



Cosimulación Geoestadística:
Aplicación a la Generación de Mapas de Precipitaciones
en una Región Modelo de Colombia

Autor: Diego Felipe Samudio Pulido

Director: Dr. Javier Vásquez

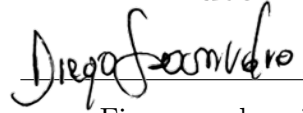
Tesis presentada para optar por el grado de
Magister en Geomática

Facultad de Ciencias Astronómicas y Geofísicas
Universidad Nacional de La Plata


La Plata
Diciembre de 2020

Cosimulación Geoestadística:
Aplicación a la Generación de Mapas de Precipitaciones
en una Región Modelo de Colombia

Autor

 Diego Samudio Diego Samudio
Firma y aclaración

Director

 Javier Vásquez Javier Vásquez
Firma y aclaración

Miembros del jurado:

Dr. Guillermo Berri

Dra. Amalia Meza

Dr. Pablo Facundo Orte

Presidente del jurado


Dr. Guillermo Berri

Firma y aclaración

*A mis padres, siempre presentes, siempre a mi lado,
a mi compañera de vida y camino, mi fuerza y apoyo incondicional,
a mi familia, fuente de motivación en todo momento...*

Agradecimientos

Un sentido y especial agradecimiento a todas las personas que tuvieron relación con la investigación. Al director, al jurado, a profesores y colegas que aportaron opiniones, retroalimentación y conceptos profesionales y que siempre manifestaron interés en el tema, el avance y los resultados del trabajo.

Resumen

La presente investigación propone una metodología para obtener mapas de precipitaciones mensuales, a partir del modelamiento geoestadístico de la relación entre la precipitación y el Índice de Vegetación Diferencial Normalizado (NDVI - Normalized Difference Vegetation Index) en una región de la Cordillera Oriental de Los Andes, ubicada principalmente sobre el Altiplano Cundiboyacense en Colombia.

Utilizando datos de precipitaciones mensuales para el periodo comprendido entre marzo y octubre de 2016, los cuales fueron medidos en 56 estaciones meteorológicas del Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia (IDEAM), y productos satelitales MODIS (Moderate-Resolution Imaging Spectroradiometer) enfocados al estudio de índices de vegetación (MODIS VI Products), se aplicó la técnica geoestadística de Co-Simulaciones Secuenciales Gaussianas (Sequential Gaussian Co-Simulation - CO-SGSim).

La precipitación, conocida únicamente en las ubicaciones de las estaciones meteorológicas dispersas en la zona de estudio, constituye la variable principal o “hard data”. Por otra parte, el NDVI, que se conoce en cualquier lugar de la zona donde la vegetación terrestre predomina como cobertura de terreno, representa la variable secundaria o “soft data”, y dado que cada pixel de una imagen equivale a una medición, su densidad de muestreo es alta. El método de Co-Simulaciones Secuenciales Gaussianas permite simular espacialmente valores de una variable con información escasa, a partir de su relación con otra variable que cuenta con información abundante.

Durante la investigación se aplicaron diferentes técnicas de adquisición y tratamiento de información aplicables a los Sistemas de Información Geográfica (SIG) y el Procesamiento Digital de Imágenes (PDI), las cuales permitieron preparar los datos para su posterior enfoque geoestadístico. Los productos MODIS utilizados son los denominados MOD13A3 (sensor TERRA) y MYD13A3 (sensor AQUA) de la Colección 6 del catálogo MODIS VI, dichos productos se componen de archivos raster de NDVI mensual con resolución espacial de un kilómetro. Dentro de su procesamiento, se realizaron actividades de reproyección y ajuste de extensión con la herramienta MODIS Reprojection Tools (MRT), y extracción de los parámetros de calidad MODLAND QA Bits, VI Usefulness, Possible Shadow y Land/Water

Mask por medio de la descomposición de la banda VI Quality con rutinas ejecutadas en la herramienta LDOPE (Land Data Operational Products Evaluation). Además, con la herramienta Model Builder del software ArcGIS, se diseñó un flujo de procesamiento, en el que se aplicaron filtros basados en la información de calidad de cada imagen, con el objetivo de establecer valores de NDVI óptimos y libres de ruido espectral en cada uno de los meses. En dicho flujo, los parámetros de calidad mencionados, junto con la información de las bandas NDVI y Pixel Reliability fueron consideradas, evaluando la información raster mediante la aplicación de geoprosos de álgebra de mapas.

En la fase exploratoria de datos, se estudiaron medidas estadísticas de tendencia, forma y distribución, y se caracterizó la variabilidad espacial de ambas variables analizando la distribución geográfica, ubicación y concentración de valores en la región. Mientras en los diferentes meses la lluvia tuvo la constante de agrupamiento hacia valores bajos, que se localizaron principalmente en la zona centro y norte de la región de estudio, el NDVI, presentó agrupamientos hacia los valores altos, localizando los mayores valores en la zona sur y centro de la región. Con mapas de distribución de valores, fue notorio el hecho de que en zonas donde se presentaron valores altos para las precipitaciones también se presentaron valores altos para el NDVI, y lo mismo ocurrió, con los valores bajos y medios en cada uno de los meses.

Los histogramas de ambas variables fueron transformados, logrando distribuciones estandarizadas, necesarias para el modelamiento de la variabilidad espacial en la etapa de análisis estructural y la posterior aplicación de simulaciones y co-simulaciones secuenciales gaussianas. En el software S-GEMS se obtuvieron variogramas muestrales omnidireccionales y en orientaciones de 0, 45, 90 y 135 grados, con los que se estableció que las lluvias presentan un comportamiento totalmente anisotrópico, y el NDVI un comportamiento isotrópico a escala local, y geoméricamente anisotrópico a escala regional. Con estas consideraciones se realizaron ajustes a sentimiento de modelos teóricos de semivarianza para cada uno de los meses.

Se definió una grilla de simulación con tamaño de celda acorde al tamaño de pixel, en la cual se espera, para poder completar la co-simulación, que todas las posiciones cuenten con valores de la variable secundaria, pero dado que en el proceso de filtrado

se descartaron píxeles dependiendo de las condiciones de calidad, estas posiciones debieron completarse con procesos de Simulaciones Secuenciales Gaussianas (SGSim) aplicadas a la variable NDVI, donde se generaron 30 realizaciones por mes, que luego fueron promediadas. Finalmente, se procedió a ejecutar las co-simulaciones entre las precipitaciones y el resultado de la simulación del NDVI utilizando el método de Markov I, en el que se involucran el variograma de la variable principal y el coeficiente de correlación entre ambas variables. Con 50 realizaciones por mes y su respectivo promedio se obtuvieron mapas que densifican la lluvia en la región de estudio. Tales mapas tienen como característica que las heterogeneidades de los valores de la precipitación se resaltan, y que la distribución original de los datos anti-transformados se respeta replicando su histograma junto con sus propiedades estadísticas en el resultado de la co-simulación.

Abstract

The present research proposes a methodology to obtain maps of monthly rainfall, based on the geostatistical modeling of the relationship between precipitation and Normalized Differential Vegetation Index (NDVI) in an area of “Cordillera Oriental de Los Andes”, which is mainly located in the “Altiplano Cundiboyacense” zone in Colombia.

Using monthly rainfall data from March to October 2016, that were measured in 56 meteorological stations of the “Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia” (IDEAM), and MODIS (Moderate Resolution Image Spectroradiometer) satellite products focused on the study of Vegetation Indexes (MODIS VI Products), the geostatistical technique of Gaussian Sequential Co-Simulations (CO-SGSim) was applied.

Precipitation, known only in dispersed weather stations within the study area, establishes the main variable or “hard data”. On the other hand, NDVI, which is known anywhere in the area where land vegetation predominates as land cover, represents the secondary variable or “soft data”, and because each pixel of an image is equivalent to a measure, its sampling density is high. Gaussian Sequential Co-Simulations method allows spatially simulating values of a scarce information variable, based on its relationship with a variable which has extensive information.

During the investigation, different techniques for information acquiring and processing, applicable to Geographic Information Systems (GIS) and Digital Image Processing (DIP) were used. Such techniques allowed data preparing for subsequent geostatistical approach. The MODIS products used are the so-called MOD13A3 (for TERRA sensor) and MYD13A3 (for AQUA sensor) from MODIS VI catalog - Collection 6, these products are composed of monthly NDVI raster files with one kilometer spatial resolution. Within its processing, reprojection and extent adjustment activities were carried out with the MODIS Reprojection Tools (MRT) application, also, extraction of quality parameters like MODLAND QA Bits, VI Usefulness, Possible Shadow and Land / Water Mask through decomposition of the VI Quality band were done with LDOPE (Land Data Operational Products Evaluation) routines. In addition, with the Model Builder tool of ArcGIS software, a processing workflow was designed, and there, filters based on the quality information of each image were applied with aim of establishing optimal NDVI values and free spectral noise images for each month. In this workflow, mentioned quality

parameters, along with information of NDVI and Pixel Reliability bands were considered while raster information was evaluated through application of map algebra geoprocesses.

In exploratory data analysis, trend, form and statistical distribution measures were studied, spatial variability of both variables was characterized by analyzing the geographical distribution, location and concentration of values in zone. While along different months rain had constant grouping towards low values, which were located mainly in the central and north zone of the study area, NDVI, presented clusters towards high values, locating the highest values in the south and in the central zone. With maps of values distribution, was evident the fact that, in areas where high values were presented for rainfall, high values were also presented for NDVI, and so happened, with low and medium values in each month.

Histograms of both variables were transformed, achieving standardized distributions which are necessary for modeling spatial variability during the structural analysis stage, and the subsequent application of Gaussian Sequential Simulations. In S-GEMS software, omnidirectional sample variograms were obtained as well as directional variograms in specific orientations of 0, 45, 90 and 135 degrees, with which was established that rainfall presents a totally anisotropic behavior, and NDVI has an isotropic behavior at local scale, and a geometrically anisotropic behavior at regional scale. With these considerations, visual adjustments were made to yield suited theoretical variogram models for each month.

A simulation grid with cell dimensions according to pixel size was defined, in order to complete the co-simulation, it is expected, that all positions in grid have values of secondary variable, but since the filtering process discarded pixels depending on the quality conditions, these positions had to be completed with Gaussian Sequential Simulation processes applied to NDVI variable. There were computed 30 realizations per month, which then were averaged. Finally, Gaussian Sequential Co-Simulations between rainfall and simulation result of NDVI were carried out using Markov I method, in which variogram of main variable and correlation coefficient between both variables are involved. With 50 realizations per month and their respective average, maps that densify the rain in the study region were obtained. Such maps have as features, that precipitation values stand out heterogeneities, and that resulting histogram of back-transformed data honors original data distribution and replicates its statistical properties in co-simulation result.

Índice

Introducción	1
1. Objetivos	3
1.1. Objetivo General	3
1.2. Objetivos Específicos	3
2. Marco Teórico	4
2.1. Precipitación	4
2.2. Índices de Vegetación	5
2.3. Relación Precipitaciones - NDVI	6
2.4. Geoestadística	7
I Análisis Exploratorio	8
II Análisis Estructural	9
Variograma muestral	12
Modelos teóricos de variograma	18
III Estimación	19
Polígonos de proximidad	21
Media espacial local	21
Triangulación	22
Vecinos más cercanos	22
Métodos basados en distancia	23
Kriging	25
IV Simulaciones Condicionales	28
3. Materiales y Métodos	36
3.1. Zona de Estudio	36
3.2. Registros de Lluvia	37
3.3. Productos MODIS NDVI	38
3.3.1. Procesamiento de Imágenes Satelitales MODIS	42
3.4. Análisis Geoestadístico	44
4. Resultados y Discusión	46
4.1. Fase Exporatoria	46
4.1.1. Estadística Descriptiva	46
4.1.2. Caracterización Espacial de las Variables	51
4.1.3. Transformación del Histograma	54

4.2. Análisis Estructural	56
4.3. Simulación Gaussiana del NDVI	76
4.4. Co-Simulación Gaussiana entre las Precipitaciones y el NDVI .	80
4.5. Influencia de la Topografía en Resultado de las Precipitaciones	85
Conclusiones	87
Anexos	88
A. Procesamiento Detallado de Productos MODIS NDVI	88
A.1. Pre-procesamiento	89
A.2. Filtrado de Calidad	91
A.2.1. Descomposición Bandas de Calidad	91
A.2.2. Álgebra de Mapas	93
B. Medidas Estadísticas	98

Índice de figuras

1.	Gráficos esquemáticos de correlograma y covariograma	11
2.	Gráfico típico del variograma experimental	13
3.	Tolerancias del vector de separación	14
4.	Elementos del variograma	16
5.	Ejemplo de variogramas direccionales con igual valor de varianza a diferentes distancias	17
6.	Ejes de máxima y mínima continuidad	17
7.	Modelos teóricos de variogramas	19
8.	Representación del concepto de estimación para el valor de una variable en el espacio	21
9.	Interpretación gráfica de métodos de estimación local	23
10.	Estimación por distancia inversa ponderada	25
11.	Comparación conceptual entre simulación y estimación	29
12.	Interpretación gráfica del resultado de simulación y estimación	29
13.	Comparación de mapas resultantes con estimación y simulación	30
14.	Histogramas obtenidos con simulaciones	31
15.	Estimación de función condicional	33
16.	Flujo esquemático de la simulación secuencial gaussiana	35
17.	Localización de la zona de estudio en el territorio colombiano	36
18.	Ubicación de las 56 estaciones meteorológicas en la zona de estudio	40
19.	Flujo de trabajo para procesamiento de imágenes MODIS	43
20.	Flujo del análisis geoestadístico	45
21.	Histogramas de las precipitaciones	47
22.	Histogramas del NDVI	48
23.	Box plot de las precipitaciones	49
24.	Box plot del NDVI	50
25.	Distribución de los valores de precipitación mensual por estación en la zona de estudio	52
26.	Distribución de los valores de NDVI mensual en la zona de estudio	53
27.	Transformación de histogramas para las precipitaciones	54
28.	Transformación de histogramas para el NDVI	55
29.	Distribución espacial de los datos originales y transformados. Ejemplo para mes de mayo	56

30.	Variogramas omnidireccionales de las precipitaciones para los meses de marzo y abril	59
31.	Variogramas omnidireccionales de las precipitaciones para los meses de mayo y junio	60
32.	Variogramas omnidireccionales de las precipitaciones para los meses de julio y agosto	61
33.	Variogramas Omnidireccionales de las precipitaciones para los meses de septiembre, y octubre	62
34.	Variogramas direccionales y ajuste de modelos de semivarianza para las precipitaciones en los meses de marzo y abril	63
35.	Variogramas direccionales y ajuste de modelos de semivarianza para las precipitaciones en los meses de mayo y junio	64
36.	Variogramas direccionales y ajuste de modelos de semivarianza para las precipitaciones en los meses de julio y agosto	65
37.	Variogramas direccionales y ajuste de modelos de semivarianza para las precipitaciones en los meses de septiembre y octubre .	66
38.	Variogramas omnidireccionales para el NDVI para los meses de marzo y abril	67
39.	Variogramas omnidireccionales para el NDVI para los meses de mayo y junio	68
40.	Variogramas omnidireccionales para el NDVI para los meses de julio y agosto	69
41.	Variogramas omnidireccionales para el NDVI para los meses de septiembre y octubre	70
42.	Variogramas direccionales y ajuste de modelos de semivarianza para el NDVI en los meses de marzo y abril	71
43.	Variogramas direccionales y ajuste de modelos de semivarianza para el NDVI en los meses de mayo y junio	72
44.	Variogramas direccionales y ajuste de modelos de semivarianza para el NDVI en los meses de julio y agosto	73
45.	Variogramas direccionales y ajuste de modelos de semivarianza para el NDVI en los meses de septiembre y octubre	74
46.	Variogramas direccionales de las precipitaciones	75
47.	Variogramas direccionales del NDVI	75
48.	Resultados de mapas e histogramas de la simulación gaussiana del NDVI - Realizaciones 1, 10, 20 y 30 mes de marzo	78

49.	Mapas e histogramas promedio de la simulación gaussiana del NDVI transformado y anti-trasformado para el mes de septiembre	79
50.	Resultados de mapas e histogramas de la co-simulación gaussiana de la precipitación - Realizaciones 1, 15, 30 y 50 mes de septiembre	81
51.	Mapas e histogramas promedio de la co-simulación gaussiana de la precipitación transformada y anti-transformada para el mes de septiembre	82
52.	Mapa resultado de la co-simulación de las precipitaciones para los meses de marzo, abril, mayo y junio	83
53.	Mapa resultado de la co-simulación de las precipitaciones para los meses de julio, agosto, septiembre, octubre	84
54.	Comparación entre la precipitación resultante de la co-simulación, la topografía y el NDVI para el mes de abril	86
55.	Flujo de trabajo para procesamiento de imágenes MODIS . . .	88
56.	Estructura de parámetros para el pre-procesamiento de las imágenes MODIS	90
57.	Configuración de la interfaz de MODIS Reprojection Tools . .	90
58.	Modelo de geoprocésamiento para el filtrado del NDVI	94
59.	Toolbox del modelado cartográfico para el álgebra de mapas .	95
60.	Resultado mapas NDVI	97

Índice de tablas

1.	Recursos SIG para la selección de estaciones	38
2.	Listado de 56 estaciones meteorológicas y valor mensual de precipitación para el periodo marzo - octubre de 2016	39
3.	Cátalogo de productos MODIS NDVI	41
4.	Capas de información de los productos MOD13A3 y MYD13A3	42
5.	Archivos MODIS adquiridos para la investigación	43
6.	Medidas de tendencia, forma y distribución de las precipitaciones para el periodo marzo - octubre de 2016	47
7.	Medidas de tendencia, forma y distribución del NDVI para el periodo marzo - octubre de 2016	48
8.	Rango de valores del dataset Pixel Reliability	89
9.	Parámetros de la banda VI Quality	92
10.	Parámetros de calidad a extraer de la banda VI Quality	92

Nomenclatura

CO-SGSim: Co-Simulación Secuencial Gaussiana (Sequential Gaussian Co-simulation)

IDEAM: Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia

LDOPE: Land Data Operational Products Evaluation

MOD13A3: Productos MODIS enfocados al estudio de la vegetación, de generación mensual con resolución espacial de un kilómetro, obtenidos por el sensor TERRA

MODIS VI Products: Productos MODIS para el monitoreo y seguimiento de la vegetación

MODIS: Moderate-Resolution Imaging Spectroradiometer

MRT: MODIS Reprojection Tools

MYD13A3: Productos MODIS enfocados al estudio de la vegetación, de generación mensual con resolución espacial de un kilómetro, obtenidos por el sensor AQUA

NDVI: Índice de Vegetación Diferencial Normalizado (Normalized Difference Vegetation Index)

PDI: Procesamiento Digital de Imágenes

SGSim: Simulación Secuencial Gaussianas

SIG: Sistemas de Información Geográfica

Introducción

La disponibilidad de agua para usos con propósitos domésticos, de agricultura, o de generación energética, entre otros, requiere la cuantificación de la precipitación pluvial [Yarleque *et al*, 2007]. Su medición constituye la base para estudios enfocados al uso y control del agua [Aparicio, 1992] y su estimación es fundamental para la seguridad alimentaria, bajo el monitoreo de la agricultura y la producción pastoril [Rojas *et al*, 2011].

El monitoreo y la evaluación de la vegetación en la superficie terrestre es una de las aplicaciones más comunes en teledetección [Di *et al*, 1994]. Información obtenida con sensores remotos acerca de la vegetación, es ampliamente aprovechada para el estudio del crecimiento y vigor de la vegetación terrestre, y para el seguimiento de las dinámicas forestales [Xue & Su, 2017]. En dichas actividades, la respuesta espectral que presenta la vegetación, puede ser realizada mediante indicadores robustos como los Índices de Vegetación, los cuales se obtienen con la combinación de dos o más bandas de una imagen satelital (generalmente la roja e infrarroja) [Didan *et al.*, 2015]. De ellos, el NDVI (Normalized Difference Vegetation Index) es uno de los más utilizados, ya que suele emplearse como un buen indicador de las condiciones y la productividad de la vegetación, sirviéndose de la capacidad que esta tiene para absorber radiación fotosintéticamente activa [Wang *et al*, 2014].

La relación entre las precipitaciones y el NDVI se ha estudiado teniendo en cuenta patrones espaciales y temporales en distintas partes del mundo con distintos tipos de vegetación. Uno de sus enfoques ha sido estimar la cuantificación de la lluvia a partir de dicho índice de vegetación: Di *et al* (1991) desarrollaron un modelo matemático que relaciona la respuesta de la vegetación a los eventos de lluvia, monitoreando las precipitaciones a través del NDVI en la región de Sandhills, Nebraska (EEUU). Yarleque *et al* (2007) proponen la cuantificación de la proporcionalidad y periodicidad de la precipitación a partir de datos de NDVI para reconstruir datos de precipitaciones en series de tiempo en el Departamento de Puno en Perú. Rojas *et al* (2011), evaluaron la calidad de las estimaciones de la lluvia en zonas de África, basándose en la relación del NDVI con las precipitaciones, y establecieron que dicho índice de vegetación es un buen indicador para detectar exceso o déficit en la estimación de la pluviosidad a escala continental y regional.

La predicción, estimación y simulación de los fenómenos espaciales es el objetivo principal de la Geoestadística, rama de la estadística enfocada al contexto geográfico [Giraldo, 2002]. Analizando la continuidad espacial, que es una característica esencial de los fenómenos espaciales, es posible determinar la estructura de correlación espacial de una variable [Isaaks & Srivastava, 1989], dicha estructura de correlación, proporciona una serie de medidas cuantitativas con las que pueden estimarse o simularse valores de la variable en cualquier lugar de la región de interés, ya que la misma solo se conoce en ubicaciones dispersas donde se han realizado mediciones [Zhang, 2011]. Con la técnica de simulación de funciones aleatorias se pretende simular realísticamente un fenómeno, reproduciendo su variabilidad [Díaz Viera, 2002], y como resultado se tiene un modelo condicionado a los datos experimentales en el que las propiedades estadísticas y variabilidad espacial originales de los datos, es replicada [Rossi & Deutsch, 2014].

El presente estudio propone una metodología para generar mapas de distribución de precipitaciones basado en la relación lluvia - NDVI. Para ello se utilizan diversos procedimientos del ámbito de la Geomática concernientes al tratamiento de datos y se da un enfoque geoestadístico en el que se aplica la técnica de Co-Simulaciones Secuenciales Gaussianas para obtener mapas de precipitaciones mensuales.

1. Objetivos

1.1. Objetivo General

Construir mapas de distribuciones espaciales de la precipitación basados en la correlación espacial entre la lluvia y el NDVI.

1.2. Objetivos Específicos

1. Describir la variabilidad espacial de las precipitaciones y el NDVI en la zona de estudio.
2. Modelar el comportamiento de la lluvia y el NDVI en la región de estudio a partir del entendimiento de sus estructuras de continuidad espacial.
3. Definir una metodología de trabajo que incluya fases de adquisición, preparación y procesamiento de insumos, aplicable al estudio de las precipitaciones y su relación con la respuesta que la vegetación presenta a los eventos de lluvia.
4. Generar un flujo de trabajo cuyo propósito sea procesar productos MODIS-NDVI, para obtener una imagen libre de toda contaminación apropiada para el análisis de la relación entre la lluvia y el NDVI.

2. Marco Teórico

2.1. Precipitación

La precipitación se considera la principal fuente de agua en la superficie terrestre, por lo tanto, sus mediciones representan una importante base para estudios relacionados con el uso y control del agua [Aparicio, 1992]. Como precipitación, se considera a toda forma de humedad originada en las nubes que llega a la superficie terrestre, así, eventos como granizadas o nevadas representan formas del fenómeno de precipitación al igual que la lluvia [Chereque 1989].

La cantidad de lluvia que cae se determina con pluviómetros, que constan de un recipiente cilíndrico en cuya área transversal, se descarga el agua captada por un embudo dispuesto en la parte superior. El área de captación del embudo suele ser 10 veces mayor que el área del recipiente, de modo que por cada milímetro de lluvia se deposite un centímetro en el recipiente. Esto permite realizar lecturas directas de la altura de lluvia caída, la cual generalmente se expresa en milímetros y se consulta diariamente a la misma hora. [Aparicio, 1992]. Las mediciones también pueden efectuarse con pluviógrafos, que son instrumentos semejantes a los pluviómetros pero proporcionan registros continuos de las precipitaciones, por lo cual su lectura se realiza con intervalos de tiempo más espaciados. Además el agua es conducida a un depósito que oscila alrededor de un pivote y a partir del movimiento oscilatorio se registra un gráfico que indica la intensidad [Chereque 1989].

Las alturas de lluvia que caen en un sitio específico suelen diferir a las que caen en otros sitios incluso siendo cercanos. Los instrumentos de medición registran la lluvia caída de manera puntual, lo que significa que las mediciones corresponden con la lluvia caída exactamente en el lugar donde se encuentran instalados los instrumentos. Sin embargo, para distintos estudios y proyectos de ingeniería se hace necesario conocer el comportamiento medio de la lluvia asociado a una región de interés, para lo cual se dispone de distintas técnicas en hidrología [Aparicio, 1992].

2.2. Índices de Vegetación

La información que se adquiere acerca de la vegetación en percepción remota, se logra usando sensores pasivos (satélites) que obtienen el valor de la reflectancia de las ondas electromagnéticas de las cubiertas vegetales. Tal valor de reflectancia, depende de las características químicas y morfológicas de los órganos y las hojas de las plantas [Xue & Su, 2017].

Con el propósito de categorizar la información sobre la vegetación dispuesta en una imagen satelital se emplean índices de vegetación, que son indicadores obtenidos al aplicar operaciones matemáticas a los valores numéricos de los píxeles de una imagen, usando dos o más bandas espectrales de una misma escena [Muñoz, 2013]. Dichos índices, se fundamentan en la experiencia de observación de firmas espectrales típicas en las hojas, la cual ha determinado, que los pigmentos fotosintéticamente activos de las plantas generan una alta absorción de la energía que se refleja en la porción visible del espectro, mientras la radiación perteneciente a la franja del infrarrojo cercano, resulta dispersa a causa de una muy baja absorción. Los índices de vegetación, representan medidas del contraste entre las bandas roja e infrarroja cercana [Didan *et al.*, 2015].

De forma generalizada, los índices usan el cociente de la reflectancia entre la banda del rojo y la del infrarrojo cercano, ya que el contraste o diferencia en la reflectancia que presenta la clorofila en ambas bandas suele ser grande [Muñoz, 2013]. Cuando la diferencia entre la banda Roja y la Infrarroja Cercana es máxima, la vegetación es abundante, y por lo tanto, cuando esta diferencia resulta mínima la vegetación es escasa. [Didan *et al.*, 2015].

El Índice de Vegetación Diferencial Normalizado (NDVI) es el índice más conocido y utilizado, su respuesta es sensible a la vegetación verde incluso en áreas de cobertura vegetal escasa. Caracteriza el crecimiento o el vigor de la cubierta vegetal, lo que permite el estudio de su estructura y fotosíntesis tanto a nivel global como local [Xue & Su, 2017]. Su cálculo es una relación estandarizada entre la reflectancia del rojo y el infrarrojo cercano proporcionando un rango de valores entre -1 y 1. Se obtiene con la siguiente expresión:

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (1)$$

Donde:

NIR es la reflectancia en el canal infrarrojo cercano; *Red* es la reflectancia en el canal rojo.

El NDVI tiene la ventaja de reducir el ruido que puede ser introducido por algunas bandas a causa del ángulo solar, las nubes y su sombra, y la atenuación atmosférica [Didan *et al.*, 2015]. Además de minimizar efectos topográficos, proporciona una escala lineal de medición directa, en la que cero representa el valor aproximado donde hay ausencia de vegetación, valores bajos (cerca de 0.1) indican presencia de rocas, arena o nieve, valores entre 0.2 a 0.3 indican áreas con pasturas o escasa presencia de arbustos y a partir de que se alcanzan estos valores se tienen coberturas vegetales como praderas, cultivos, o forestaciones. Valores negativos corresponden con superficies sin ningún tipo de vegetación [Muñoz, 2013].

2.3. Relación Precipitaciones - NDVI

Tanto la lluvia como el NDVI son variables periódicas y proporcionales, ambos procesos están ligados a dinámicas de distinta índole como la acción humana y los fenómenos naturales. Mientras las precipitaciones tienen influencia sobre el estado de la vegetación, esta última también incide sobre el ciclo hidrológico. [Gómez, 2003]. La variación de factores climáticos como la precipitación y la temperatura tienen una marcada influencia en la variación del NDVI. Se debe a que inciden de manera directa en el balance de agua, y por ende, en la humedad del suelo, vital para el desarrollo de las plantas [Wang *et al.*, 2014]. La humedad del suelo, cuya fuente natural principal son las precipitaciones, contribuye en gran medida al crecimiento de la vegetación. Al analizar las condiciones generales que presenta la cobertura vegetal a través de índices de vegetación, el NDVI también puede utilizarse como un indicador útil de la humedad del suelo, especialmente en grandes extensiones de área donde no es posible conocer la humedad in situ de manera representativa [Di *et al.*, 1994]. El NDVI ha sido utilizado para estimar las precipitaciones en áreas donde no se cuenta con registros de lluvia, entendiendo la oferta de agua como uno de los principales controles ambientales que determinan la productividad primaria [De La Casa & Ovando 2006]. En el estudio de su relación con la lluvia se han encontrado tanto relaciones lineales como no lineales [Rojas *et al.*, 2011]

2.4. Geoestadística

La geoestadística consiste en la aplicación de métodos probabilísticos a variables regionalizadas [Delfiner & Chilès, 1999]. Su objetivo principal es la estimación, predicción y simulación de los fenómenos espaciales y su continuidad [Giraldo, 2002]. Al igual que en estadística clásica, se estudia la distribución estadística de un conjunto de datos muestreados, pero al tratarse de fenómenos espaciales, los métodos geoestadísticos permiten adicionar a esta distribución la correlación espacial que existe entre ellos [Zhang, 2011].

La continuidad espacial es una característica fundamental de muchos fenómenos naturales, la geoestadística describe dicha continuidad adaptando técnicas de regresión clásicas que permiten sacar provecho de su estudio [Isaaks & Srivastava, 1989], entendiendo, que si un fenómeno es continuo en la región que se estudia, la variable de interés puede ser medida en cualquier ubicación [Giraldo, 2002].

G. Matheron definió la geoestadística en 1962 como “la aplicación de la teoría de las Variables Regionalizadas a la estimación de procesos o fenómenos geológicos en el espacio”. Una variable regionalizada, en un concepto básico, es una variable que puede tener distintos valores en distintas localizaciones [Pircz & Deutsch, 2014], y en una definición más formal, es una variable aleatoria z definida en un punto del espacio x , expresada como una función $z(x)$ la cual presenta una estructura de correlación espacial [Díaz Viera, 2002]. Dichas variables aleatorias pueden representar la magnitud de una variable ambiental medida en un conjunto de coordenadas de la región de estudio [Giraldo, 2002].

Una variable aleatoria $z(\underline{x})$, puede corresponder a cada punto x que tenga dominio en el espacio, siendo así, el conjunto de variables aleatorias espacialmente distribuidas será una función aleatoria $Z(\underline{x})$. La variable regionalizada, estará constituida por la función espacial discreta Z' que se obtiene al tomar una muestra de dicha función aleatoria. [Díaz Viera, 2002]. En el proceso estocástico que define la variable regionalizada, el vector aleatorio expresado en la Ecuación 2 está definido por su función de distribución conjunta, si esa función de distribución se mantiene igual al aplicar una traslación definida por el vector h a los puntos (Ecuación 3), se establece que el proceso es estacionario, lo cual, en términos probabilísticos corresponde con la Ecuación 4 [Giraldo, 2002].

$$\vec{Z}(x) = [Z(X_1), Z(X_2), \dots, Z(X_n)] \quad (2)$$

$$\vec{Z}(x) = [Z(X_1 + h), Z(X_2 + h), \dots, Z(X_n + h)] \quad (3)$$

$$P\{Z(X_1) < z_1, \dots, Z(X_n) < z_n\} = P\{Z(X_1 + h) < z_1, \dots, Z(X_n + h) < z_n\} \quad (4)$$

La implicación física de una función aleatoria estacionaria, es que el proceso será homogéneo y se repetirá a sí mismo en el espacio [Delfiner & Chilès, 1999]. Así, todas las parejas de una variable aleatoria separadas una distancia específica (vector h) tienen la misma función de distribución conjunta [Isaaks & Srivastava, 1989]. Estableciendo el valor esperado de la función aleatoria como se muestra en la Ecuación 5, es decir, igualándolo a la media (m), se tendrá un valor finito y constante para todo punto en el dominio [Giraldo, 2002]. A tal condición se le conoce como estacionariedad de segundo orden y corresponde con el primer momento de la función aleatoria.

$$E[Z(x)] = m \quad (5)$$

Por lo general, un estudio geoestadístico se compone de distintas fases que comprenden (I) una examinación de los datos y sus características, conocida como análisis exploratorio, (II) la interpretación y modelamiento de su estructura de correlación espacial, denominada análisis estructural, y (III) la estimación y obtención de valores en zonas no muestreadas, llamada predicción o estimación. Dichas etapas se exponen a continuación:

I Análisis Exploratorio

En geoestadística, el análisis exploratorio tiene como propósito la obtención de información, en principio desconocida acerca del conjunto de datos que se estudia, con el apoyo de procedimientos estadísticos convencionales. La información inferida en esta fase, es fundamental para la toma de decisiones en etapas posteriores de un estudio geoestadístico [Díaz Viera, 2002]. La etapa se considera, como el punto de partida para el proceso de entendimiento de los datos y sus características más relevantes. [Isaaks & Srivastava, 1989].

Durante el análisis exploratorio se computan cálculos de estadística convencional univariada y bi-varida y se analizan gráficos como histogramas, box plot, o dispersogramas, entre otros. Con el apoyo de estos elementos, es posible detectar aspectos característicos de los datos, por ejemplo, una mezcla de poblaciones podría identificarse si se obtiene un histograma bimodal, y en dado caso, sería conveniente estudiar las poblaciones por separado [Delfiner & Chilès, 1999]. Al igual que en estadística clásica, todos los elementos descriptivos de análisis permiten identificar variabilidad, forma y observaciones atípicas en los datos. Adicionalmente, tratándose de datos georeferenciados, los valores de coordenadas pueden ser involucrados en los análisis, por ejemplo, de un gráfico de dispersión podría concluirse según la nube de puntos, sobre el comportamiento de los valores medios de la variable en la región, o si los datos podrían presentar alguna tendencia [Giraldo, 2002].

