

Análisis Geoestadístico de la Temperatura de la Superficie Terrestre de Croacia, 13 de Marzo de 2008



Daniela Marín López - 20142025027

Ana María Rojas Cantor - 20142025036

Carol Tatiana Chicuazuque Gutiérrez - 20142025105

**Análisis Geoestadístico de la Temperatura
de la Superficie Terrestre de Croacia,
13 de Marzo de 2008**

Primera Entrega

Presentado Por.

Daniela Marín - 20142025027

Ana María Rojas Cantor - 20142025036

Carol Tatiana Chicuazuque Gutiérrez - 20142025105

Presentado A.

Carlos Eduardo Melo Martínez

**Universidad Distrital Francisco José De Caldas
Facultad de Ingeniería
Ingeniería Catastral y Geodesia
Geoestadística
Bogotá, D. C.
02 de Mayo de 2018**

Tabla de Contenido

Lista de Ilustraciones	3
Lista de Tablas	4
1. Introducción	5
2. Justificación	5
3. Objetivos	5
3.1. Objetivo general	5
3.2. Objetivo específico	5
4. Estado del arte	6
5. Marco teórico	17
5.1. Funciones de correlación espacial	17
5.2. Anamorfosis gaussiana	21
5.3. Anisotropía	22
6. Descripción de la base de datos	23
7. Análisis Exploratorio	24
Espacialización de los datos	24
Estadísticos de tendencia central	25
7.1 Análisis de normalidad	26
7.2 Análisis de tendencia	27
7.3 Análisis de anisotropía a los residuos	28
8. Modelamiento del semivariograma	29
9. Referencias	32

Lista de Ilustraciones

<i>Ilustración 1.</i> Departamento de Boyacá - Estaciones meteorológicas.	6
<i>Ilustración 2.</i> Distribución de Frecuencia de las Temperaturas de 1997 y 2001.	7
<i>Ilustración 3.</i> Distribución de Frecuencia Transformada de la Temperatura de 2001.	7
<i>Ilustración 4.</i> Análisis de variación espacial (semivariograma) aplicado a la Temperaturas de 1997 y 2001.	7
<i>Ilustración 5.</i> Distribución de frecuencia de Precipitaciones de 1997 y 2001	7
<i>Ilustración 6.</i> Distribución frecuencia T-97.	8
<i>Ilustración 7.</i> Semivariograma T-97.	8
<i>Ilustración 8.</i> Predicción Kriging Disyuntivo T-97	8
<i>Ilustración 9.</i> Error Kriging Disyuntivo T-97.	8
<i>Ilustración 10.</i> Semivariograma T-2001.	8
<i>Ilustración 11.</i> Tipo de Sector T-2001	8
<i>Ilustración 12.</i> Semivariograma P-97 y P-2001.	9
<i>Ilustración 13.</i> Variación espacial T-97 y T-2001.	9
<i>Ilustración 14.</i> Variación espacial P-97 y P-2001(Kriging Ordinario).	9
<i>Ilustración 15.</i> Mapa de 779 estaciones base meteorológicas con observaciones diarias de NOAA NCDC en California (1980-2003)	10
<i>Ilustración 16.</i> Variables climáticas para Mar de PRISM: media anual de 30 años (1961-90) de (a) precipitación total, (b) temperaturas máximas y (c) temperaturas mínimas.	11
<i>Ilustración 17.</i> Mapas de predicción utilizando el método de interpolación única para interpolar datos diarios de estaciones base .	11
<i>Ilustración 18.</i> Semivariogramas experimentales exponenciales (línea continua) y empíricos (puntos) para los valores diarios de precipitación (izquierda), temperatura máxima (medio) y temperatura mínima (derecha) el 26 de marzo de 2001	12
<i>Ilustración 19.</i> Mapas de predicción utilizando el método de impresión climática para integrar PRISM con la interpolación diaria de la estación base.	12
<i>Ilustración 20.</i> Mapa de localización para la comparación a gran escala de métodos de interpolación solo e impresión climática con estaciones base.	13
<i>Ilustración 21.</i> Comparación del ajuste del método de Kriging variando el tamaño de celda y el semivariograma	15
<i>Ilustración 22.</i> Variación del ajuste de acuerdo con el método de interpolación espacial utilizado	16
<i>Ilustración 23.</i> Campos de precipitación obtenidos (para el evento registrado en 13-05-1996).	16
<i>Ilustración 24.</i> a) Modelo Esférico b) Modelo Exponencial c) Modelo Gaussiano	20
<i>Ilustración 25.</i> Construcción de la gráfica de la anamorfosis gaussiana	21
<i>Ilustración 26.</i> Semivariogramas experimentales en dos direcciones fundamentales.	22
<i>Ilustración 27.</i> Ejemplo bidimensional de anisotropía geométrica.	22
<i>Ilustración 28.</i> Ejemplo tridimensional de anisotropía zonal.	23
<i>Ilustración 29.</i> Distribución de las estaciones.	23
<i>Ilustración 30.</i> Distribución espacial 3D de la temperatura el día 13 de marzo de 2008.	24

<i>Ilustración 31.</i> Distribución espacial 2D de la temperatura el día 13 de marzo de 2008.	25
<i>Ilustración 32.</i> Estadísticos de tendencia Central	25
<i>Ilustración 33.</i> Gráficos para la detección de la normalidad en la variable Temperatura de la Superficie Terrestre	26
<i>Ilustración 34.</i> Tests de Normalidad aplicados a la variable Temperatura de la Superficie Terrestre	26
<i>Ilustración 35.</i> Gráficos para la detección de la normalidad en la variable Temperatura de la Superficie Terrestre Transformada.	26
<i>Ilustración 36.</i> Tests de Normalidad aplicados a la variable Temperatura de la Superficie Terrestre Transformada	27
<i>Ilustración 37.</i> Primera Regresión Lineal realizada. Salida software R	27
<i>Ilustración 38.</i> Regresión Lineal seleccionada. Salida software R.	28
<i>Ilustración 39.</i> Tests de Normalidad aplicados a los residuos del modelo seleccionado.	28
<i>Ilustración 40.</i> Evaluación de la Anisotropía en los residuos Salida Software R.	28
<i>Ilustración 41.</i> Cálculo de la Distancia a utilizar para el modelado del semivariograma	29
<i>Ilustración 42.</i> Semivariograma experimental robusto.	29
<i>Ilustración 43.</i> Semivariograma experimental mediana.	29
<i>Ilustración 44.</i> Semivariograma experimental clásico.	29
<i>Ilustración 45.</i> Semivariograma experimental media recortada.	29
<i>Ilustración 46.</i> Modelo LST experimental robusto.	30
<i>Ilustración 47.</i> Modelo LST experimental mediana.	30
<i>Ilustración 48.</i> Modelo LST experimental clásico.	30
<i>Ilustración 49.</i> Modelo LST experimental media recortada.	30
<i>Ilustración 50.</i> Valores del AIC para los semivariogramas teóricos	30
<i>Ilustración 51.</i> Semivariogramas Teóricos Exponenciales ajustados por OLS,WLS, ML, RML (izquierda). Valores AIC para semivariogramas Teóricos Exponenciales (derecha).	31
<i>Ilustración 52.</i> Semivariograma Teórico Exponencial estimado mediante RML	31

