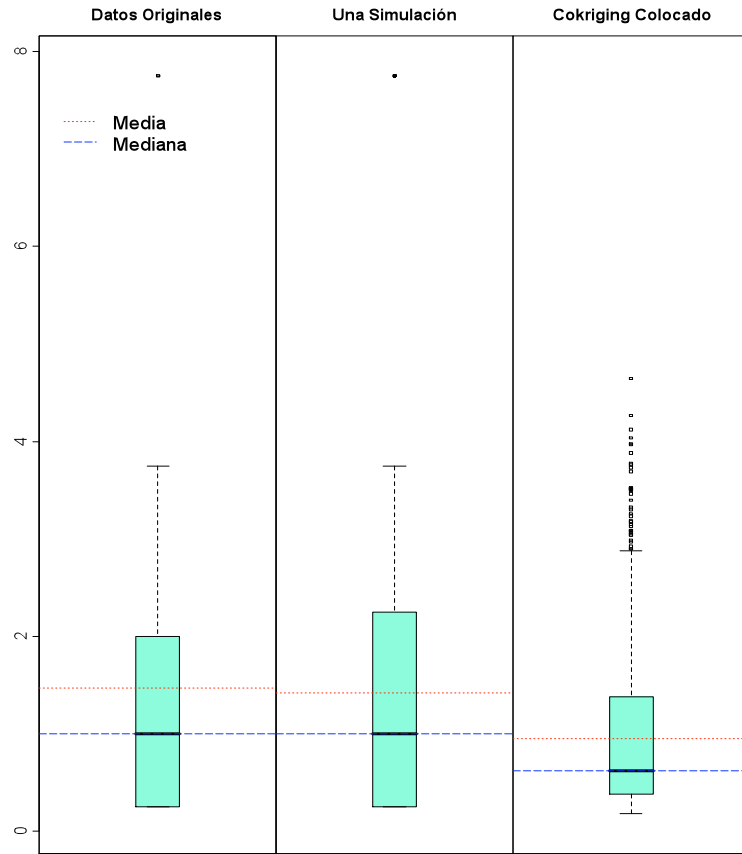


Figura 10.8: Gráficos de caja de los resultados bivariados usando Cokriging Colocado

Tabla 10.6: Modelos para resultados bivariados usando Cokriging Colocado

Modelos	Variable	nugget	meseta	Alcance
Esférico	Datos Originales	0.16	0.98	19
Esférico	Una Simulación	0.05	1.1274	20.8477
Esférico	Cokriging Colocado	0	1.2689	20.1029

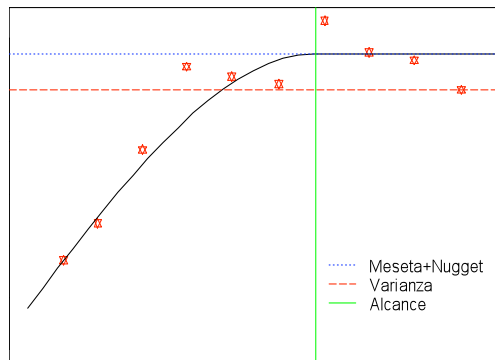


Figura 10.9: Semivariogramas de los resultados bivariados usando Cokriging Colocado

Observando de la Tabla 10.1 a 10.6 y de la Figura 10.1 a 10.9, se comprueba que las simulaciones reproducen de manera adecuada la estadística de los datos de los pluviómetros en términos de mínimo, máximo, valor medio, varianza, histograma y variograma, mientras que las estimaciones no. Por lo que la simulación sería la herramienta ideal para modelar el fenómeno bajo estudio ya que reproduce mucho mejor el comportamiento del mismo.

10.2 Conclusiones

Cuando se tienen redes de monitoreo suficientemente densas como es el caso de la Ciudad de México, el uso del Kriging Ordinario produce estimaciones bastante precisas de la distribución espacial de la precipitación en sentido general. Sin embargo el Cokriging Ordinario con un modelo de correogionalización lineal usando mediciones de pluviómetros y radar produce estimaciones mucho más precisas y detalladas de la variabilidad espacial.

Los métodos Kriging y Cokriging Colocado son menos precisos que el Cokriging Ordinario pero son computacionalmente más eficientes, por lo que estos dos métodos podrían ser usados en sistemas de estimación en tiempo real aunque habría que tomar en cuenta el sesgo que introducen factores externos como lo es la presencia de viento.

Se pudo constatar que las simulaciones estocásticas espaciales usando el método Secuencial Gaussiano tanto para el caso univariado como para los bivariados, a diferencia de los métodos de estimación, reproducen de manera adecuada la estadística de los datos de los pluviómetros en términos de mínimo, máximo, valor medio, varianza, histograma y variograma, etc. Basados en este marco teórico se pudieran generar las distribuciones de la precipitación en tormentas artificiales si se especificaran las propiedades estadísticas que se esperarían en las mismas. Por lo que sería la herramienta ideal para estudiar la variabilidad espacial de dicho fenómeno.

Desde el punto de vista práctico, se debe remarcar que la aplicación de los procedimientos geoestadísticos deben realizarse apegados a las metodologías y verificando los supuestos que los sustentan para que su desempeño sea cercano al óptimo. En caso contrario nos exponemos a obtener resultados que pudieran ser incorrectos o aberrantes.