En el contexto espacial, recursos como un mapa de curvas de nivel o un gráfico 2D con etiquetas de valores, logran resaltar visualmente las propiedades de los datos respecto a su distribución geográfica, y aunque los valores aparentemente no parezcan tener ubicaciones aleatorias, valores altos tienden a estar cercanos a otros valores altos, mientras valores bajos tienden a la cercanía con otros valores bajos. De ahí pueden detectarse anomalías por valores agrupados, pero también, es posible determinar que existen casos atípicos en los que un solo valor alto puede estar rodeado de valores bajos y viceversa [Isaaks & Srivastava, 1989]. Los mapas y gráficos también pueden indicar la existencia de inconsistencias en el conjunto original de datos, ya que no contar datos depurados desde la etapa inicial es una situación común en los estudios geoestadísticos. Muestreo excesivo en alguna zona específica respecto al resto de la región de estudio, o la existencia de errores (groseros) en los valores de las coordenadas, son ejemplos de posibles errores a detectar en la etapa exploratoria [Delfiner & Chilès, 1999].

II Análisis Estructural

Las características estructurales de los datos espaciales se analizan teniendo en cuenta particularidades de las variables regionalizadas como que el fenómeno estudiado es generalmente único e irreproducible, que está definido en un dominio 2D o 3D, que suele ser muy complejo para descripciones determinísticas precisas, y que es conocido por la información que proporcionan los puntos de muestreo distribuidos en la región. Entender

cuestionamientos sobre el comportamiento de la variabilidad espacial en direcciones y escalas específicas, la continuidad estadística esperada del fenómeno, o lo que indica un valor en un punto respecto a sus vecinos, entre muchos otros planteamientos, hacen parte de la etapa de análisis estructural [Delfiner & Chilès, 1999].

La dependencia espacial entre las mediciones de la variable puede ser cuantificada utilizando funciones de correlación espacial basándose en el supuesto de que existe estacionariedad [Giraldo, 2002]. Para una función aleatoria estacionaria que es caracterizada por su media (Ecuación 5), se puede considerar la covarianza ($C(h)$) como se indica en la Ecuación 6 (segundo momento de la función), y expresarla de manera estandarizada en términos del coeficiente de correlación $\rho(h)$, mediante la Ecuación 7, en la que $C(0)$ representa la varianza.

$$C(h) = E \{ [Z(x) - m] * [Z(x + h) - m] \} \quad (6)$$

$$\rho(h) = \frac{C(h)}{C(0)} \quad (7)$$

Dado que h representa un vector de separación entre las posiciones (x) y ($x + h$), se puede aseverar que las funciones son totalmente dependientes de la distancia y la dirección [Delfiner & Chilès, 1999]. El vector de separación h suele también denominarse como “rezago” o “lag”.

Isaaks & Srivastava (1989) explican cómo la continuidad espacial puede ser analizada en la etapa exploratoria, con gráficos conocidos como dispersogramas de distancia (“h-scatterplot”), y cómo la relación entre el vector de separación y la función de covarianza define el covariograma, mientras la relación entre el mismo vector y el coeficiente de correlación, define el correlograma durante la etapa de análisis estructural. Pyrcz y Deutch (2014) definen el dispersograma de distancia como un gráfico cruzado entre los valores de “cabeza” y “cola” (head and tail) del vector de separación y detallan como los principios de estacionariedad y las relaciones entre la funciones de covarianza, correlación y semivarianza pueden ser interpretadas apoyándose en este tipo de gráficos, en los que cada lag representa la distancia, incremento, o espaciado entre dispersogramas sucesivos y su magnitud debe ser acorde al patrón de muestreo que se tenga.

El covariograma y correlograma son indicadores de cómo la covarianza y correlación varían en función del vector de separación, ambos tienen un comportamiento similar: decrecen a medida que la distancia de separación aumenta. Haciendo un símil con una serie de dispersogramas de distancia, se tiene que ambas funciones tienden a disminuir a medida que la nube de puntos se hace más ancha [Isaaks & Srivastava, 1989]. En la Figura 1, se observan a manera de ejemplo, gráficos típicos de covariograma (arriba) y correlograma (abajo), con una tendencia a que las dos funciones sean decrecientes a mayor distancia de separación.

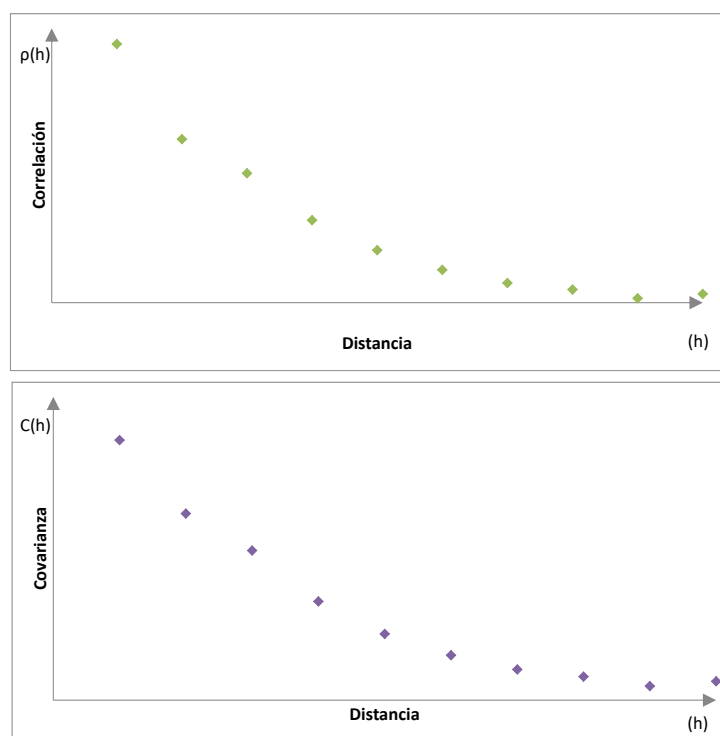


Figura 1: Gráficos esquemáticos de correlograma (arriba) y covariograma (abajo)

Una función aleatoria intrínseca (IRF - Intrinsic Random Function) es una función aleatoria cuyos incrementos son estacionarios de segundo orden [Delfiner & Chilès, 1999]. En este caso, la varianza de la variable regionalizada es finita [Giraldo, 2002]. A ese momento de segundo orden de

la función aleatoria, se le conoce como variograma y se expresa en términos probabilísticos según indica la Ecuación 8.

$$2\gamma(h) = E[(Z(x+h) - Z(x))^2] \quad (8)$$

El variograma es una herramienta que describe estadísticamente, la forma en que los valores en dos puntos difieren, a medida que la separación entre ellos aumenta. Constituye un instrumento simple para relacionar la incertidumbre con la distancia desde una observación, cuantificando la disimilitud de valores de la variable en las posiciones x y $(x+h)$ [Delfiner & Chilès, 1999]. Por efectos de practicidad el variograma es la más utilizada de las funciones de correlación. En los estudios geoestadísticos suele recurrirse directamente a ella prefiriéndola respecto a las otras funciones, ya que su cálculo no requiere de parámetros como sucede con el covariograma y correlograma [Giraldo, 2002]. Aunque las tres funciones se consideran equivalentes para caracterizar la correlación entre dos puntos como se observa en la Ecuación 9, el variograma suele ser una medida más robusta frente a cambios locales de la media. [Rossi & Deutsch, 2014].

$$C(h) = C(0) * \rho(h) = C(0) - \gamma(h) \quad (9)$$

Analizando la continuidad espacial, es habitual encontrar que muchos fenómenos pueden presentar mayor continuidad en alguna dirección concreta a comparación de otras [Isaaks & Srivastava, 1989]. Si el patrón de continuidad espacial es igual en todas direcciones entonces se tendrá isotropía, sin embargo, esto solo sucede ocasionalmente, siendo común encontrar variogramas que exponen fenómenos anisotrópicos [Rossi & Deutsch, 2014]. Como se detalla en la siguiente sección, analizar variogramas muestrales en direcciones concretas, una vez que se cuente con los parámetros estructurales generales, permite establecer conclusiones sobre las condiciones de anisotropía que pueda presentar el fenómeno estudiado.

Variograma muestral

La función de semivarianza para una distancia de separación permite establecer el variograma muestral (también conocido como variograma experimental) con el cálculo de la diferencia cuadrada promedio de los valores separados [Pircz & Deutsch, 2014]. Su cálculo y gráfico típico se indican en la Ecuación 10 y Figura 2 respectivamente:

$$\gamma(h) = \frac{1}{2n} * \sum_{i=1}^n (Z(x_i + h) - Z(x_i))^2 \quad (10)$$

Donde:

$\gamma(h)$ es la función de semivarianza.

$Z(x)$ es el valor de la variable en un sitio x .

$Z(x + h)$ es otro valor de la variable a una distancia h de x .

n es el número de parejas $Z(x)$ y $Z(x + h)$ separados una distancia h .

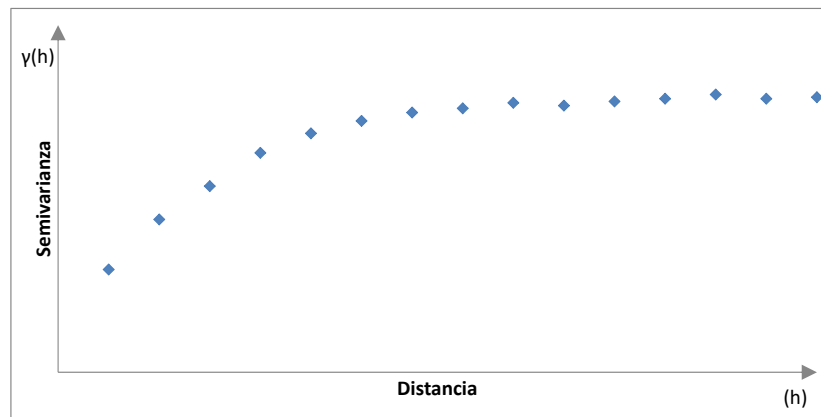


Figura 2: Gráfico típico del variograma experimental

Al igual que la media y la varianza, en un proceso estacionario, el variograma no depende de la localización absoluta de cada dato. Se le considera una medida de la variabilidad que aumenta su valor si el muestreo se hace más disímil [Rossi & Deutsch, 2014]. Debido a que en la realidad, los muestreos pueden llegar a ser muy irregulares, se acostumbra a utilizar intervalos de distancia de modo que el variograma experimental corresponda con un valor de distancia promedio entre puntos y no con una distancia fija, entonces, se totalizará un número de parejas presentes por intervalo [Giraldo, 2002]. El uso de tolerancias en distancia y también en ángulo, proporcionará a los rezagos espaciales medidas razonables para que el número de parejas considerado en cada distancia y dirección sea suficiente para el cálculo del variograma. Dichas tolerancias definirán los sectores de acción de cada vector de separación [Rossi & Deutsch, 2014]. El gráfico de la Figura 3 ilustra los conceptos de tolerancias en distancia y ángulo para un vector de

separación. La zona sombreada representa la región de acción del vector, la cual se construye a partir de la longitud, dirección, tolerancias, y ancho de banda. Cualquier dato muestral ubicado en dicha zona será emparejado con la observación (x, y) desde la que se origina el vector.

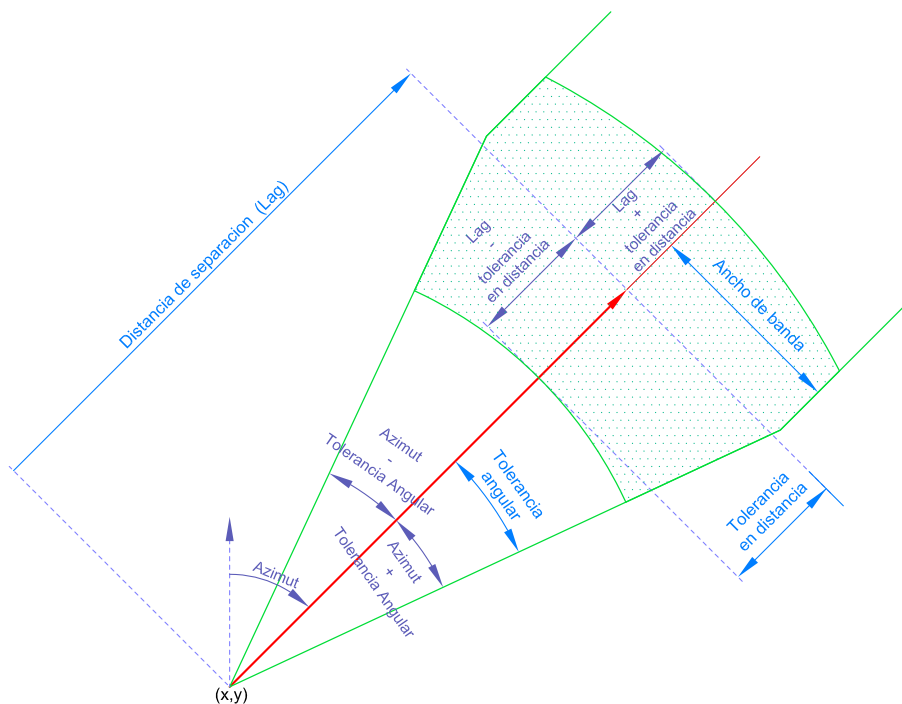


Figura 3: Tolerancias del vector de separación

Para lograr medidas confiables de la variabilidad espacial versus la distancia y la dirección, en la construcción del variograma experimental se debe tener en cuenta: el entendimiento y visualización de los datos en la perspectiva que corresponde, la utilización de un sistema de coordenadas apropiado (generalmente el mismo que se usará para modelar), y la consideración de valores extremos así como una posible transformación de los datos [Rossi & Deutsch, 2014].

Un único variograma que combine todas las posibles direcciones tendrá una tolerancia angular suficientemente grande, por lo cual, en el vector de separación será relevante la magnitud y no la dirección. En ese caso se tendrá un variograma omnidireccional, que en una definición simple, puede

asemejarse al promedio de los variogramas individuales para cada dirección, aunque no se trate de un promedio estricto. Su interpretación, no obedece a que se tenga la misma continuidad espacial en todas las direcciones, pero es el punto de partida para establecer los parámetros necesarios para un variograma experimental, y dado que la dirección no tiene importancia en un variograma de este tipo, la prioridad debe apuntarse a encontrar los parámetros óptimos de la distancia, es decir el incremento en cada rezago y la tolerancia, con los que se logra la estructura de variograma más clara [Isaaks & Srivastava, 1989].

El máximo de rezagos a utilizar debe buscarse de modo que el valor máximo de la distancia de los rezagos sea menor que la mitad de la máxima distancia posible entre un par de puntos. Distancias mayores a esta condición, excluirán puntos del cálculo y por lo tanto se tendrá un variograma con pocos puntos que no resultará válido. Generalmente la tolerancia en distancia es la mitad del valor del rezago, en algunas situaciones puntuales como cuando se dispone de pocos datos se consideran valores mayores. Los valores de tolerancia angular e intervalo de distancia se definen en un proceso de prueba y error. Variogramas ruidosos pueden ser el resultado de la utilización valores pequeños de tolerancia, a su vez, valores de tolerancia muy grandes, producirán un variograma que tienda a ser promediado por lo que puede terminar siendo impreciso. Las tolerancias de distancia y ángulo deben procurarse lo más pequeñas posibles, siempre que se logre un variograma estable y una buena definición de la continuidad direccional [Rossi & Deutsch, 2014].

Las características más importantes de un variograma se detallan a continuación y se ilustran en la Figura 4:

- **Meseta:** Es el límite del variograma cuando la distancia h tiende a infinito, puede definirse como la varianza máxima y representa la cota superior del variograma.
- **Rango:** Es el valor del intervalo de distancia a partir del cual el variograma no se incrementa, se interpreta que la dependencia entre un par de observaciones deja de existir cuando este intervalo es alcanzado, por lo que su definición es comparable a la de una zona de influencia.
- **Efecto Pepita:** Caracteriza una discontinuidad del variograma en el origen que puede significar una escala de variabilidad en intervalos de

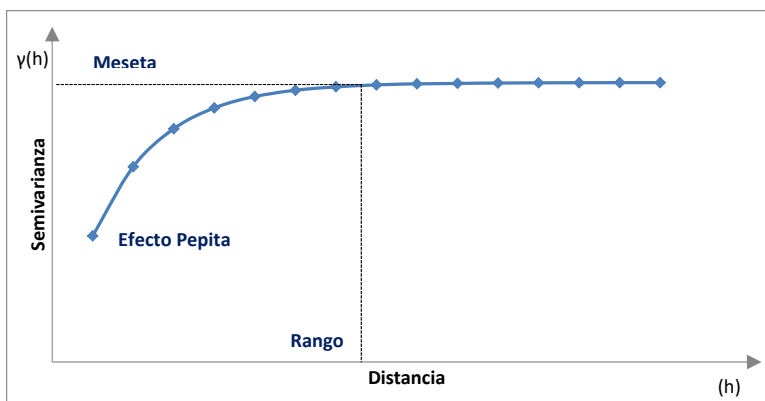


Figura 4: Elementos del variograma

distancias menores a los considerados. Su presencia puede atribuirse a errores de medición o de escala.

Una vez se considera que el variograma omnidireccional es aceptable, se procede al estudio de la continuidad espacial en direcciones puntuales generando variogramas direccionales, con los que no solo se establecerá si el proceso es o no isotrópico, sino también bajo qué condiciones anisotrópicas se comporta la variabilidad [Isaaks & Srivastava, 1989].

La anisotropía es geométrica cuando la meseta (es decir la varianza) se alcanza a diferentes distancias en diferentes direcciones. Por otra parte, cuando la meseta esperada no se alcanza en ninguna de las distancias consideradas para el cálculo del variograma, la anisotropía será zonal [Rossi & Deutsch, 2014]. La práctica común para detectar si existe anisotropía geométrica consiste en analizar mínimo cuatro direcciones, calcular variogramas en cada una de ellas, y decidir sobre su presencia con la estimación de los rangos [Díaz Viera, 2002]. Para un mismo valor de semivarianza, que generalmente es la meseta, de cada variograma direccional puede obtenerse su intercepto con el valor de distancia, y graficando dichos valores de distancia en un diagrama de rosas, se conformará una elipse con la que podrán identificarse las direcciones de máxima y mínima continuidad espacial que representan los ejes de anisotropía geométrica [Isaaks & Srivastava, 1989]. La Figura 5 muestra un gráfico ilustrativo de variogramas que se analizan en cuatro direcciones diferentes, cada uno trazado con un color: 0 grados en celeste, 45 grados en verde, 90 grados

en púrpura, y 135 grados en rojo. Para cada variograma, se identifica la distancia a la que se alcanza el mismo valor de semivarianza (C). Posteriormente, cada valor de longitud se gráfica en su respectiva orientación y con su correspondiente color, lo que permite construir esquemáticamente el diagrama de rosas de la Figura 6a y deducir que los ejes de máxima y mínima continuidad se dan en las direcciones de 0 y 90 grados respectivamente, y que dichos ejes conforman la elipse de anisotropía de la figura 6b indicando que se tendrá una mayor variabilidad hacia el norte o hacia el sur.

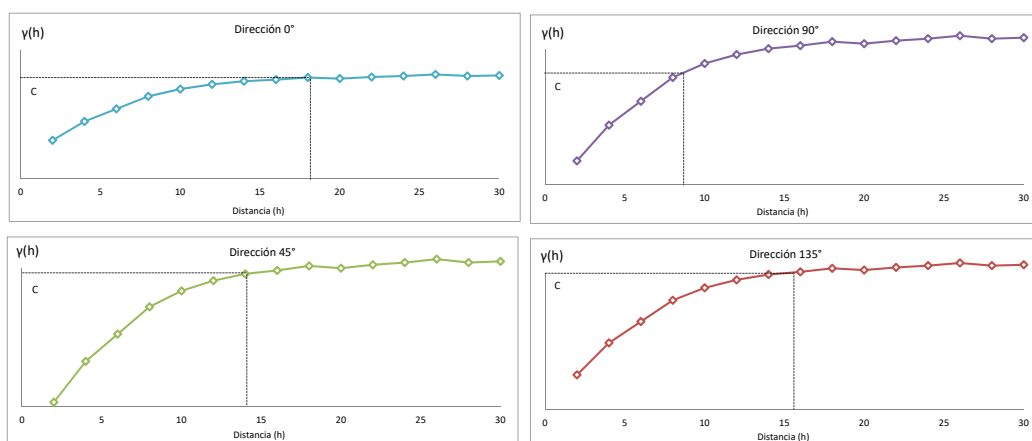


Figura 5: Ejemplo de variogramas direccionales con igual valor de varianza a diferentes distancias

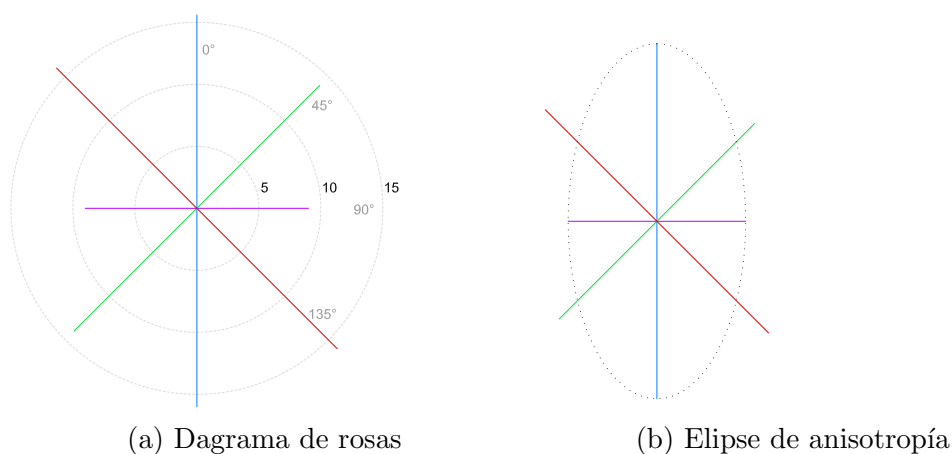


Figura 6: Ejes de máxima y mínima continuidad

Modelos teóricos de variograma

Dado que el variograma experimental se calcula para intervalos de distancia específicos, modelos teóricos de variograma son ajustados de modo que la estructura de correlación espacial observada en el variograma experimental se generalice a cualquier distancia. Así, se tendrá el fundamento que requieren las técnicas de modelamiento y predicción espacial como el kriging, las cuales demandan que en la región de estudio, dicha estructura de correlación se conozca en cualquier distancia posible. Existe gran variedad de modelos teóricos que pueden ser utilizados para ajustar el variograma experimental, en todos ellos la meseta, el rango y el efecto pepita son parámetros comunes [Giraldo, 2002].

Modelos paramétricos tradicionales de variograma teóricos son ajustados a los puntos del variograma muestral con funciones específicas definidas positivamente. Los modelos más utilizados son el esférico, exponencial y gaussiano. [Pircz & Deutsch, 2014]. Forman parte del grupo de modelos de transición acotados, su cálculo se detalla a continuación y en la Figura 7 se observa una comparación gráfica entre ellos:

- **Esférico:**

$$\gamma(h) = \begin{cases} 1,5(h/a) - 0,5(h/a)^3, & \text{si } h \leq a. \\ 1, & \text{de otro modo.} \end{cases} \quad (11)$$

- **Exponencial:**

$$\gamma(h) = 1 - \exp(-3h/a) \quad (12)$$

- **Gaussiano:**

$$\gamma(h) = 1 - \exp(-3(h/a)^2) \quad (13)$$

Donde:

a : es el rango, h : es la distancia

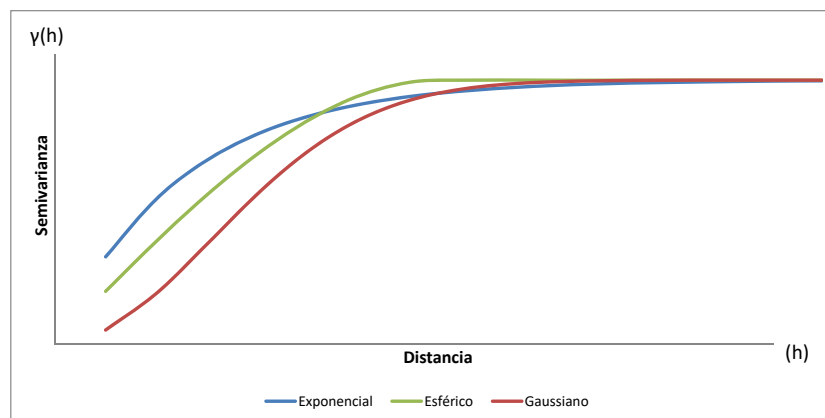


Figura 7: Modelos teóricos de variogramas

Los modelos de variograma en general, son sistemáticamente ajustados a la meseta teórica. Pueden ser construidos por más de una estructura de semivarianza anidando las funciones. De ser así, se tiene una suma definida positiva donde cada estructura aporta a la suma de la varianza, lo que garantizará que las ecuaciones en el kriging puedan ser resueltas y que su varianza sea positiva. Por simplicidad muchas veces se opta por utilizar solamente una estructura, aunque el uso de más una puede ser útil para propósitos específicos como cuantificar la anisotropía zonal [Pircz & Deutsch, 2014]. La posibilidad de usar más de una estructura obedece a que cualquier combinación lineal de modelos positivos de variograma, con coeficientes positivos, es también un modelo positivo, siendo cada estructura del modelo anidado un término de la combinación lineal. Usar un modelo anidado requiere evaluar cuales de los modelos teóricos representan mejor la forma general del variograma experimental considerando aspectos como la meseta, el comportamiento cerca al origen, y que la combinación de modelos de distintos tipos puede ser empleada [Isaaks & Srivastava, 1989].

III Estimación

En general, los muestreos son una pequeña representación de la realidad, por ende, se debe recurrir a modelos matemáticos que puedan predecir los valores de las variables en sitios donde por distintos motivos no han sido medidas. Tales modelos, usualmente se apoyan en los valores de los datos

y permiten inferir de manera determinística, los atributos de la variable en aquellas posiciones que no cuentan con muestreo. Dichas estimaciones proporcionan valores que en promedio son lo más cercano posible a los valores desconocidos, es decir a la realidad, priorizando los datos localmente, lo que resulta útil para detectar tendencias [Rossi & Deutsch, 2014].

En ciencias de la tierra, uno de los problemas más comunes, es obtener un mapa de una variable regionalizada a partir de datos de muestreo limitados. Históricamente se recurrió a prácticas como la construcción de mapas de contornos calculados manualmente, los cuales proporcionaban información para la observación de tendencias, incertidumbre, o condiciones de anisotropía, entre otras características. Los primeros algoritmos de computación se basaron en el principio de cálculo de dichos mapas análogos y con ellos se empezaron a obtener mapas suavizados de los que podía interpretarse la misma información. [Pircz & Deutsch, 2014]. El mapa de contornos, es un ejemplo de la representación alternativa de las muestras dispersas observadas en posiciones puntuales, como también los son los modelos digitales de elevación (DEM), o modelos de superficie, entre muchas otras posibilidades. Para lograr estas representaciones, se recurre a la interpolación espacial, que consiste en predecir valores exactos de atributos en ubicaciones no muestreadas, a partir de aquellos puntos donde se realizó una medición. Basándose en la ley de la Geografía, en la que: “Todo está relacionado con todo, pero las cosas cercanas se relacionan más que las distantes”, distintos métodos de interpolación pueden ser utilizados para la predicción espacial, en los que la definición del concepto de “cerca” es la principal diferencia entre ellos [O’Sullivan & Unwin, 2003]. Por ejemplo, para estimar un valor localmente, se puede recurrir a combinaciones lineales de los datos medidos, a las que asignan ponderaciones determinadas por la distancia entre el punto a estimar y las muestras cercanas, y también por el posible agrupamiento espacial que tengan las ubicaciones de dichas muestras [Isaaks & Srivastava, 1989].

La Figura 8 representa esquemáticamente una región donde una variable ha sido medida en posiciones conocidas, simbolizadas por puntos en escala de color con valores entre 1 y 10, y posiciones que no cuentan con mediciones en las que se desea estimar el valor de la variable, representadas por puntos de color negro. A modo de concepto y sin involucrar el modelamiento geoestadístico en esta instancia, a continuación se explican algunos métodos de estimación tradicionales junto con su interpretación gráfica:

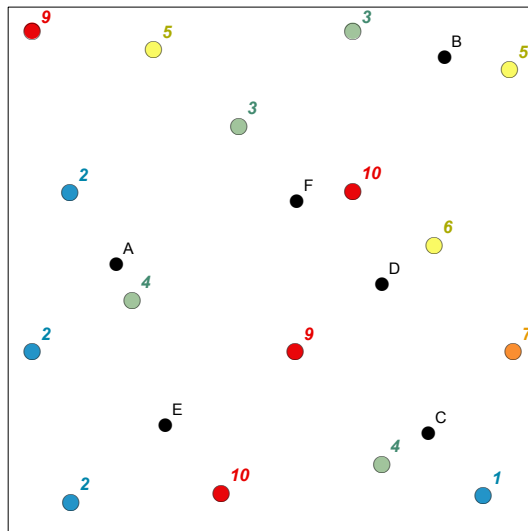


Figura 8: Representación del concepto de estimación para el valor de una variable en el espacio

Polígonos de proximidad

El método de estimación por polígonos de proximidad consiste en construir un área de influencia para cada observación, definida por el trazo de bisectrices perpendiculares que unen la observación central con el resto de observaciones. Cualquier ubicación que se localice dentro del polígono de influencia tomará el valor de la observación central, que además corresponderá con la observación más cercana [Isaaks & Srivastava, 1989]. Aunque el método es simple, puede tener saltos abruptos entre polígonos adyacentes con lo cual pueden presentarse discontinuidades. Es un método útil para variables categóricas [O'Sullivan & Unwin, 2003]. En la Figura 9a, los polígonos de proximidad se han trazado asignando a cada observación su correspondiente área.

Media espacial local

El método de la media espacial local, a diferencia de los polígonos, considera varias observaciones en lugar de asignar todo el peso a una sola. Las observaciones que considera son aquellas que se encuentren dentro de un radio fijo desde la ubicación donde se quiere hacer la estimación y con el promedio de todas ellas se obtiene el valor estimado. La desventaja es que

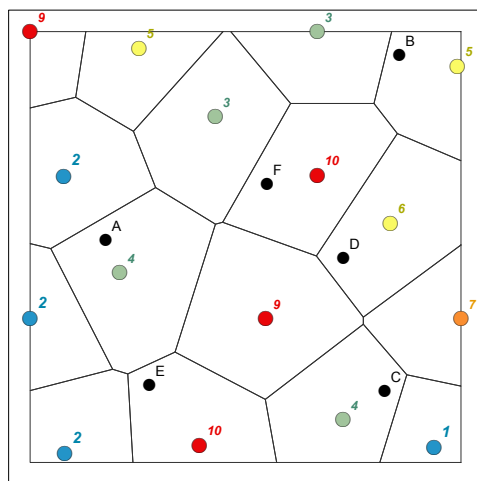
las observaciones abruptamente quedan por dentro o por fuera del cálculo, con lo cual, dependiendo de las distancias, podría perderse la continuidad o negarse la posibilidad de obtener una superficie representativa de la región que se estudia [O'Sullivan & Unwin, 2003]. En el ejemplo de la Figura 9b, se tiene una distancia fija desde la posición E, los puntos dentro de la zona radial sombreada son los seleccionados para calcular el promedio.

Triangulación

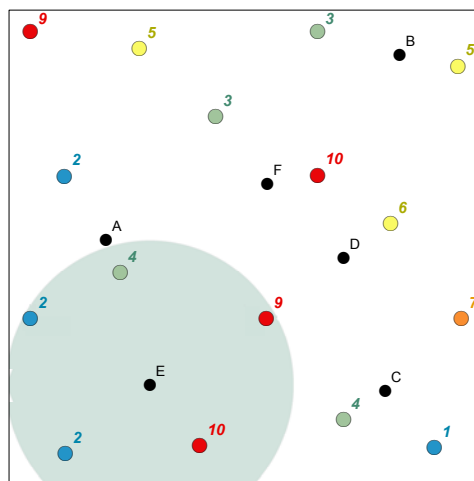
Discontinuidades entre puntos adyacentes pueden ser evitadas con el método de triangulación, el cual ajusta un plano construido por tres puntos observados que rodean la ubicación a estimar. Con la ecuación del plano ($z = ax + by + c$) se construye un sistema de tres ecuaciones donde a partir de las coordenadas (x, y) y el valor de la variable (z) se obtiene una nueva ecuación del plano que pasa por los tres puntos muestrales seleccionados. Sustituyendo en ella los valores de las coordenadas de la ubicación donde se está interpolando, se obtendrá el valor estimado [Isaaks & Srivastava, 1989]. En la Figura 9c, la posición F pretende ser estimada por triangulación construyendo un triángulo entre sus observaciones cercanas.

Vecinos más cercanos

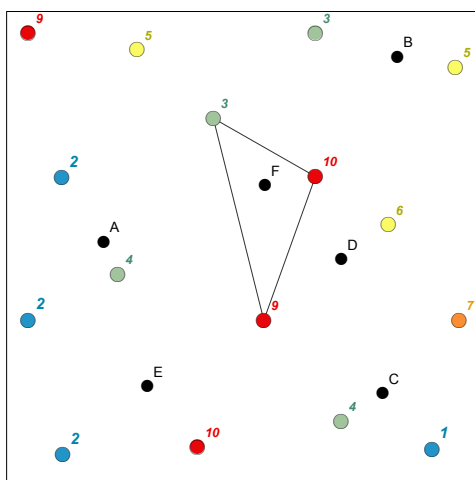
Similar a la media espacial local, el método de vecinos más cercanos estima un valor, promediando una cantidad fija de los puntos de muestreo que se encuentren más próximos. Cualquier ubicación que no cuente con valores observados puede tener el mismo número de vecinos cercanos que otra, con lo cual, todas las ubicaciones deseadas podrían ser estimadas. Sin embargo, debe tenerse en cuenta que el resultado es influenciado por la densidad de muestreo [O'Sullivan & Unwin, 2003]. En el ejemplo de la Figura 9d, desde la posición B se han definido los tres vecinos más cercanos para su estimación.