Lista de Tablas

<i>Tabla 1.</i> Comparación de rendimiento de los modelos IO (Interpolación) y CI (Impresión Climática) para precipitación temperaturas máximas y mínimas.	12
<i>Tabla 2.</i> Parámetros del semivariograma para precipitación diaria	14
<i>Tabla 3.</i> Comparación de metodologías de interpolación.	15
<i>Tabla 4.</i> Correlación del método de Kriging empleando diferentes semivariogramas.	16
<i>Tabla 5.</i> Descripción base de datos	24

1. Introducción

El estudio de fenómenos con correlación espacial, por medio de métodos geoestadísticos, surgió a partir de los años sesenta, especialmente con el propósito de predecir valores de las variables en sitios no muestreados. La geoestadística es solo una de las áreas del análisis de datos espaciales, por lo que es importante reconocer cuando la información georreferenciada es susceptible de ser analizada por medio de esta metodología (Henao 2015)[5]. Por otra parte, en el estudio de información georreferenciada, la primera etapa que se debe cumplir es la del análisis exploratorio de datos (AED). Este busca identificar localización, variabilidad, forma, observaciones extremas y correlación ya que permiten establecer si algunos supuestos necesarios para la aplicación de la teoría geoestadística son válidos o para definir qué procedimiento de predicción es el más conveniente (Henao 2015)[5].

Luego, es necesario determinar la dependencia espacial entre los datos medidos de una variable. Por ello se hace necesario el ajuste de modelos que generalicen lo observado en el semivariograma experimental a cualquier distancia. El autor cita a Samper y Carrera (1990)[10], ya que presentan una discusión respecto a las características y condiciones que éstos deben cumplir. En general dichos modelos pueden dividirse en no acotados y acotados (Warrick et al., 1986[15]; citado por Henao 2015)[5]. Los acotados garantizan que la covarianza de los incrementos es finita, por lo cual son ampliamente usados cuando hay evidencia de que presentan buen ajuste.

A continuación, se presentan los registros de la temperatura media terrestre de Croacia referente al día 13 de marzo de 2008, los cuales fueron sometidos a un tratamiento estadístico mediante pruebas de normalidad, tendencia y anisotropía, que fueron evaluadas previamente para finalmente realizar el ajuste del variograma. El procedimiento se llevó a cabo en el software R.

2. Justificación

El cambio climático se ha convertido en una gran preocupación que afecta a todo el globo debido a las consecuencias que traerá a generaciones futuras. Este, implica el incremento de condiciones extremas que se traducen en sequías, olas de calor, inundaciones, entre otros (Stott et al.[11] & Jones et al., 2004 [6]; citado por Arango, 2012). Para nuestro caso, la implementación del tratamiento estadístico (análisis exploratorio) y técnicas geoestadísticas, permiten estimar el comportamiento de la Temperatura a partir de 158 registros de estaciones meteorológicas localizadas a lo largo de Croacia.

3. Objetivos

3.1. Objetivo general

- ❖ Realizar el ajuste del variograma a través del modelamiento de la variación espacial para los registros de la temperatura media terrestre diaria de Croacia.

3.2. Objetivo específico

- ❖ Elaborar un análisis exploratorio mediante pruebas (normalidad, tendencia y anisotropía), observando y valorando el comportamiento de los registros para la respectiva elección del modelo.

4. Estado del arte

Los cambios ocasionados en el patrón climático, afectan no solo el comportamiento meteorológico sino también la forma de vida de todo ser que habita. Sin embargo, la temperatura y la humedad son las principales variables que afectan directamente los procesos físicos y biológicos. Para ello, se emplean pruebas en lo que respecta al análisis exploratorio y modelamiento de semivariogramas. A continuación, se presenta un resumen de algunos artículos publicados que hablan al respecto.

“Análisis Tendencial de la variación climática, temperatura y precipitación espacio-temporal del departamento de Boyacá (Colombia).”

La aparición de las temperaturas mensuales extremas se han convertido en uno de los factores más preocupantes para el estudio del cambio climático vinculado con el calentamiento global de nuestro tiempo que no es más que el aumento progresivo de la temperatura medida del aire el cual puede alterar el patrón climático global haciendo que el ciclo hidrológico se altere. La zona de estudio se localiza en el departamento de Boyacá, representada en 23.189 Km² respecto al país, cuenta con 123 Municipios y 10 Corregimientos distribuidos en 12 provincias.

Para el estudio se tomaron 29 estaciones meteorológicas distribuidas a lo largo del departamento y que además son monitoreadas por el IDEAM. En la *Ilustración 1*, se observa la cartografía utilizada que corresponde a escala 1:1.000.000 y las 29 estaciones georeferenciadas.



Ilustración 1. Departamento de Boyacá - Estaciones meteorológicas. Fuente: Socha (2014)[13].

Análisis Exploratorio

A continuación se realizó un análisis estadístico descriptivo para la variable Temperatura del año 1997 y 2001; en la *Ilustración 2*, se puede observar que los datos para el año 1997 presentaron un comportamiento asimétrico de manera natural, sesgado hacia la izquierda con cola positiva, con un valor medio de 16.247 y una mediana de 15.833, encontrándose una mínima diferencia entre estas mediciones, lo cual es aceptable para trabajar con los datos sin transformación ya que aunque no presentan una distribución simétrica no están lejos de serlo, por otra parte, se observa que los datos, presentan un comportamiento asimétrico de cola positiva hacia la izquierda, con un valor medio de 16.11 y una mediana de 16.016, encontrándose una diferencia del 0.094 entre estas mediciones de tendencia central. De igual manera se detectó que existe un periodo muy amplio en donde no existe información. Luego se aplicó una transformación Box-Cox con un parámetro de 0.8, como resultado se obtuvo nuevas medidas de centralización obtenidas (Media= 10.235; Mediana= 10.246),

Ilustración 3. Por otra parte, con aplicación de dicha transformación se logró que el 97% de los datos se ajustarán, es decir, se los datos se comportan de manera casi simétrica.

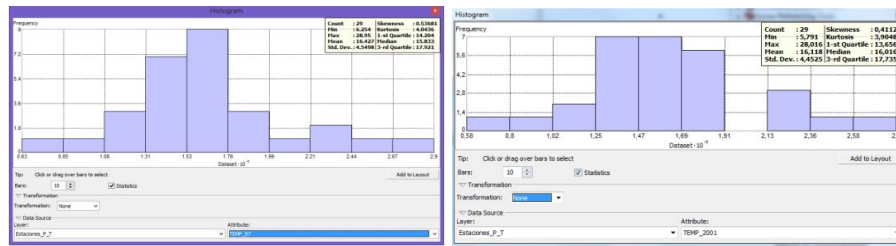


Ilustración 2. Distribución de Frecuencia de las Temperaturas de 1997 y 2001. Fuente: Socha (2014)[13].