Referencias

1. Akaike, H., 1977, An entropy maximization principle, Applications of Statistics, P.R. Krishnaiah, ed., North Holland, Amsterdam.
2. Anhert, P., W. Krajewski y E. Johnson, "Kalman filter estimation of radar-rainfall field bias". Memorias XXIII Conferencia de radar meteorológico, Amer. Meteor. Soc., 1986, pp. 33-37.
3. Armstrong, M. y P. Delfiner, "Towards a more robust variogram: case study on coal", Note n-671, CGMM, Fontainebleau, Francia, 1980.
4. Calheiros, R. and I. Zawadzki, "Reflectivity-rain rate relationships for radar hydrology in Brazil", J. Clim. Appl. Met, 26, 1987, pp.118-132.
5. Cassiraga, E. F. and J. J. Gómez-Herrández, "Improved Rainfall estimation by integration of radar data: A geostatistical approach". Soares et. al. (eds.), GeoENV I: Geostatistics for Environmental Applications, Kluwer Academic Publishers, 1997, pp. 363-374.
6. Chilès, P. J. and Delfiner P., 1999, Geoestatics: Modeling Spatial Uncertainty, Wiley, New York
7. Clark I. and W. V. Harper , 2000, "Practical Geostatistics 2000".
8. Cressie, N. y J. Hawkins, "Robust estimation of the variogram", Math. Geol. 12(2), 1980, pp. 115-126.
9. Cressie, N., 1985, Fitting variogram models using weighted least squares, J. Math. Geol., 17, 693-702.
10. Creutin, J. D., G. Delrieu and T. Lebel, "Rain measurement by raingage-radar combination: A geostatistical approach". J. Atmos. Oceanic Tecnol. 5(1), 1988, pp.102-115.
11. David M., 1988, "Handbook Of Applied Advanced Geostatistical Ore Reserve Estimation".
12. Deutsch C. V. and Journel A. G. 1998, Geostatistical Software Library and User's Guide, Oxford University Press.

13. Diaz V. M., 2002, *Geoestadística Aplicada*, Instituto de Geofísica, UNAM.
14. Diggle P. J. and Riveiro Jr., 2007, *Model-Based Geoestadistics*, Springer.
15. Fedorov, V.V., 1989, *Kriging and other estimators of spatial field characteristics*, Atmospheric Environment.
16. Goovaerts, P., *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York, 1997.
17. Hildebrand, P.H., Towery, N. y M.R. Snell, "Measurements of convective mean rainfall over small areas using high density raingages and radar", *J. Appl. Met.* 18, 1979, pp. 1316-1326.
18. Hohn M. E., 1998, "Geostatistics and Petroleum Geology", Second Edition, Kluwer Academic Publishers, 235 pp.
19. Isaaks, E. H. and R. M. Srivastava, 1989, "An Introduction to Applied Geostatistics", Oxford University Press.
20. Journel, A. G. y Ch. J. Huijbregts, 1978, "Mining Geostatistics", Academic Press, New York, 590 pp.
21. Journel, A. G., 1989, "Fundamentals of Geostatistics in Five Lessons", *Math. Geol.* 18(1), 119-140.
22. Kitanidis, P. K., 1997, "Introduction to Geostatistics: Applications in Hydrogeology", Cambridge University Press, 249 pp.
23. Krajewski, W.F., 1987, "Cokriging radar-rainfall and rain gage data", *J. Geophys. Res.*, 92, pp. 9571-9580.
24. Matheron, G., 1963, *Principles of Geostatistics*.
25. Omre, H., 1984, *The variogram and its estimation*, *Geostatistics for Natural Resources Characterization*, Verly et al. (Eds.), NATO ASI Series, Reidel, Hingham, Mass., Vol. 1, 107-125.
26. R Development Core Team. 2006. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
27. Ripley, B. D., 1987, *Stochastics Simulation*, Wiley, New York.

28. Rosenfeld, D., D.B. Wolff and E. Amitai, "The window probability matching method for rainfall measurements with radar", J. Appl. Met. 33, 1994, pp.682-693.
29. Samper, F.J. y J. Carrera, 1990, "Geoestadística, aplicaciones a la hidrogeología subterránea", Centro Internacional de Métodos Numéricos en Ingeniería, Barcelona, 484 pp.
30. Seo, D. J., W. F. Krajewski and D. S. Bowles, "Stochastic interpolation of rainfall data from rain gages and radar using Cokriging. 1. Design of experiments". Water Resources Research, 26(3), 1990a, pp. 469-477.
31. Seo, D. J., W. F. Krajewski, A. Azimi-Zonooz and D. S. Bowles, "Stochastic interpolation of rainfall data from rain gages and radar using Cokriging. 2. Results". Water Resources Research, 26(5), 1990b, pp. 915-924.
32. Wackernagel, H., Multivariate Geostatistics: An Introduction with Applications. 2nd ed. (completely revised), Springer, Berlin, 1998, 291 pp.

Anexos

A1. Datos de pluviómetro

UTM X en Km	UTM Y en Km	Pluviómetro (mm)	UTM X en Km	UTM Y en Km	Pluviómetro (mm)
485.303	2162.682	1.00	491.597	2138.323	0.25
484.253	2157.150	2.50	504.199	2136.108	0.25
487.403	2154.937	1.75	497.899	2133.895	1.00
490.552	2153.830	0.50	498.949	2131.682	3.75
489.502	2151.601	0.50	480.038	2133.906	2.00
480.054	2157.142	1.50	486.341	2132.793	2.75
481.104	2154.828	2.25	482.135	2129.478	7.75
483.203	2152.713	2.50	485.282	2122.836	0.75
475.855	2149.399	1.75	489.495	2136.111	0.50
486.353	2148.284	1.75	490.542	2128.365	1.25
482.155	2149.394	1.00	497.898	2129.469	7.75
492.651	2147.174	0.25	477.955	2157.144	1.75
494.751	2149.386	0.25	494.751	2165.983	0.25
477.955	2144.973	2.75	473.755	2148.298	1.50
472.694	2140.554	0.75	466.407	2144.991	0.50
473.739	2137.233	2.25	485.303	2152.711	3.50
481.095	2139.437	2.75	491.602	2149.387	0.25
482.154	2144.968	1.25	488.451	2146.070	0.50
486.348	2141.645	1.25	471.654	2147.195	0.25
471.645	2141.662	1.25	469.543	2140.559	1.00
468.489	2138.348	0.50	467.433	2135.030	0.25
462.178	2133.934	0.25	470.588	2137.238	0.50
474.787	2135.019	1.25	500.000	2140.533	0.25
472.682	2132.809	0.25	503.149	2123.937	0.25
485.294	2137.220	2.50	472.686	2135.022	0.25