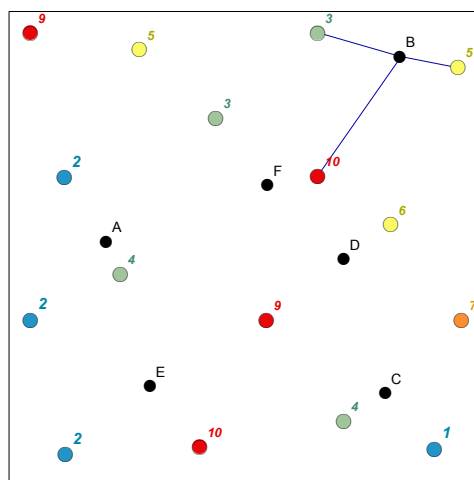
(a) Polígonos de proximidad



(b) Media espacial local



(c) Triangulación



(d) Vecinos cercanos

Figura 9: Interpretación gráfica de métodos de estimación local

Métodos basados en distancia

Métodos basados en distancias proporcionan una mejora en la estimación, ya que en lugar de atribuir el mismo peso a todas las observaciones consideradas, se asignan pesos a la distancia que separa el punto a estimar de cada una de las muestras. Las ponderaciones se asignan de modo

que el peso sea inversamente proporcional a la distancia de separación [Isaaks & Srivastava, 1989]. Para valores de distancia grandes, es decir donde el punto de control (observación) está distante, se recibe un peso pequeño, mientras que para puntos de control a distancias cortas el peso es alto. Así, las ubicaciones más cercanas cobran mayor importancia en el cálculo de una media local. Los pesos toman valores entre 0 y 1, y se expresan como función de la distancia, su estandarización indica que su sumatoria es igual a 1 [O'Sullivan & Unwin, 2003]. Cada peso se obtiene con la Ecuación 14.

$$w_{ij} = \frac{1/d_{ij}}{\sum_{i=1}^n 1/d_{ij}} \quad (14)$$

Donde:

w_{ij} es el peso asignado a la distancia de separación d_{ij} , entre la posición i (observación) y la posición j (a estimar); n es el número de pesos a obtener.

Con base en la formula, los pesos también pueden asignarse de manera proporcionalmente inversa a cualquier potencia de la distancia. En comparación, serán similares a medida que la potencia tienda a cero y disimilares a medida que tienda a infinito. La elección es arbitraria aunque comúnmente se eleva al cuadrado [Isaaks & Srivastava, 1989]. En Sistemas de Información Geográfica, los métodos de distancias ponderadas son frecuentemente utilizados para generar superficies de interpolación o mapas de contornos. Para ello se define una cuadrícula de interpolación, y para cada punto (nodo) de la cuadrícula se estima su correspondiente valor. Debe tenerse en cuenta que hay aspectos clave que influirán en el resultado, como lo fino o basto que se especifique el espaciado de la cuadrícula; la forma en que se defina la vecindad de búsqueda, la cual puede estar dada por distancias fijas (radios de búsqueda) o un número de definido de vecinos cercanos; y la valoración del peso elevando la distancia a alguna potencia [O'Sullivan & Unwin, 2003]. La estimación en un punto sin observaciones, se obtiene según la Ecuación 15

$$\hat{z}_j = \sum_{i=1}^n w_{ij} z_i \quad (15)$$

Donde:

\hat{z}_j es la estimación de la variable en la posición j ; w_{ij} el peso asignado a la distancia entre la ubicación a estimar y la observada; z_i el valor observado para la variable en la posición i .

La Figura 10 muestra un ejemplo de cómo se establecen los pesos se según el valor de la distancia a las cuatro observaciones más próximas:

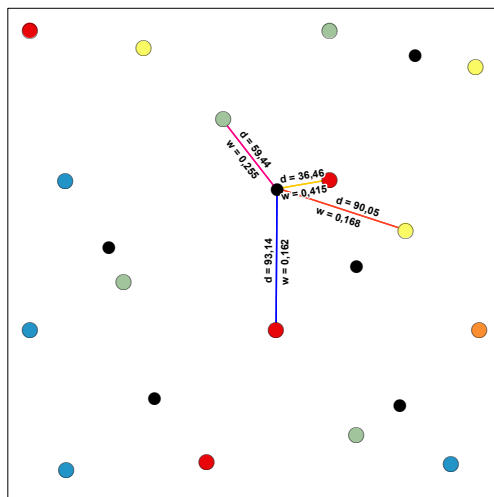


Figura 10: Estimación por distancia inversa ponderada

Métodos de interpolación como los anteriormente descritos, no consideran la correlación espacial de los datos, y justo allí radica el potencial de los métodos geoestadísticos, donde la correlación espacial es incorporada en la estimación junto con las relaciones de proximidad. La estimación por kriging, se basa en los datos muestrales y en un modelo de variograma que representa con precisión dicha correlación espacial [Zhang, 2011].

Kriging

El kriging recopila un conjunto de procedimientos aplicables en predicción espacial cuya característica fundamental es minimizar el error medio cuadrático de la predicción [Giraldo, 2002]. Como técnica de estimación local, proporciona el mejor estimador lineal insesgado de una variable desconocida que está siendo estudiada [Díaz Viera, 2002]. Es “Lineal”, porque su estimación es una combinación lineal ponderada de los datos disponibles (Ecuación 16), “insesgado” porque el error medio intenta ser igualado a cero (Ecuación 17), y “mejor” porque se enfoca en minimizar la varianza de los errores (Ecuación 18) [Isaaks & Srivastava, 1989]. Su metodología está basada en calcular los pesos que minimicen el error esperado

de la varianza, los cuales son condicionados para que sumen uno, lo que garantiza el insesgamiento (Ecuación 19), y por lo tanto la varianza del predictor es igual a la varianza de la variable [Rossi & Deutsch, 2014].

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z_i \quad (16)$$

$$E[Z^*(x_0)] = E[Z(x_0)] \quad (17)$$

$$\min \{Var [Z(x_0) - Z^*(x_0)]\} \quad (18)$$

$$\sum \lambda_i = 1 \quad (19)$$

Donde:

$Z^*(x_0)$ es el valor estimado para la variable en una posición donde no se tiene medición, λ es el peso asignado al valor Z de la variable para la posición i en la que fue medida.

En los distintos tipos de kriging que existen, las principales diferencias entre los métodos consisten en asumir si el dominio de la media es local o estacionario condicionando el conjunto de pesos, por ejemplo, en el kriging simple, se busca minimizar la varianza del error sin restricciones en los pesos y la media es asumida como una constante conocida que proviene del muestreo. El kriging ordinario, condiciona que los pesos sumen uno, por lo cual no es necesario asumir un valor para la media. El kriging universal (conocido también como kriging con tendencia) considera un modelo de función aleatoria que expresa una suma de tendencias y un residual [Rossi & Deutsch, 2014]. Los distintos tipos de kriging pueden ser clasificados según el tipo de estimador, el soporte de la medición de los datos, o su parametrización. A continuación se destacan clasificaciones para los tipos de kriging más comunes [Díaz Viera, 2002]:

1. Según tipo de estimador:

- **Lineal:** Kriging Simple, Kriging Ordinario, Kriging Universal, Kriging Residual.
- **No Lineal:** Kriging Disyuntivo, Kriging Indicador, Kriging Probabilístico.

2. Según soporte de la medición de los datos:

- *Puntual*
- *En bloques*

3. Según parametrización:

- ***Paramétrico:*** Kriging Multigaussiano, Kriging Disyuntivo, Kriging Lognormal.
- ***No Paramétrico:*** Kriging Simple, Kriging Ordinario, Kriging Universal, Kriging Residual, Kriging Indicador, Kriging Probabilístico.

El kriging ordinario, propone que el valor de la variable de interés Z en una posición x_0 donde no fue medida, puede estimarse con una combinación lineal de la variable medida en las n posiciones x_1, x_2, \dots, x_n , es decir, con los valores de las n variables aleatorias $Z(x_1), Z(x_2), \dots, Z(x_n)$ como se indica en la Ecuación 20, en la que λ representa las ponderaciones (pesos) de los valores originales [Giraldo, 2002]:

$$Z^*(x_0) = \lambda_1 Z(x_1) + \lambda_2 Z(x_2) + \dots + \lambda_n Z(x_n) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (20)$$

La varianza de la estimación, a la que se le conoce como varianza de kriging, proporciona una medida del error asociado con el estimador. Depende directamente de las ubicaciones y no de los valores de los datos [O'Sullivan & Unwin, 2003]. Será más grande a medida que más redundancia de datos se tenga; más pequeña, si las estimaciones se encuentran más próximas a las muestras; e igual a la varianza ($C(0)$) si todos los datos están muy distantes para recibir algún peso, por lo que la estimación se asume conocida igualándola a la media local de la forma $Z^*(x_0) = m(x_0)$. La varianza de la estimación depende directamente del modelo de covarianza o variograma [Pircz & Deutsch, 2014]. Considerando un estimador lineal como el de la Ecuación 16, su error de varianza estará definido por la Ecuación 21.

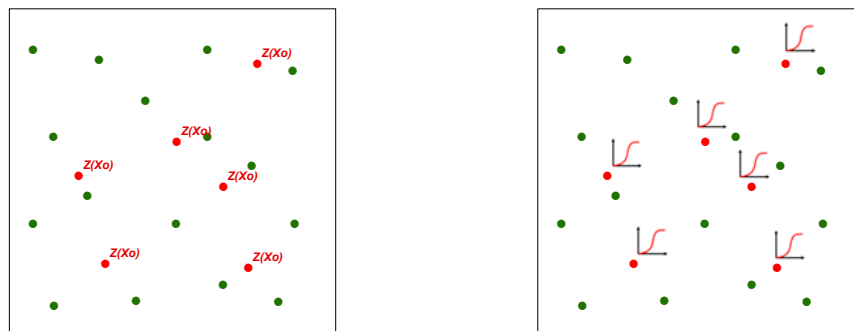
$$\sigma_E^2(x_0) = E [(Z^*(x_0) - Z(x_0))^2] \quad (21)$$

Como característica de la predicción espacial y del kriging ordinario se tiene que en la práctica, el error medio y varianza del error no se conocen hasta que se intente realizar una estimación, por lo tanto, para que el error medio sea igualado a cero y la varianza del error pueda ser minimizada, se recurre a trabajar con el error medio y la varianza de error del modelo que ha sido construido para los datos asignando pesos a las ubicaciones de muestreo cercanas, calculando los pesos en función de la distancia entre los puntos muestreados y la ubicación a estimar [Isaaks & Srivastava, 1989].

IV Simulaciones Condicionales

Teniendo en cuenta que un fenómeno espacial puede ser representado mediante una función aleatoria $Z(x)$, y que generalmente la información que se conoce de dicha función (caracterizada por las mediciones) es limitada, las estimaciones espaciales pueden resultar insuficientes y algo imprecisas, incluso recurriendo a estimadores como kriging, en especial, para ciertas aplicaciones que requieren cálculos detallados de las dispersiones. Una opción para estimar la realidad con suficiente detalle es simularla en base a un modelo [Díaz Viera, 2002]. Las simulaciones condicionales geoestadísticas son una herramienta útil para obtener representaciones realistas de la variabilidad espacial de un fenómeno, se basan en la teoría de los métodos de Montecarlo, y permiten evaluar el impacto de la incertidumbre geográfica en los resultados del modelamiento de procesos complejos [Delfiner & Chilès, 1999]. A partir de ellas, se construyen modelos que reproducen bajo el condicionamiento de los datos observados, el histograma y las medidas de continuidad espacial modeladas originalmente, logrando respetar las características de la variable espacial [Rossi & Deutsch, 2014].

Mientras el objetivo de la estimación es obtener en cada punto x un valor estimado $Z^*(x)$ tan cercano como sea posible del valor real desconocido, el de la simulación, es reproducir las propiedades estadísticas de la función aleatoria $Z(x)$ [Díaz Viera, 2002]. Tal situación se muestra en la Figura 11 donde los puntos verdes representan datos observados, y los rojos, puntos en los que se desea obtener un valor. La estimación se obtiene según la Ecuación 22 y la simulación con la Ecuación 23, donde la probabilidad estimada (P^*) de que la variable (Z), medida en una posición (x_0), es menor o igual que un umbral de la variable condicionado a n datos ($z|(n)$), y está determinada por la función de distribución acumulada (F^*).



(a) Estimación geoestadística

(b) Simulación geoestadística

Figura 11: Comparación conceptual entre simulación y estimación

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i * Z(x_i) \quad (22)$$

$$P^* \{Z(x_0) \leq z | (n)\} = F^*(x_0; z | (n)) \quad (23)$$

A diferencia de la estimación, en la simulación se preservan los valores extremos evitando resultados suavizados, y aunque los datos también se priorizan localmente, la variabilidad original se replica, permitiendo además, evaluar el nivel de incertidumbre [Rossi & Deutsch, 2014]. Adicionalmente, las fluctuaciones en escalas locales son también reproducidas, por lo que se logran mejores representaciones de las heterogeneidades, manteniendo consistencia con los datos observados [Delfiner & Chilès, 1999]. La Figura 12 ilustra conceptualmente el resultado obtenido con estimación y su comparación con el que se obtiene usando simulaciones. Posteriormente, la figura 13 muestra como ejemplo un mapa obtenido con cada método.

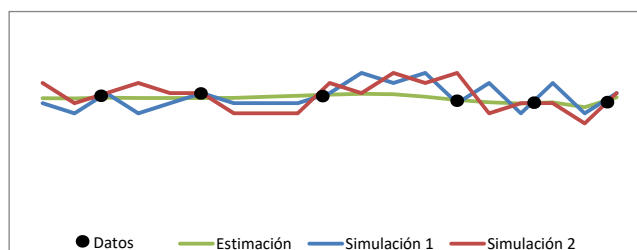
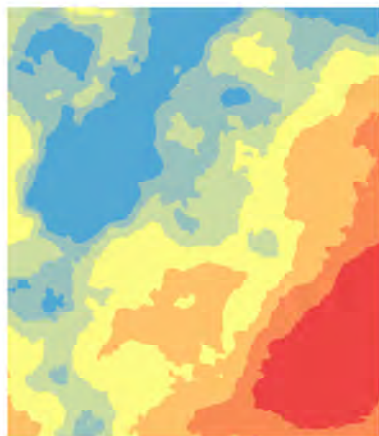
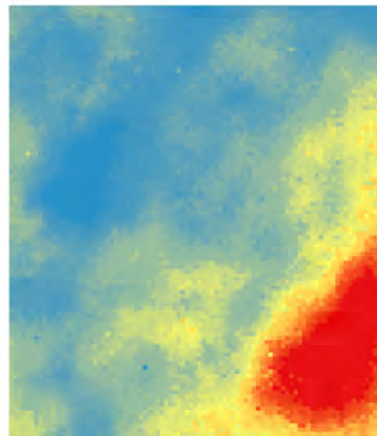


Figura 12: Interpretación gráfica del resultado de simulación y estimación



(a) Resultado estimación



(b) Resultado simulación

Figura 13: Comparación de mapas resultantes con estimación y simulación

En la simulación se busca construir una realización de una función aleatoria espacial que tenga el mismo valor esperado, covariograma y función de covarianza que la función aleatoria original. De ahí, pueden inferirse las propiedades estadísticas esperadas para la función aleatoria, y luego, dichas propiedades pueden seguirse replicando con nuevas realizaciones simuladas que sean estadísticamente equivalentes, lo que significa que la función de distribución de probabilidad será igual tanto para la función aleatoria como para cada función simulada [Díaz Viera, 2002]. Este procedimiento define la técnica de simulaciones condicionales, donde la realización es seleccionada aleatoriamente del subconjunto de realizaciones que coinciden con los puntos muestreados [Delfiner & Chilès, 1999].

Cada realización está condicionada a los datos experimentales ya que en estas ubicaciones, los valores simulados deben coincidir con los valores observados, y de acuerdo con el kriging los valores interpolados en los puntos de observación son iguales a los valores medidos [Díaz Viera, 2002]. Una simulación, no debe interpretarse como la realidad, sino como una de sus posibles versiones entre muchas otras. Dado que pueden tenerse infinitas versiones de la función aleatoria, se espera que algunas de ellas tengan los mismos valores que los que tiene la variable en los puntos donde se han realizado mediciones, y por lo tanto la variable regionalizada sea

fielmente representada [Delfiner & Chilès, 1999]. El modelo de simulación debe exponer acertadamente la media, la varianza, y la proporción de valores altos y bajos, entre otras medidas estadísticas, así como características del variograma tales como la anisotropía y el efecto pepita relativo [Rossi & Deutsch, 2014]. Mientras los métodos de estimación proporcionan un único mapa como resultado, la simulación puede producir varios mapas de igual probabilidad y propiedades de distribución de los datos, generados con el mismo modelo de correlación espacial que es usado en kriging [Zhang, 2011]. En la Figura 14 en la parte izquierda, se representa una variable espacial junto con su histograma, y en la parte derecha, se representa el resultado de cinco realizaciones cuyos histogramas honran la distribución original de los datos.

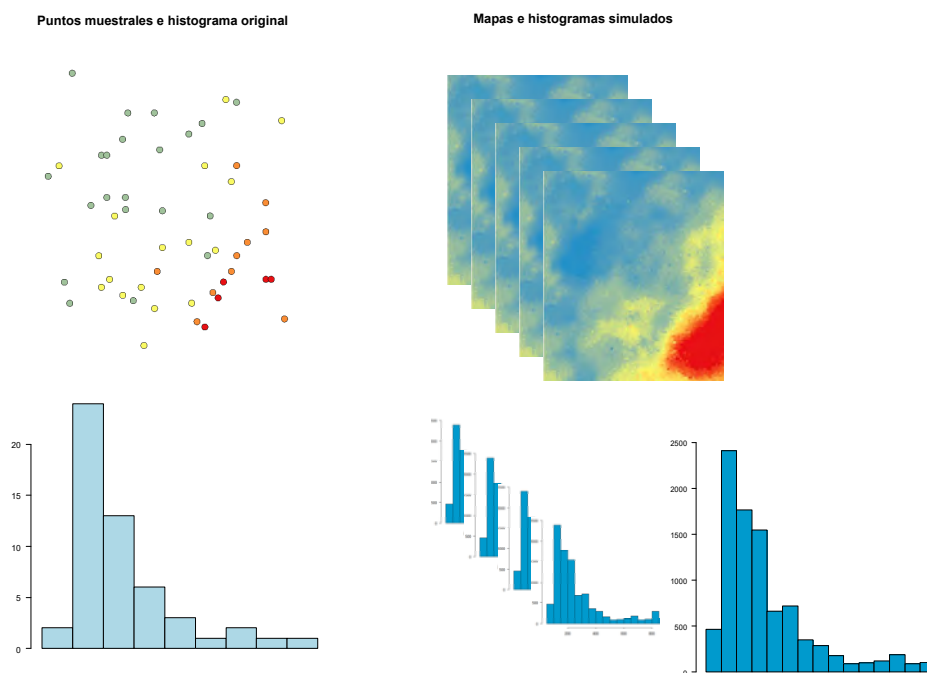


Figura 14: Histogramas obtenidos con simulaciones

Las simulaciones condicionales son construidas en grillos suficientemente finos para proporcionar un número suficiente de nodos (celdas) dentro del tamaño de bloque de interés. Los modelos en este tipo de simulaciones derivan en un conjunto de realizaciones para cada nodo,

donde cada realización, que es equiprobable por construcción, describe el modelo de incertidumbre y a su vez proporciona la función de distribución condicional acumulada. La caracterización del rango de posibles valores por nodo puede lograrse con un número de realizaciones de entre 20 y 50 [Rossi & Deutsch, 2014].

La Simulación Secuencial Gaussiana (Sequential Gaussian Simulation - SGSim) es uno de los algoritmos más utilizados en simulaciones geoestadísticas. Se basa en la suposición de un modelo de función aleatoria, en el que todas las distribuciones condicionales son gaussianas. Los valores observados se incorporan a la grilla de simulación como datos de condicionamiento, y los valores simulados se obtienen estimando con kriging simple, la distribución condicional de cada nodo, donde la media es el valor estimado, y la varianza es la varianza de la estimación. Aplicando un generador de números aleatorios, (generalmente simulación de Montecarlo), se genera un número entre 0 y 1, y se toma el valor de su correspondiente cuantil en la distribución condicional. Los valores simulados se van incluyendo en la grilla, y se utilizan en las simulaciones de los nodos subsiguientes [Rossi & Deutsch, 2014].

La Figura 15 muestra en concepto lo que sucede en un nodo de una grilla de simulación cuando se pretende obtener un valor simulado: con los datos tanto originales como previamente simulados (si los hay), que están dentro de la vecindad de búsqueda, se estima el valor usando kriging simple. En el espacio gaussiano, puede asumirse que la media de la distribución de los datos es igual al valor estimado, y que su varianza es igual a la varianza del kriging, entonces, la función de distribución condicional puede ser construida. Con el método de Montecarlo se obtiene un número aleatorio para un valor de probabilidad en dicha distribución condicional estimada, y su correspondiente intercepto con el valor de la variable representará el valor simulado.

Realizar simulaciones secuenciales gaussianas, implica que los datos deban transformarse a la distribución normal standard (aunque no se limita solo a este tipo de distribución). Un fundamento para llevar los datos al campo gaussiano, es que el teorema del límite central indica que la adición secuencial de residuos aleatorios para obtener valores simulados tiende a una distribución gaussiana. El uso de una distribución de tipo no gaussiano, induce a que los valores simulados sean incorrectos a nivel global debido la forma de la distribución [Rossi & Deutsch, 2014]. La estimación mediante kriging se aplica a los datos estandarizados, usando un modelo de variograma

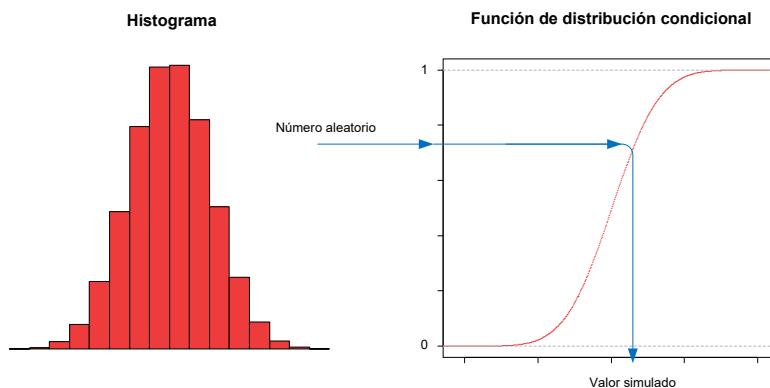


Figura 15: Estimación de función condicional

construido con los datos transformados, y luego de obtener los valores simulados, se regresan a su escala original de valores [Díaz Viera, 2002]. El procedimiento se puede aplicar a la co-simulación de varias funciones aleatorias no independientes, lo que produce simulaciones que coinciden con la covarianza así como con la distribución espacial [Delfiner & Chilès, 1999].

Autores como Pyrcz & Deutsch (2013); Delfiner & Chilès (1999); Díaz Viera (2002), exponen mediante diagramas, el flujo a seguir en un estudio geostatístico en el que se aplica la simulación secuencial gaussiana. En la Figura 16 se observa un esquema general de la metodología para un proceso de este tipo con conceptos de dichas publicaciones. Rossi & Deutsch (2014) definen 14 etapas a completar como se indica a continuación:

1. Realizar el análisis exploratorio completo, incluyendo variografía y definición de dominio.
2. Analizar la existencia de tendencias y en dado caso removerlas.
3. Transformar los datos originales a la forma Normal Score obteniendo su distribución gaussiana.
4. Generar los modelos de variograma de la variable transformada.
5. Aleatoriamente es definida la ruta para la simulación de cada nodo en el dominio.

6. Con kriging simple, estimar la distribución condicional de cada nodo a simular. El valor estimado es el promedio, y la varianza del kriging es la varianza de la distribución.
7. Obtener el valor simulado para el nodo accediendo aleatoriamente a la distribución condicional.
8. Incorporar el valor simulado como dato para el condicionamiento, de modo que también sea considerado en simulaciones posteriores.
9. Completar la simulación de todos los nodos.
10. Verificar que los histogramas simulados sean gaussianos, y que los valores obtenidos con la simulación reproduzcan el modelo de variograma.
11. Transformar los datos a la escala original de valores.
12. Reincorporar la tendencia en caso de que se haya removido.
13. Verificar que el histograma resultante de la transformación a la escala de valores originales, se asemeje al histograma de los datos originales.
14. Comprobar que el modelo proporcione una distribución espacial razonable.

Múltiples variables también pueden ser consideradas para obtener simulaciones: una forma es obtener una simulación conjunta de la información primaria y secundaria, y otra es considerar una variable primaria que condiciona tanto la información primaria como la secundaria, en cuyo caso, la información secundaria puede representar una variable correlacionada con la información primaria. En ambos casos, las Co-Simulaciones Secuenciales Gaussianas (CO-SGSim) es uno de los métodos más utilizados ya que su aplicación resulta simple y similar a la simulación univariada. También requiere la transformación de los datos a la distribución normal estándar y un proceso de recorrido aleatorio de los nodos de la grilla considerando en cada cálculo estimaciones previas, donde la distribución condicional se construye por co-kriging. Si se especifica una jerarquía para las variables (en lugar de un procedimiento conjunto), esta es tenida en cuenta en el condicionamiento, y en dado caso se puede implementar el proceso con distintas opciones: co-kriging completo, co-kriging colocalizado, o una mejor aproximación basada en modelos de Markov Bayes [Rossi & Deutsch, 2014].

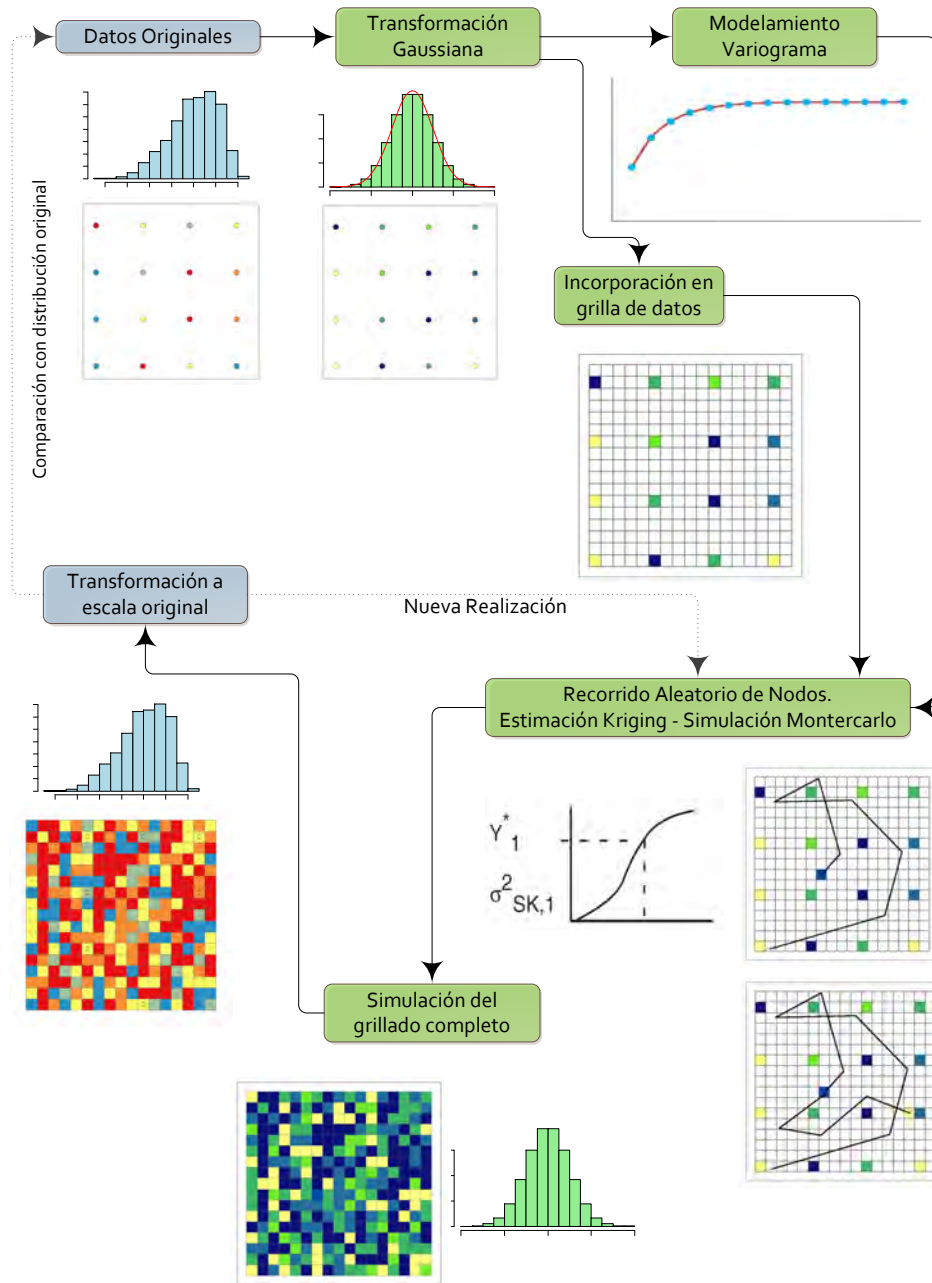


Figura 16: Flujo esquemático de la simulación secuencial gaussiana

3. Materiales y Métodos

3.1. Zona de Estudio

La zona de estudio comprende varios municipios localizados en la Cordillera Oriental de los Andes. Los municipios forman las provincias o subregiones denominadas Almeidas, Guavio y Ubaté en el departamento de Cundinamarca, y Márquez, Neira y Oriente en el departamento de Boyacá. Hay varios municipios que pertenecen a la región conocida como el Altiplano Cundiboyacense y el resto conforman regiones aledañas en cada departamento. La topografía de la zona se caracteriza por tener relieves de altiplanicies que combinan terrazas y valles interandinos alcanzando altitudes entre 2500 y 2800 metros sobre el nivel del mar. La localización de la región puede observarse en la Figura 17



Figura 17: Localización de la zona de estudio en el territorio colombiano

Se considera para la región un clima promedio de 15°C con temperaturas que oscilan entre 0 y 24°C y una precipitación anual que varía entre 500 y 1000 mm, existiendo en ambos casos variaciones entre municipios. En la región predominan coberturas del suelo como bosques, pastos, cultivos y una gama de combinaciones entre ellas donde predomina la formación

vegetal “bosque seco montano bajo” mayormente en la zona del altiplano. Las principales actividades económicas de la zona son la agricultura y la ganadería.

3.2. Registros de Lluvia

Los datos de precipitación utilizados, hacen referencia al parámetro “Valor Total Mensual de Precipitación” que es obtenido por el Instituto de Hidrología, Meteorología, y Estudios Ambientales de Colombia (IDEAM) mediante el registro de la precipitación en su red nacional estaciones meteorológicas. El periodo de estudio corresponde con el año 2016, aunque una vez que se adquirió e inspeccionó la información, se acotó a los 8 meses comprendidos entre marzo y octubre. La elección final de estos meses obedece a factores como su correspondencia con la temporada habitual de lluvias en la región, descartando los meses de enero y febrero por tener mínima pluviosidad, y a la disponibilidad de información distribuida uniformemente en la región según la ubicación de las estaciones, descartando los últimos meses del año porque varias de ellas no contaban con el registro del parámetro.

Una serie de aspectos con las posibles estaciones meteorológicas a utilizar se tuvo en cuenta para elegir las que resultaban aptas para la investigación. Debido a que en la aplicación geoestadística los datos de precipitaciones y NDVI se relacionan en cada par de puntos donde ambas variables son medidas, es de vital importancia, que las estaciones se ubiquen en lugares donde se obtengan valores reales de los índices de vegetación, por lo tanto, de la red de estaciones del IDEAM se eligieron las estaciones que se encontraran localizadas en zonas de cobertura vegetal predominante, y se omitieron aquellas que se localizan en centros urbanos o cursos de agua, de esta forma puede reducirse el posible ruido al momento de correlacionar las dos variables. Adicionalmente se buscó que las estaciones elegidas estuvieran activas en el periodo de estudio y que fueran de tipo meteorológico. Dichos criterios de selección se aplicaron con rutinas de consultas espaciales y alfanuméricas propias de los Sistemas de Información Geográfica donde se contó con los recursos indicados en la Tabla 1.

Como resultado 56 estaciones fueron elegidas para el análisis, y para ellas se adquirió el valor de precipitación total mensual. Si bien la información es pública y de libre acceso, la solicitud se debe realizar mediante un protocolo

Tabla 1: Recursos SIG para la selección de estaciones

Objeto	Capa	Función
Red de Estaciones IDEAM	Puntos	Selección alfanumérica de tipo y estado de estación. Selección espacial dentro de cobertura vegetal. Selección espacial fuera de centros poblados
Centros Poblados	Polígonos	Barrera para excluir estaciones.
Cobertura de la Tierra	Polígonos	Selección alfanumérica de coberturas vegetales.
Servicio Web Imagen Satelital	Mapa Base	Verificación de resultados y apoyo en toma de decisión.

vía web, esto implica ingresar al sitio de internet dispuesto para el trámite de solicitudes de la entidad, y estando allí, por cada estación que se elige, luego de buscarla por nombre o código, se indica el parámetro de interés que se está tramitando, en este caso denominado “Precipitación Total (mm)”, con periodicidad mensual, y con un rango de fechas que va desde el 01 de enero de 2016 hasta 31 de diciembre de 2016. De esta forma una solicitud queda compuesta por varias estaciones (y para este caso, el mismo parámetro, periodo y rango de fechas). La información que es devuelta en un archivo plano por correo electrónico, se estructura desde su formato de reporte crudo a una tabla que consolide la información, y de esta manera, se integran los datos de código, nombre, ubicación, y el valor de precipitación total por mes para cada una de las estaciones. La estructuración final de la información solicitada para las 56 estaciones elegidas puede apreciarse en la Tabla 2 mientras su distribución en la zona de estudio en el mapa de la Figura 18.

A la capa de puntos de estaciones se incorporaron los valores de precipitaciones totales del periodo de estudio, construyendo una tabla de atributos donde para cada estación meteorológica pudiera conocerse su posición (coordenadas X, Y), y el valor de precipitación total mensual de cada uno de los meses analizados. Tal estructura de información define la variable precipitaciones que sería tratada como variable primaria (hard data) en el análisis geoestadístico.