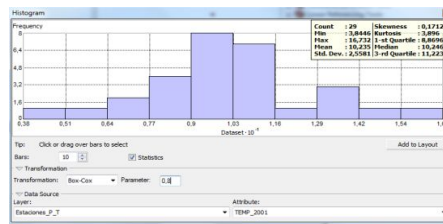


Ilustración 3. Distribución de Frecuencia Transformada de la Temperatura de 2001. Fuente: Socha (2014)[13].

En las **Ilustración 4.** se observa que los datos presentan anisotropía debido a que no existe una distribución normal de los puntos en todas las direcciones, por lo que la concentración es noroeste – sureste.

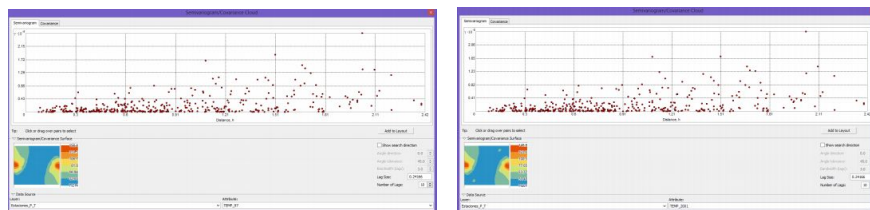


Ilustración 4. Análisis de variación espacial (semivariograma) aplicado a la Temperaturas de 1997 y 2001. Fuente: Socha (2014)[13].

Luego, se procede realizar un análisis a la variable precipitación tanto de 1997 como 2001. En la **Ilustración 5.** se observa que los datos presentan asimetría, sesgada hacia la izquierda con un valor medio de 113.86 y una mediana de 80.158, sin embargo, los datos se trabajaron sin ningún tipo de transformación ya que al aplicar las diferentes transformaciones no se obtenía ninguna distribución normal y más bien deformaba la muestra. Respecto al año 2001, se observa que los datos presentan un comportamiento asimétrico, sesgado hacia la izquierda con un valor medio de 122.23 y una mediana de 87.5, se realizó el mismo tratamiento que a los datos de 1997.

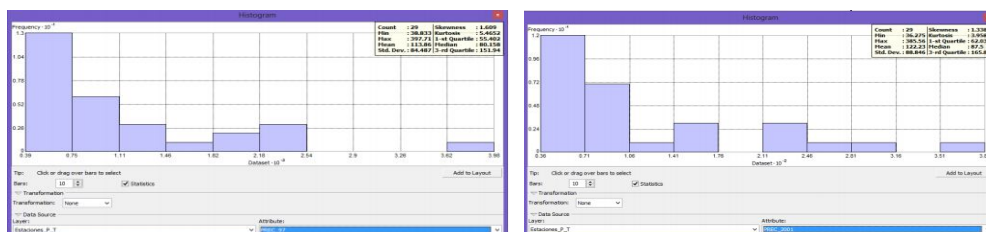


Ilustración 5. Distribución de frecuencia de Precipitaciones de 1997 y 2001. Fuente: Socha (2014)[13].

Método interpolación

Para la generación de los Mapas de tendencia de la variable de Temperatura y teniendo en cuenta que la muestra se clasificó como paramétrica se utilizó el Método de Kriging Disyuntivo ya que es el único método que maneja el software (ArcGis) para los datos con distribución paramétrica. Este método requiere de hacer fuertes premisas asumiendo que todos los pares de datos vienen en una distribución normal bivariada. Para la generación de los Mapas de tendencia de la variable de Precipitación y teniendo en cuenta que la muestra se clasificó como No Paramétrica se utilizó el Método de Kriging Ordinario ya que es un buen método para analizar datos no paramétrico. A continuación se presentan los resultados:

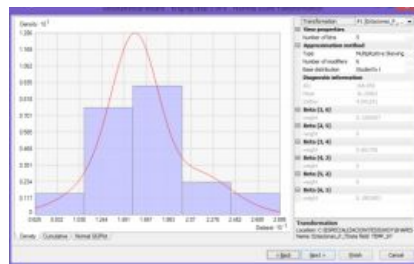


Ilustración 6. Distribución frecuencia T-97.
Fuente: Socha (2014)[13].

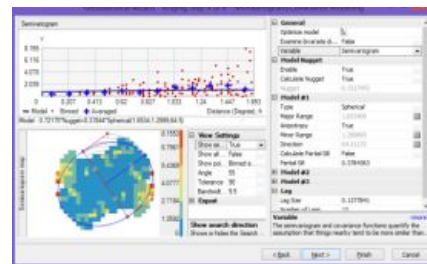


Ilustración 7. Semivariograma T-97.
Fuente: Socha (2014)[13].

En la **Ilustración 6**, se observa que los datos presentaron una distribución casi simétrica por lo cual se aplicó un Kriging Disyuntivo, para llevar a cabo la cuantificación de la estructura espacial se procede a calcular el semivariograma, **Ilustración 7**, del cual se determinó que es anisotrópico, es decir, las propiedades no son iguales en todas las direcciones. Para la variable de temperatura se utilizó el modelo esférico, con anisotropía verdadera y con una distribución de 4 sectores.

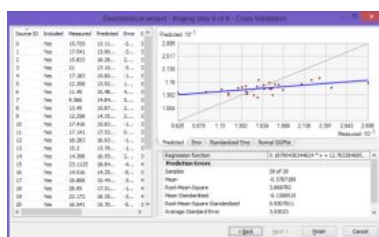


Ilustración 8. Predicción Kriging Disyuntivo T-97.
Fuente: Socha (2014)[13].

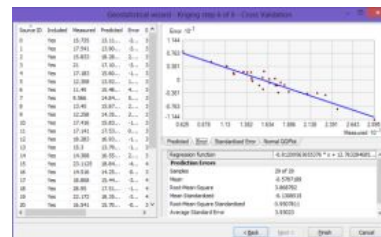


Ilustración 9. Error Kriging Disyuntivo T-97.
Fuente: Socha (2014)[13].

En la **Ilustración 8**, se observa la superficie interpolada a partir de valores interpolados y la **Ilustración 9**, los errores estándar de predicción producidos a partir de los valores interpolados.

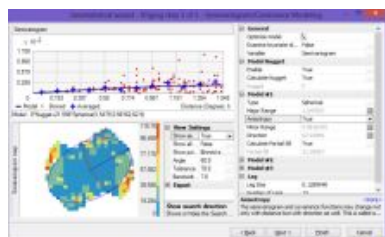


Ilustración 10. Semivariograma T-2001.
Fuente: Socha (2014)[13].

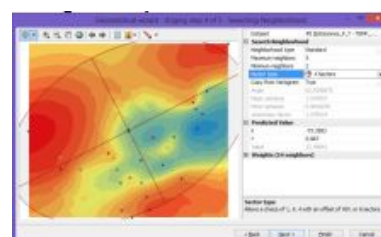


Ilustración 11. Tipo de Sector T-2001.
Fuente: Socha (2014)[13].

En la **Ilustración 10**, los datos presentaron una distribución asimétrica con sesgamiento a la izquierda por lo cual se le aplicó el método paramétrico de interpolación Kriging Disyuntivo con un método de transformación de datos Box-cox, con parámetro 0.8. En la **Ilustración 11**, se observa el cálculo del semivariograma con el cual se determinó que es anisotrópico, es decir, las propiedades no son iguales en todas las direcciones. Para el semivariograma se utilizó un modelo esférico con anisotropía verdadera y con una distribución de 4 sectores.