A2. Programa en R

Realiza el cálculo de las estimaciones univariadas

```
require(sp)
require(geoR)
require(nortest)
require(sgeostat)
require(gstat)
require(outliers)
require(fields)
require(moments)
require(lawstat)
require(reshape)
require(distributions)
Originales<-as.data.frame(read.table("C:\\TESIS\\Tesis\\pluv.txt", header= TRUE))
Originalesgeo<-read.geodata("C:\\TESIS\\Tesis\\pluv.txt", header= TRUE, coords.col= 1:2,
data.col=3) # Carga los datos
summary(Originalesgeo)
var(Originalesgeo$data)
sd(Originalesgeo$data)
minhis<-min(Originalesgeo$data)
maxhis<-max(Originalesgeo$data)
difhis<-maxhis-minhis+0.2
numclases<-10
tamint<-difhis/numclases
valorhis<-0
vectorhis<-0
for (i in 1:numclases)
{
valorhis[i]<-minhis-0.1+i*tamint
vectorhis<-c(minhis-0.1, valorhis)
}
orden<-matrix(c(2,3,1),ncol=1,nrow=3,byrow=T)
```



```

div<-layout(orden, widths= 9, heights=c(2, 0.8, 3.5), TRUE)
layout.show(div)
par(mar=c(5,6,0,3))
plot(Originalesgeo$data, c(1:50), pch= 22, xlim=c(min(vectorhis),max(vectorhis)), col="black",
bg="aquamarine", ylab="Numero de Observación", xlab= "Valores de Pluviometro", cex.lab =
1.8, cex.axis=1.2)
abline(v = mean(Originalesgeo$data), lty = 3, col = "Red" )
abline(v = median(Originalesgeo$data), lty = 5, col= "Blue")
legend(5, 15, c("Media", "Mediana"), cex= 1.5, col= c("Red", "Blue"), lty=c(3, 5), bty="n")
par(mar=c(0,6,1,3))
histograma<-hist(Originalesgeo$data, breaks= vectorhis, xaxt="n", freq = FALSE, col=
"aquamarine3", ylab = "F. Relativas", cex.lab = 1.8, cex.axis=1.2, main = "")
abline(v = mean(Originalesgeo$data), lty = 3, col = "Red" )
abline(v = median(Originalesgeo$data), lty = 5, col= "Blue")
par(mar=c(0,6,0,3))
plot(0,0,type="n", xlim=c(min(vectorhis),max(vectorhis)), ylim=c(0,1.54), xaxt='n', yaxt='n',
xlab="",ylab="")
boxplot(Originalesgeo$data, range=1.5, notch=TRUE, ylim=c(min(vectorhis),max(vectorhis)),
horizontal= TRUE, col= "aquamarine1", pch= 22, axes= FALSE, add=TRUE )
abline(v = mean(Originalesgeo$data), lty = 3, col = "Red" )
abline(v = median(Originalesgeo$data), lty = 5, col= "Blue")
orden<-matrix(1,ncol=1,nrow=1,byrow=T)
div<-layout(orden, TRUE)
layout.show(div)
par(mar=c(5,5,5,5))
qqnorm(Originalesgeo$data, xlab = "Cuantiles teóricos", ylab="Cuantiles muestrales", cex.lab =
1.8, cex.axis=1.2 )
qqline(Originalesgeo$data)
orden<-matrix(c(1,2),ncol=2,nrow=1,byrow=T)
div<-layout(orden, heights=7.5,widths=c(4.5, 4.5), TRUE)
layout.show(div)
par(mar=c(4.5,4,6,1))

```

```

plot(Originalesgeo$data~Originalesgeo$coord[,1], pch= 22, col="black", bg="aquamarine",
ylab="Datos Pluviometro", xlab= "UTM X (Km)", cex.lab = 1.4, cex.axis=1.2)
abline(h = mean(Originalesgeo$data), lty = 5, col = "red" )
lines(loess.smooth(Originalesgeo$coord[,1], Originalesgeo$data))
legend(460, 7.5, "Media", cex= 1.5, col= "red", lty= 5, bty="n")
par(mar=c(4.5,4,6,1))
plot(Originalesgeo$data~Originalesgeo$coord[,2], pch= 22, col="black", bg="aquamarine",
ylab="Datos Pluviometro", xlab= "UTM Y (Km)", cex.lab = 1.4, cex.axis=1.2)
abline(h = mean(Originalesgeo$data), lty = 5, col = "Red" )
lines(loess.smooth(Originalesgeo$coord[,2], Originalesgeo$data))
legend(2153, 7.5, "Media", cex= 1.5, col= "Red", lty= 5, bty="n")
shapiro.test(Originalesgeo$data)
ad.test(Originalesgeo$data)
cvm.test(Originalesgeo$data)
chisq.out.test(Originalesgeo$data, variance= var(Originalesgeo$data))
outlier(Originalesgeo$data)
grubbs.test(Originalesgeo$data)
symmetry.test(Originalesgeo$data, option = c("mgg.test"))
symmetry.test(Originalesgeo$data, option = c("cabilio.masaro.test"))
symmetry.test(Originalesgeo$data, option = c("mira.test"))
Transformacion<-cbind(Originales, transformados=log(Originalesgeo$data))
comb<-cbind(Transformacion$UTM_X_en_Km, Transformacion$UTM_Y_en_Km,
Transformacion$transformados)
Transformaciongeo<-as.geodata(comb,coords.col= 1:2, data.col=3)
summary(Transformaciongeo)
var(Transformaciongeo$data)
sd(Transformaciongeo$data)
coordinates(Transformacion)<--~UTM_X_en_Km+UTM_Y_en_Km
minhis<-min(Transformaciongeo$data)
maxhis<-max(Transformaciongeo$data)
difhis<-maxhis-minhis+0.2
numclases<-10
tamint<-difhis/numclases