3.3. Productos MODIS NDVI

El proyecto satelital MODIS (Moderate-Resolution Imaging Spectroradiometer), incluye los satélites TERRA y AQUA que proporcionan imágenes de la superficie entera del globo cada 1 a 2 días. Las 36 bandas espectrales que contienen los sensores, proporcionan imágenes de alta sensibilidad radiométrica con resoluciones espaciales de 250, 500 y 1000

Tabla 2: Listado de 56 estaciones meteorológicas y valor mensual de precipitación para el periodo marzo - octubre de 2016

CODIGO	NOMBRE	LATITUD	LONGITUD	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT
21205700	GUASCA	4.879861	-73.868111	40.2	-	82.3	82.2	91.6	71.6	48.7	64.1
21205740	SILOS	5.117722	-73.701417	15.4	120.9	112.7	167.1	171.9	129.4	101.8	59
21206010	LORETOKI HDA	5.089028	-73.80275	30.1	77.6	48	41.7	49.8	54.2	-	-
21200620	PISCIS	5.079167	-73.696861	23.2	135.1	218.4	380.5	241.8	206.5	155.7	73.4
21200780	POTRERO LARGO	4.929222	-73.780472	37.2	175.5	110.1	178.7	204.2	112.1	54.6	57.8
21201190	LAGUNITAS	5.214528	-73.90725	211.3	220.3	61.6	42.3	38	32.6	115.6	82.1
21201610	SAN ISIDRO	4.850289	-73.890722	58.1	127.3	78.8	57.2	68.2	78.7	44	67.3
21201650	STA ROSITA	5.115917	-73.757389	20.6	102.7	67.8	56.5	52.3	70.4	43.1	75.6
24010070	LETICIA	5.303194	-73.70975	55.6	113.8	70.9	23.3	45.1	32.7	27.3	93.2
24010140	CUCUNUBA	5.251028	-73.77075	94.5	94.4	57.7	3.8	2.8	25	48	65.4
21200160	PANONIA	5.057972	-73.734333	39.6	176.4	117.4	179.6	191.4	171	123.5	120.8
21201180	GUANQUICA	5.184278	-73.941111	41.9	157.4	81.2	57	35	50.1	92.3	58.6
21201220	STA CRUZ DE SIECHA	4.784278	-73.870806	33.5	155.8	164.2	177.5	117.7	113.4	152.6	58.2
21201640	VILLAPINZON GJA	5.26375	-73.590861	10.4	127.8	80.8	45.2	77.4	52	50.8	98.6
24010150	SIMIACACA	5.511306	-73.863694	84.9	137.6	131.3	30.5	36	11.7	42.3	128.6
24010170	GUACHETA	5.385889	-73.691056	95.4	142.6	58.7	15.9	21.7	21.1	41.1	114.4
24011080	CUCUNUBA 1	5.248	-73.7525	102.6	106.2	31.3	16.2	32.9	23.4	43.1	60.6
35060020	SUEVA	4.81	-73.707167	56.2	211.3	189	169.3	161.2	198.4	127.8	68.4
35060050	JUNIN	4.792222	-73.668722	13	105.4	136	100.1	-	31.4	-	-
35060120	VEGA LA SAN JUAN	4.723306	-73.470556	-	418.9	375.1	276.2	338.7	376.1	363.4	163.6
35060140	PALOMAS LAS	4.704917	-73.444889	255	646	605.5	350.8	505	573.5	532	316.5
35060160	POTRERITOS	4.828806	-73.769278	73	217.1	184.1	249.7	223.1	179.3	105	63.9
35060200	AMOLADERO EL	4.857972	-73.745389	66.3	230.9	207.3	249	222.6	120.3	114.1	57.7
35060220	GLORIA LA	4.815694	-73.419778	110.1	302.7	291.8	427.5	415	360.5	229	136.6
35060240	GAMA	4.764278	-73.609917	81.5	215.7	214	144.6	194.1	236.7	163.9	113.9
35060300	MUNDO NUEVO	4.852528	-73.391944	160.5	563.5	456	429.5	433	445	393.5	207
35070030	TURMEQUE	5.31775	-73.496361	10.1	122	71.5	77.5	122.4	61.1	88.7	-
35070060	QUINCHOS LOS	5.219556	-73.347917	61.5	313.4	329	241.4	330.7	131.8	213	77.3
35070170	NAZARETH	4.733333	-73.2	28	322	536	539	536	422	482	253
35070190	CHIVOR	4.886361	-73.366736	164.1	442.5	-	392	445.4	358.5	401.5	-
35070310	TEATINOS	5.422822	-73.375778	70.2	210.8	140.6	182.4	137.8	134.9	90.9	-
35070470	ESC LOS MOLINOS	4.950417	-73.410167	105.9	187	256.1	-	309.4	-	-	-
35070490	HATO GRANDE HDA	4.959389	-73.584361	74.7	262.5	363.9	374.1	591.6	330.1	208.5	238.7
35070550	VALLE GRANDE	5.056611	-73.426917	49.2	154.9	168.8	193.1	267.7	177.5	-	-
35075010	NUEVO COLON	5.352694	-73.453778	13.1	134.5	113.7	130.6	128	-	-	-
35075030	VILLA LUISA	5.422222	-73.349417	37.1	112.5	112.8	-	-	-	-	-
35080080	PIEDRA CAMPANA	4.860389	-73.234083	83.4	778.3	735.9	518.9	478.4	753.4	549.8	-
35080110	ALTO MUCENO	5.005139	-73.258611	171.6	463.7	530.4	539.4	621	422.8	-	-
35085020	RONDON	5.358417	-73.203611	82.1	274.7	-	-	-	-	-	-
35060130	MINAS LAS	4.675167	-73.482139	248.9	-	399.5	295	334.7	-	389	179.8
35060150	TEMBLADARES	4.704861	-73.665278	-	-	-	287.9	337	377.8	255.8	151.8
35060170	TASAJERAS	4.885694	-73.600361	74	354.6	241.2	185.9	263.1	188.9	203.1	212.3
35060210	CLARAVAL	4.65025	-73.641444	59	258	260.9	218.1	238.5	292.3	220.6	103.4
35060230	TRES ESQUINAS	4.781417	-73.489056	158.5	203	-	-	-	-	-	-
35060250	STA ROSA DE UBALA	4.796778	-73.407194	199	678.2	496.9	482.1	444.7	564.8	386.8	292.8
35065010	GACHETA	4.830389	-73.646472	57	189.2	159.8	123.2	190.4	168.4	124.5	-
35070020	VENTAQUEMADA	5.383056	-73.602889	-	165.6	110.6	112.2	105.8	102.1	85.3	-
35070050	UMBITA	5.219111	-73.444556	28.6	179.8	117.9	87.8	149.9	105.1	101.8	-
35070070	CHINAVITA	5.164861	-73.36425	76	207.6	187.7	171.9	253.9	160.1	159	-
35070110	GUAYATA LA GRANJA	4.976083	-73.499306	48.7	230.2	165	165.5	234.9	171	152.6	-
35070180	STA MARIA	4.860778	-73.25675	171	843.5	619.3	635.8	448	640.5	594.1	532.6
35070230	MACHETA GJA AGROP	5.075111	-73.579417	67.3	122.8	99.8	108.7	305.2	182.6	122.4	76.8
35070320	MARTOTA	4.936472	-73.436083	18.6	161.3	40.5	145.5	57.6	38.8	29.8	15.6
35070520	POMARROSOS LOS	4.931417	-73.350361	179.8	339.6	459.3	521.2	456.5	-	-	-
35075040	INST AGR MACANAL	4.974361	-73.31675	163.1	356.6	417.1	413.9	436.2	335.6	-	-
35080100	PTE FORERO	5.101389	-73.251417	82	296.9	335.6	281.3	412.6	228.6	-	-

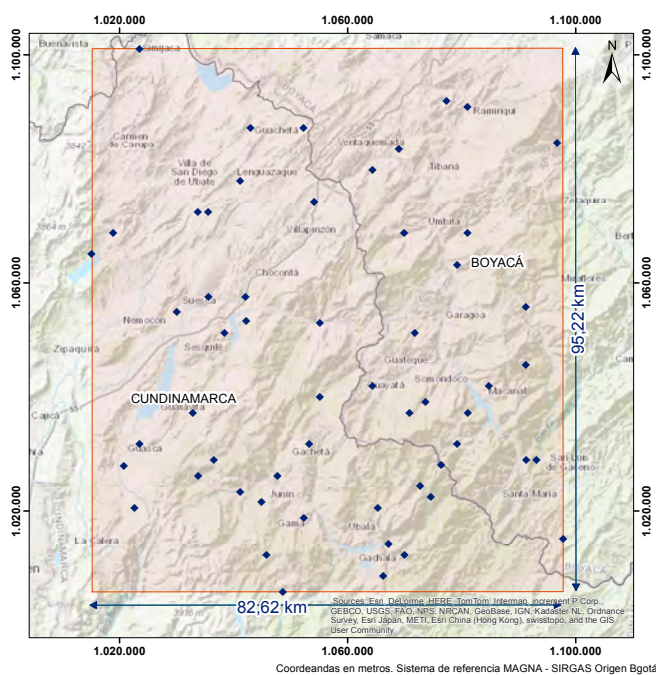


Figura 18: Ubicación de las 56 estaciones meteorológicas en la zona de estudio

metros en teselas de $10^\circ \times 10^\circ$ la cuales conforman una grilla de cobertura global. [Barona *et al.*, 2006]. Una gran variedad de productos estándar¹, con diferentes propósitos y aplicaciones, derivan de la misión satelital. Uno de estos productos hace parte del grupo de productos terrestres (Land Products) y contiene información sobre índices de vegetación (MODIS cuenta con un set de datos de cobertura global que ha estado monitoreando el NDVI desde 1981).

Los productos de índices de vegetación MODIS (MODIS Vegetation Index Products) proporcionan series de datos espaciotemporales con el propósito de monitorear continuamente la vegetación terrestre a nivel global mediante los índices NDVI Y EVI (Enhanced Vegetation Index). Cuadrículas raster (imágenes) son generadas con distintas resoluciones espaciales, tanto en intervalos mensuales como cada 16 días, para suministrar información que describa el estado de la vegetación en la superficie terrestre [Didan *et al.*, 2015]. Dicha información compone el catálogo de productos

¹el listado de productos puede ser consultado en el sitio oficial de MODIS: <https://modis.gsfc.nasa.gov/data/dataproduct/>

mostrado en la Tabla 3.

Tabla 3: Catálogo de productos MODIS NDVI

Nomenclatura	Sensor	Resolucion Temporal	Resolución Espacial
MOD13Q1: 16-day 250m VI	TERRA	16 días	250 metros
MOD13A1: 16-day 500m VI	TERRA	16 días	500 metros
MOD13A2: 16-day 1km VI	TERRA	16 días	1 kilómetro
MOD13A3: Monthly 1km VI	TERRA	Mensual	1 kilómetro
MOD13C1: 16-day 0.05deg VI	TERRA	16 días	0.05 grados
MOD13C2: Monthly 0.05deg VI	TERRA	Mensual	0.05 grados
MYD13Q1: 16-day 250m VI	AQUA	16 días	250 metros
MYD13A1: 16-day 500m VI	AQUA	16 días	500 metros
MYD13A2: 16-day 1km VI	AQUA	16 días	1 kilómetro
MYD13A3: Monthly 1km VI	AQUA	Mensual	1 kilómetro
MYD13C1: 16-day 0.05deg VI	AQUA	16 días	0.05 grados
MYD13C2: Monthly 0.05deg VI	AQUA	Mensual	0.05 grados

En el listado se observa el nombre del producto, sensor, intervalo de tiempo de adquisición de información y resolución espacial. El prefijo MOD13 hace referencia a productos MODIS VI (Vegetación Index) obtenidos por el sensor TERRA, mientras el prefijo MYD13 hace lo propio para los que se obtienen mediante el sensor AQUA. La familia de productos para ambos sensores tiene una cobertura para las porciones terrestres del globo generada en teselas (tiles) de aproximadamente 1200 x 1200 kilómetros (en el ecuador) con el sistema de proyección cartográfico sinusoidal. Los productos MODIS se disponen en el formato estándar HDF-EOS (Hierarchical Data Format-Earth Observing System) donde cada archivo está compuesto por dos estructuras separadas de información. Por un lado los Datasets Científicos (Scientific Data Sets - SDS) son capas de información raster que almacenan el dato a nivel del pixel. Por otro, los metadatos subdivididos a su vez en tres conjuntos: (I) Estructural que describe el contenido del archivo, (II) núcleo que describe proyección y nombre de cuadrícula, y (III) de archivo que describe características como fechas, estadísticas de calidad e indicadores de uso. En los metadatos los atributos pueden ser globales o específicos de producto. [Didan *et al.*, 2015].

Para el presente estudio se eligieron productos MODIS NDVI de la colección 6 denominados serie MOD13A3 para el sensor TERRA y MYD13A3 para el sensor AQUA. Las imágenes de dichas series son productos de índices de vegetación logrados en periodos mensuales con resolución espacial de 1 kilómetro. Cada producto consta de un archivo raster en formato HDF

compuesto por 11 capas de información (Datasets científicos), entre las cuales destacan dos capas de índices de vegetación (NDVI y EVI), capas de cada banda espectral original (Rojo, Azul, Infrarrojo Cercano, Infrarrojo Medio), y capas que proporcionan información acerca de la calidad y las condiciones con que los datos fueron adquiridos (Pixel Reliability, VI Quality). El tipo de dato, rango de valores y formato de las capas de información de los productos MOD13A3 y MYD13A3 se referencia en la documentación MODIS como se observa en la Tabla 4.

Tabla 4: Capas de información de los productos MOD13A3 y MYD13A3

Science Data Set	Units	Data type	Valid Range	Scale factor
1km monthly NDVI	NDVI	int16	-2000, 10000	0.0001
1km monthly EVI	EVI	int16	-2000, 10000	0.0001
1km monthly VI Quality	Bits	uint16	0, 65534	NA
1km monthly red reflectance (Band 1)	Reflectance	int16	0, 10000	0.0001
1km monthly NIR reflectance (Band 2)	Reflectance	int16	0, 10000	0.0001
1km monthly blue reflectance (Band 3)	Reflectance	int16	0, 10000	0.0001
1km monthly MIR reflectance (Band 7)	Reflectance	int16	0, 10000	0.0001
1km monthly view zenith angle	Degree	int16	-9000, 9000	0.01
1km monthly sun zenith angle	Degree	int16	-9000, 9000	0.01
1km monthly relative azimuth angle	Degree	int16	-18000, 18000	0.01
1km monthly pixel reliability	Rank	int8	0, 3	NA

Las imágenes de las series MOD13A3 y MYD13A3 utilizadas para el periodo de estudio se descargaron de la plataforma Earth Explorer del Servicio Geológico de Estados Unidos (USGS). Se adquirieron 16 imágenes en total, que comprenden un par de imágenes (TERRA Y AQUA) por cada mes del periodo de estudio y corresponden con la nomenclatura de localización h10 v08. El listado de imágenes utilizadas se indica en la Tabla 5.

3.3.1. Procesamiento de Imágenes Satelitales MODIS

El procesamiento de imágenes MODIS se compone de distintas etapas donde la información de las imágenes adquiridas es llevada a distintos niveles derivando nuevos productos necesarios para el análisis. La primera etapa es la de pre-procesamiento, que consiste en la preparación de archivos raster como insumos ajustados a las condiciones del proyecto. La segunda, es la de filtrado, que se divide en la descomposición de bandas que proporcionen información necesaria para evaluar la calidad, y en el álgebra de mapas, que calcula los píxeles óptimos con la información de calidad y en base a ella

Tabla 5: Archivos MODIS adquiridos para la investigación

Archivo	Sensor	Fecha Imagen	Mes Observación
MOD13A3.A2016092.h10v08.006.2016130120001.hdf	TERRA	01/04/2016	Marzo
MYD13A3.A2016092.h10v08.006.2016140121404.hdf	AQUA	01/04/2016	
MOD13A3.A2016122.h10v08.006.2016166144049.hdf	TERRA	01/05/2016	Abril
MYD13A3.A2016122.h10v08.006.2016165132507.hdf	AQUA	01/05/2016	
MOD13A3.A2016153.h10v08.006.2016200100443.hdf	TERRA	01/06/2016	Mayo
MYD13A3.A2016153.h10v08.006.2016188131234.hdf	AQUA	01/06/2016	
MOD13A3.A2016183.h10v08.006.2016229085309.hdf	TERRA	01/07/2016	Junio
MYD13A3.A2016183.h10v08.006.2016223021056.hdf	AQUA	01/07/2016	
MOD13A3.A2016214.h10v08.006.2016263152028.hdf	TERRA	01/08/2016	Julio
MYD13A3.A2016214.h10v08.006.2016250034938.hdf	AQUA	01/08/2016	
MOD13A3.A2016245.h10v08.006.2016292070143.hdf	TERRA	01/09/2016	Agosto
MYD13A3.A2016245.h10v08.006.2016282023653.hdf	AQUA	01/09/2016	
MOD13A3.A2016275.h10v08.006.2016322042312.hdf	TERRA	01/10/2016	Septiembre
MYD13A3.A2016275.h10v08.006.2016314154034.hdf	AQUA	01/10/2016	
MOD13A3.A2016306.h10v08.006.2016340092612.hdf	TERRA	01/11/2016	Octubre
MYD13A3.A2016306.h10v08.006.2016348141824.hdf	AQUA	01/11/2016	

define la variable índice de vegetación. El diagrama de flujo de la Figura 19 muestra la metodología general de procesamiento digital de imágenes. El componente detallado de cada etapa puede consultarse en el anexo A, donde se explican los recursos, productos de información, y análisis derivados del procesamiento.

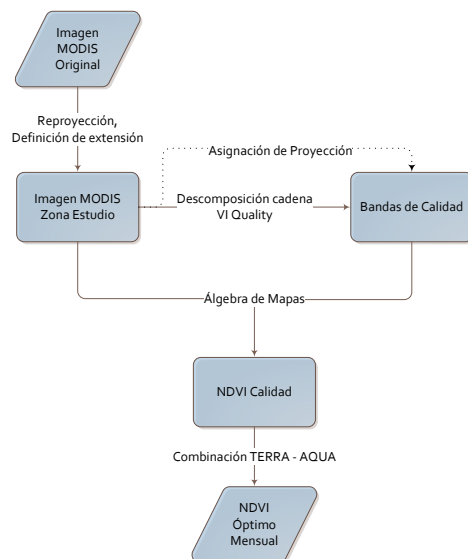


Figura 19: Flujo de trabajo para procesamiento de imágenes MODIS

3.4. **Ánalysis Geoestadístico**

En el área de estudio, las precipitaciones tienen baja densidad de información, mientras el índice de vegetación cuenta con información abundante. Ambas variables son conjuntos de puntos de los que se conoce en cada posición, el valor de la precipitación en milímetros y el NDVI (adimensional) para cada uno de los meses del periodo de estudio. Partiendo de esos conjuntos de datos, se realizaron los procedimientos necesarios para el desarrollo del estudio geoestadístico, que buscaba obtener mapas de distribución de la precipitación, con base en la respuesta espectral de la vegetación a los eventos de lluvia a través del NDVI usando la técnica de co-simulaciones gaussianas.

Para poder establecer una relación entre la lluvia como hard data (variable principal) y el NDVI como soft data (variable secundaria), y obtener mapas a través de co-simulaciones gaussianas, se debe cumplir una serie de condiciones con el propósito de controlar los resultados y no introducir ruido durante el procesamiento: en primer lugar, debe tenerse presente que el procedimiento consiste en dar valores a las celdas de una grilla de simulación para la variable de hard data, entonces, resulta conveniente que la variable de soft data tenga valores en todas las celdas de dicha cuadrícula, para lograr esto, se debe respetar el espaciamiento de los datos, en este caso representado por el tamaño de pixel, por lo tanto, mantener el mismo tamaño de espaciado en el diseño de la cuadrícula, evitará en lo posible, celdas sin valor de soft data. Adicionalmente, hay que tener en consideración que algunos píxeles han sido descartados durante el proceso de filtrado, lo que implica presencia de celdas sin valor de la variable secundaria, y por lo tanto, estas casillas deben completarse. Con un procedimiento geoestadístico completo, aplicado a la variable NDVI de manera univariada, se dan valores a las celdas vacías. El proceso involucra etapas exploratorias y estructurales, para que finalmente se calculen simulaciones gaussianas logrando una grilla promedio de cada mes, de acuerdo con sus realizaciones. Los promedios mensuales, representarán la grilla completa de soft data que servirá para relacionar ambas variables mediante la co-simulación.

Para la variable lluvia también se llevan a cabo las etapas exploratoria y estructural, y finalmente se relacionan ambas variables mediante el coeficiente de correlación, lo que posteriormente permitirá obtener las co-simulaciones entre las precipitaciones y el resultado geoestadístico del índice de vegetación.

En el diagrama de la Figura 20 puede observarse el flujo completo del análisis, donde se detalla como la co-simulación gaussiana se completa una vez que las etapas exploratoria, y estructural se llevan cabo individualmente en cada variable. El siguiente capítulo muestra el resultado obtenido en cada fase de análisis.

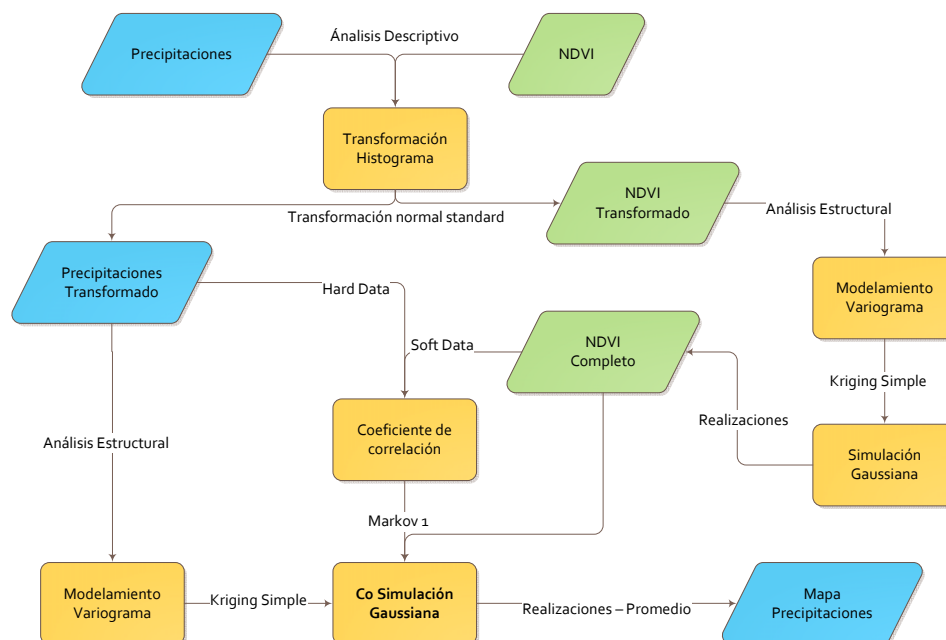


Figura 20: Flujo del análisis geoestadístico

4. Resultados y Discusión

4.1. Fase Exporatoria

4.1.1. Estadística Descriptiva

Para cada una de las variables se analizaron sus estadísticos en cada mes del periodo de estudio (marzo a octubre de 2016). Los gráficos y cálculos necesarios se obtuvieron en el software R-Studio, y a modo de resumen, en la Tabla 6 pueden observarse los resultados para las precipitaciones y en la Tabla 7 para el NDVI. Cabe recordar que los registros de lluvia corresponden con medidas en milímetros y el NDVI es adimensional. Sin embargo, para este último debe aplicarse un factor de escala de 0.0001 si se desea ubicar sus valores en la escala convencional de -1 a 1, los valores originales MODIS varían entre -2.000 y 10.000.

Según los valores obtenidos de las medidas estadísticas, las precipitaciones tienen heterogeneidad alta, pues en todos los meses el coeficiente de variación es cercano o mayor al 75 %. Los valores de asimetría positiva en los 8 meses y kurtosis cercana a 3 en la mayoría de ellos, indican mayor concentración de los valores bajos y distribuciones claramente asimétricas a la derecha. Dichas distribuciones se observan en los histogramas de la Figura 21 donde se identifican patrones de agrupamiento parecidos a lo largo del periodo de estudio. Los máximos y mínimos muestran que los rangos de valores varían bastante entre los meses como puede corroborarse en el box plot de la Figura 23 donde también se aprecia para algunos meses valores atípicos altos. En general se identifica que la mayor concentración de los datos corresponde con valores cercanos a los 150 y 200 mm de acuerdo con los rangos intercuartílicos y los histogramas. En el caso de los valores bajos existe mayor variación en la comparación entre meses aunque en ningún caso se superan ampliamente los 100 milímetros como límite de agrupación. En otras palabras, hay variaciones entre los meses para el valor del Q1 (25 % de los datos) y similitudes en el valor del Q3 (75 %) como se observa en el box plot. Tal situación evidencia predominancia de valores medios y bajos en las precipitaciones, y señala que cuando se presentan valores altos estos se distancian del grueso de la distribución en la región de estudio.

Tabla 6: Medidas de tendencia, forma y distribución de las precipitaciones para el periodo marzo - octubre de 2016

Medida/Mes	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre
Mínimo	10,1	77,6	31,3	3,8	2,8	11,7	27,3	15,6
Máximo	255	843,5	735,9	635,8	621	753,4	594,1	532,6
Promedio	82,88	253,75	225,03	218,78	242,46	206,65	183,46	129,7
Mediana	67,3	203	164,6	178,1	222,85	168,4	124	93,2
Varianza	3906,25	30506,12	30569,03	27188,71	28036,15	31698,24	23984,72	10267,77
Desviación Estándar	62,5	174,66	174,84	164,89	167,44	178,04	154,87	101,33
Desviación Mediana	30,2	75,2	88,65	106,5	124,15	114,2	77,55	34,2
Asimetría	1,08	1,74	1,09	0,73	0,43	1,16	1,18	2,06
Kurtosis	3,31	5,53	3,23	2,48	2,07	3,7	3,27	7,91
Coef. Var. Promedio(%)	75,41	68,83	77,7	75,37	69,06	86,16	84,41	78,13
Coef. Var. Mediana(%)	44,87	37,04	53,86	59,8	55,71	67,81	62,54	36,7

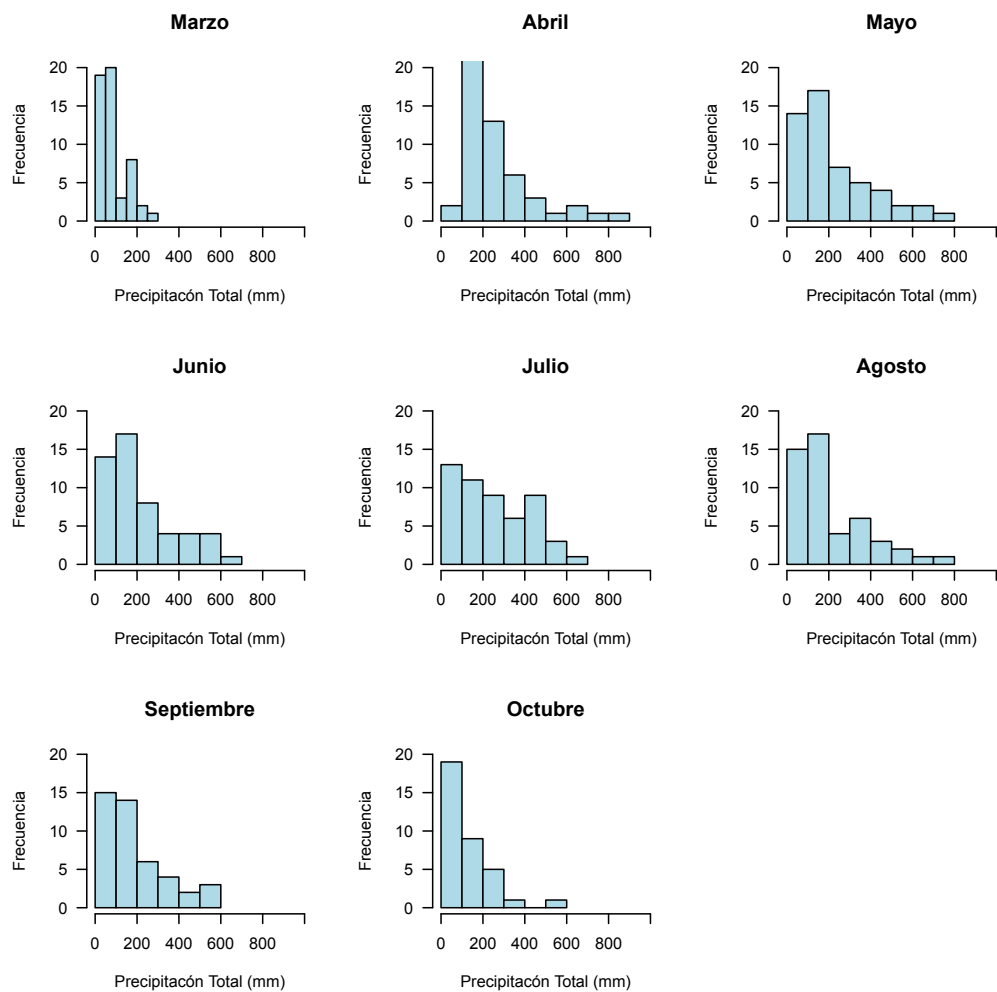


Figura 21: Histogramas de las precipitaciones

Tabla 7: Medidas de tendencia, forma y distribución del NDVI para el periodo marzo - octubre de 2016

Medida/Mes	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre
Mínimo	2500	3180	3700	1850	1779	3378	3345	3723
Máximo	9450	9367	9382	9170	9207	9359	9436	9482
Promedio	7025,99	7538,63	7607,85	7471,57	7385,70	7451,98	7474,46	7630,86
Mediana	7136	7681	7734	7633	7579	7657	7679	7764
Varianza	1188514,24	527090,52	435362,43	753528,16	798824,81	747740,68	707213,72	478186,08
Desviación Estándar	1090,19	726,01	659,82	868,06	893,77	864,72	840,96	691,51
Desviación Mediana	780	390	321	516	495	444	389	349
Asimetría	-0,55	-1,14	-1,37	-1,14	-1,19	-1,26	-1,30	-1,21
Kurtosis	2,88	5,23	6,41	5,14	4,71	4,65	4,91	5,57
Coef. Var. Promedio(%)	15,52	9,63	8,67	11,62	12,10	11,60	11,25	9,06
Coef. Var. Mediana(%)	10,93	5,08	4,15	6,76	6,53	5,80	5,07	4,50

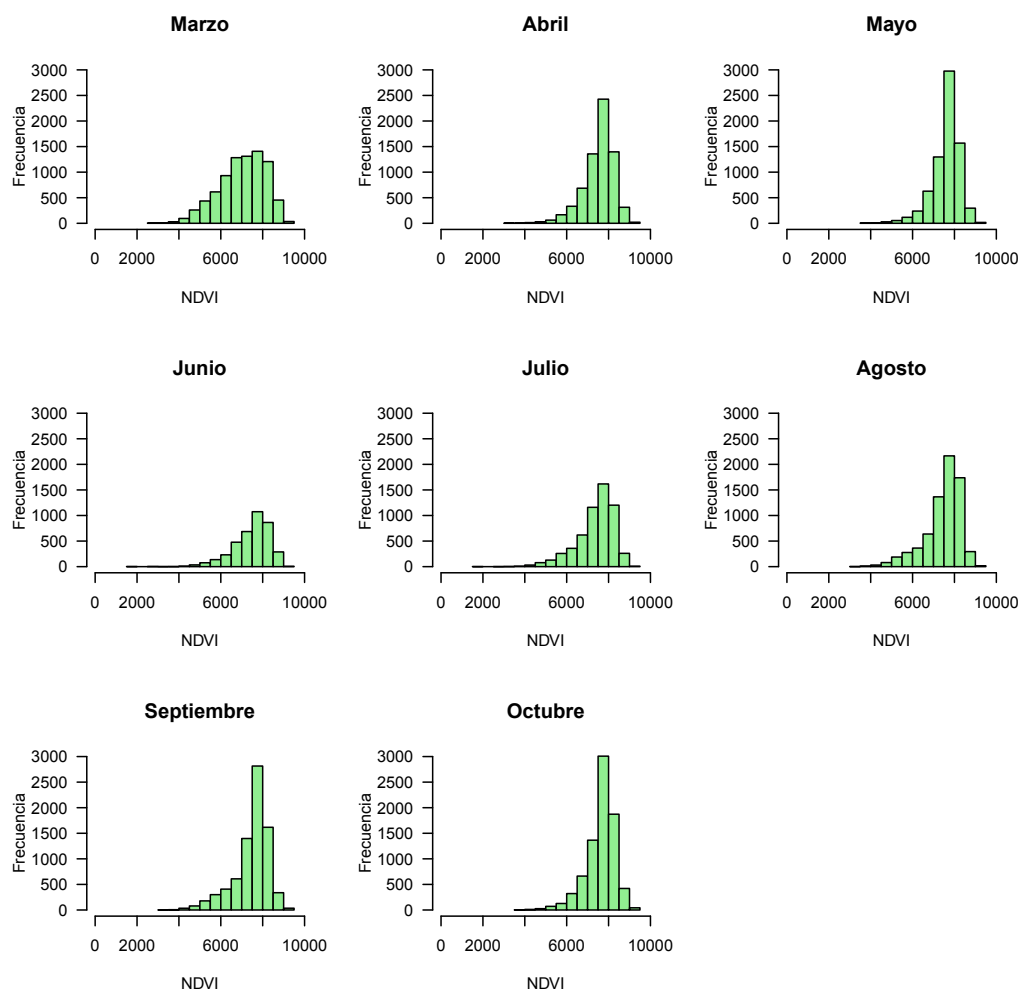


Figura 22: Histogramas del NDVI

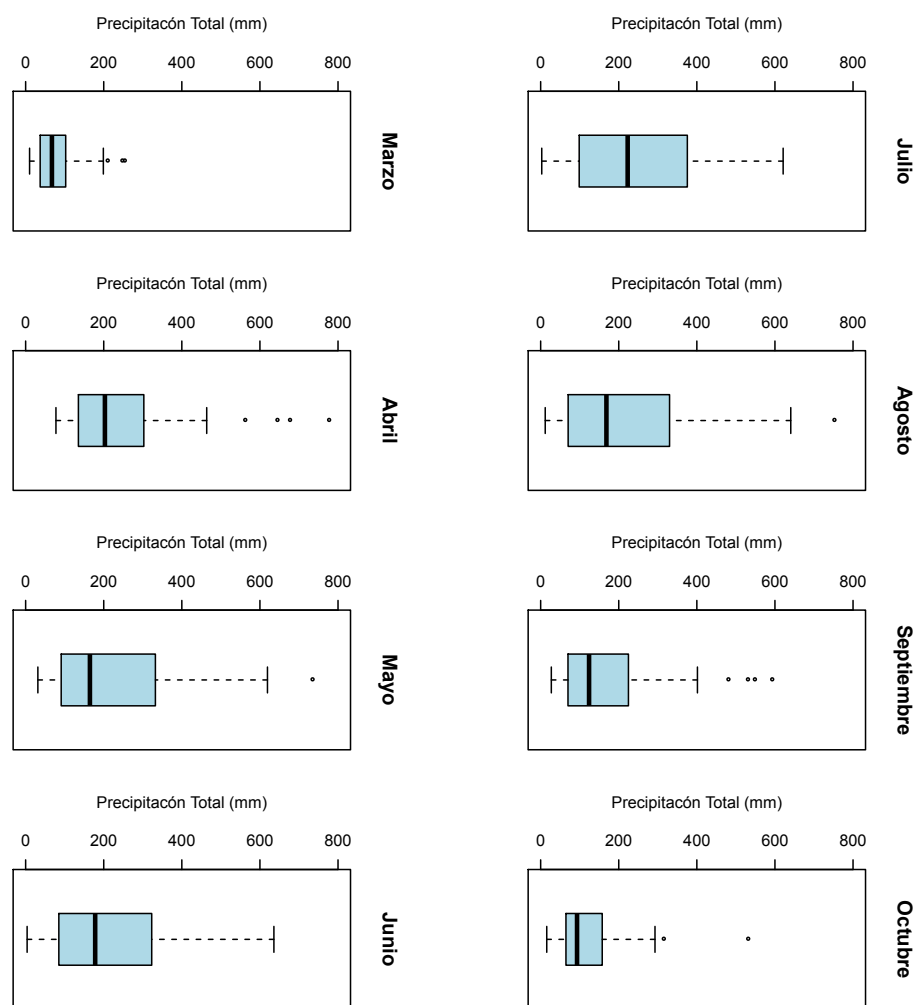


Figura 23: Box plot de las precipitaciones

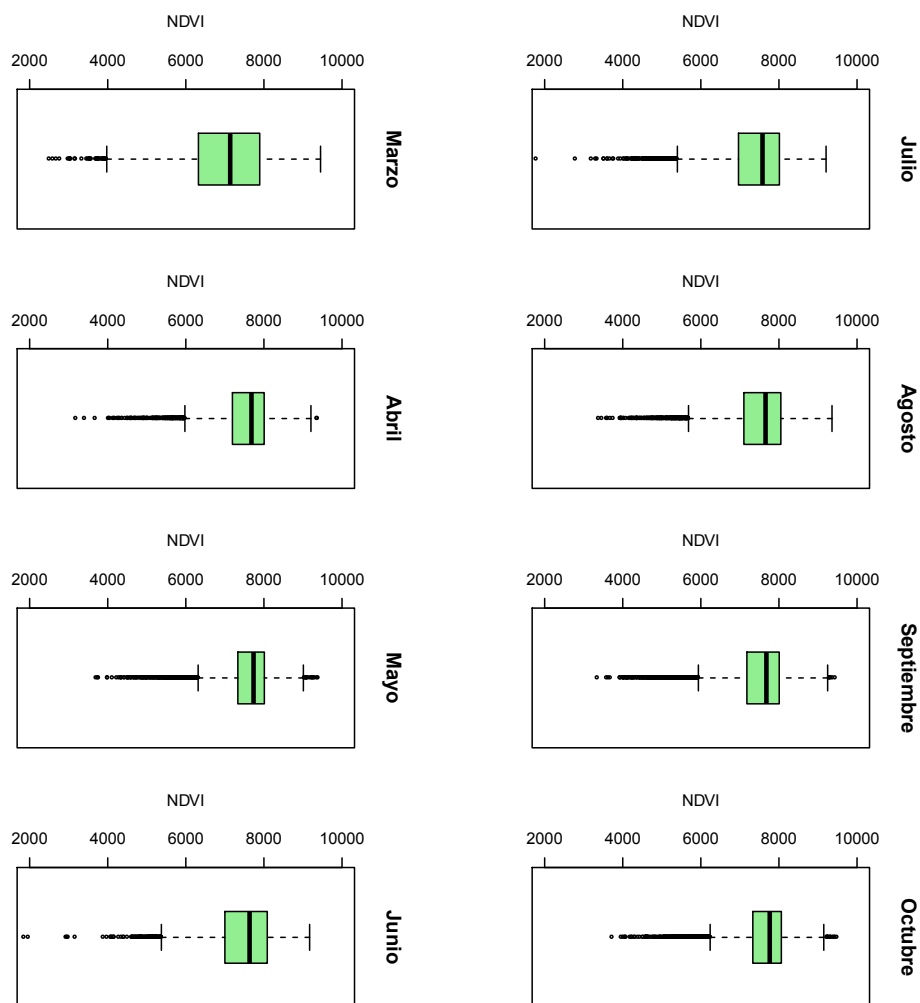


Figura 24: Box plot del NDVI

En el caso del índice de vegetación, las medidas indican que la variabilidad de los datos es relativamente homogénea según valores del coeficiente de variación siendo inferiores o cercanos al 15%. La asimetría negativa junto con los valores de kurtosis, indican tendencias de concentración hacia los valores altos y distribuciones asimétricas a la izquierda. Los histogramas (Figura 22) señalan un comportamiento similar a lo largo de los meses, mientras que medidas como los máximos, mínimos y promedios muestran que no existe variación grande entre los rangos de valores mensuales. En el

box plot de la Figura 24 se observa la existencia de valores atípicos bajos en cada uno de los meses pero debe tenerse en cuenta que los conjuntos de datos están compuestos por cientos de miles de píxeles, dichos valores generalmente se encuentran por debajo de 5000. Tener este valor como dato atípico bajo, junto con un patrón de distribución donde los valores altos son predominantes sugiere que en general, puede considerarse que la vegetación en la zona de estudio se encuentra en entornos favorables para su vigorosidad y desarrollo. Adicionalmente no hay diferenciaciones marcadas entre los rangos intercuartílicos donde se tienen valores que rondan 6000 en el Q1 y 8000 en el Q3.