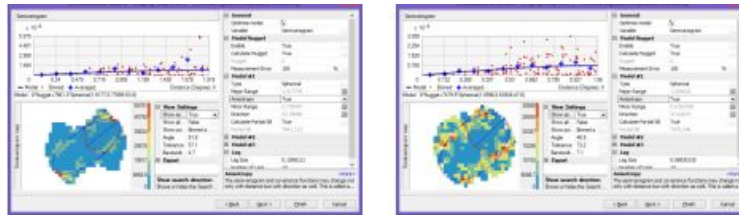


Ilustración 12. Semivariograma P-97 y P-2001. Fuente: Socha (2014)[13].

En la **Ilustración 12**, los datos de las variables de precipitación del año 1997 y 2001 presentaron un comportamiento asimétrico, se usó el método No Paramétrico de interpolación Kriging Ordinario sin ningún tipo de transformación para no alterar la distribución y el comportamiento de los datos. Para la variable de precipitación del año 97 y 2001, se realizó un semivariograma empírico sin aplicarle ningún tipo de transformación; para el cálculo del semivariograma de la P-97 y P-2001 se determinó que son anisotrópicas, es decir, que las propiedades no son iguales en todas las direcciones y para la variable de precipitación se utilizó un modelo esférico, con anisotropía verdadera y con una distribución de 4 sectores.

Resultado Mapas de Tendencia

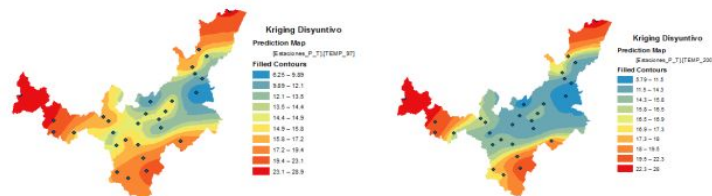


Ilustración 13. Variación espacial T-97 y T-2001 Fuente: Socha (2014)[13].

En la **Ilustración 13**, con el fin de obtener los mayores ajustes y resultados a la hora de realizar dicha interpolación para los datos de 1997, se puede analizar que los niveles más altos de temperatura se presentan en las zonas extremas del departamento y se condensan las temperaturas más bajas en el centro del Departamento evidenciando un crecimiento paulatino de la temperatura hacia sus exteriores. Por otra parte, para los datos de 2001, se puede analizar que los niveles de temperatura se comportan de igual manera que los datos de 1997, sin embargo, más del 50% corresponde a temperaturas bajas, es decir, reemplaza las temperaturas medias de 1997 con temperaturas bajas.

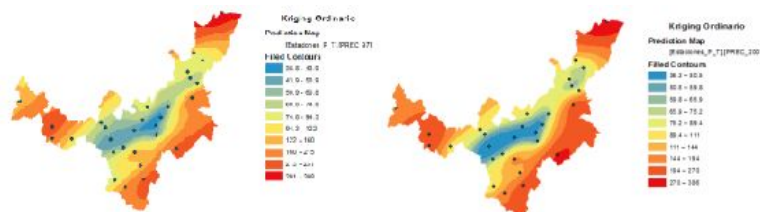


Ilustración 14. Variación espacial P-97 y P-2001(Kriging Ordinario) Fuente: Socha (2014)[13].

En la **Ilustración 14**, respecto a datos del año 1997 se puede analizar que los niveles más altos de precipitación se presentan en las zonas extremas del departamento y se condensan las precipitaciones más bajas en el centro del Departamento evidenciando un crecimiento paulatino de la variable hacia sus exteriores, sin embargo, para el año 2001 se comporta de manera similar que en el caso anterior, debido a que dicha comparación es casi imperceptible con la precipitación del año 1997 pues los cambios de precipitación son muy similares, aumentan o disminuyen pero no se evidencia algún cambio abrupto en el nivel de la variable.

El análisis de la correlación espacial nos confirma la relación e influencia entre las variables precipitación y temperatura. El resultado del análisis realizado a las 29 estaciones meteorológicas en el Departamento de Boyacá se evidencian de forma espacial mediante los mapas de tendencia los cuales nos permiten ver de forma clara el comportamiento de la variable tanto espacialmente como temporalmente ya que al hacer la comparación entre los dos años de la misma variable se pueden evidenciar claramente los cambios presentes o no en los dos periodos de tiempo analizados.

“Mapeo Climatológico asistido de la Precipitación y Temperatura Diaria.”

Este documento presenta un método para mapear las condiciones climáticas diarias que integra una red de observaciones de punto de estación base con mapas de clima de promedio a largo plazo. El método está diseñado para utilizar datos de estación base fácilmente accesibles y un conjunto de datos de clima espacial. El Centro Nacional de Datos Climáticos (NCDC) de la Administración Nacional Oceánica y Atmosférica (NOAA) proporciona datos meteorológicos diarios para una red de estaciones meteorológicas cooperativas y de primer orden en California. Estas estaciones meteorológicas fueron descargadas y mapeadas en un sistema de información geográfica (GIS) para California. **Ilustración 15**, la base de datos contiene 24 años de parámetros climáticos diarios (1980-2003) en 779 ubicaciones de puntos.

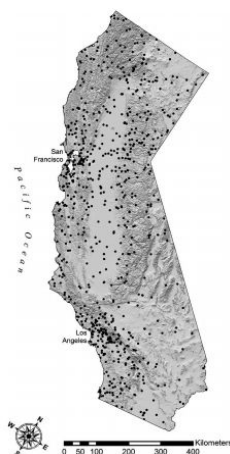


Ilustración 15. Mapa de 779 estaciones base meteorológicas con observaciones diarias de NOAA NCDC en California (1980-2003). Fuente: Richard et al. (2005)[9].

Para los propósitos de este estudio, los datos se consideraron generalmente confiables sin control de calidad adicional. Las redes y fuentes meteorológicas adicionales podrían aumentar la densidad de la estación, pero no se incluyeron para minimizar la complejidad en el proceso de construcción de la base de datos.

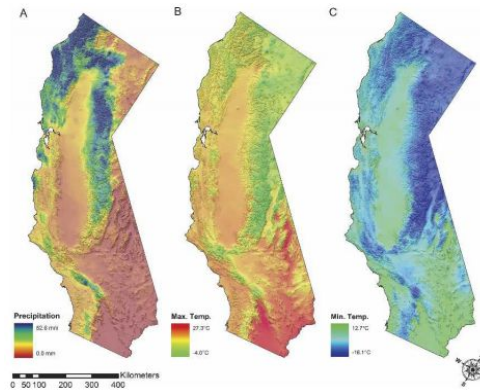


Ilustración 16. Variables climáticas para Mar de PRISM: media anual de 30 años (1961-90) de (a) precipitación total, (b) temperaturas máximas y (c) temperaturas mínimas. Fuente: Richard et al. (2005)[9].