```

```

valorhis<-0
vectorhis<-0
for (i in 1:numclases)
{
valorhis[i]<-minhis-0.1+i*tamint
vectorhis<-c(minhis-0.1, valorhis)
}
orden<-matrix(c(2,3,1),ncol=1,nrow=3,byrow=T)
div<-layout(orden, widths= 9, heights=c(2, 0.8, 3.5), TRUE)
layout.show(div)
par(mar=c(5,6,0,3))
plot(Transformaciongeo$data, c(1:50), pch= 22, xlim=c(min(vectorhis),max(vectorhis)),
col="black", bg="aquamarine", ylab="Numero de Observación", xlab= "Valores de Pluviometro",
cex.lab = 1.8, cex.axis=1.2)
abline(v = mean(Transformaciongeo$data), lty = 3, col = "Red" )
abline(v = median(Transformaciongeo$data), lty = 5, col= "Blue")
legend(5, 15, c("Media", "Mediana"), cex= 1.5, col= c("Red", "Blue"), lty=c(3, 5), bty="n")
par(mar=c(0,6,1,3))
histograma<-hist(Transformaciongeo$data, breaks= vectorhis, xaxt="n", freq = FALSE, col=
"aquamarine3", ylab = "F. Relativas", cex.lab = 1.8, cex.axis=1.2, main = "")
abline(v = mean(Transformaciongeo$data), lty = 3, col = "Red" )
abline(v = median(Transformaciongeo$data), lty = 5, col= "Blue")
par(mar=c(0,6,0,3))
plot(0,0,type="n", xlim=c(min(vectorhis),max(vectorhis)), ylim=c(0,1.54), xaxt='n', yaxt='n',
xlab="",ylab="")
boxplot(Transformaciongeo$data, range=1.5, notch=TRUE,
ylim=c(min(vectorhis),max(vectorhis)), horizontal= TRUE, col= "aquamarine1", pch= 22, axes=
FALSE, add=TRUE )
abline(v = mean(Transformaciongeo$data), lty = 3, col = "Red" )
abline(v = median(Transformaciongeo$data), lty = 5, col= "Blue")
orden<-matrix(1,ncol=1,nrow=1,byrow=T)
div<-layout(orden, TRUE)
layout.show(div)

```

```

par(mar=c(5,5,5,5))
qqnorm(Transformaciongeo$data, xlab = "Cuantiles teóricos", ylab="Cuantiles muestrales",
cex.lab = 1.8, cex.axis=1.2)
qqline(Transformaciongeo$data)
orden<-matrix(c(1,2),ncol=2,nrow=1,byrow=T)
div<-layout(orden, heights=7.5, widths=c(4.5, 4.5), TRUE)
layout.show(div)
par(mar=c(4.5,4,6,1))
plot(Transformaciongeo$data ~Transformaciongeo$coord[,1], pch= 22, col="black",
bg="aquamarine", ylab="Datos Transformados", xlab= "UTM X (Km)", cex.lab = 1.4,
cex.axis=1.2)
abline(h = mean(Transformaciongeo$data), lty = 5, col = "red" )
lines(loess.smooth(Transformaciongeo$coord[,1], Transformaciongeo$data))
legend(460, 2, "Media", cex= 1.2, col= "red", lty= 5, bty="n")
par(mar=c(4.5,4,6,1))
plot(Transformaciongeo$data ~ Transformaciongeo$coord[,2], pch= 22, col="black",
bg="aquamarine", ylab="Datos Transformados", xlab= "UTM Y (Km)", cex.lab = 1.4,
cex.axis=1.2)
abline(h = mean(Transformaciongeo$data), lty = 5, col = "Red" )
lines(loess.smooth(Transformaciongeo$coord[,2], Transformaciongeo$data))
legend(2155, 2, "Media", cex= 1.2, col= "Red", lty= 5, bty="n")
shapiro.test(Transformaciongeo$data)
ad.test(Transformaciongeo$data)
cvm.test(Transformaciongeo$data)
chisq.out.test(Transformaciongeo$data, variance=var(Transformaciongeo$data))
grubbs.test(Transformaciongeo$data)
outlier(Transformaciongeo$data)
symmetry.test(Transformaciongeo$data, option = c("mgg.test"))
symmetry.test(Transformaciongeo$data, option = c("cabilio.masaro.test"))
symmetry.test(Transformaciongeo$data, option = c("mira.test"))
DistMax<-max(Transformaciongeo$coord[,1])-min(Transformaciongeo$coord[,1])
Distancia<-DistMax/20
ordenados<-sort(Transformaciongeo$coord[,1])

```

```

i<-0
diferencia<-0
diferencias<-0
for(i in 1:length(ordenados)-1)
{
diferencia[i]<-abs(ordenados[i]-ordenados[i+1])
diferencias<-c(diferencia)
}
min.dif<-min(diferencias)
if (Distancia >= min.dif)
{lags<-round(Distancia)
if (lags <= Distancia)
{lags<-round(Distancia)+1}
else
{lags<-round(Distancia) }
}
if (Distancia < min.dif)
{lags<-round(min.dif)
if (lags <= min.dif)
{lags<-round(min.dif)+1}
else
{lags<-round(min.dif)}
}
variograma<-variog(Transformaciongeo, breaks = c(seq(0, lags*10, lags)), trend="cte", lambda=
1, estimator.type="classical", nugget.tolerance=0, direction= "omnidirectional", pairs.min=1)
orden<-matrix(1,ncol=1,nrow=1,byrow=T)
div<-layout(orden, TRUE)
layout.show(div)
par(mar=c(5,5,5,5))
plot(variograma$u,variograma$v, type="p", pch=11, col= "red", xlim=c(0,max(variograma$u)+.1
), ylim=c(0,max(variograma$v)+.1 ), xlab="Distancias (Km)", ylab="Semivarianzas", cex.lab =
1.8, cex.axis=1.2)
abline(h= variograma$var.mark, lty= 5, lwd=2, col = "darkgreen")