4.1.2. Caracterización Espacial de las Variables

En el contexto espacial, se observa un patrón de ubicación repetitivo en los distintos meses para las precipitaciones: los valores altos son localizados en la zona suroriental del área que es donde se tiene la mayor elevación del terreno y donde en algunos meses dichos valores representan datos atípicos. También es constante que los valores medios se presenten hacia la parte central, y los valores bajos hacia la parte noroccidental de la zona de estudio, situación que se aprecia en la Figura 25.

Por otra parte, los valores del índice de vegetación también cuentan con tendencias de agrupación y zonificación concentrando los más altos en la zonas sur y centro, mientras valores medios mezclados con bajos se ubican en la parte noroccidental en una proporción mucho menor. El mapa de la Figura 26 muestra la distribución espacial de los valores del NDVI.

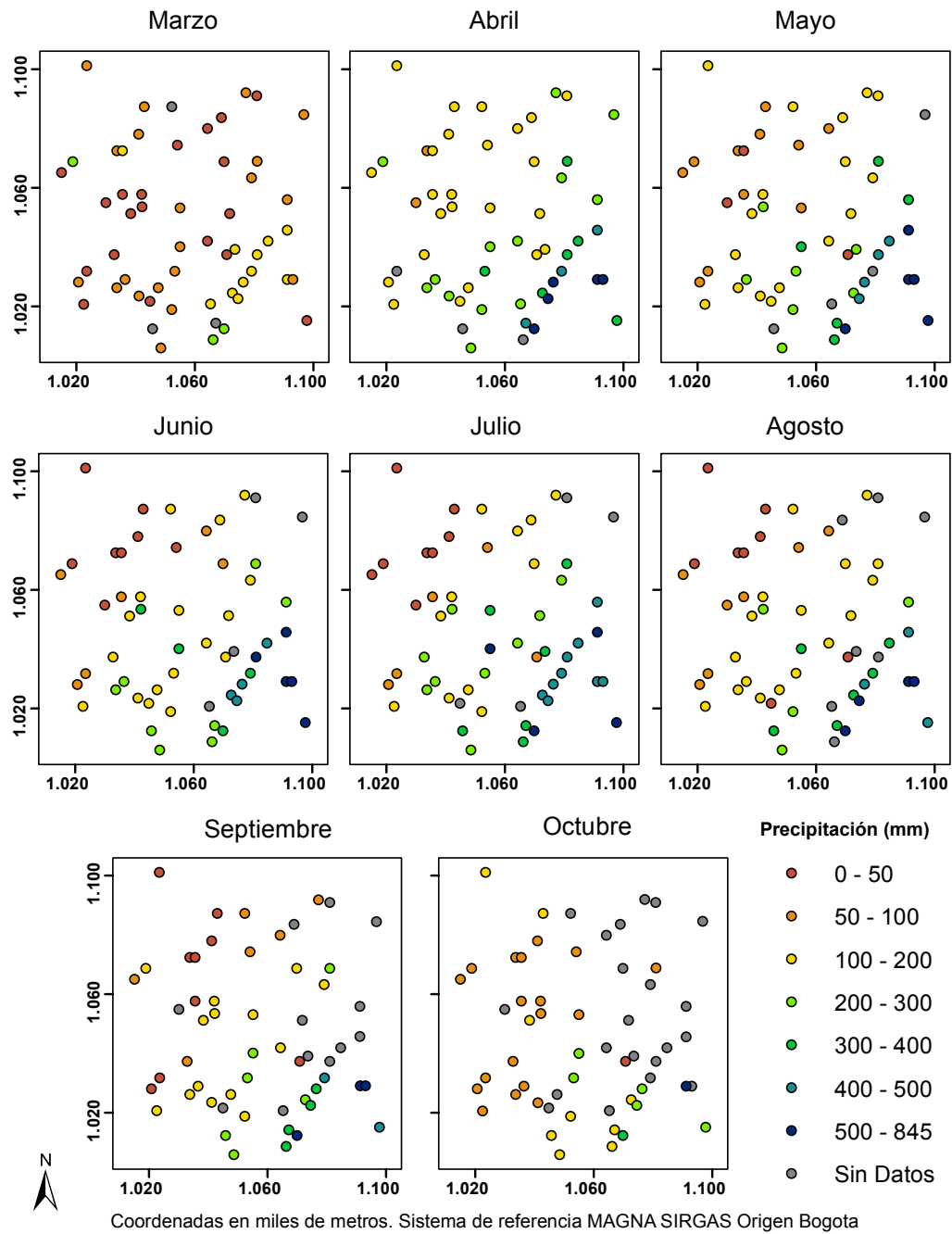


Figura 25: Distribución de los valores de precipitación mensual por estación en la zona de estudio

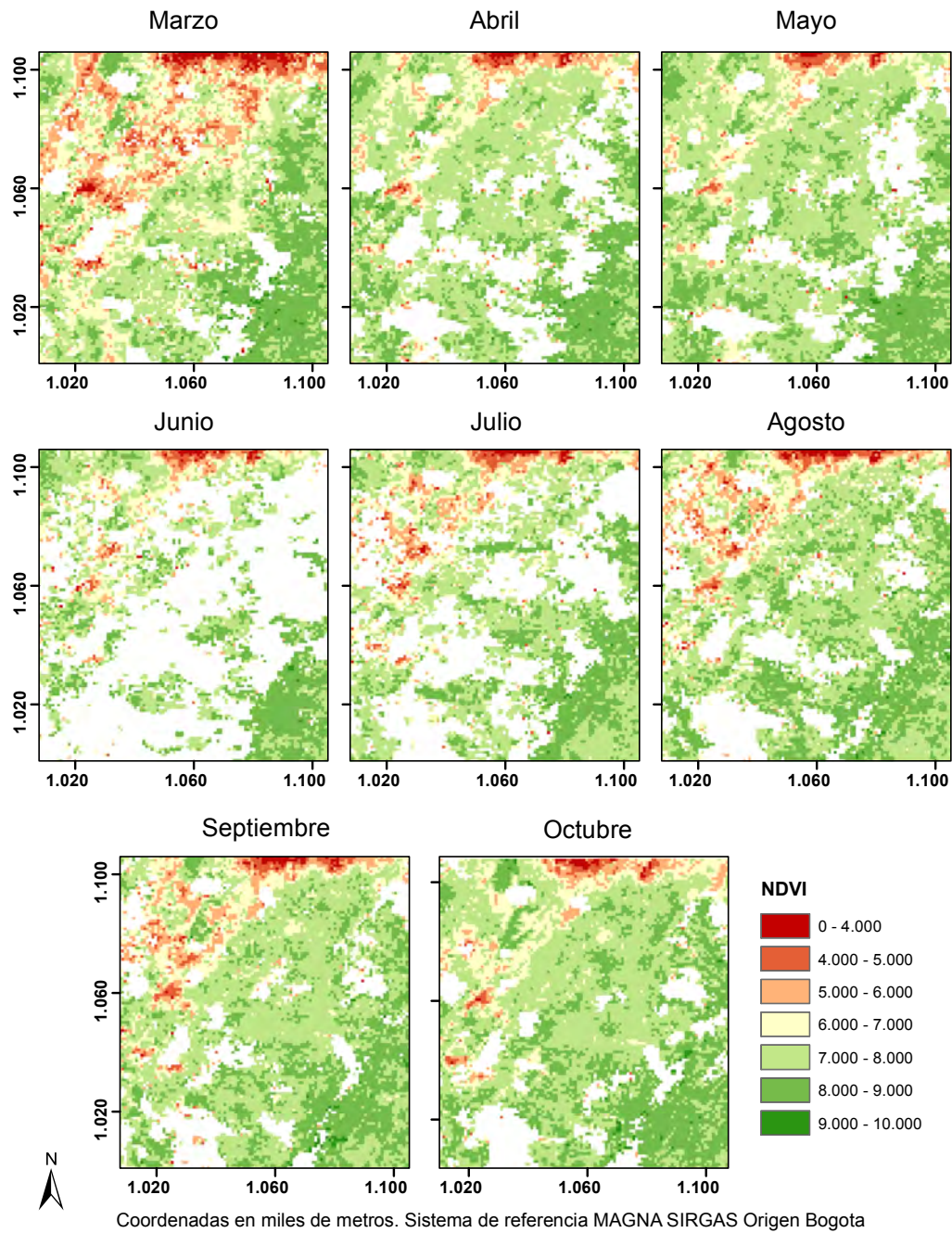


Figura 26: Distribución de los valores de NDVI mensual en la zona de estudio

4.1.3. Transformación del Histograma

Una condición para obtener las simulaciones secuenciales gaussianas, es que los datos sigan distribuciones en el campo gaussiano. Ya que en la realidad esto no sucede, los datos originales son transformados y como resultado, se obtienen distribuciones estandarizadas con valor de promedio igual a cero y de varianza igual a uno, las cuales siguen la forma gaussiana. El resultado de las transformaciones aplicadas se puede observar en los histogramas de las Figuras 27 (precipitaciones) y 28 (NDVI).

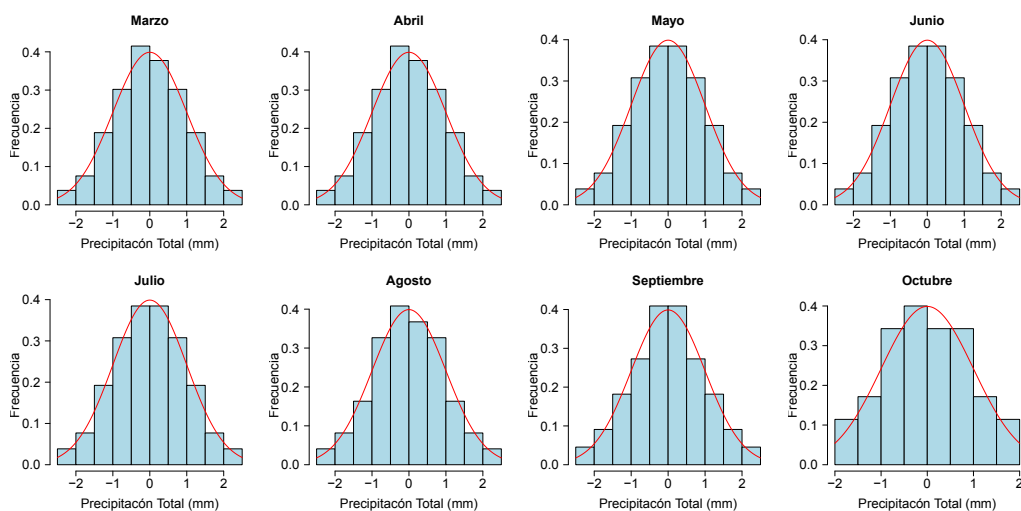


Figura 27: Transformación de histogramas para las precipitaciones

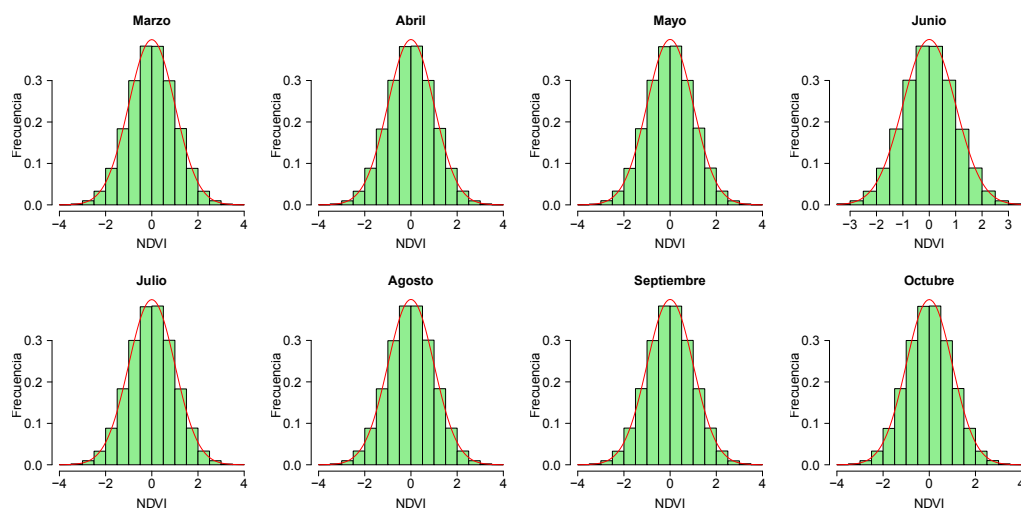


Figura 28: Transformación de histogramas para el NDVI

Adicionalmente, el hecho de que cada elemento original del conjunto de datos le corresponda ahora un valor equivalente transformado, también tiene un impacto en el contexto geográfico como se ilustra en la Figura 29. Allí se observa la distribución espacial de las variables originales y transformadas tomando como ejemplo el mes de mayo. Los mapas transformados se convierten en el escenario de modelamiento de la estructura de correlación espacial de las variables que permiten completar los procesos de SGSIM (NDVI) y CO-SGSIM (precipitaciones vs NDVI).

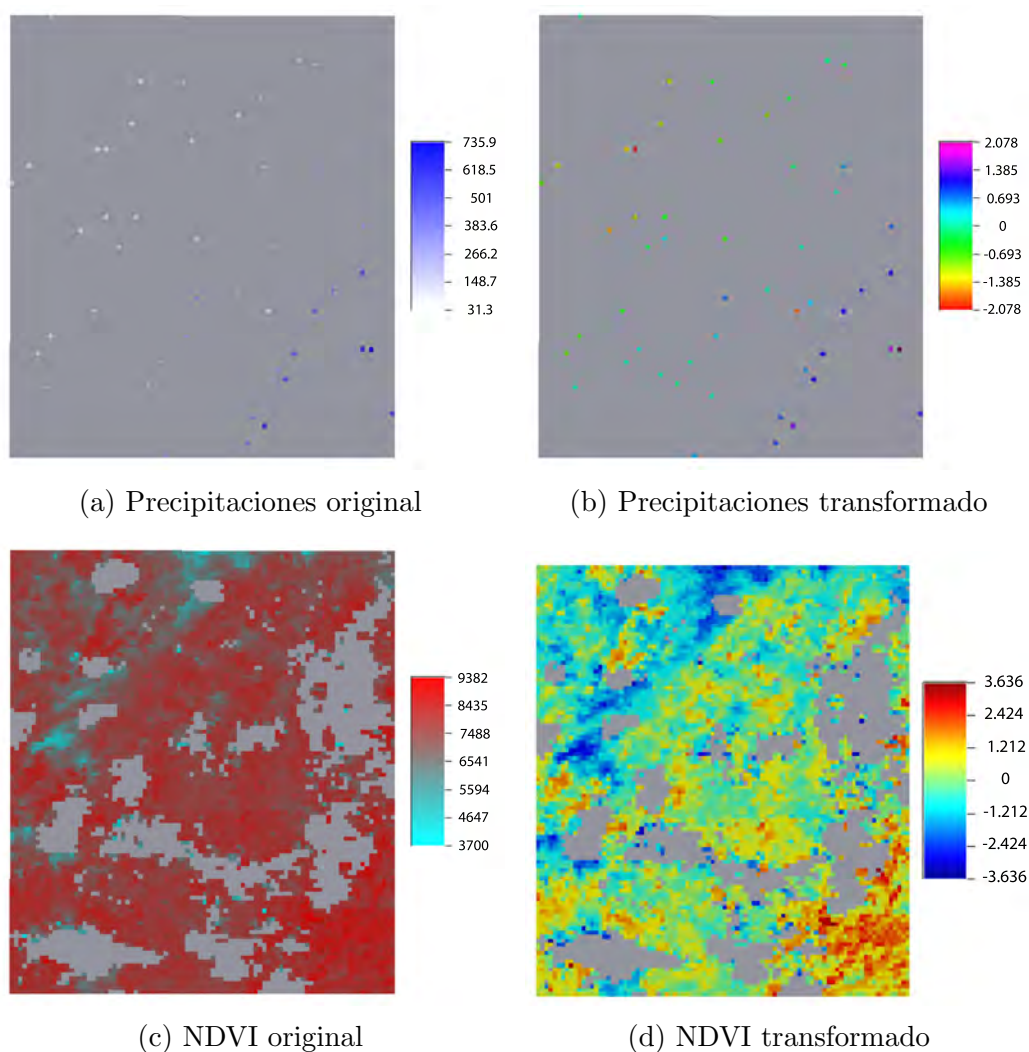


Figura 29: Distribución espacial de los datos originales y transformados. Ejemplo para mes de mayo

4.2. Análisis Estructural

Con cada mes transformado, se procedió a la generación de variogramas empíricos tanto omnidireccionales como unidireccionales, y posteriormente, se definieron modelos de ajuste en los que puedan apoyarse los procesos de simulación. Tanto la obtención de la mejor estructura de variograma como el

ajuste a ella de un modelo teórico se completó usando el software S-GEMS.

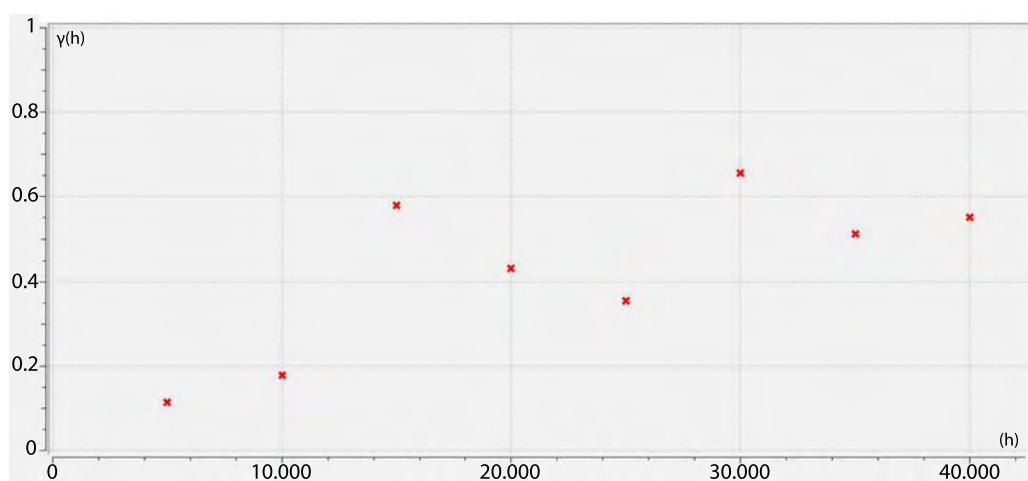
Las precipitaciones presentaron estructuras de comportamiento en el espacio en las que cada mes tiene sus particularidades y similitudes respecto a los de más. Por lo general, las precipitaciones se modelaron hasta distancias de 40.000 y 50.000 metros, ya que hasta este valor de distancia se encontró una estructura estable de correlación, el valor también equivale aproximadamente a la mitad del largo del rectángulo que define la zona de estudio. Los rezagos se variaron dependiendo del mes con valores entre los 4.000 y 5.000 metros siendo necesarios para el modelamiento cantidades de 8, 10 y 12 rezagos dependiendo del caso. En las Figuras 30, 31, 32, y 33 se detallan los variogramas omnidireccionales obtenidos en la investigación.

Para la construcción de variogramas unidireccionales se usaron los mismos parámetros de distancia y separación que se usaron para construir los variogramas empíricos omnidireccionales. La estructura de correlación espacial fue analizada en direcciones específicas de 0, 45, 90 y 135 grados y el resultado (Figuras 34, 35, 36 y 37) muestra comportamientos totalmente independientes entre las direcciones, con lo cual se puede afirmar que el proceso es totalmente anisotrópico descartando isotropía a cualquier escala. La Figura 46 hace más evidente este hecho, en ella se observa como ejemplo para el mes de abril, los cuatro variogramas superpuestos y de estos se concluye que no hay relación entre la dirección y la asociación espacial de los datos. Volviendo a las Figuras 34 a 37, se observa también que la dirección de 90 grados es donde generalmente la semivarianza se mantiene creciente hasta una mayor distancia de separación si se compara con las otras direcciones, y que la dirección de 45 grados es aquella donde a menor distancia los valores de semivarianza se establecen relativamente constantes. Adicionalmente, en el mismo gráfico, se pueden apreciar los modelos teóricos de semivarianza utilizados para cada mes en las precipitaciones, donde se visualiza el ajuste en simultáneo para las cuatro direcciones, y sus parámetros de configuración se eligieron dando prioridad a la dirección que presenta la mayor estabilidad.

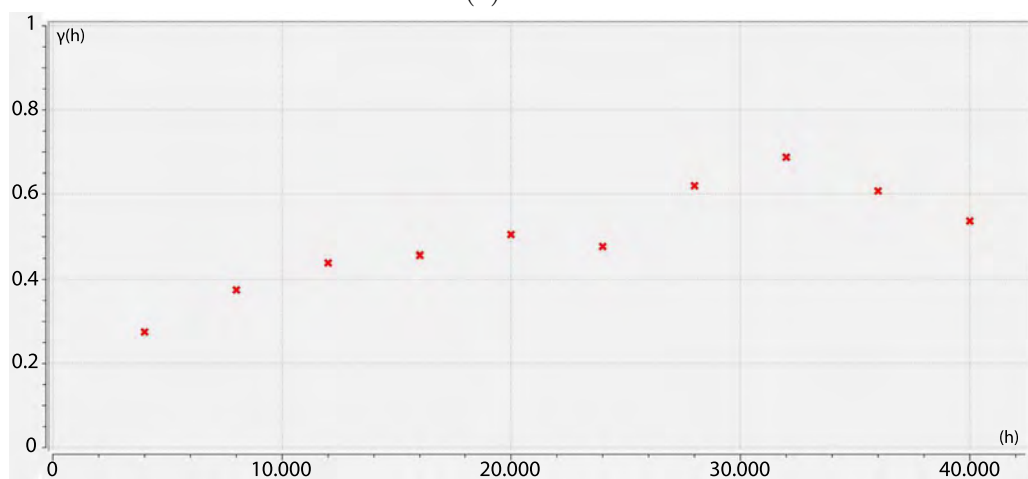
En cuanto al NDVI, se obtienen estructuras de correlación más robustas debido a la gran cantidad de información que existe para la generación de los variogramas empíricos tal como se aprecia en las Figuras 38, 39, 40 y 41. En este caso, se usaron separaciones de 2.000 metros con 20 rezagos, lo que proporciona el cálculo hasta los 40.000 metros al igual que se hizo con las precipitaciones. Una particularidad de estos variogramas es que la semivarianza siempre se mantiene creciente, pero a partir de cierta distancia

el aumento pasa a ser menor.

Analizando los variogramas direccionales a 0, 45, 90 y 135 grados (Figuras 42, 43, 44 y 45), es posible concluir varias situaciones. En la mayoría de meses, hasta distancias de 20.000 metros se observa que todas las direcciones presentan un comportamiento bastante similar, lo que da indicios de un proceso isotrópico a escalas a pequeñas y medias. Una vez alcanzada esta distancia, el comportamiento empieza a variar pero presentando un patrón que sugiere asociación entre las direcciones 0, 90 y 135. Tal situación sugiere que el proceso a gran escala puede ser geoméricamente anisotrópico. El hecho es más fácil de comprender en la Figura 46, que superpone los variogramas direccionales usando como ejemplo el mes de agosto. Como puede observarse en las Figuras 42 a 45, no existe algún variograma direccional que se destaque por presentar una estructura más estable que el resto, por lo que en el proceso de ajuste a sentimiento del modelo no se priorizó ninguna dirección sobre las otras, en el gráfico se puede ver que cada modelo de semivarianza logra un acople igual de aceptable en todas las orientaciones. Se deduce que en general, el NDVI presenta una variación mínima de sus valores cuando más proximidad hay entre ellos, y una variación mayor mientras más separación exista entre los mismos, lo cual sugiere coberturas vegetales homogéneas distribuidas en la región.

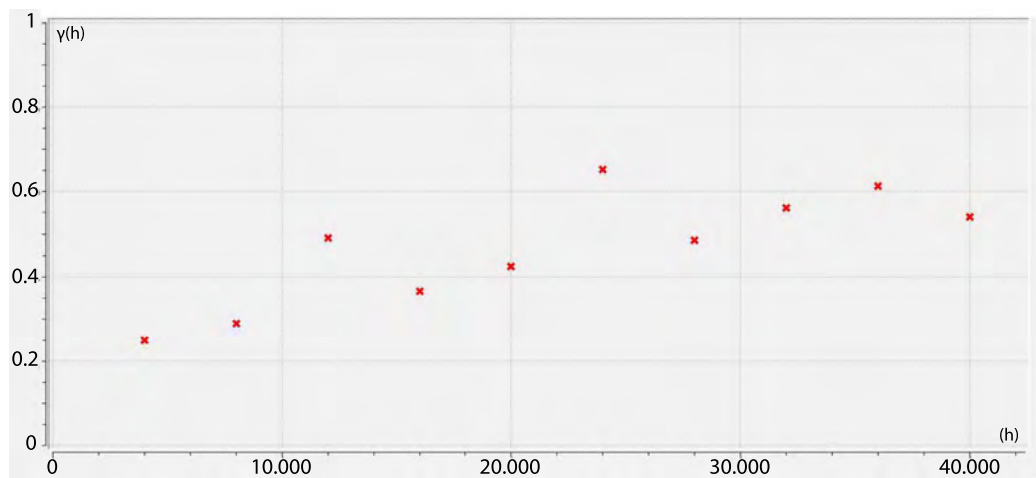


(a) Marzo

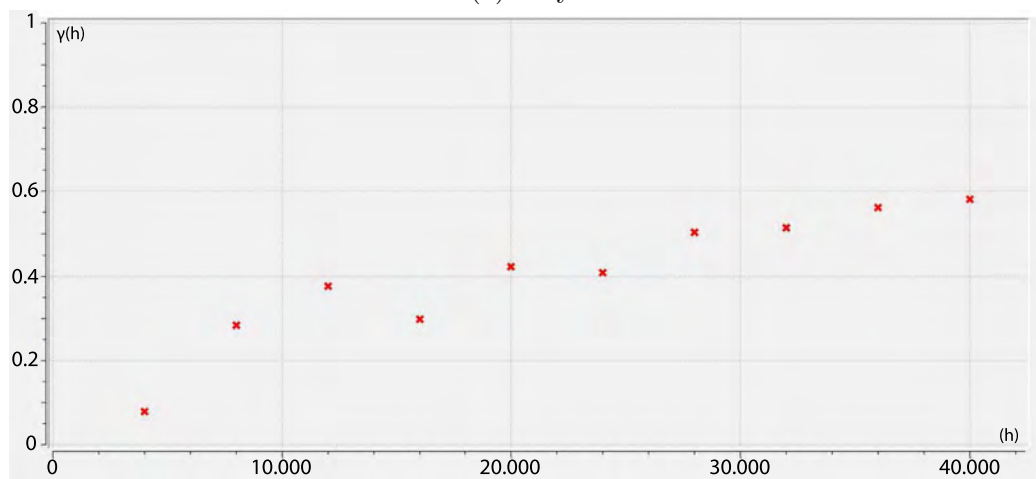


(b) Abril

Figura 30: Variogramas omnidireccionales de las precipitaciones para los meses de marzo y abril

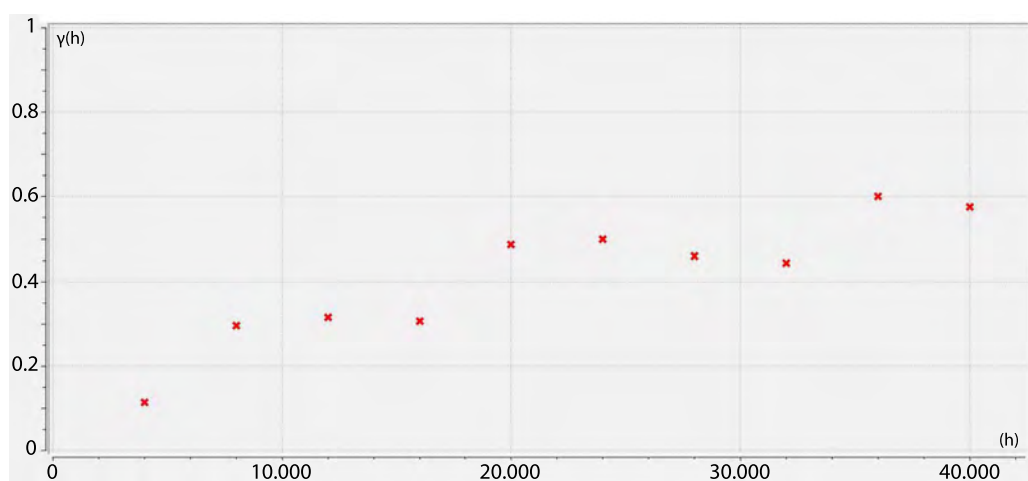


(a) Mayo

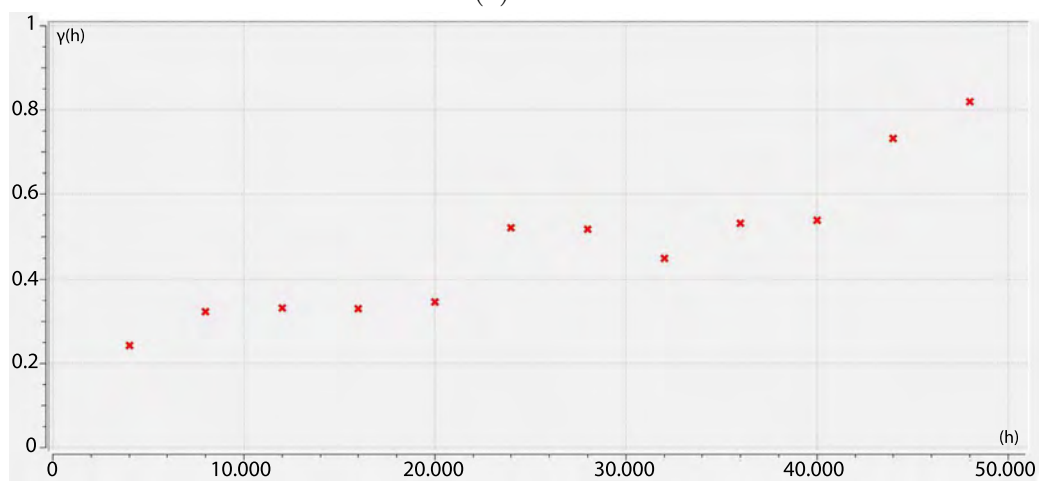


(b) Junio

Figura 31: Variogramas omnidireccionales de las precipitaciones para los meses de mayo y junio