PRISM utiliza 30 años de observaciones climáticas de estaciones base meteorológicas junto con datos digitales del terreno y otros factores ambientales para interpolar la variabilidad climática espacialmente en un paisaje **Ilustración 16**. El método PRISM supone que la elevación es uno de los factores más importantes que controlan los patrones de temperatura y humedad del paisaje, y utiliza la regresión lineal para estimar la variabilidad climática dentro de las orientaciones topográficas locales, o facetas. Se incorporan otros factores ambientales utilizando la ponderación de regresión diferencial de los puntos de datos de la estación base. El peso combinado de una estación es una función de elevación, proximidad costera, aspecto, relieve local y estratificación vertical de la masa de aire. PRISM captura la influencia de cuerpos de agua grandes, terrenos complejos e inversiones atmosféricas para determinar la temperatura y la humedad, incluidos los efectos de la lluvia. Estos factores son especialmente importantes en California, donde el clima varía considerablemente en distancias cortas.

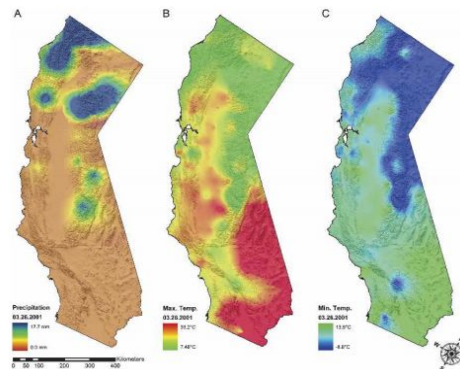


Ilustración 17. Mapas de predicción utilizando el método de interpolación única para interpolar datos diarios de estaciones base para (a) precipitación, (b) temperaturas máximas y (c) temperaturas mínimas: 26 de marzo de 2001. Fuente: Richard et al. (2005)[9].

En la **Ilustración 17**, se observan los mapas producidos espacialmente de precipitación y temperaturas máximas y mínimas, para ello se utilizó kriging ordinario permitiendo interpolar las observaciones puntuales de la red de estaciones base meteorológicas. Las estimaciones del Krigeado para una variable distribuida espacialmente en cualquier ubicación no monitoreada se calculan como un promedio ponderado de los valores conocidos de un conjunto circundante de puntos muestreados. Los pesos Kriging se derivan de un modelo estadístico de correlación espacial expresado como semivariogramas que caracterizan la dependencia espacial y la estructura en los datos como se observó en la **Ilustración 18**.

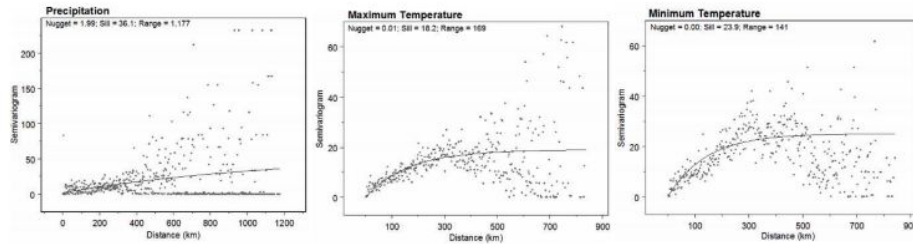


Ilustración 18. Semivariogramas experimentales exponenciales (línea continua) y empíricos (puntos) para los valores diarios de precipitación (izquierda), temperatura máxima (medio) y temperatura mínima (derecha) el 26 de marzo de 2001. Fuente: Richard et al. (2005)[9].

Kriging supone que las observaciones son realizaciones de un proceso estocástico estacionario y que las predicciones tienen errores estándar y probabilidades asociadas a ellas. Aunque el ajuste de la interpolación está limitado por el ajuste del semivariograma estimado a las semivarianzas reales de los datos del punto muestreado generalmente se sostiene que las predicciones espaciales son robustas con respecto a la especificación incorrecta del modelo del semivariograma (Cressie 1993 [2]; Goovaerts 1997 [4]; citado por Richard et al. 2005, p. 1504)[9].

Resultados

En la **Ilustración 19**, se observan los mapas predictivos finales a partir de la *Impresión Climática*, esto permite que los mapas PRISM proporcionen una “huella espacial” para representar gradientes ambientales en ubicaciones no muestreadas para las que no existen datos de estaciones base, por lo que se combinan con el resultado de un proceso de interpolación kriging para incorporar las relaciones ambientales a dichos mapas.

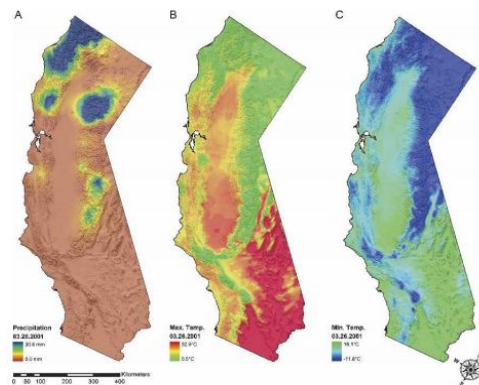


Ilustración 19. Mapas de predicción utilizando el método de impresión climática para integrar PRISM con la interpolación diaria de la estación base para (a) precipitación, (b) temperaturas máximas y (c) temperaturas mínimas: 26 de marzo de 2001. Fuente: Richard et al. (2005)[9].

Tabla 1. Comparación de rendimiento de los modelos IO (Interpolación) y CI (Impresión Climática) para precipitación, temperaturas máximas y mínimas. Fuente: Richard et al. (2005)[9].

	r^2, ρ		MAE		MER	
	IO	CI	IO	CI	IO	CI
Precipitación	0.59; <0.0001	0.64; <0.0001	2.8 mm	2.5 mm	-0.04	0.01
Max. Temperatura	0.85; <0.0001	0.92; <0.0001	2.7 °C	2.0 °C	-0.07	-0.38
Min. Temperatura	0.79; <0.0001	0.89; <0.0001	2.2 °C	1.7 °C	-0.09	-0.41

a. *Precipitación*

Interpolación: El análisis de regresión mostró que el enfoque de interpolación sólo predijo el 59% de la variabilidad en los valores diarios observados de precipitación con una ligera tendencia a sobrestimar. Según la **Tabla 1**. Los cálculos de MAE (error medio absoluta) indicaron que las predicciones de precipitación diaria estuvieron, en promedio, dentro de los $\pm 2,8$ mm de observado. El número de estaciones base de entrada disponibles varió de día a día debido a datos incompletos, que van desde 166 a 234. La prueba del coeficiente de correlación de Spearman mostró una tendencia débil de MAE reducido con más estaciones.

Impresión climática: El análisis de regresión mostró que el modelo de precipitación con impronta climática predijo el 64% de la variabilidad en los valores diarios observados con una ligera tendencia a sobrestimar. Los cálculos MAE indicaron que el modelo predice la precipitación diaria, en promedio, dentro de ± 2.5 mm de observado.

b. *Temperatura*

Interpolación: Los modelos de solo interpolación de temperatura máxima y mínima predijeron el 85% y el 79% de la variación en los valores diarios observados, respectivamente. En promedio, la temperatura máxima y mínima están ligeramente sobrestimadas. Los cálculos de MAE mostraron que las predicciones de temperatura máxima de los modelos de interpolación solamente fueron, en promedio, 2.7°C de las temperaturas mínimas observadas, y predichas fueron 2.2°C de observadas.