```

```

legend(20, 0.2, "Varianza", cex= 1.5, col= "darkgreen", lty= 5, lwd=2, bty="n")
variogramaD<-variog4(Transformaciongeo, breaks = c(seq(0, lags*10, lags)), trend="cte",
lambda= 1, estimator.type="classical", nugget.tolerance=0, max.dist= lags*10, direction = c(0,
45, 90, 135), tolerance= 22.5, unit.angle = "degrees", pairs.min=1)
plot(variogramaD, xlab="Distancias (Km)", ylab="Semivarianzas", cex.lab=1.5,
omnidirectional=FALSE, same.plot=TRUE, legend = FALSE, type="b", pch=c(3,2,1,8), xlab="",
ylab="", col=c("darkorange", "darkblue", "darkgreen","darkviolet"), lty=c(4,2,3,1), lwd=2,
cex.axis=1.2)
abline(h= variograma$var.mark, lty= 5, lwd=2, col = "darkgreen")
legend(20, 0.2, "Varianza", cex= 1.5, col="darkgreen" , lwd=2, lty= 5, bty="n")
legend(0, 1.5, c("0°","45°","90°","135°"), cex= 1.5, pch=c(3,2,1,8),col=c("darkorange", "darkblue",
"darkgreen","darkviolet"), lty=c(4,2,3,1), lwd=2, bty="n")
max.pepita<-max(variograma$v)
for (i in 1:length(variograma$v))
{ if (variograma$v[i] == max.pepita)
{ rango<-variograma$u[i]
rango<-trunc(rango) }
}
modelo_Exp<-variofit(variograma, ini = c(variograma$var.mark, rango), cov.model
="exponential" , fix.nug=FALSE)
modelo_Exp
modelo_Sph<-variofit(variograma, ini = c(variograma$var.mark, rango), cov.model ="spherical" ,
fix.nug=FALSE)
modelo_Sph
modelo_Gau<-variofit(variograma, ini = c(variograma$var.mark, rango), cov.model ="gaussian"
, fix.nug=FALSE)
modelo_Gau
plot(variograma$u,variograma$v, type="p", pch=11, col= "red", xlim=c(0,max(variograma$u)),
ylim=c(0,max(variograma$v)), xlab="Distancias (Km)", ylab="Semivarianzas", cex.lab = 1.5,
cex.axis=1.2)
lines(modelo_Exp, lty=1, lwd=2.5, col="darkorange")
lines(modelo_Sph, lty=2, lwd=2.5, col="darkblue")
lines(modelo_Gau, lty=3, lwd=2.5, col="darkgreen")

```

```

abline(h= variograma$var.mark, lty= 5, col = "Red")
legend(20, 0.6, c("Exponencial", "Esférico", "Gaussiano"), col= c("darkorange", "darkblue",
"darkgreen"), lty= c(1,2,3), lwd=2.5, bty="n", cex=1.5)
legend(0, 1.13, "Varianza", bty="n", cex=1.5)
SCerror<-min(c(modelo_Exp$value, modelo_Sph$value, modelo_Gau$value))
if (SCerror==modelo_Exp$value)
{ modelo<-modelo_Exp
modelog<-"Exp"
}
if (SCerror==modelo_Sph$value)
{ modelo<-modelo_Sph
modelog<-"Sph"
}
if (SCerror==modelo_Gau$value)
{ modelo<-modelo_Gau
modelog<-"Gau"
}
modeloele<-variofit(variograma, ini =c(modelo$cov.pars[1], modelo$cov.pars[2]), cov.model =
modelo$cov.model, nugget = modelo$nugget, fix.nug= TRUE)
modeloele
plot(variograma$u,variograma$v, type="p", pch=11, col= "red", xlim=c(0,max(variograma$u)),
ylim=c(0,max(variograma$v)), xlab="Distancias (Km)", ylab="Semivarianzas", cex.lab = 1.5,
cex.axis=1.2)
lines(modeloele)
abline(h= modeloele$cov.pars[1]+modeloele$nugget, lty= 3, col = "Blue")
abline(h= variograma$var.mark, lty= 5, col = "Red")
abline(v= modeloele$cov.pars[2], lty = 1, col = "Green")
legend(20, 0.4, c("Meseta+Nugget", "Varianza", "Alcance"), col= c("Blue", "Red", "Green"), lty=
c(3, 5, 1), bty="n", cex=1.5)
modeloele<-variofit(variograma, ini =c(modelo_Sph$cov.pars[1], modelo_Sph$cov.pars[2]),
cov.model = modelo_Sph$cov.model, nugget = 0.06, fix.nug= TRUE)
modeloele

```

```

plot(variograma$u,variograma$v, type="p", pch=11, col= "red", xlim=c(0,max(variograma$u)),
ylim=c(0,max(variograma$v)), xlab="Distancias (Km)", ylab="Semivarianzas", cex.lab = 1.5,
cex.axis=1.2)
lines(modeloele)
abline(h= modeloele$cov.pars[1]+modeloele$nugget, lty= 3, col = "Blue")
abline(h= variograma$var.mark, lty= 5, col = "Red")
abline(v= modeloele$cov.pars[2], lty = 1, col = "Green")
legend(20, 0.4, c("Meseta+Nugget", "Varianza", "Alcance"), col= c("Blue", "Red", "Green"), lty=
c(3, 5, 1), bty="n", cex=1.5)
validacion<-xvalid(Transformaciongeo, coords = Transformaciongeo$coords, data =
Transformaciongeo$data, model = modeloele, variog.obj = variograma)
medErrorEst<-mean(validacion$std.error)
varErrorEst<-var(validacion$std.error)
minhis<-min(validacion$std.error)
maxhis<-max(validacion$std.error)
difhis<-maxhis-minhis+0.2
numclases<-9
tamint<-difhis/numclases
valorhis<-0
vectorhis<-0
for (i in 1:numclases)
{
valorhis[i]<-minhis-0.1+i*tamint
vectorhis<-c(minhis-0.1, valorhis)
}
min1<-min(validacion$predicted)
min2<-min(validacion$data)
max1<-max(validacion$predicted)
max2<-max(validacion$data)
min3<-min(min1, min2)
max3<-max(max1, max2)
orden<-matrix(c(1,3,1,4,1,2), ncol=2,nrow=3,byrow=T)
div<-layout(orden, widths=c(6.2,6),heights=c(2,1,4), TRUE)