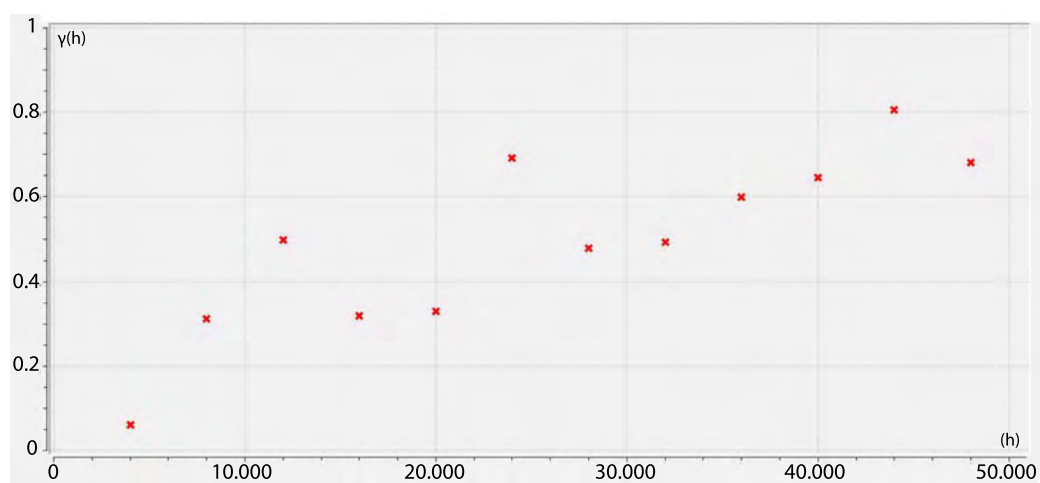


(a) Julio

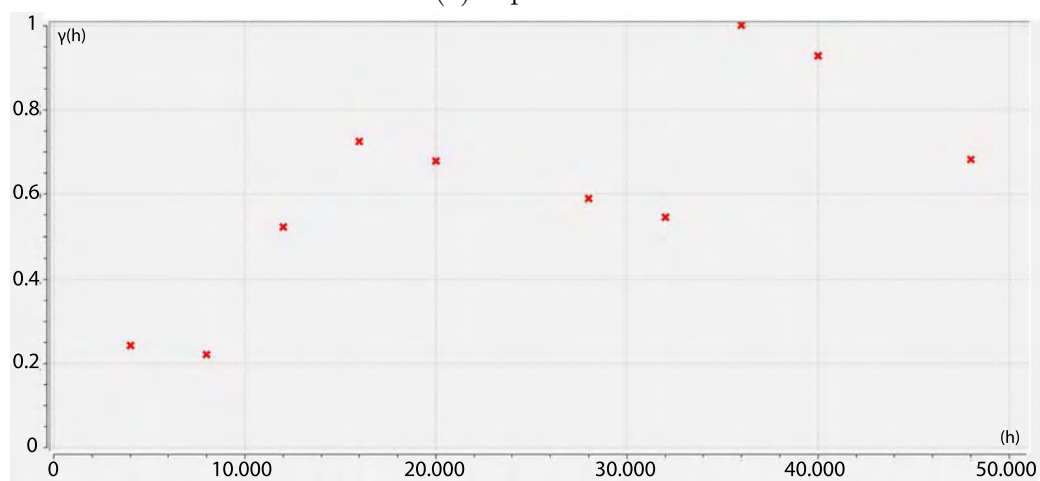


(b) Agosto

Figura 32: Variogramas omnidireccionales de las precipitaciones para los meses de julio y agosto

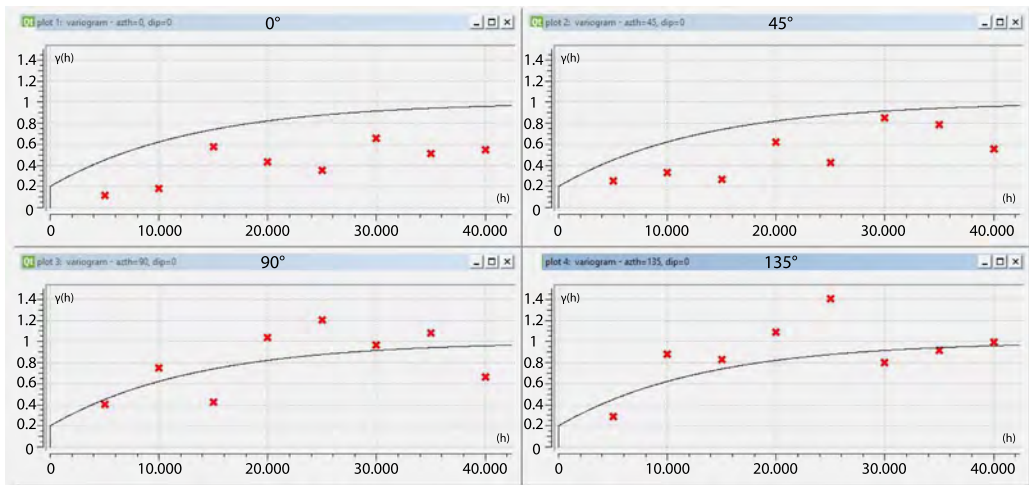


(a) Septiembre

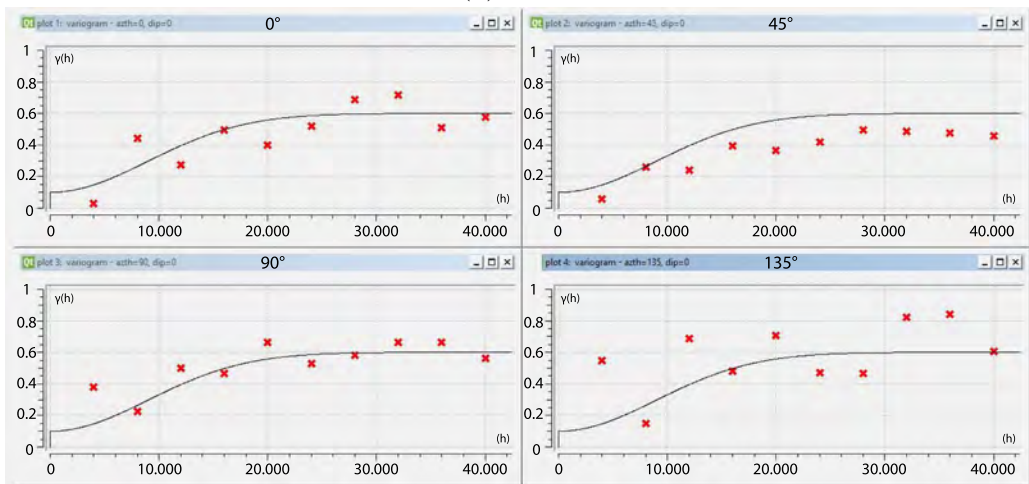


(b) Octubre

Figura 33: Variogramas Omnidireccionales de las precipitaciones para los meses de septiembre, y octubre

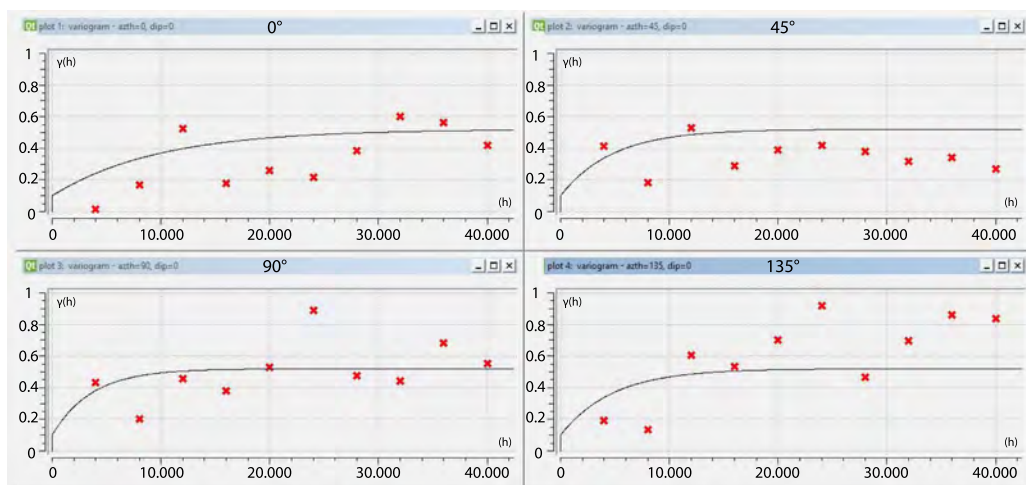


(a) Marzo

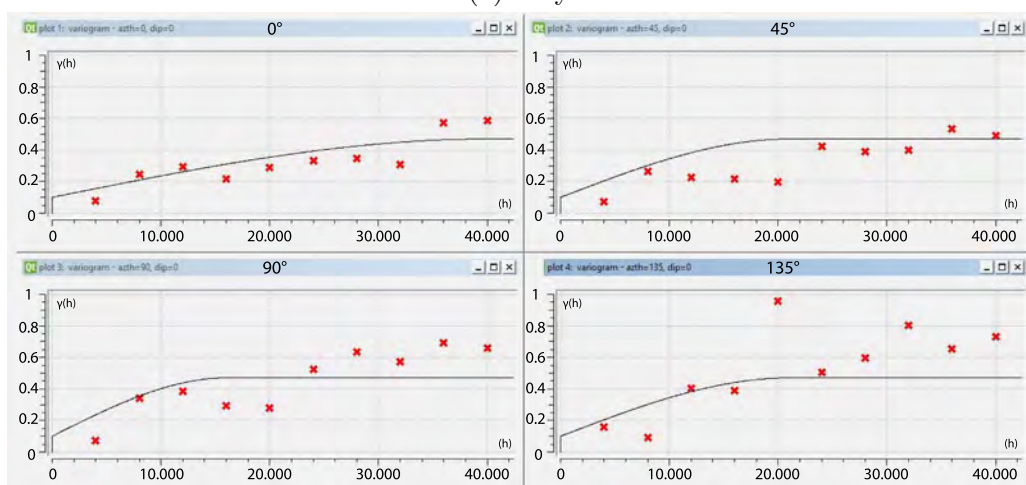


(b) Abril

Figura 34: Variogramas direccionales y ajuste de modelos de semivarianza para las precipitaciones en los meses de marzo y abril

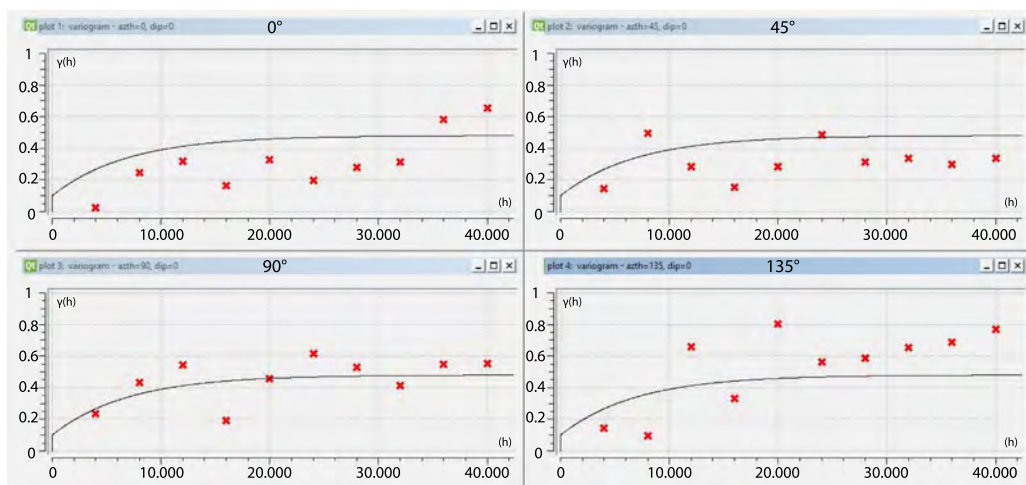


(a) Mayo

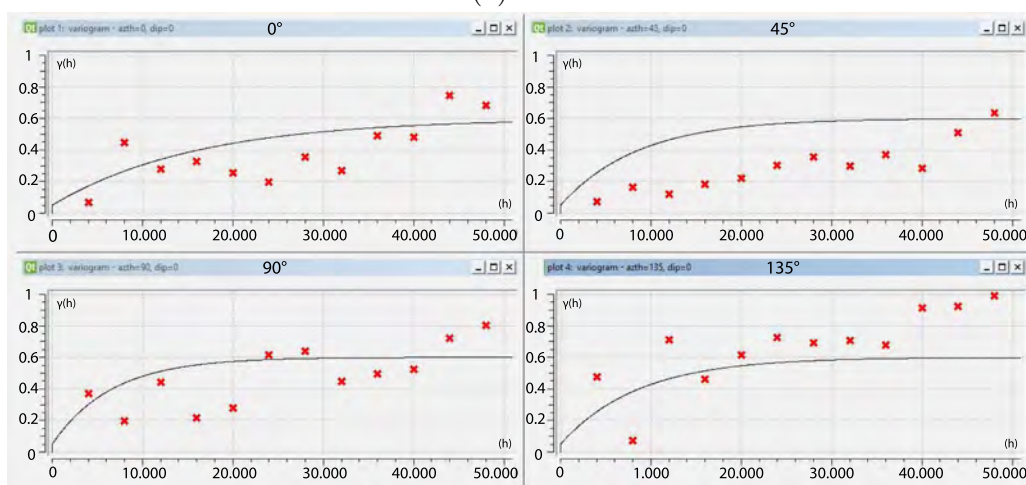


(b) Junio

Figura 35: Variogramas direccionales y ajuste de modelos de semivarianza para las precipitaciones en los meses de mayo y junio

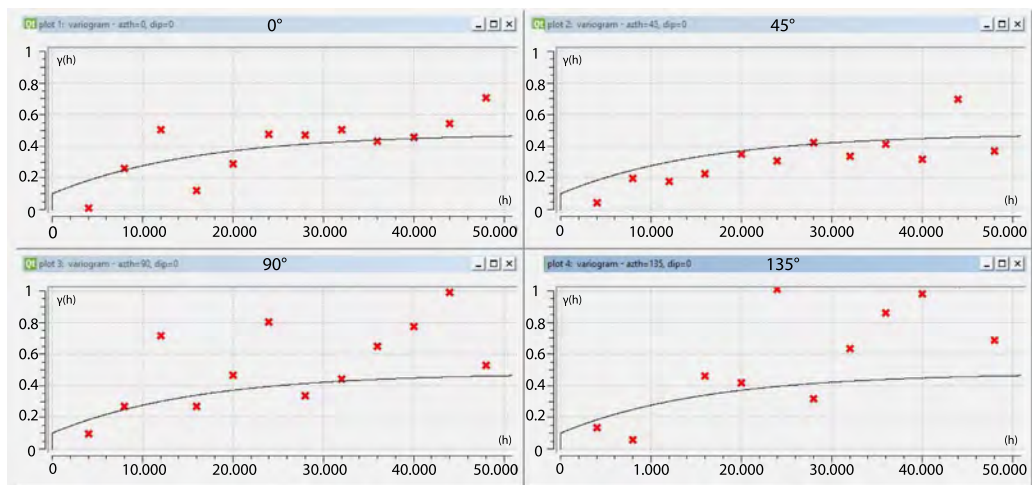


(a) Julio

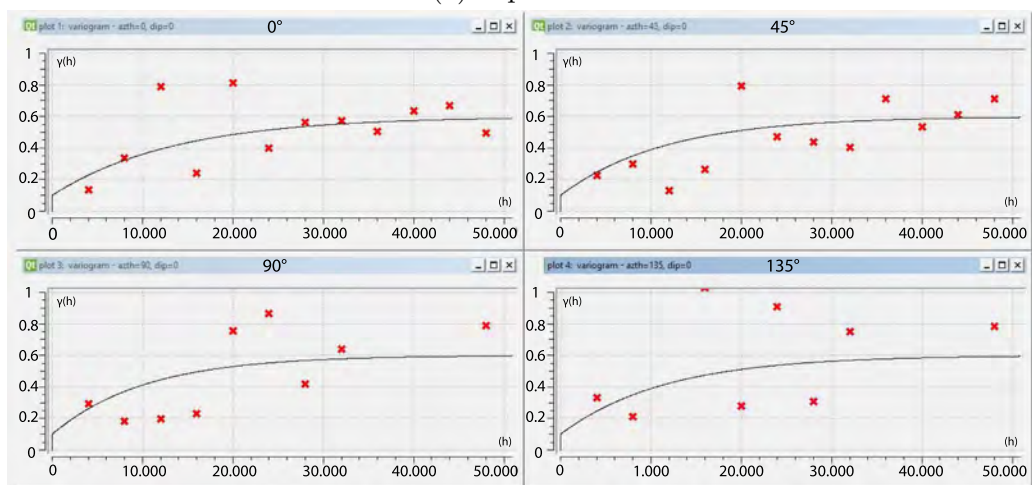


(b) Agosto

Figura 36: Variogramas direccionales y ajuste de modelos de semivarianza para las precipitaciones en los meses de julio y agosto

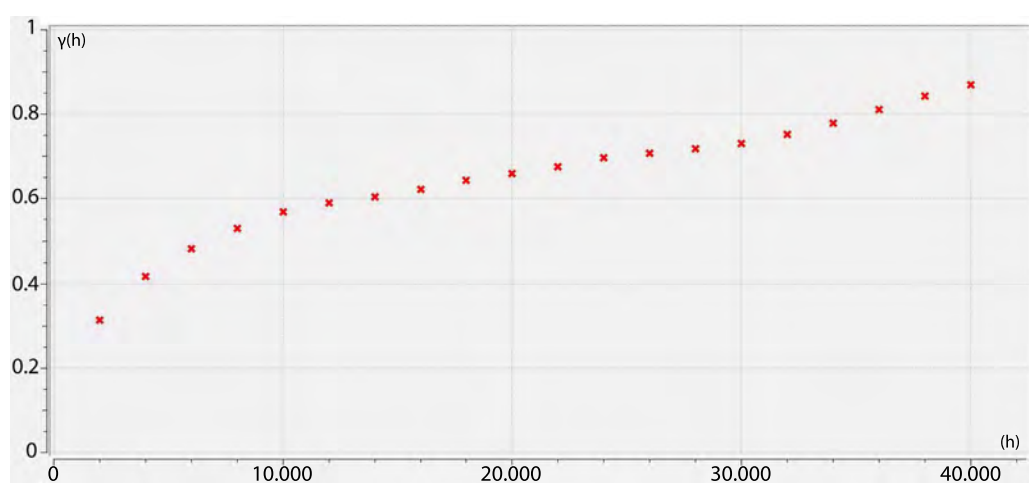


(a) Septiembre

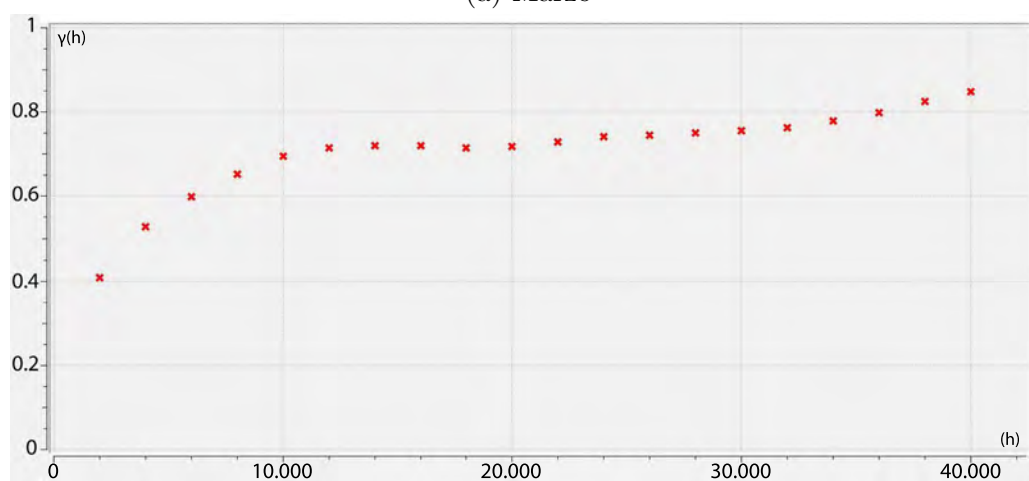


(b) Octubre

Figura 37: Variogramas direccionales y ajuste de modelos de semivarianza para las precipitaciones en los meses de septiembre y octubre

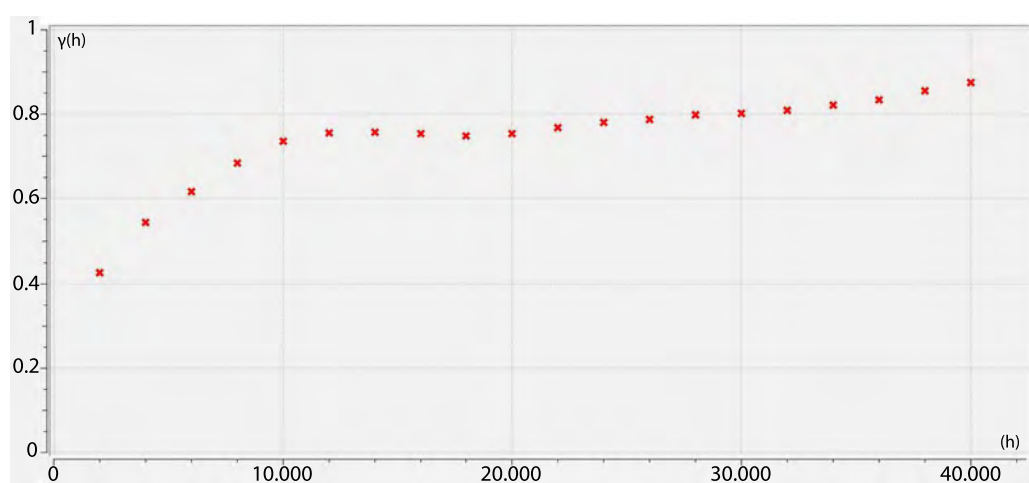


(a) Marzo

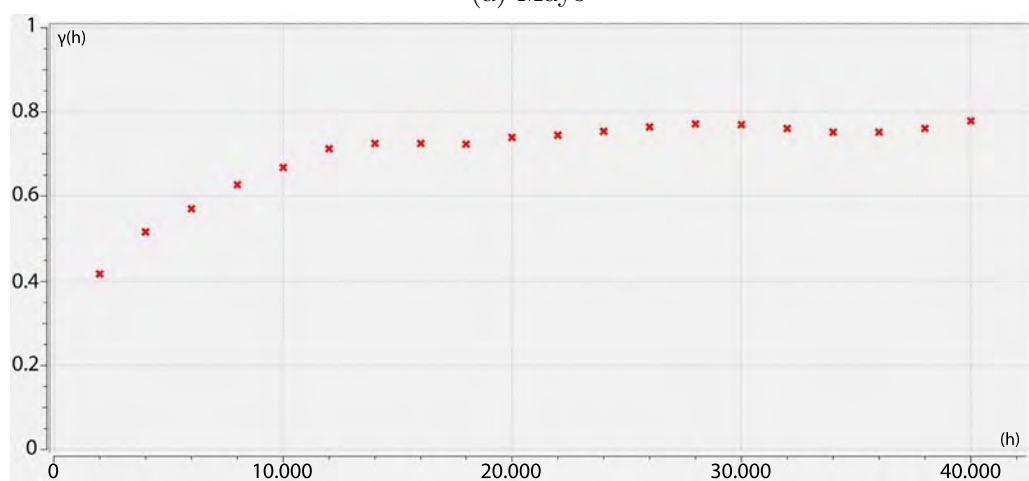


(b) Abril

Figura 38: Variogramas omnidireccionales para el NDVI para los meses de marzo y abril

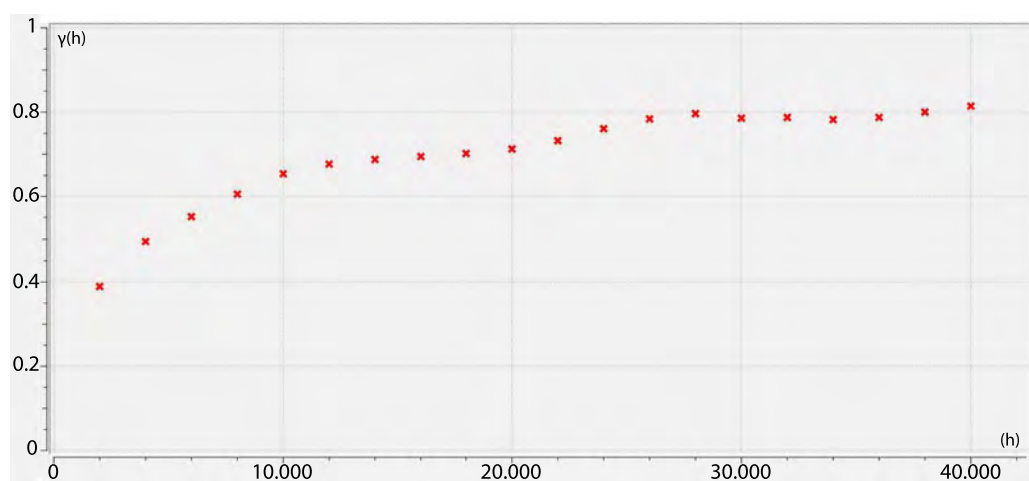


(a) Mayo

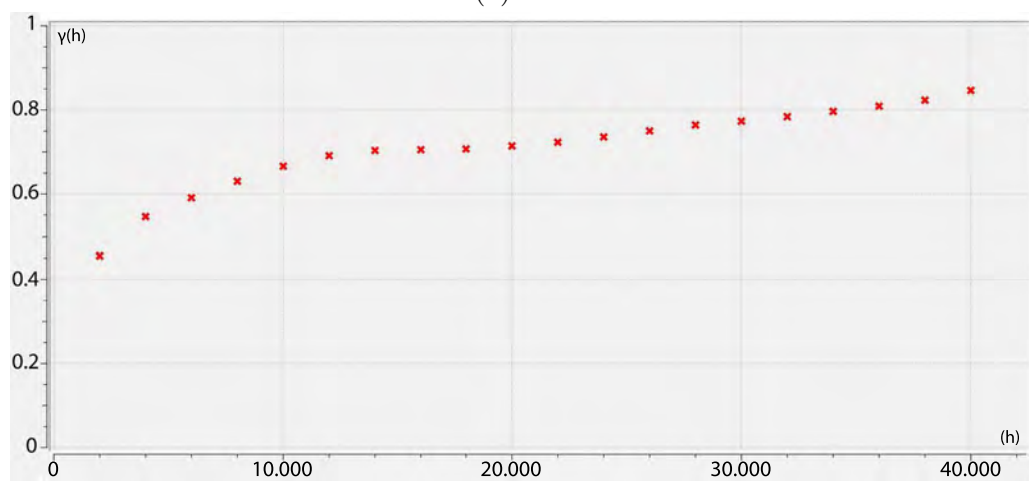


(b) Junio

Figura 39: Variogramas omnidireccionales para el NDVI para los meses de mayo y junio

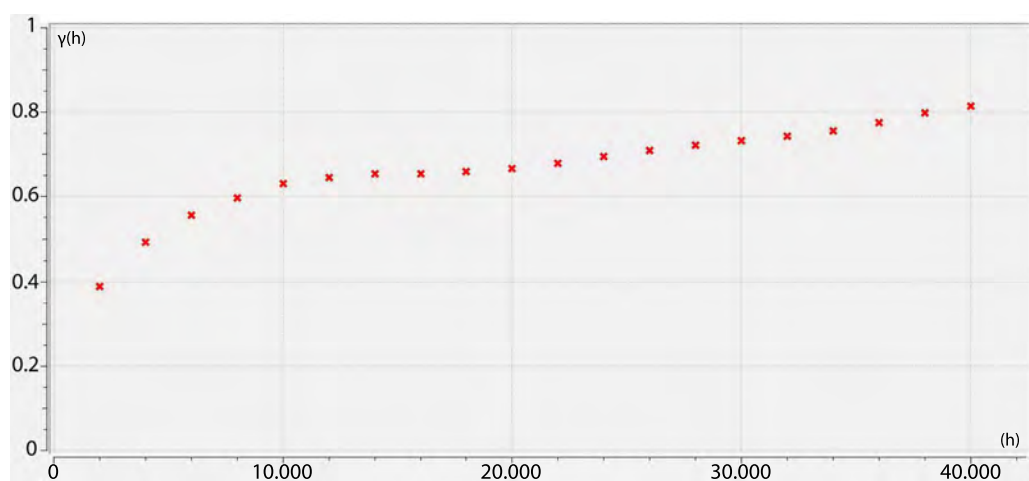


(a) Julio

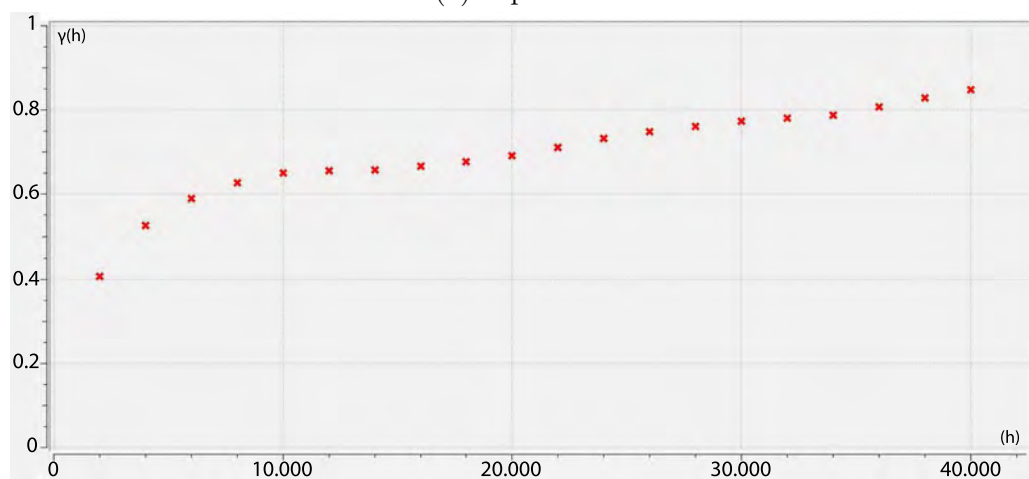


(b) Agosto

Figura 40: Variogramas omnidireccionales para el NDVI para los meses de julio y agosto

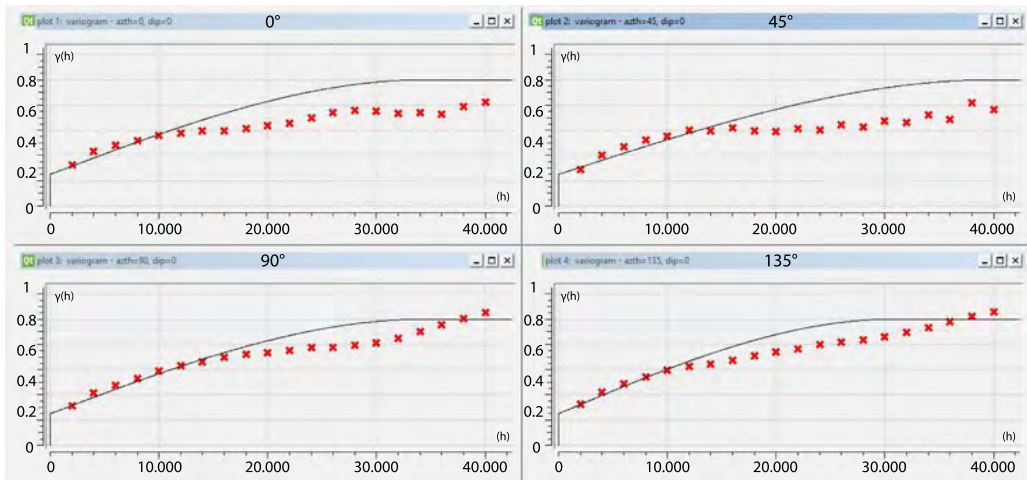


(a) Septiembre

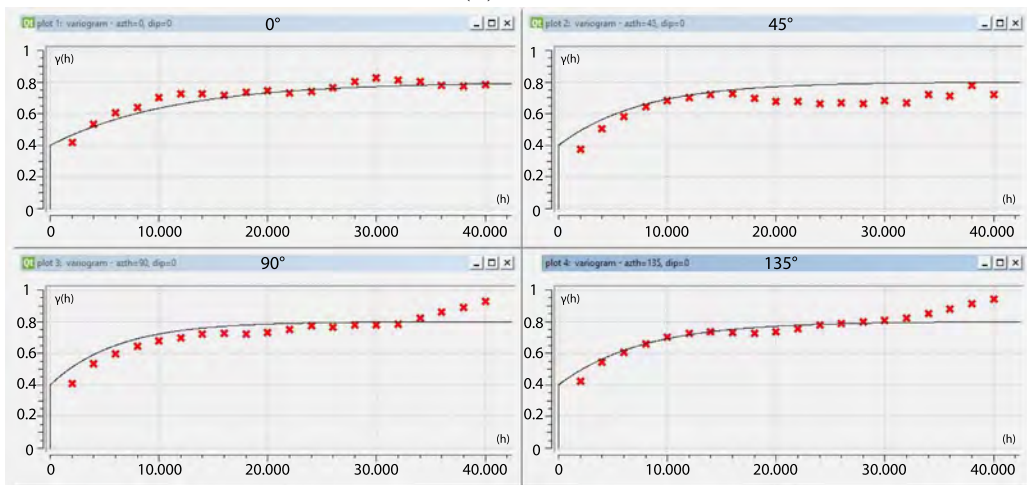


(b) Octubre

Figura 41: Variogramas omnidireccionales para el NDVI para los meses de septiembre y octubre