Respecto a lo modelos de impresión climática **Tabla 1**. para temperaturas máximas y mínimas predijeron 92% y 89% de la variación en los valores diarios observados, respectivamente. El MAE mostró que las predicciones de temperatura máximas estaban, en promedio, dentro de 2.0°C de las temperaturas mínimas observadas, y predichas estaban dentro de 1.7°C de lo observado. Las distribuciones de error absoluto para las temperaturas máximas y mínimas fueron similares.

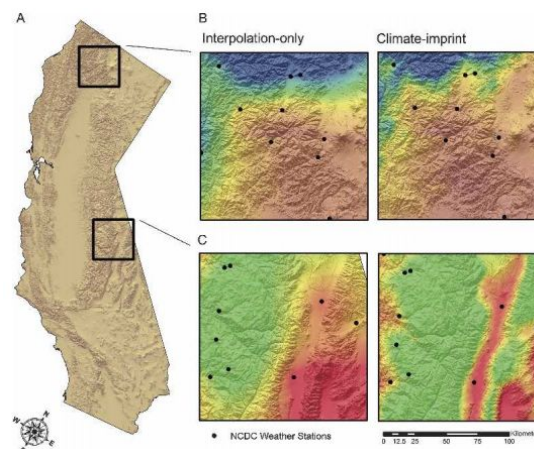


Ilustración 20. Mapa de localización para la comparación a gran escala de métodos de interpolación solo e impresión climática con estaciones base para la estimación de (b) precipitación diaria en las gamas de la costa norte y (c) temperatura máxima en el este de Sierra Nevada: 26 de marzo de 2001. Fuente: Richard et al. (2005)[9].

“Análisis de la Distribución e Interpolación espacial de lluvias en Bogotá, Colombia”

En la ciudad de Bogotá el registro de la precipitación se realiza en estaciones de medición puntual, mediante instrumentos de medición discreta o continua. En la práctica es necesario conocer su variación espacial en un área de estudio, para lo cual deben construirse campos de precipitación empleando metodologías de interpolación tales como el método de Kriging, el método IDW (Inverse Distance Weighted). Las estaciones seleccionadas para el desarrollo del presente estudio son de tipo meteorológica especial (ME), pluviométrica (PM), climatológica principal (CP), pluviográfica (PG) y climatológica ordinaria (CO). Los registros hacen parte de la base de datos de las entidades como: Corporación Autónoma Regional (CAR), Instituto de Hidrología, Meteorología y Estudios ambientales (IDEAM) y la Empresa de Acueducto y Alcantarillado de Bogotá (EAAB). En resumen se consideraron 77 estaciones con precipitación diaria y 37 estaciones con precipitación sub-horaria.

Se ha establecido que para la existencia de una red pluviométrica óptima en zonas urbanas se debe contar al menos con un pluviómetro cada 10 a 20 km², mientras que algunos autores argumentan que la densidad de estaciones debe ser de 6 km² por estación (World Meteorological Organization[16]; citado por Vargas et al. 2011, p. 152)[14]. Para la generación y comparación de los campos de precipitación se emplearán los métodos de interpolación de IDW (distancia inversa ponderada), Kriging.

Análisis Estructural de los Datos

A nivel diario se realizó el análisis estructural de una serie de días en los que se presentaron valores altos de precipitación, comprendidos en la segunda temporada de lluvia de Bogotá. El software utilizado para dicho análisis es el GS+, que se distribuye libremente en la red en el sitio: <http://www.gammadesign.com/>. Se estimaron los semivariogramas experimentales con el fin de elegir el modelo teórico que mejor representara el fenómeno. En la **Tabla 2**, se presentan los parámetros de los semivariogramas obtenidos a partir de la precipitación diaria.

Tabla 2. Parámetros del semivariograma para precipitación diaria

Zona	Fecha	Modelo	Alcance (m)	R²
Casco Urbano 50 estaciones	21/09/1995	Gaussiano	1260	0.04
	22/09/1995	Gaussiano	65900	0.59
	24/09/1995	Gaussiano	4190	0.70
	30/09/1995	Esférico	1680	0.18
Todas 77 estaciones	21/09/1995	Gaussiano	78900	0.66
	22/09/1995	Gaussiano	13920	0.97
	24/09/1995	Esférico	7800	0.51
	30/09/1995	Exponencial	7500	0.46

En la **Tabla 2** se observa la variación del modelo de ajuste y del orden de magnitud del alcance entre días consecutivos. Al realizar el análisis estructural para un mismo día con condiciones diferentes de número de estaciones, se observa que el mejor desempeño se obtuvo para el día 22 de septiembre de 1995. En la estimación de los semivariogramas del análisis estructural se encontró que no existe una dirección predominante en los semivariogramas direccionales. Para este caso se recomienda utilizar semivariogramas isotrópicos para el ajuste a los modelos teóricos.

Con el fin de disminuir la incertidumbre de los valores interpolados en las fronteras de un campo de precipitación calculado (interpolado a partir de registros en superficie), se procede a disponer estaciones localizadas fuera de la zona de interés. De esta manera se logra que al momento de hacer la interpolación se cuente con más información, para luego hacer el “recorte” de la zona de estudio. Para

el caso de la validación cruzada, el parámetro de comparación empleado en este análisis fue la sumatoria del valor absoluto del error entre el valor estimado y el valor registrado, para cada uno de los 731 días modelados, en las 5 estaciones consideradas (estaciones seleccionadas). La selección de la mejor metodología se realizó de dos formas, la primera de ellas en función de la frecuencia de éxito (número de días de éxito/número total de días); y la segunda, seleccionando el menor error acumulado. Ambos criterios se aplicaron en todos los días registrados y en las estaciones definidas. Los resultados de esta validación cruzada se presentan en la **Tabla 3.** para los diferentes tamaños de celda (Δx) considerados, estandarizando los valores acumulados del error dividiéndolos entre el número de días (mm/día).

Tabla 3. Comparación de metodologías de interpolación. Fuente: Vargas et al. (2011)[14].

ΔX (m)	Frecuencias		Error	
	Método	\sum Error	Método	\sum Error
100	IDW	5.37	IDW	5.37
200	Kriging*	5.63	Kriging*	5.63
250	V4	8.03	IDW	5.46
500	IDW	5.33	IDW	5.33

Con base en los resultados consignados en la **Tabla 3.**, se observa que con el método IDW se obtiene la menor diferencia entre los valores estimados y los registrados. En general puede apreciarse que con el método V4 (Método MATLAB) se obtienen diferencias sustanciales, aspecto que se evidencia en las altas sumatorias de error obtenidas, para los dos criterios considerados.