```



```

layout.show(div)
par(mar=c(6.5,5,6.5,1))
plot(validacion$predicted~validacion$data, xlim= c(min3, max3), ylim = c(min3, max3),
xlab="Estimados", ylab="Reales", pch= 22, col="black", bg="aquamarine", cex.lab = 1.8,
cex.axis=1.2, main = "Reales Vs. Estimados", cex.main = 2, pty = "s")
abline(a= 0, b= 1)
par(mar=c(4,5,0,2))
plot(validacion$std.error, c(1:50), pch= 22, xlim=c(min(vectorhis),max(vectorhis)), col="black",
bg="aquamarine", ylab="Numero de Observación", xlab= "Valores de Pluviometro", cex.lab =
1.8, cex.axis=1.2)
abline(v = mean(validacion$std.error), lty = 3, col = "Red" )
abline(v = median(validacion$std.error), lty = 5, col= "Blue")
legend(1.5, 15, c("Media", "Mediana"), col= c("Red", "Blue"), cex= 1.5, lty=c(3, 5), bty = "n")
par(mar=c(0,5,1.5,2))
histograma<-hist(validacion$std.error, breaks= vectorhis, xaxt="n", freq = FALSE, col=
"aquamarine3", ylab = "F. Relativas", cex.lab = 1.8, cex.axis=1.2, main = "Errores", cex.main =
2)
abline(v = mean(validacion$std.error), lty = 3, col = "Red" )
abline(v = median(validacion$std.error), lty = 5, col= "Blue")
par(mar=c(0,5,0,2))
plot(0,0,type="n", xlim=c(min(vectorhis),max(vectorhis)), ylim=c(0,1.54), xaxt='n', yaxt='n',
xlab="", ylab="")
boxplot(validacion$std.error, notch=TRUE, range=1.5, ylim=c(min(vectorhis),max(vectorhis)),
horizontal= TRUE, col= "aquamarine1", pch= 22, axes= FALSE, add=TRUE )
abline(v = mean(validacion$std.error), lty = 3, col = "Red" )
abline(v = median(validacion$std.error), lty = 5, col= "Blue")
minx<-min(Transformaciongeo$coord[,1])
maxx<-max(Transformaciongeo$coord[,1])
miny<-min(Transformaciongeo$coord[,2])
maxy<-max(Transformaciongeo$coord[,2])
local<-pred_grid(c(minx,maxx),c(miny,maxy), by = 1)
localizacion<-as.data.frame(local)
coordinates(localizacion) <- ~Var1+Var2

```

```

modelocal<- vgm(modeloele$cov.pars[1], "Sph", modeloele$cov.pars[2], modeloele$nugget)
kriging<-krige(transformados~1, Transformacion, localizacion, model = modelocal, nmax= Inf,
nmin = 0, maxdist= lags*10, indicators = FALSE)
Kriging<-as.data.frame(kriging)
# En las siguiente dirección guardara los resultados
write.table(Kriging, file = "C:\\\\TESIS\\Tesis\\ KrigingOrd.xls", sep = " ", row.names = FALSE,
col.names = TRUE)
localest<-as.data.frame(cbind(local, exp(kriging$var1.pred)))
estimados<-as.geodata(localest,coords.col= 1:2, data.col=3)
summary(estimados)
var(estimados$data)
sd(estimados$data)
imagenest<- as.image(estimados$data , x= estimados$coords)
orden<-matrix(1,ncol=1,nrow=1,byrow=T)
div<-layout(orden, TRUE)
layout.show(div)
par(mar=c(5,5,5,5))
imaest<- image.smooth( imagenest$z, dx=1, dy=1)
image.plot( x= imagenest$x, y= imagenest$y, z = imaest, xlab="Coordenadas X",
ylab="Coordenadas Y", col = rainbow(15) )
contour(x= imagenest$x, y= imagenest$y, z = imaest, add=TRUE)
points(Originalesgeo$coords, pch= 16, col = "white")
localdesv<-as.data.frame(cbind(local, sqrt((kriging$var1.var))))
desviacion<-as.geodata(localdesv,coords.col= 1:2, data.col=3)
summary(desviacion)
var(desviacion$data)
sd(desviacion$data)
imagendesv<- as.image(desviacion$data , x= desviacion$coords)
imadesv<- image.smooth( imagendesv$z, dx=1, dy=1)
image.plot( x= imagendesv$x, y= imagendesv$y, z = imadesv, xlab="Coordenadas X",
ylab="Coordenadas Y", col = rainbow(15) )
contour(x= imagendesv$x, y= imagendesv$y, z = imadesv, add=TRUE)
points(Originalesgeo$coords, pch= 16, col = "white")