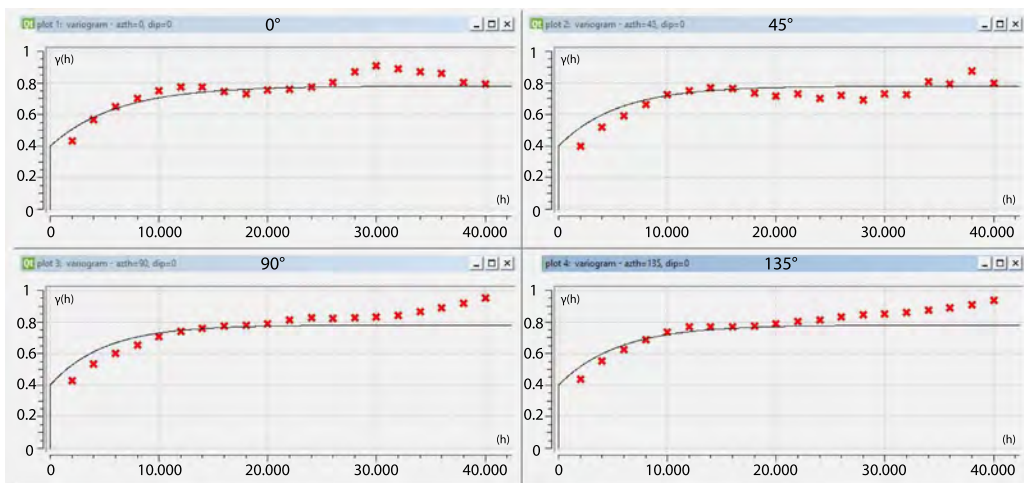


(a) Marzo

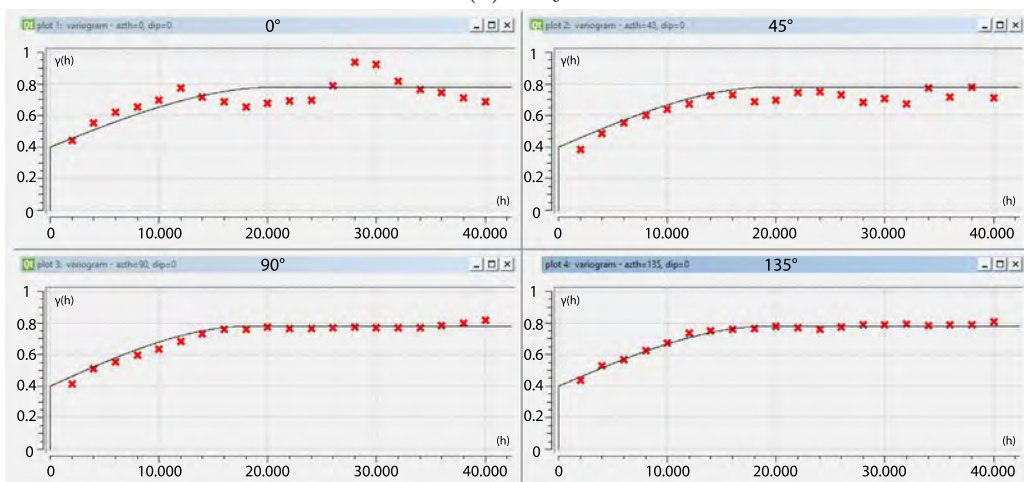


(b) Abril

Figura 42: Variogramas direccionales y ajuste de modelos de semivarianza para el NDVI en los meses de marzo y abril

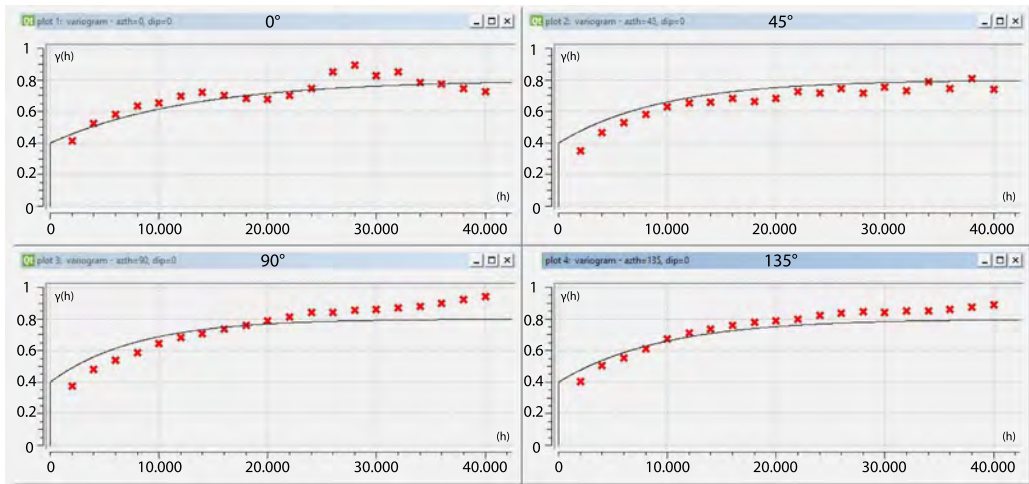


(a) Mayo

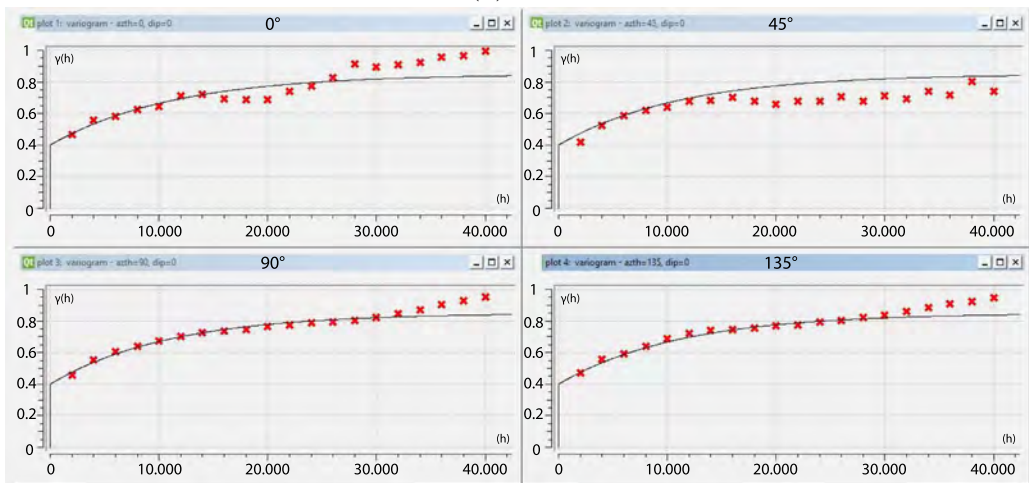


(b) Junio

Figura 43: Variogramas direccionales y ajuste de modelos de semivarianza para el NDVI en los meses de mayo y junio

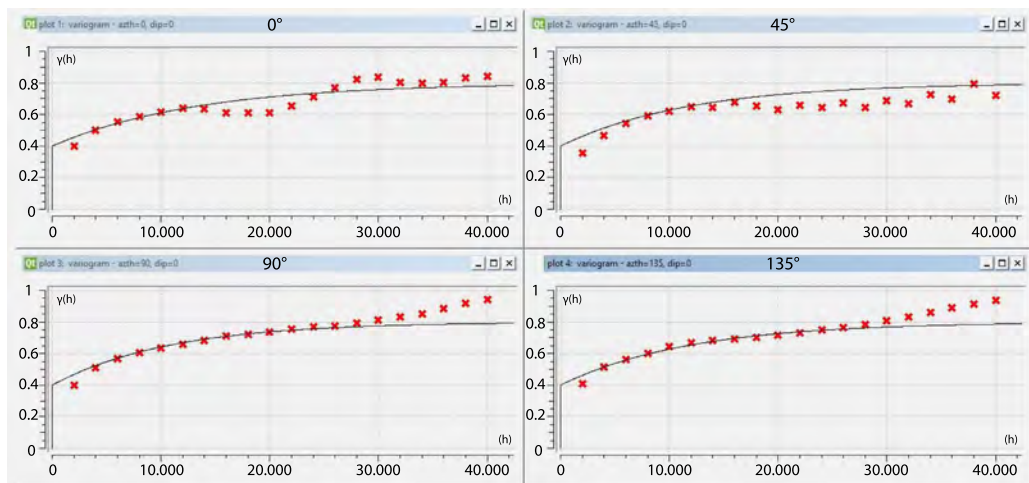


(a) Julio

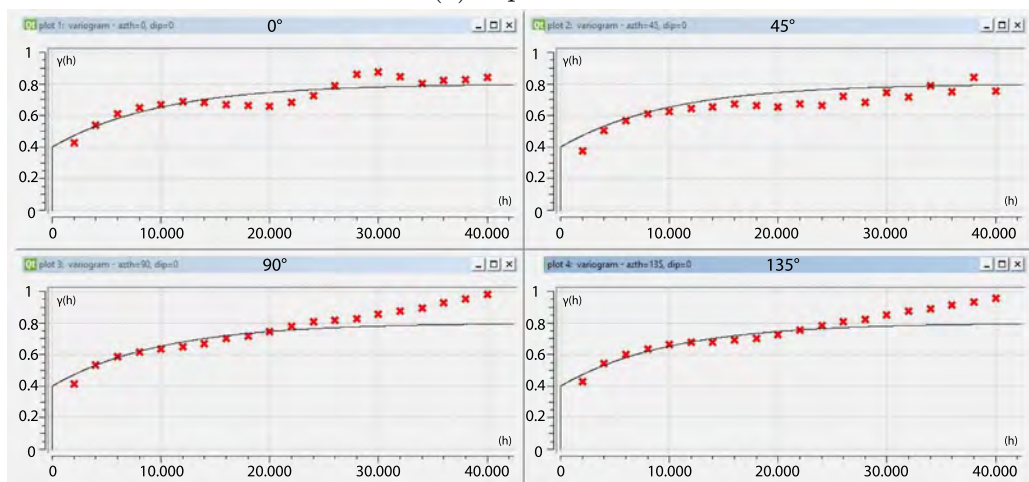


(b) Agosto

Figura 44: Variogramas direccionales y ajuste de modelos de semivarianza para el NDVI en los meses de julio y agosto



(a) Septiembre



(b) Octubre

Figura 45: Variogramas direccionales y ajuste de modelos de semivarianza para el NDVI en los meses de septiembre y octubre

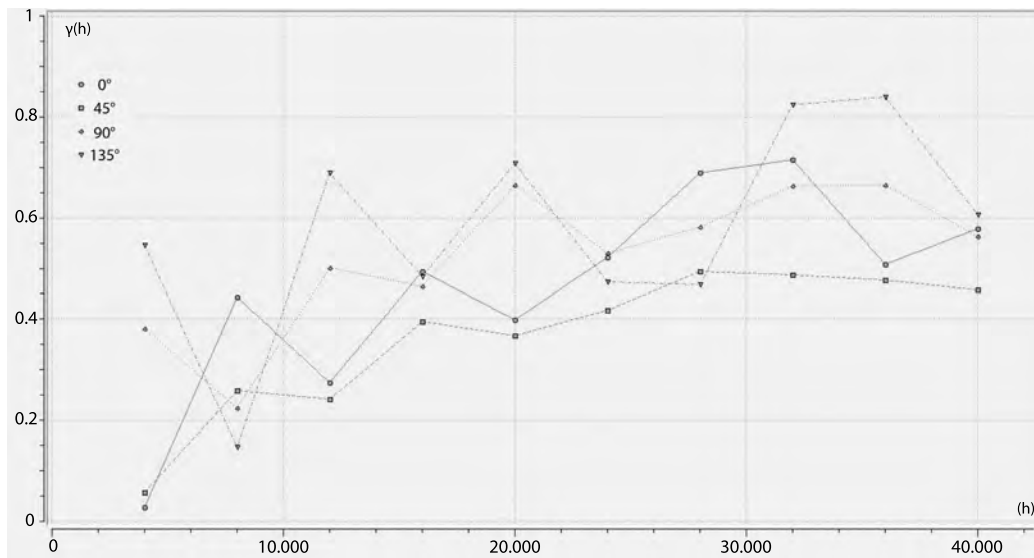


Figura 46: Variogramas direccionales de las precipitaciones

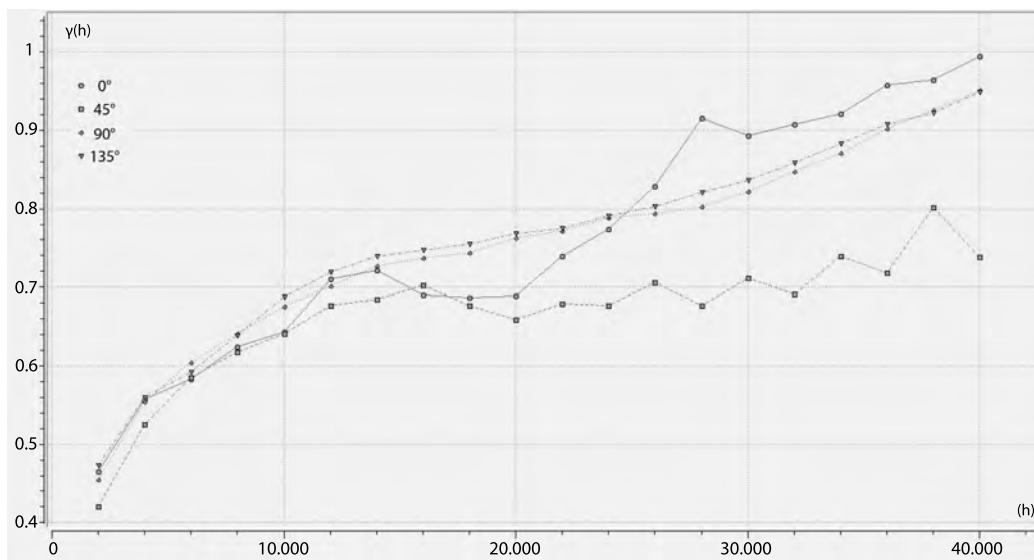


Figura 47: Variogramas direccionales del NDVI

4.3. Simulación Gaussiana del NDVI

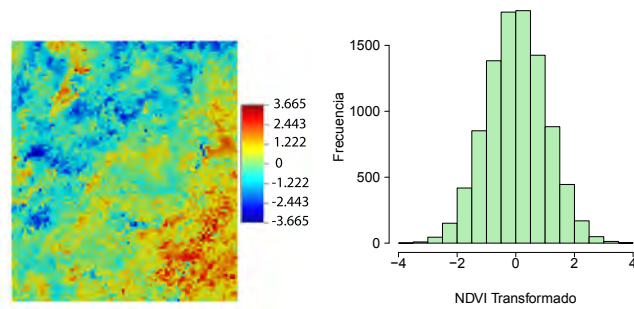
Con las estructuras de correlación modeladas, es posible densificar valores de las variables en zonas no muestreadas, y así replicar el comportamiento de la variable objetivo en toda la zona de estudio. Sin embargo, antes de densificar la precipitación, debe tenerse presente que a pesar de que el NDVI cuenta con información abundante de muestreo, la grilla de simulación no contendrá valores en todas sus posiciones cuando dicha variable se incorpore en ella, esto se debe a los píxeles rechazados en el proceso de filtrado. Por tal motivo, los valores faltantes de NDVI deben completarse y para ello se recurre a procesos de simulación gaussiana, que permitan generar valores de soft data en la totalidad de la cuadrícula, evitando posible ruido cuando ambas variables se correlacionen espacialmente.

Teniendo en cuenta el tamaño del pixel y las características del espaciamiento de los datos, se definió la grilla de simulación con dimensiones de 83.520 por 96.200 metros conformada por 90 filas y 104 columnas, y tamaños de celda de 928 y 925 metros respectivamente. En el diseño de la grilla, se busca que cada celda contenga en su interior, el centro de pixel que le corresponde por ubicación, ya que otros tamaños pueden ser inconsistentes: si son muy grandes, suavizan la distribución espacial de la variable, y si son muy pequeños, las celdas entre píxeles sucesivos no capturan valores, por ende, un gran número de ellas quedaría sin información.

A la grilla de simulación, se incorpora la variable NDVI en su versión transformada para cada uno de los meses. Luego se procede a ejecutar la simulación gaussiana indicando el mes a calcular y su respectiva estructura de correlación espacial representada por el modelo de variograma ajustado. Con el método de estimación de kriging simple, se corrieron 30 realizaciones por mes, constatando que sus histogramas reflejaran el comportamiento gaussiano. La Figura 48 muestra como ejemplo para el mes de marzo, las realizaciones 1, 10, 20 y 30, con sus correspondientes histogramas. Como puede observarse, cada realización tiene un histograma que respeta la distribución de los datos de entrada para el proceso de simulación, es decir, el histograma producto de la transformación gaussiana.

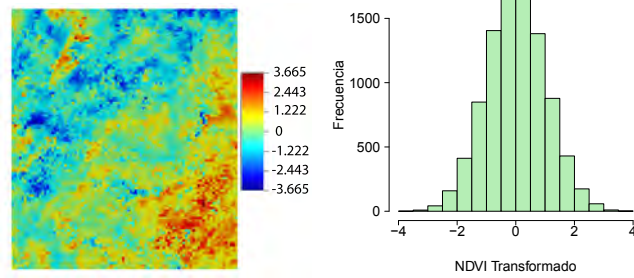
Finalmente, las 30 realizaciones de cada mes son promediadas y el mapa resultante del promedio se regresa a la escala original de valores con un proceso de anti-transformación, en el que la distribución buscada la define el histograma del mes en su versión original, y por lo tanto se espera la réplica

de su comportamiento. En la parte superior de la Figura 49 se muestra para el mes de septiembre, el mapa promedio de las 30 realizaciones y el resultado de su anti-transformación que lo lleva de vuelta a la escala de medición inicial, los respectivos histogramas se observan en la parte inferior. Con la corroboración de que el histograma de la anti-transformación se asemeja en forma y proporción de valores al histograma original del correspondiente mes, se consideran concluidos el análisis y la preparación de la variable NDVI, habiendo constituido lo que serán las entradas de la variable secundaria para la co-simulación.



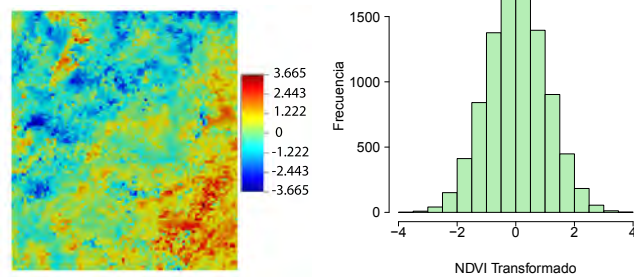
(a) Mapa 1

(b) Histograma 1



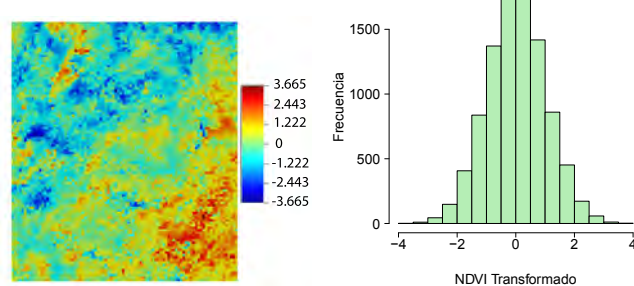
(c) Mapa 10

(d) Histograma 10



(e) Mapa 20

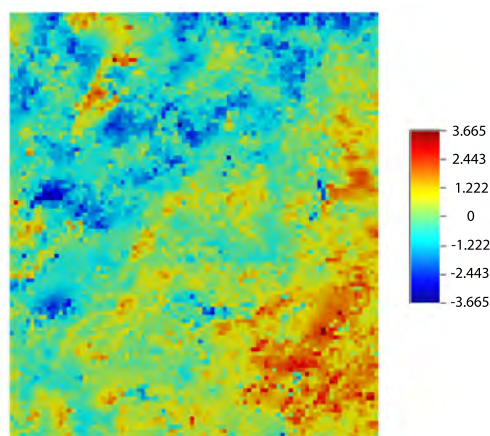
(f) Histograma 20



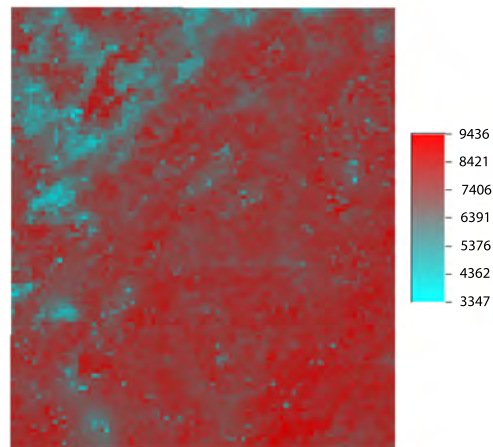
(g) Mapa 30

(h) Histograma 30

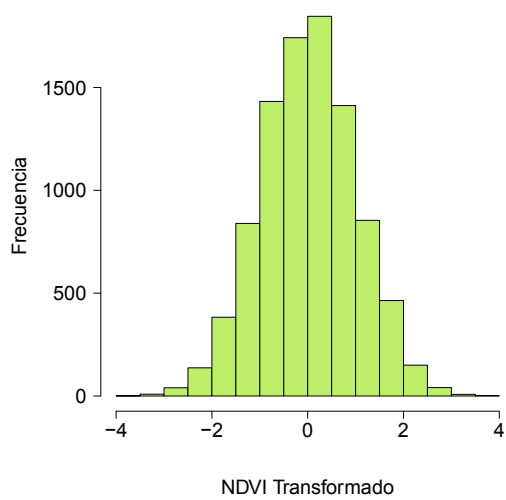
Figura 48: Resultados de mapas e histogramas de la simulación gaussiana del NDVI - Realizaciones 1, 10, 20 y 30 mes de marzo



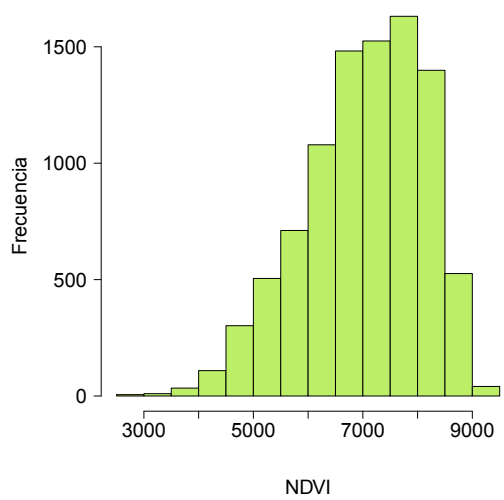
(a) Mapa transformada



(b) Mapa anti-transformada



(c) Histograma transformada



(d) Histograma anti-transformada

Figura 49: Mapas e histogramas promedio de la simulación gaussiana del NDVI transformado y anti-transformado para el mes de septiembre

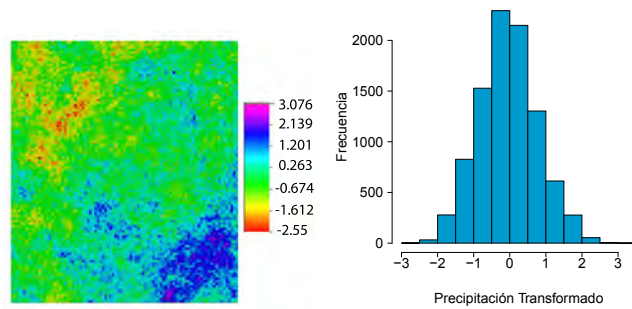
4.4. Co-Simulación Gaussiana entre las Precipitaciones y el NDVI

Para cada mes, las precipitaciones transformadas a la forma gaussiana establecen el conjunto de hard data, su estructura de correlación espacial está definida por los modelos de variograma que se ajustaron en la etapa de análisis estructural. El índice de vegetación, considerado también en la forma gaussiana, establece el soft data, y gracias a la simulación secuencial gaussiana aplicada, ha poblado toda zona de estudio. Como paso siguiente, se calcula el coeficiente de correlación de cada mes para todas las posiciones donde se conoce información de las dos variables, es decir, en las ubicaciones de las estaciones meteorológicas.

Con la información anteriormente descrita fue posible calcular las co-simulaciones secuenciales gaussianas y construir la distribución espacial de la lluvia basada en su relación con el índice de vegetación. El proceso utilizó como método de estimación kriging simple, y como método de co-kriging el modelo de Markov tipo I, que tiene como parámetros el variograma del hard data y el coeficiente de correlación entre el hard y el soft data. Cada co-simulación comprendió la generación de 50 realizaciones, de ellas, se ilustran como ejemplo en la Figura 50, las número 1, 15, 30 y 50 para el mes de septiembre. Allí se corrobora cómo cada uno de los histogramas es fiel a la distribución obtenida con la transformación gaussiana de los datos.

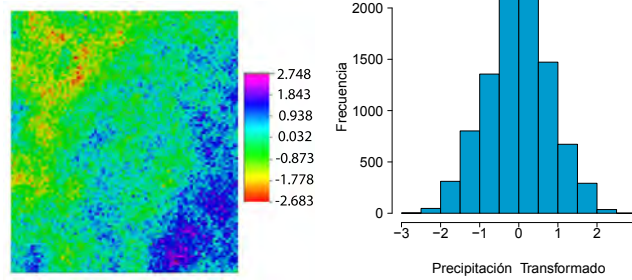
El mapa promedio de las 50 realizaciones puede observarse en la parte superior de la Figura 51, su histograma se anti-transforma para volver los datos a sus valores originales, y se logra la generación de mapas de la distribución de las precipitaciones en la región de estudio como lo muestra el gráfico en la parte derecha. Los respectivos histogramas se observan en la parte inferior de la misma figura, y como puede apreciarse, el histograma final en la escala original de valores sigue la misma distribución que presentaron las precipitaciones en el mes de septiembre.

Los mapas finales son mostrados en las en las Figuras 52 y 53 donde se observa la misma tendencia de agrupación detectada en la etapa exploratoria densificada a la totalidad de la región analizada. Con variaciones entre los meses, el patrón que se observa es que valores altos están concentrados en la parte suroriental de la zona y progresivamente hay disminución hacia las zonas centro y norte teniendo los más bajos en la zona noroccidental.



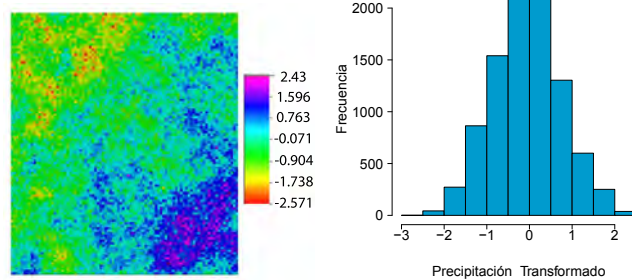
(a) Mapa 1

(b) Histograma 1



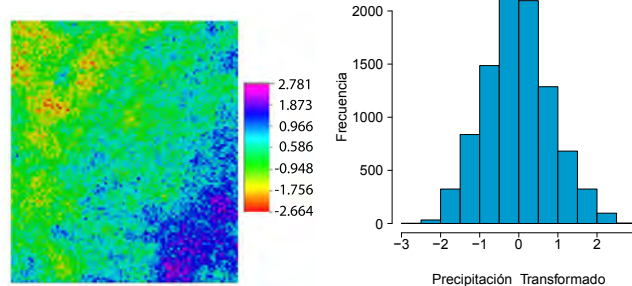
(c) Mapa 15

(d) Histograma 15



(e) Mapa 30

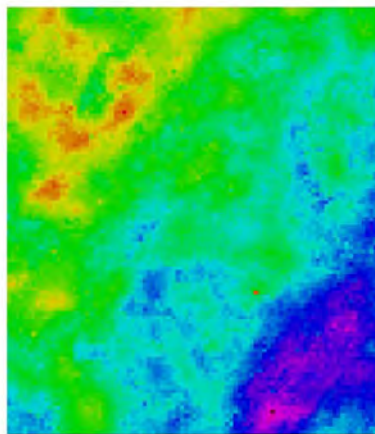
(f) Histograma 30



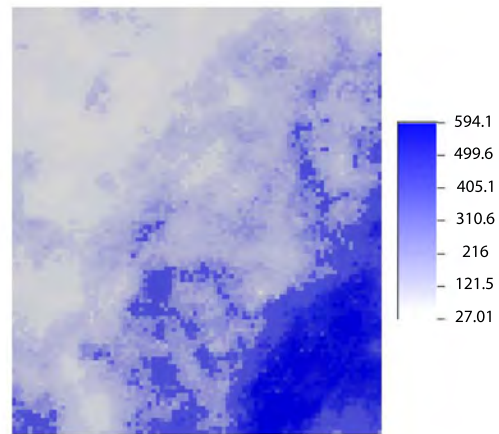
(g) Mapa 50

(h) Histograma 50

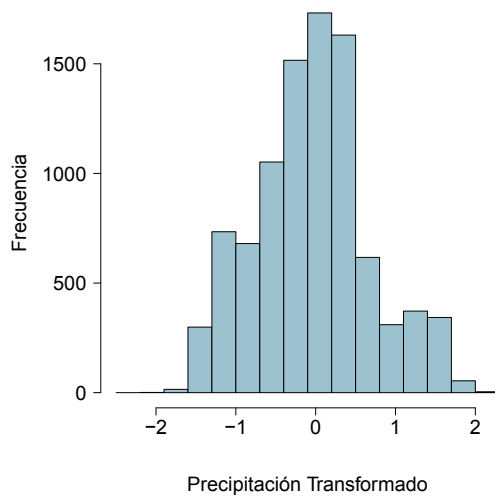
Figura 50: Resultados de mapas e histogramas de la co-simulación gaussiana de la precipitación - Realizaciones 1, 15, 30 y 50 mes de septiembre



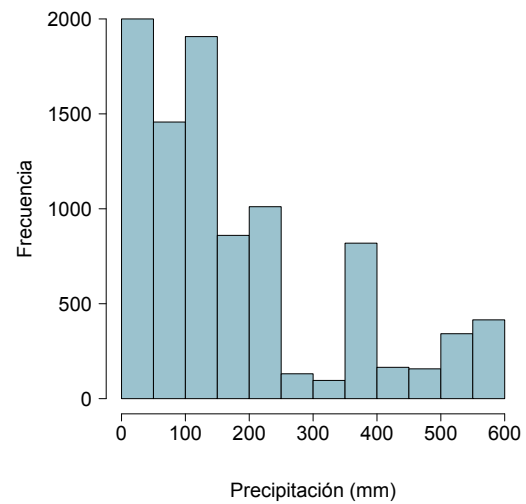
(a) Mapa transformada



(b) Mapa anti-transformada



(c) Histograma transformada



(d) Histograma anti-transformada

Figura 51: Mapas e histogramas promedio de la co-simulación gaussiana de la precipitación transformada y anti-transformada para el mes de septiembre

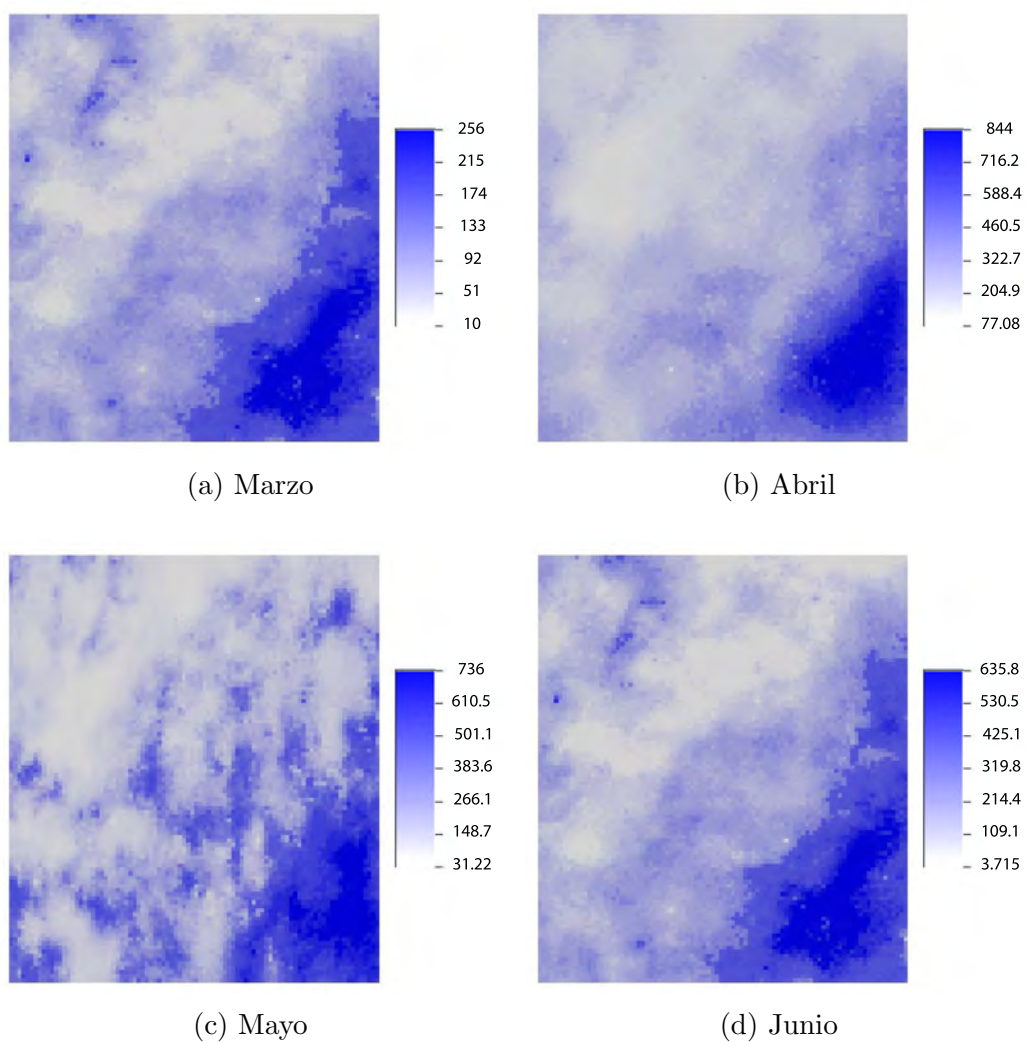


Figura 52: Mapa resultado de la co-simulación de las precipitaciones para los meses de marzo, abril, mayo y junio

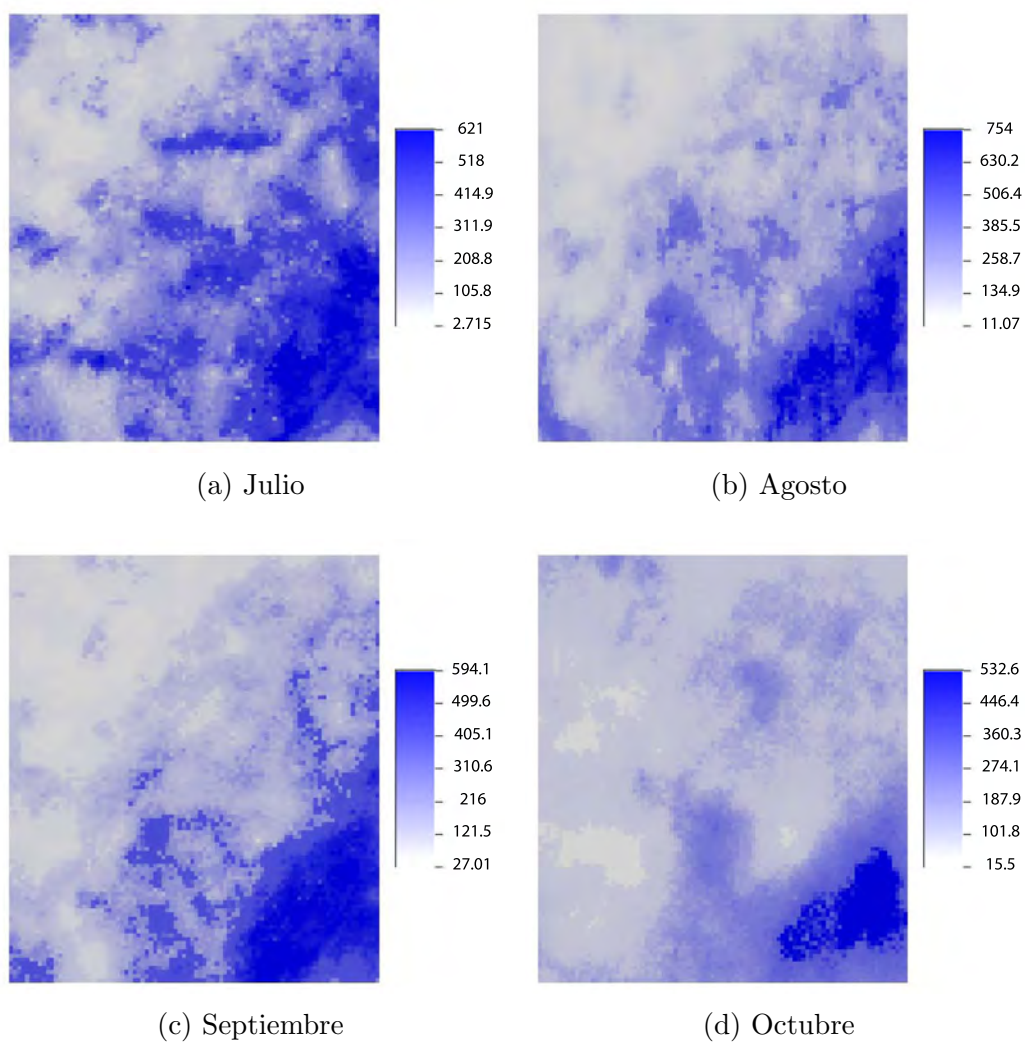


Figura 53: Mapa resultado de la co-simulación de las precipitaciones para los meses de julio, agosto, septiembre, octubre

4.5. Influencia de la Topografía en Resultado de las Precipitaciones

De los distintos mapas de co-simulación de las precipitaciones, puede apreciarse una trayectoria diagonal que ubica los valores más bajos en la parte noroccidental, y llega con los más altos a la parte suroriental. Si esta distribución de valores se analiza teniendo en cuenta las condiciones topográficas de la región, es notorio que el aumento en los valores de precipitación, se empieza a dar en aquellas zonas donde se transita desde los relieves de altiplanicie hacia la cordillera y por lo tanto se van alcanzando mayores pendientes: por ejemplo, en la zona media que se dan pendientes moderadas, los valores de precipitación pasan de rangos bajos a medios, distribuyéndose en áreas que sin ser tan extensas, presentan una mayor progresión en la forma que varía el terreno. A su vez, los valores altos, se dan donde mayor pendiente de terreno se tiene, evidenciando que la topografía está estrictamente relacionada con el comportamiento de las precipitaciones y la vegetación.