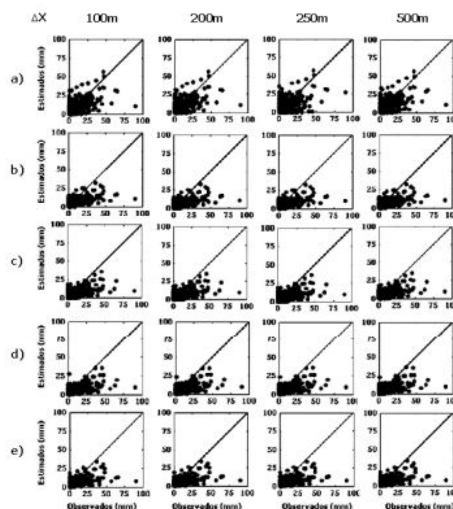


Ilustración 21. Comparación del ajuste del método de Kriging variando el tamaño de celda y el semivariograma. a) lineal, b) gaussiano fijo, c) gaussiano ajustable, d) esférico ajustable y e) exponencial ajustable. Fuente: Vargas et al. (2011)[14].

En la **Ilustración 21.** se observan los diagramas de dispersión, obtenidos a partir del método de Kriging, entre los valores de precipitación estimados y los valores registrados (en las 5 estaciones empleadas para la validación cruzada) para cada uno de los tamaños de celda considerados, comparando adicionalmente, los diferentes semivariogramas empleados. La propiedad “ajustable” y “Fijo” hacen referencia a si se realizó el ajuste de los parámetros de ese modelo teórico para cada día, o si se consideran invariantes en el tiempo. En la **Tabla 4.** se presentan los coeficientes de correlación obtenidos en los diagramas de dispersión de la **Ilustración 21.**

Tabla 4. Correlación del método de Kriging empleando diferentes semivariogramas. Fuente: Vargas et al. (2011)[14].

Semivariograma	ΔX (m)			
	100	200	250	500
Lineal	0.705	0.703	0.578	0.709
Gaussiano fijo	0.702	0.702	0.701	0.701
Gaussiano ajustable	0.688	0.688	0.689	0.689
Esférico ajustable	0.696	0.698	0.699	0.699
Exponencial ajustable	0.688	0.687	0.688	0.687

De la **Tabla 4.** se puede evidenciar que el mejor modelo teórico de semivariograma obtenido para la metodología de Kriging es el lineal, lo cual indica que el ajuste de los parámetros de un semivariograma en cada instante del período considerado no mejora sustancialmente su desempeño. Tomando los resultados obtenidos por el método de Kriging empleando un semivariograma lineal y los otros métodos considerados en el presente trabajo se obtuvieron los diagramas de dispersión que se presentan en la **Ilustración 22**

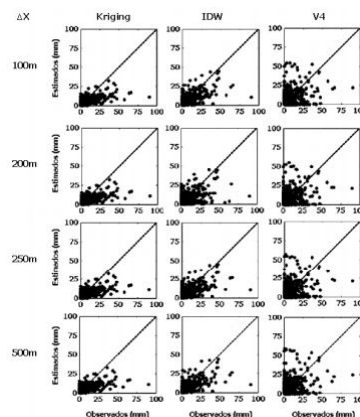


Ilustración 22. Variación del ajuste de acuerdo con el método de interpolación espacial utilizado. Fuente: Vargas et al. (2011)[14].

Generación de campos de Precipitación

En la **Ilustración 23.** se observa los campos de precipitación obtenidos para cada una de las metodologías, para este caso la interpolación con IDW representa con mayor aptitud los lugares con altas precipitaciones o centros de tormenta, cabe anotar que estos sitios son aproximados al lugar geográfico de la estación que registró el valor de precipitación. La interpolación realizada con el método de Kriging (Modelo Lineal) suaviza las regiones donde se observaban los centros de tormenta y respecto a la interpolación con V4, se observa que las precipitaciones interpoladas superan las de los modelos anteriores en casi el doble. De otro lado también se observa una zona de gran extensión con precipitación igual a cero.

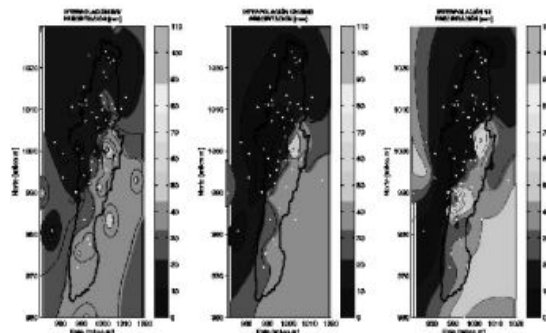


Ilustración 23. Campos de precipitación obtenidos (para el evento registrado en 13-05-1996).
Fuente: Vargas et al. (2011)[14].

Discusión y Conclusión

Se observa que el método IDW es el que mejor representa el comportamiento espacial de la precipitación debido a que se obtienen los menores errores absolutos en la comparación de los valores interpolados con los observados. Para el caso de la interpolación con el método de Kriging, se observó que en una serie de registros consecutivos (precipitación), no existe un modelo teórico de semivariograma que se mantenga en el tiempo y muestre un buen desempeño, sin embargo el semivariograma lineal es el que genera menor error. También se observó que el modelo que genera mayores sobreestimaciones de precipitación es el V4, seguido por el método de Kriging. El modelo IDW no genera sobreestimaciones de la precipitación, sin embargo, produce concentraciones bastante fuertes en los puntos de medición.

5. Marco teórico

5.1. Funciones de correlación espacial

5.1.1. Covariograma y correlograma

La función de covarianza muestral entre parejas de observaciones que se encuentran separadas por una distancia h , es calculada haciendo uso de la fórmula clásica de covarianza muestral

$$C(h) = COV(Z(x+h), Z(x)) = \frac{\sum_{i=1}^n (Z(x+h) - m)(Z(x) - m)}{n} \quad (1)$$

$$C(h) = \frac{\sum_{i=1}^n (Z(x+h) \cdot Z(x))}{n} - m^2$$

Donde m hace referencia al valor promedio en todos los puntos de la zona de estudio y n el número de parejas que cumplen con la distancia establecida. Tomando este comportamiento como estacionario y usando la varianza muestral para estimar la varianza de la variable, se puede definir de la siguiente manera el correlograma muestral. (Henaar 2015)[5].

$$r(h) = \frac{COV(Z(x+h), Z(x))}{S_{x+h} \cdot S_x} = \frac{C(h)}{S_x^2} = \frac{C(h)}{C(0)} \quad (2)$$

5.1.2. Variograma

La definición del variograma se da a través de un proceso estacionario intrínsecamente, de manera que

$$2\gamma(h) = Var(Z(s_i) - Z(s_j)) \quad (3)$$

Donde se tiene que, s_i, s_j pertenece a D y el vector de distancia entre las localizaciones se define como $h = s_i - s_j$. Adicionalmente es posible deducir que $\gamma(h) = \gamma(-h)$ y $\gamma(0) = 0$. El variograma debe cumplir necesariamente con dos condiciones,

1. Ser una función condicionalmente definida negativa.

$$\sum_{i=1}^n \sum_{j=1}^n \varphi_i \varphi_j 2\gamma(s_i - s_j) = -2 \sum_{i=1}^n \sum_{j=1}^n \varphi_i \varphi_j \text{Cov}(Z(s_i), Z(s_j)) = -\text{Var} \left(\sum_{i=1}^n \varphi_i Z(s_i) \right) \leq 0 \quad (4)$$