```

Realiza el cálculo de las simulaciones univariadas

```
require(sp)
require(geoR)
require(nortest)
require(sgeostat)
require(gstat)
require(outliers)
require(fields)
require(moments)
require(lawstat)
require(reshape)
require(distributions)
Originales<-as.data.frame(read.table("C:\\TESIS\\Tesis\\pluv.txt", header= TRUE))
Originalesgeo<-read.geodata("C:\\TESIS\\Tesis\\pluv.txt", header= TRUE, coords.col= 1:2,
data.col=3)
summary(Originalesgeo)
var(Originalesgeo$data)
sd(Originalesgeo$data)
orden<-sort_df(Originales, vars="Pluv_mm")
cuantos<-c(seq(1:length(Originales$Pluv_mm)))
ordenados<-cbind(cuantos, orden)
Transforma<-cbind(ordenados, transpluv =((ordenados$cuantos)/length(cuantos))-
(0.1/length(cuantos)))
Transformacion<-cbind(Transforma, transformados=invNorm(Transforma$transpluv,0,1))
comb<-cbind(Transformacion$UTM_X_en_Km, Transformacion$UTM_Y_en_Km,
Transformacion$transformados)
Transformaciongeo<-as.geodata(comb,coords.col= 1:2, data.col=3)
summary(Transformaciongeo)
var(Transformaciongeo$data)
sd(Transformaciongeo$data)
coordinates(Transformacion)<-~UTM_X_en_Km+UTM_Y_en_Km
DistMax<-max(Transformaciongeo$coord[,1])-min(Transformaciongeo$coord[,1])
Distancia<-DistMax/20
```

```

ordenados<-sort(Transformaciongeo$coord[,1])
i<-0
diferencia<-0
diferencias<-0
for(i in 1:length(ordenados)-1)
{
  diferencia[i]<-abs(ordenados[i]-ordenados[i+1])
  diferencias<-c(diferencia)
}
min.dif<-min(diferencias)
if (Distancia >= min.dif)
{lags<-round(Distancia)
  if (lags <= Distancia)
    {lags<-round(Distancia)+1}
  else
    {lags<-round(Distancia) }
}
if (Distancia < min.dif)
{lags<-round(min.dif)
  if (lags <= min.dif)
    {lags<-round(min.dif)+1}
  else
    {lags<-round(min.dif) } }
variograma<-variog(Transformaciongeo, breaks = c(seq(0, lags*10, lags)), trend="cte", lambda=
1, estimator.type="classical", nugget.tolerance=0, direction= "omnidirectional", pairs.min=1)
orden<-matrix(1,ncol=1,nrow=1,byrow=T)
div<-layout(orden, TRUE)
layout.show(div)
par(mar=c(5,5,5,5))
plot(variograma$u,variograma$v, type="p", pch=11, col= "red", xlim=c(0,max(variograma$u)+.1
), ylim=c(0,max(variograma$v)+.1 ), xlab="Distancias (Km)", ylab="Semivarianzas", cex.lab =
1.8, cex.axis=1.2)
abline(h= variograma$var.mark, lty= 5, lwd=2, col = "darkgreen")

```

```

legend(20, 0.2, "Varianza", cex= 1.5, col= "darkgreen", lty= 5, lwd=2, bty="n")
max.pepita<-max(variograma$v)
for (i in 1:length(variograma$v))
  { if (variograma$v[i] == max.pepita)
    {rango<-variograma$u[i]
    rango<-trunc(rango)
    } }
modelo_Sph<-variofit(variograma, ini = c(variograma$var.mark, rango), cov.model ="spherical" ,
fix.nug=FALSE)
modelo_Sph
modeloele<-variofit(variograma, ini =c(modelo_Sph$cov.pars[1], modelo_Sph$cov.pars[2]),
cov.model = modelo_Sph$cov.model, nugget = 0, fix.nug= TRUE)
modeloele
plot(variograma$u,variograma$v, type="p", pch=11, col= "red", xlim=c(0,max(variograma$u)),
ylim=c(0,max(variograma$v)), xlab="Distancias (Km)", ylab="Semivarianzas", cex.lab = 1.5,
cex.axis=1.2)
lines(modeloele)
abline(h= modeloele$cov.pars[1]+modeloele$nugget, lty= 3, col = "Blue")
abline(h= variograma$var.mark, lty= 5, col = "Red")
abline(v= modeloele$cov.pars[2], lty = 1, col = "Green")
legend(20, 0.4, c("Meseta+Nugget", "Varianza", "Alcance"), col= c("Blue", "Red", "Green"), lty=
c(3, 5, 1), bty="n", cex=1.5)
validacion<-xvalid(Transformaciongeo, coords = Transformaciongeo$coords, data =
Transformaciongeo$data, model = modeloele, variog.obj = variograma)
medErrorEst<-mean(validacion$std.error)
medErrorEst
varErrorEst<-var(validacion$std.error)
varErrorEst
minhis<-min(validacion$std.error)
maxhis<-max(validacion$std.error)
difhis<-maxhis-minhis+0.2
numclases<-9
tamint<-difhis/numclases

```

```

valorhis<-0
vectorhis<-0
for (i in 1:numclases)
{
  valorhis[i]<-minhis-0.1+i*tamint
  vectorhis<-c(minhis-0.1, valorhis)
}
vectorhis
min1<-min(validacion$predicted)
min2<-min(validacion$data)
max1<-max(validacion$predicted)
max2<-max(validacion$data)
min3<-min(min1, min2)
max3<-max(max1, max2)
orden<-matrix(c(1,3,1,4,1,2), ncol=2,nrow=3,byrow=T)
div<-layout(orden, widths=c(6.2,6),heights=c(2,1,4), TRUE)
layout.show(div)
par(mar=c(6.5,5,6.5,1))
plot(validacion$predicted~validacion$data, xlim= c(min3, max3), ylim = c(min3, max3),
xlab="Estimados", ylab="Reales", pch= 22, col="black", bg="aquamarine", cex.lab = 1.8,
cex.axis=1.2, main = "Reales Vs. Estimados", cex.main = 2, pty = "s")
abline(a= 0, b= 1)
par(mar=c(4,5,0,2))
plot(validacion$std.error, c(1:50), pch= 22, xlim=c(min(vectorhis),max(vectorhis)), col="black",
bg="aquamarine", ylab="Numero de Observación", xlab= "Valores de Pluviometro", cex.lab =
1.8, cex.axis=1.2)
abline(v = mean(validacion$std.error), lty = 3, col = "Red" )
abline(v = median(validacion$std.error), lty = 5, col= "Blue")
legend(1.5, 15, c("Media", "Mediana"), col= c("Red", "Blue"), cex= 1.5, lty=c(3, 5), bty = "n")
par(mar=c(0,5,1.5,2))
histograma<-hist(validacion$std.error, breaks= vectorhis, xaxt="n", freq = FALSE, col=
"aquamarine3", ylab = "F. Relativas", cex.lab = 1.8, cex.axis=1.2, main = "Errores", cex.main =
2)