En la zona donde se dan los valores de precipitación generalmente más bajos, se observa una menor variación de la altitud, debido a la predominancia de las planicies en altura, las cuales son características de la región, allí también se presentan valores bajos del NDVI y se tiende a la sequía. Esto puede explicarse por qué los valores de precipitación bajos, se distribuyen con menor densidad en comparación con los medios y altos, prolongando su presencia hacia áreas más extensas cuya variación topográfica es menor. Así mismo, los valores altos tienden a estar concentrados en menores proporciones de terreno en las que justamente, la pendiente presenta variaciones abruptas y se tiene mayor concentración de vegetación según los valores del NDVI.

Una relación directa entre las condiciones de topografía escarpada y los altos valores de precipitación que es constante en los meses analizados, puede interpretarse en la región suroriental donde se alcanza un filo con altas pendientes que inicia el tránsito hacia el piedemonte llanero por el oriente. Allí los valores de precipitación que suelen ser los más altos, se distribuyen a ambos lados de la vertiente y se posicionan en las zonas de topografía más pronunciadas.

La figura 54 muestra una comparación tomando el mes de abril, entre el resultado de la co-simulación, el modelo de terreno SRTM (Shuttle Radar Topography Mission) de la región, y los valores de NDVI. Allí se pueden

detectar las relaciones espaciales descritas, y observar la correspondencia de valores y sus variaciones según la localización confirmando la estrecha interacción entre la lluvia, la vegetación, y la topografía.

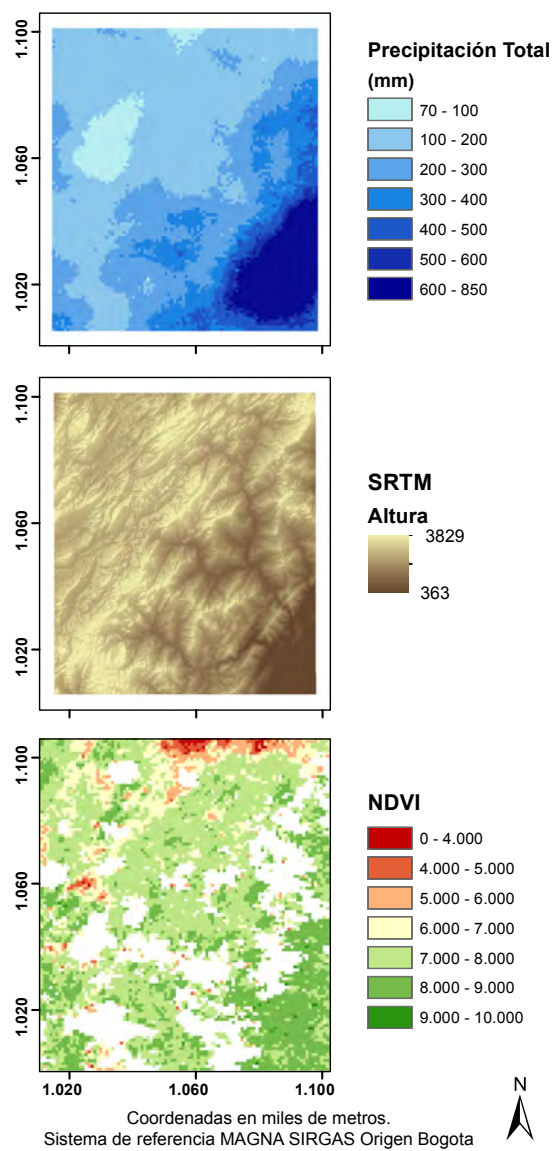


Figura 54: Comparación entre la precipitación resultante de la co-simulación, la topografía y el NDVI para el mes de abril

Conclusiones

Se generaron mapas de la distribución mensual de las precipitaciones en la región a partir de la relación lluvia - NDVI. Dichos mapas respetan las propiedades estadísticas y variabilidad espacial original observada en los datos de las estaciones meteorológicas, logrando una densificación apropiada de la variable de interés.

Los mapas de precipitaciones obtenidos exponen la continuidad espacial de la lluvia en la región, permiten detectar patrones de agrupamiento, y destacan variaciones de la lluvia a escala local.

Las precipitaciones tuvieron una alta heterogeneidad en sus datos, presentando distribuciones que principalmente concentraron hacia los valores bajos. Por su parte el NDVI presentó distribuciones relativamente homogéneas con distribuciones que se centraron en valores altos.

Geográficamente se evidencia una correspondencia entre valores altos y bajos de ambas variables, tanto para los datos experimentales como para los resultados de la co-simulación: en la zona suroriental se localizaron los valores más altos, en la zona central valores medios, y en la zona noroccidental los más bajos.

Se estableció por medio de variogramas direccionales que las lluvias presentan un comportamiento totalmente anisotrópico en la zona de estudio, mientras el NDVI un comportamiento isotrópico a escala local, y geoméricamente anisotrópico a escala regional.

Se diseñó un flujo de procesamiento para productos de índices de vegetación MODIS (MODIS VI) que comprende etapas de extracción de parámetros de calidad y aplicación de filtros en base a ellos. Con él se obtienen archivos raster de NDVI listos para analizar, y aprovechando la abundancia de información disponible de datos de NDVI, el procedimiento podría aplicarse a otras variables que tengan relación con este indicador.

Se propuso un método que involucra distintas técnicas del ámbito de la geomática, para generar mapas a partir de una variable escasamente muestreada, y su relación con otra muestreada densamente.

Anexos

A. Procesamiento Detallado de Productos MODIS NDVI

En el capítulo de metodología se mencionó que las imágenes MODIS fueron procesadas en distintas etapas para lograr la representación de la variable NDVI. En el presente anexo se detallan las metodologías y recursos utilizados en cada una de las fases. Cabe recordar que el flujo general se basa en el diagrama que ilustra la Figura 55 expuesta también en el capítulo metodológico.

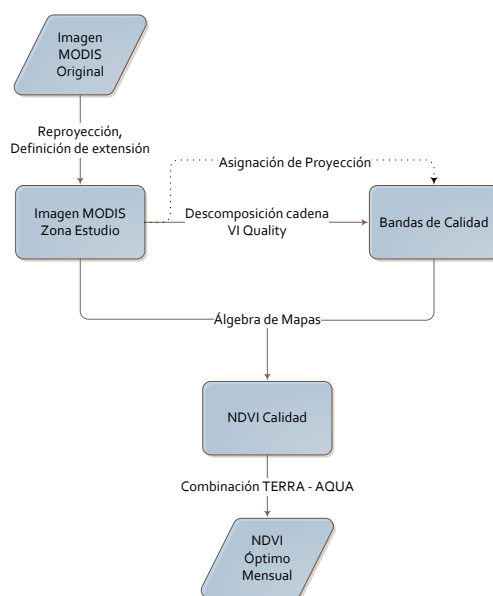


Figura 55: Flujo de trabajo para procesamiento de imágenes MODIS

La evaluación de calidad es un análisis totalmente basado en la información a nivel del píxel. En ella interesa el dataset Pixel Reliability que proporciona por medio de un indicador generalizado, información para interpretación directa acerca de las condiciones de calidad donde las características del píxel dependen del valor que se tenga dentro del rango indicado en la Tabla 8. También se considera el dataset VI Quality

que suministra un valor decimal para convertir al formato binario de 16 bits y conformar una cadena de derecha a izquierda construida por los nueve parámetros listados en la Tabla 9, lo que demanda procesamiento, interpretación detallada y permite realizar una evaluación específica de algún parámetro de interés.

Tabla 8: Rango de valores del dataset Pixel Reliability

Pixel Reliability Rank	Summary QA	Description
-1	Fill/No data	Not processed
0	Good data	Use with confidence
1	Marginal data	Useful, but look at other QA information
2	Snow/Ice	Target covered with snow/ice
3	Cloudy	Target not visible, covered with cloud

A.1. Pre-procesamiento

La fase inicial de procesamiento de las imágenes MODIS consistió en generar para cada imagen un nuevo archivo raster acorde a las necesidades del proyecto. Con el software MODIS Reprojection Tools (MRT) en un procedimiento conjunto, las imágenes fueron re proyectadas al sistema de coordenadas geográficas WGS84 y su extensión fue reducida para que coincidiera con la de la zona de estudio, las 11 bandas originales de cada imagen fueron preservadas al igual que su formato. Con los archivos raster generados en esta etapa, se lograron productos que permiten un uso más conveniente de los datos para su aprovechamiento en SIG, y una reducción en espacio de almacenamiento y tiempo de procesamiento.

El procedimiento se lleva a cabo configurando en la interfaz de MRT, los distintos valores de extensión, proyección, método de remuestreo, y la indicación de los archivos raster de entrada (imagen original) y salida con su respectivo formato. Para lograr una optimización en la metodología y obtener la salida raster esperada de cada imagen, se generaron archivos de parámetros usando la estructura detallada en la Figura 56 y progresivamente se fueron cargando a la aplicación MRT como lo muestra la Figura 57.

A.2. Filtrado de Calidad

La etapa de filtro de calidad consistió en definir los parámetros de calidad aceptables que debían cumplir los píxeles para que proporcionen valores óptimos del NDVI. La evaluación se basa en analizar la información de la banda de confiabilidad del píxel (Pixel Reliability), y la banda de calidad del índice de vegetación (VI Quality), es decir los datasets científicos que proporcionan información de calidad. Dependiendo de la combinación de estos valores junto con los del NDVI se eligen los píxeles de calidad óptima que representen dicha variable.

A.2.1. Descomposición Bandas de Calidad

Dado que la banda “VI Quality” dispone la información en formato binario como una cadena de caracteres contruida con 9 parámetros de calidad, se hace necesario descomponer esta cadena extrayendo los bits para que representen cada parámetro de interés en una banda individual. Así, los datos son llevados a un formato raster de codificación convencional. El procedimiento se realizó con rutinas de código ejecutadas en la herramienta LDOPE (Land Data Operational Products Evaluation) donde se tuvieron en cuenta para cada una de las imágenes reproyectadas y redefinidas en el paso anterior, los parámetros indicados en la Tabla 10.

Tabla 9: Parámetros de la banda VI Quality

Bit No	Parameter	Bit Word	Interpretation
0-1	VI Quality (Modland QA)	00	VI produced with good quality
		01	VI produced, but check other QA
		10	Pixel produced, but most probably cloudy
		11	Pixel not produced due to other reasons than clouds
2-5	VI Usefulness	0000	Highest quality
		0001	Lower quality
		0010	Decreasing quality
		0100	Decreasing quality
		1000	Decreasing quality
		1001	Decreasing quality
		1010	Lowest quality
		1100	Quality so low that it is not useful
1101	L1B data faulty		
1110	Not useful for any other reason/not processed		
6-7	Aerosol quantity	00	Climatology
		01	Low
		10	Intermediate
		11	High
8	Adjacent cloud detected	0	No
		1	Yes
9	Atmospheric BRDF Correction	0	No
		1	Yes
11-13	Land - Water Mask	000	Shallow ocean
		001	Land (Nothing else but land)
		010	Ocean coastlines and lake shorelines
		011	Shallow inland water
		100	Ephemeral water
		101	Ocean coastlines and lake shorelines
		110	Moderate or continental ocean
111	Deep ocean		
14	Possible snow/ice	0	No
		1	Yes
15	Possible shadow	0	No
		1	Yes

Tabla 10: Parámetros de calidad a extraer de la banda VI Quality

Parámetro	Dataset	Bits
Calidad del Índice de Vegetación	MODLAND QA Bits	0-1
Usabilidad del Índice de Vegetación	VI Usefulness	2-5
Posibilidad de Sombra	Possible Shadow	15
Máscara de Agua/Terreno	Land/Water Mask	11-13

Cada raster resultante de la extracción de bits de calidad en la herramienta LDOPE se genera sin sistema de referencia, sin embargo, en la

herramienta es posible asignar a una imagen el mismo sistema de referencia que posea otra. En este caso, una vez que se tuvieron descompuestos como bandas cada uno de los parámetros de calidad a analizar, se procedió a asignarles el sistema de referencia de la imagen base para la descomposición.

Con la extensión ajustada a la zona de estudio, las bandas de calidad extraídas, y la definición del sistema de proyección para todas las imágenes generadas, se obtuvo el esquema de productos necesario para ejecutar un procedimiento de álgebra de mapas que prepare en cada mes la variable NDVI.

A.2.2. Álgebra de Mapas

El Álgebra de Mapas es una técnica de procesamiento de archivos raster comúnmente utilizada en SIG, con la que se genera una capa de información raster mediante operaciones aplicadas a los valores de pixel de otras capas preexistentes. El cálculo tuvo como objetivo conservar los valores de pixel del NDVI siempre que las reglas establecidas para cada indicador de calidad se cumplieran en conjunto. Los valores esperados de la banda NDVI, Pixel Reliability, y VI Quality en los parámetros extraídos individualmente se establecieron de la siguiente forma:

- **NDVI \geq 1000:** Valores inferiores no corresponden con coberturas de vegetación, y presumen que se trata de agua (negativos) o zonas urbanas (ceranos a cero).
- **Pixel Reliability \leq 1:** Confiabilidad total (0) y aceptable bajo la revisión parámetros adicionales (1)
- **Modland QA \leq 1:** Índice de vegetación producido con buena calidad (0) y producido con calidad aceptable si así lo indica la evaluación de parámetros adicionales (1)
- **VI Usefulness \leq 10:** Valores de usabilidad: más alta (0), y escala decreciente (1-10).
- **Possible Shadow = 0:** Sin sombras
- **Land/Water Mask = 1:** Pixeles de terreno. Se omiten cuerpos de agua (\neq 1).

Con el propósito de automatizar la metodología, en el software ArcGIS mediante la herramienta Model Builder, se diseñó un modelo de geoprocresamiento para el cálculo del NDVI óptimo. En él, se establecieron como variables cada uno de los rásteres de entrada conformando un flujo de trabajo como el que se ilustra en la Figura 58. Los parámetros del geoprocreso se conforman por la imagen MODIS de 11 bandas redefinida para la zona de estudio y los cuatro indicadores de calidad extraídos de la banda VI Quality. De la imagen de 11 bandas se extraen la banda del índice de vegetación (banda 0) y la banda de confiabilidad del pixel (banda 11), y junto con las bandas de calidad se calcula el NDVI aceptable mediante la expresión de Python indicada en la Ecuación 24, su cálculo implica conservar en cada pixel del raster de salida el valor del NDVI si los valores de calidad se cumplen, o fijar su valor en cero en caso que no se cumplan. Posteriormente los valores iguales a cero se convierten a valores nulos y se obtiene un raster que contiene valores de pixel únicamente cuando existen valores del NDVI óptimo.

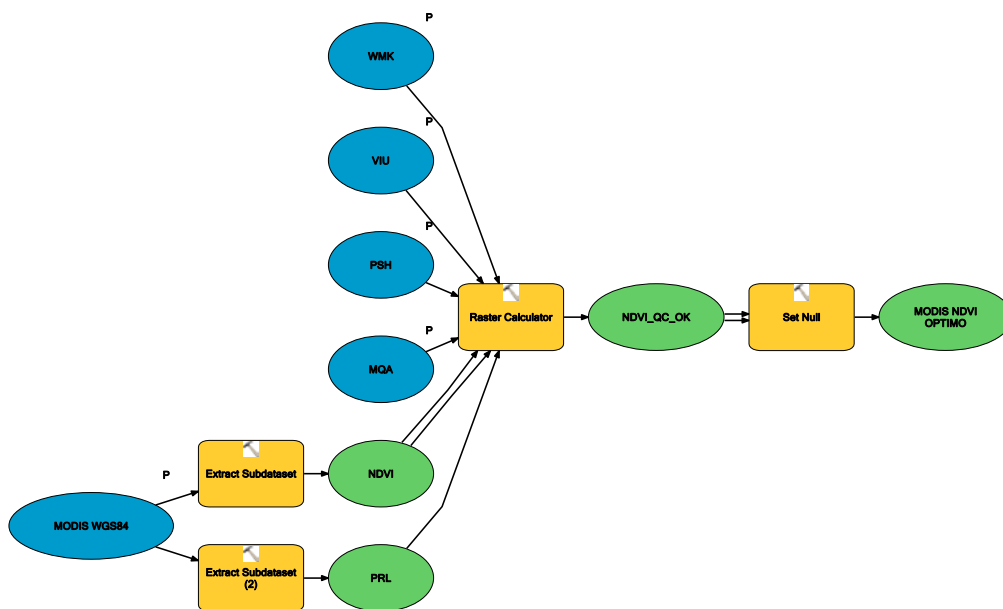


Figura 58: Modelo de geoprocresamiento para el filtrado del NDVI

$$(NDVI \geq 1000) * (PRL \leq 1) * (MQA \leq 1) * (VIU \leq 10) * (PSH == 0) * (WMK == 1) * NDVI \quad (24)$$

Para el cálculo de un mes, se indican los archivos asociados a la imagen TERRA y se ejecuta la herramienta, posteriormente el proceso se repite con los archivos asociados a la imagen AQUA del mismo mes. La interfaz en la que se indican cada uno de estos archivos se observa en la Figura 59 y corresponde con la utilización de un Toolbox personalizado en ArcGIS. Como resultado se obtiene un raster con el NDVI óptimo correspondiente al sensor TERRA, y uno correspondiente al sensor AQUA para cada uno de los meses estudiados.

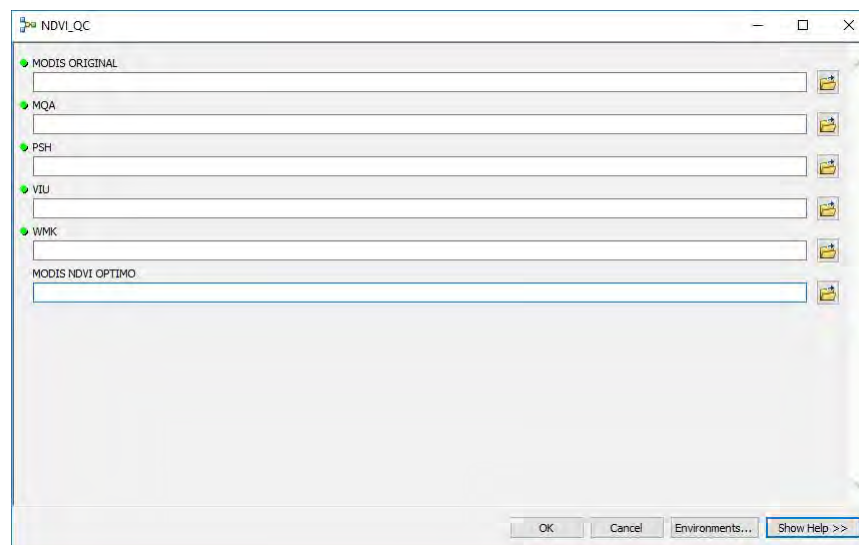


Figura 59: Toolbox del modelado cartográfico para el álgebra de mapas

Para consolidar ambos resultados en un raster único mensual, en ArcGIS se corrió un proceso de análisis de estadísticas de celda (Toolbox Cell Statistics) donde se obtuviera el valor máximo del NDVI, y teniendo presente que no todas las ubicaciones de píxeles necesariamente conservaron datos de ambos sensores, se tomó como valor definitivo el único valor disponible en esos casos. El consolidado de las imágenes filtradas proporciona la mejor cobertura posible de la zona de estudio para la variable NDVI en cada uno de los meses.

En la Figura 60 se muestra el ejemplo para un mes de los productos obtenidos en las distintas etapas del procesamiento. La gráfica muestra la escena TERRA MODIS original, la escena ajustada y re proyectada según especificaciones geográficas del proyecto, las bandas de calidad extraídas, y el resultado del álgebra de mapas del sensor TERRA que involucra los productos anteriormente mencionados. También se muestra el resultado del álgebra de mapas obtenido con los equivalentes productos AQUA (en este caso no se muestran los productos logrados en cada etapa). Por último se muestra la imagen de NDVI óptimo mensual que combina ambos sensores. Todas las imágenes están simbolizadas de modo que en cada mapa pueden identificarse los valores de interés, y en base a ellos interpretar como trabajó el filtro.

Finalmente, los datos mensuales fueron agrupados en una capa de puntos donde se incorporó cada mes como un campo de la tabla de atributos, y conociendo la posición de cada pixel se obtuvo la estructura: X, Y, y NDVI mensual, representando así la variable NDVI que constituirá la variable secundaria (soft data) en el cálculo de las co-simulaciones.

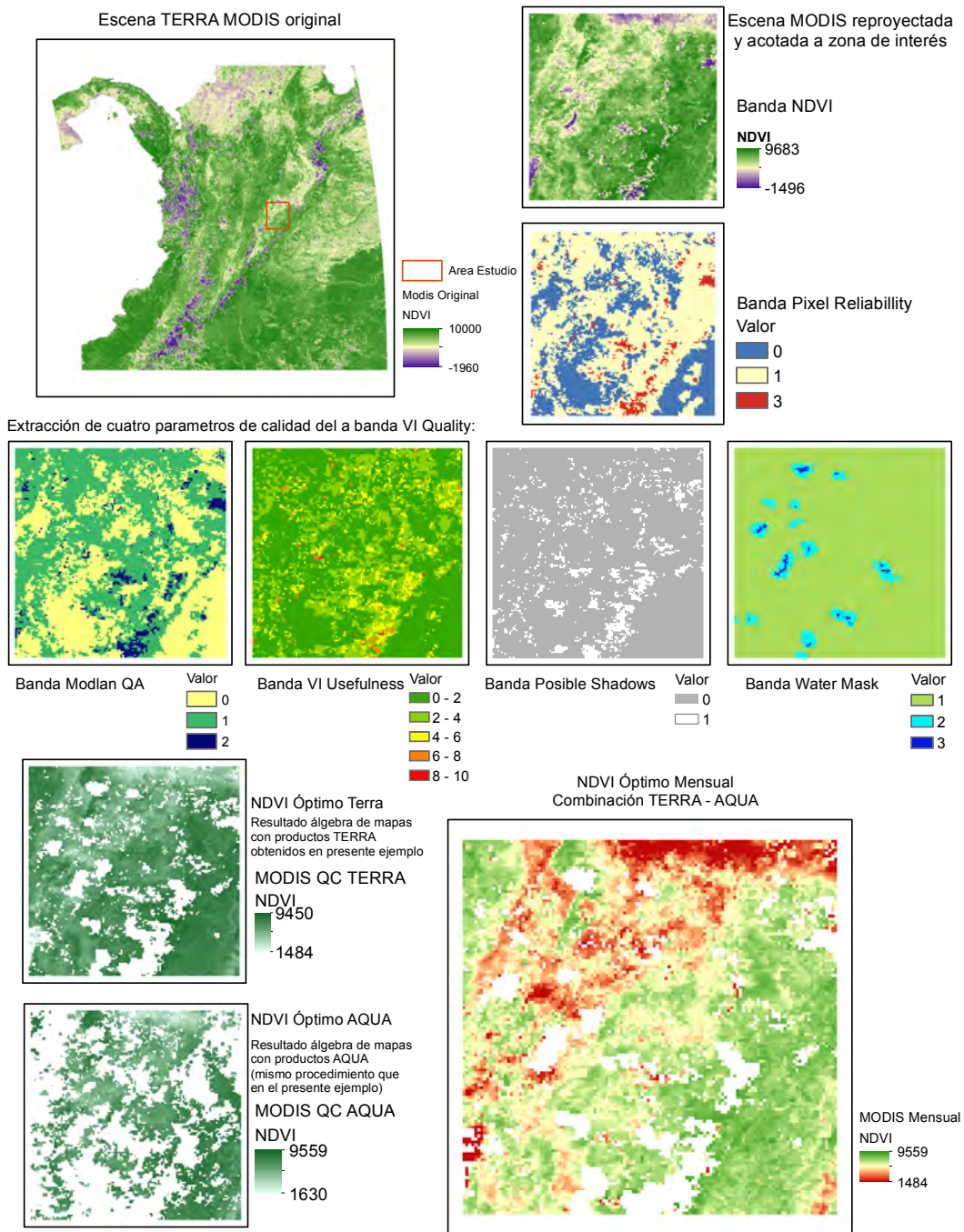


Figura 60: Resultado mapas NDVI

B. Medidas Estadísticas

A continuación, se definen brevemente y se indican las expresiones para determinar las medidas estadísticas analizadas durante la etapa exploratoria de la investigación. Los cálculos obtenidos corresponden con los valores consignados en las Tablas 6 y 7.

- **Mínimo:** valor menor del conjunto de datos, extremo más pequeño.

$$\min(x_0, x_1, \dots, x_n) \quad (25)$$

- **Máximo:** valor mayor del conjunto de datos, extremo más grande.

$$\max(x_0, x_1, \dots, x_n) \quad (26)$$

- **Promedio:** también denominado media aritmética o media. Representa el centro de gravedad o centro geométrico de la distribución. Aunque es la medida de tendencia central más común, es poco robusta, depende de los datos particulares y es altamente influenciado por valores extremos. Se obtiene sumando todos los valores de la variable y dividiendo entre el número de datos.

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n} \quad (27)$$

- **Varianza:** medida de dispersión que indica el promedio de las desviaciones respecto a la media. Las desviaciones (diferencia de cada dato con el promedio) son sumadas y elevadas al cuadrado para evitar que los valores positivos y negativos se compensen, el resultado se divide entre el número de datos. La medida no tiene las mismas unidades que el conjunto de datos original.

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n} \quad (28)$$

- **Desviación estándar:** también llamada desviación típica. Se obtiene calculando la raíz cuadrada de la varianza, lo que permite obtener una medida de dispersión en las mismas unidades que el conjunto de datos original. Evalúa la dispersión de los datos respecto a la media.

$$S = \sqrt{S^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \quad (29)$$

- **Desviación mediana:** medida de dispersión expresada por la media aritmética de las diferencias absolutas entre los valores de la variable y la mediana.

$$D_{Me} = \frac{\sum(x_i^n |x_i - Me|)}{n} \quad (30)$$

- **Coefficiente de asimetría:** medida que describe la forma geométrica de la distribución de los datos al compararla con el modelo probabilístico normal. Mide la cantidad de sesgo a la izquierda o derecha de la media en la distribución.

$$CA = \frac{n}{(n-1) * (n-2)} * \sum \left[\frac{x_i - \bar{x}}{S} \right]^3 \quad (31)$$

- **Curtois:** medida que indica el grado de anchura o estrechez de la curva de frecuencias respecto a la curva normal. Un valor positivo indica que la distribución es relativamente elevada mientras uno negativo que la distribución es relativamente plana.

$$K = \frac{M_4}{S_4} \quad (32)$$

- **Coefficiente de variación:** Medida en porcentaje, que determina el grado de dispersión de un conjunto de datos relativo a su media. Es de gran utilidad para comparar variabilidad entre poblaciones que tiene unidades de medida diferente. Se obtiene con la relación entre una medida dispersión y una de tendencia central como la desviación estándar y el promedio, o la mediana y el promedio.

$$CV_M = \frac{S}{\bar{x}} \quad (33)$$

$$CV_{Me} = \frac{Me}{\bar{x}} \quad (34)$$

Referencias

- [Aparicio, 1992] FRANCISCO JAVIER APARICIO., Fundamentos de hidrología de superficie, Grupo Noriega editores, Mexico 1992.
- [Barona *et al.*, 2006] ELIZABETH BARONA, ERNESTO GIRON, KELLY L. FEISTNER, JOHN L. DWYER, GLENN HYMAN., Método de procesamiento de imágenes MODIS para Colombia, Centro Internacional de Agricultura Tropical (CIAT), 2006.
- [Chereque 1989] WENDOR CHEREQUE MORAN., Hidrología para estudiantes de ingeniería civil, Pontificia Universidad Católica del Perú (CIAT), Lima 1989.
- [De La Casa & Ovando 2006] DE LA CASA, A., & OVANDO, G. Relación entre la precipitación e índices de vegetación durante el comienzo del ciclo anual de lluvias en la provincia de córdoba, Argentina. RIA. Revista de Investigaciones Agropecuarias, 35 (1), 67-85. 2006
- [Delfiner & Chilès, 1999] PIERRE DELFINER, JEAN-PAUL CHILÈS., Geostatistics. Modeling Spatial Uncertainty, John Willey & Sons, INC., New York, 1999.
- [Di *et al.*, 1994] LIPING DI, D. C. RUNDQUIST & LUOHENG HAN Modelling relationships between NDVI and precipitation during vegetative growth cycles., International Journal of Remote Sensing, 15:10, 2121-2136., 2007. Disponible en: <https://doi.org/10.1080/01431169408954231>
- [Díaz Viera, 2002] MARTÍN A DÍAZ VIERA., Geoestadística Aplicada, Instituto de Geofísica UNAM, México, 2002
- [Didan *et al.*, 2015] KAMEL DIDAN, ARMANDO BARRETO MUNOZ, RAMON SOLANO, ALFREDO HUETE., MODIS Vegetation Index User's Guide (MOD13 SERIES), The University of Arizona, 2015.
- [Giraldo, 2002] RAMÓN GIRALDO., Introducción a la geoestadística: Teoría y aplicación, Universidad Nacional de Colombia, Bogotá, 2002.
- [Gómez, 2003] LETICIA GÓMEZ MENDOZA., La Relación entre el Índice Normalizado de Vegetación y la Variabilidad del Clima en Oaxaca: Una

-
- Herramienta para el Manejo de Ecosistemas. Centro de Ciencias de la Atmosfera, UNAM, Mexico., 2003.
- [Isaaks & Srivastava, 1989] EDWARD H. ISAAKS, R. MOHAN SRIVASTAVA., An Introduction to Applied Geostatistics, Oxford University Press, New York, 1989.
- [Muñoz, 2013] PEDRO MUÑOZ AGUAYO., Apuntes de Teledetección: Índices de vegetación, Ministerio de Agricultura - Centro de Información de Recursos Naturales, Chile, 2013. Disponible en: <http://bibliotecadigital.ciren.cl/bitstream/handle/123456789/26389/Tema%20Indices%20de%20vegetaci%C3%B3n%2C%20Pedro%20Mu%C3%B1oz%20A.pdf?sequence=1&isAllowed=y>
- [O'Sullivan & Unwin, 2003] DAVID O'SULLIVAN, DAVID UNWIN., Geographic Information Analysis, John Wiley & Sons, New Jersey, 2003.
- [Pircz & Deutsch, 2014] MICHAEL J. PIRCZ, CLAYTON V. DEUTSCH., Geostatistical Reservoir Modeling, Oxford University Press, New York, 2014.
- [Rojas *et al*, 2011] O. ROJAS, F. REMBOLD, J. DELINCÉ & O. LÉO., Using the NDVI as auxiliary data for rapid quality assessment of rainfall estimates in Africa, International Journal of Remote Sensing, 32:12, 3249-3265. 2011 Disponible en: <http://dx.doi.org/10.1080/01431161003698260>
- [Rossi & Deutsch, 2014] MARIO E ROSSI, CLAYTON V. DEUTSCH., Mineral Resource Estimation, Springer, New York, 2014.
- [Wang *et al*, 2014] J. WANG, P. M. RICH & K. P. PRICE., Temporal responses of NDVI to precipitation and temperature in the central Great Plains, USA, International Journal of Remote Sensing, 24:11, 2345-2364., 2003. Disponible en: <https://www.tandfonline.com/doi/abs/10.1080/01431160210154812>
- [Xue & Su, 2017] JINRU XUE, BAOFENG SU., Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications, Journal of Sensors, vol. 2017, Article ID 1353691, 17 pages, 2017. Disponible en: <https://doi.org/10.1155/2017/1353691>

- [Yarleque *et al*, 2007] CHRISTIAN YARLEQUE, ADOLFO POSADAS & ROBERTO QUIROZ,. Reconstrucción de datos de precipitación pluvial en series de tiempo mediante transformadas de wavelet con dos niveles de descomposición., Centro Internacional de la Papa, Lima, 2007.
- [Zhang, 2011] YE ZHANG,. Introduction to Geostatistics Course Notes., Dept. of Geology & Geophysics. University of Wyoming., 2011.