```

```

abline(v = mean(validacion$std.error), lty = 3, col = "Red" )
abline(v = median(validacion$std.error), lty = 5, col= "Blue")
par(mar=c(0,5,0,2))
plot(0,0,type="n", xlim=c(min(vectorhis),max(vectorhis)), ylim=c(0,1.54), xaxt='n', yaxt='n',
xlab="",ylab="")
boxplot(validacion$std.error, notch=TRUE, range=1.5, ylim=c(min(vectorhis),max(vectorhis)),
horizontal= TRUE, col= "aquamarine1", pch= 22, axes= FALSE, add=TRUE )
abline(v = mean(validacion$std.error), lty = 3, col = "Red" )
abline(v = median(validacion$std.error), lty = 5, col= "Blue")
minx<-min(Transformaciongeo$coord[,1])
maxx<-max(Transformaciongeo$coord[,1])
miny<-min(Transformaciongeo$coord[,2])
maxy<-max(Transformaciongeo$coord[,2])
local<-pred_grid(c(minx,maxx),c(miny,maxy), by = 1)
localizacion<-as.data.frame(local)
coordinates(localizacion) <- ~Var1+Var2
modelocal<- vgm(modeloele$cov.pars[1], "Sph", modeloele$cov.pars[2], modeloele$nugget)
modelocal
set.seed(7)
simul1<-krige(transformados~1, Transformacion, localizacion, model = modelocal, nmax= Inf,
nmin = 0, maxdist= lags*10, nsim= 100, indicators = FALSE)
simul<-as.data.frame(simul)
write.table(simul, file = "C:\\TESIS\\Tesis\\SimulacionOrdSem7.xls", sep = " ", row.names =
FALSE, col.names = TRUE)
simulacion<-as.matrix(simul[3:102])
simulado<-pnorm(simulacion,0,1)
simulado1<-(simulado+(0.1/length(cuantos)))*length(cuantos)
vectoruno<-as.matrix(simulado1[,29])
diez<-as.matrix(simulado1[,61:70])
mediadiezt<-0
vectordiez<-0
i<-0
for(i in 1:length(diez[,1]))

```

```

{
  mediadies[i]<-mean(diez[i,])
  vectordiez<-c(mediadies)
}
cien<-as.matrix(simulado1[,1:100])
mediacien<-0
vectorcien<-0
i<-0
for(i in 1:length(cien[,1]))
{
  mediacien[i]<-mean(cien[i,])
  vectorcien<-c(mediacien)
}
cortadosim1<-round(vectoruno)
cortadosim10<-round(vectordiez)
cortadosim100<-round(vectorcien)
i<-0
for(i in 1:length(cortadosim1))
{if (cortadosim1[i]==0)
{
  cortadosim1[i]<-1
} }
i<-0
for(i in 1:length(cortadosim10))
{if (cortadosim10[i]==0)
{
  cortadosim10[i]<-1
} }
i<-0
for(i in 1:length(cortadosim100))
{if (cortadosim100[i]==0)
{
  cortadosim100[i]<-1
}
}

```



```

} }
minx<-min(simul[,1])
maxx<-max(simul[,1])
miny<-min(simul[,2])
maxy<-max(simul[,2])
local<-pred_grid(c(minx,maxx),c(miny,maxy), by = 1)
localizacion<-as.data.frame(local)
coordinates(localizacion) <- ~Var1+Var2
simulacionfinal1<-as.data.frame(cbind(local,
simulado2=Transformacion$Pluv_mm[cortadosim1]))
simulaciongeo1<-as.geodata(simulacionfinal1, coords.col= 1:2, data.col=3)
summary(simulaciongeo1)
var(simulaciongeo1$data)
sd(simulaciongeo1$data)
imagensim<-as.image(simulaciongeo1$data , x= simulaciongeo1$coords)
orden<-matrix(1,ncol=1,nrow=1,byrow=T)
div<-layout(orden, TRUE)
layout.show(div)
par(mar=c(5,5,5,5))
imagsim<- image.smooth( imagensim$z, dx=1, dy=1)
image.plot( x= imagensim$x, y= imagensim$y, z = imagsim, xlab="Coordenadas X",
ylab="Coordenadas Y", col = rainbow(15) )
contour(x= imagensim$x, y= imagensim$y, z = imagsim, add=TRUE,labcex = 1, nlevels = 8)
points(Originalesgeo$coords, pch= 16, col = "white")
simulacionfinal10<-as.data.frame(cbind(local,
simulado2=Transformacion$Pluv_mm[cortadosim10]))
simulaciongeo10<-as.geodata(simulacionfinal10, coords.col= 1:2, data.col=3)
summary(simulaciongeo10)
var(simulaciongeo10$data)
sd(simulaciongeo10$data)
imagensim<-as.image(simulaciongeo10$data , x= simulaciongeo10$coords)
imagsim<- image.smooth( imagensim$z, dx=1, dy=1)

```

```

image.plot( x= imagensim$x, y= imagensim$y, z = imagsim, xlab="Coordenadas X",
ylab="Coordenadas Y", col = rainbow(15) )
contour(x= imagensim$x, y= imagensim$y, z = imagsim, add=TRUE,labcex = 1, nlevels = 8)
points(Originalesgeo$coords, pch= 16, col = "white")
simulacionfinal100<-as.data.frame(cbind(local,
simulado2=Transformacion$Pluv_mm[cortadosim100]))
simulaciongeo100<-as.geodata(simulacionfinal100, coords.col= 1:2, data.col=3)
summary(simulaciongeo100)
var(simulaciongeo100$data)
sd(simulaciongeo100$data)
imagensim<-as.image(simulaciongeo100$data , x= simulaciongeo100$coords)
imagsim<- image.smooth( imagensim$z, dx=1, dy=1)
image.plot( x= imagensim$x, y= imagensim$y, z = imagsim, xlab="Coordenadas X",
ylab="Coordenadas Y", col = rainbow(15) )
contour(x= imagensim$x, y= imagensim$y, z = imagsim, add=TRUE,labcex = 1, nlevels = 8)
points(Originalesgeo$coords, pch= 16, col = "white")

```