2. Tener un ritmo de crecimiento inferior a h^2 . (Melo, 2012)[7].

$$\lim_{h \rightarrow \infty} \frac{2\gamma(h)}{h^2} = 0 \quad (5)$$

5.1.3. Estimación del variograma y el variograma

Estimador clásico

Este estimador, propuesto por Matheron (1962); es considerado el más sencillo y su obtención se hace mediante el estimador del método de momentos. Asumiendo la hipótesis de existencia de estacionariedad intrínseca, la media del proceso constante y que los puntos de muestreo $\{s_1, \dots, s_n\}$ están localizados sobre una rejilla regular; la definición del método de momentos, viene dada por

$$2\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{(s_i, s_j) \in N(h)} (Z(s_i) - Z(s_j))^2 \quad (6)$$

Donde $N(h)$ hace referencia a todas las parejas cuya distancia es igual a h . En este proceso no hace necesario estimar la media μ .

Debido al hecho de no ser robusto, este es un estimador no paramétrico que solo se considera óptimo en aquellos casos donde se disponga de una malla regular de muestreo representativa y con una distribución normal. En ocasiones la aplicación de este estimador genera variogramas experimentales erráticos, generalmente por variaciones del caso ideal para su aplicación, como lo son, distribuciones alejadas de la normalidad, heterocedasticidad o existencia de atípicos. (Melo, 2012)

En el caso del covariograma, el estimador del método de momentos se define como

$$\hat{C}(h) = |N(h)| \sum_{(s_i, s_j) \in N(h)} (Z(s_i) - \hat{Z})(Z(s_j) - \hat{Z}) \quad (7)$$

Donde \hat{Z} es un estimador de la media μ del proceso.

Estimador robusto

Propuesto por Cressie & Hawkins en 1980, este posee mayor robustez como estimador insesgado del variograma y se define por medio de

$$2\hat{\gamma}(h) = \frac{1}{|N(h)|(0.457 + 0.494/|N(h)|)} \sum_{(s_i, s_j) \in N(h)} |Z(s_i) - Z(s_j)|^{1/2} \quad (8)$$

Donde los coeficientes cumplen con la función de garantizar la insesgadez de este estimador. (Melo, 2012)[7].

5.1.4. Modelos comunes de variogramas

La Dr. Margaret Armstrong [1], en su libro Basic Linear Geostatistics, define de la siguiente manera los modelos más comunes de variogramas

Efecto pepita

Corresponde a un modelo puramente aleatorio (ruido blanco), donde no existe correlación entre los valores sin importar la cercanía entre los mismos.

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C & |h| > 0 \end{cases} \quad (9)$$

Modelo esférico

Es el modelo más utilizado, se compone de una expresión polinómica simple y su forma coincide con lo observado generalmente; inicialmente un crecimiento casi lineal hasta cierta distancia y luego la estabilización, alcanzada en un punto cuya abscisa sea $2a\sqrt{3}$.

$$\gamma(h) = \begin{cases} C \left(\frac{3}{2} \frac{|h|}{a} - \frac{1}{2} \left(\frac{(|h|^3)}{(a^3)} \right) \right) & |h| < 0 \\ C & |h| \geq 0 \end{cases} \quad (10)$$

Modelo exponencial

El rango efectivo de este modelo se encuentra en $3a$, al ser la distancia en la cual se alcanza el 95% de su límite; este modelo tiene un comportamiento lineal para distancias pequeñas, en el caso del exponencial inicialmente presenta un crecimiento muy rápido.

$$\gamma(h) = \{C (1 - \exp(-|h|/a))\} \quad (11)$$

Modelo Gaussiano

Representa un fenómeno extremadamente continuo, donde se ha evidenciando la ocurrencia de inestabilidades numéricas cuando no se hace uso del efecto pepita. Su rango efectivo es $\sqrt{3}a$.

$$\gamma(h) = C \left(1 - \exp \left(-\frac{|h|^2}{a^2} \right) \right) \quad (12)$$

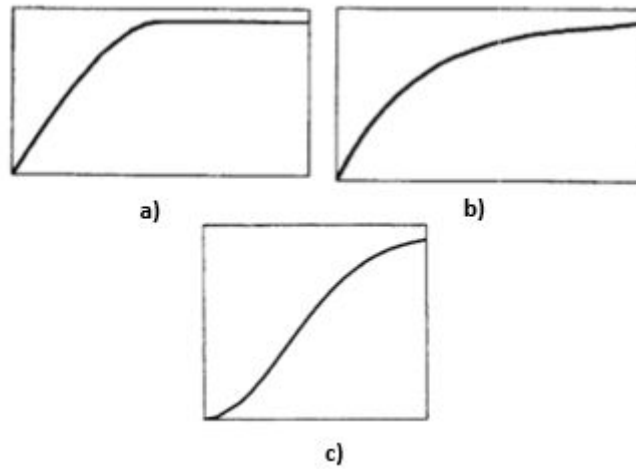


Ilustración 24. a) Modelo Esférico b) Modelo Exponencial c) Modelo Gaussiano
Tomado de (Armstrong, 1950)[1].

5.1.5. Estimación de los parámetros del variograma

Los semivariogramas estimados $\hat{\gamma}(h)$ carecen de la propiedad de ser semidefinidos positivos, lo que abre la posibilidad a que algunas predicciones espaciales generadas con estos tengan varianzas negativas. La forma más común de evitar esto es reemplazando el semivariograma empírico por un modelo paramétrico definido como $\gamma(h, v)$ el cual se aproxime a la dependencia espacial encontrada por el semivariograma empírico y a su vez cumpla con la condición de ser semidefinido positivo. El objetivo de esto será elegir el modelo que mejor se ajuste al conjunto de observaciones para ser utilizado en el proceso de predicción. (Melo, 2015)

Estimación por mínimos cuadrados

Al estimar por mínimos cuadrados ordinarios (OLS), se busca la obtención del valor de \hat{v} que minimice

$$\sum_{j=1}^n (\hat{\gamma}(h_j) - \gamma(h_j, v))^2 = [\hat{\gamma} - \gamma(v)]' [\hat{\gamma} - \gamma(v)] \quad (13)$$

Un problema presentando en la aplicación de este procedimiento es la presencia de autocorrelación en las estimaciones y la existencia de varianzas diferentes. La solución por mínimos cuadrados generalizados (GLS), busca minimizar

$$[\hat{\gamma} - \gamma(v)]' V(v)^{-1} [\hat{\gamma} - \gamma(v)] \quad (14)$$

Donde $V(v)$ es la matriz de varianzas- covarianzas de $\hat{\gamma}$, dependiendo de que el valor de v sea conocido y sus elementos pueden ser difíciles de encontrar.

Finalmente la estimación por mínimos cuadrados ponderados (WLS), se basa en minimizar

$$\sum_{j=1}^n w_j [\hat{\gamma}(h_j) - \gamma(h_j, v)]^2 = [\hat{\gamma} - \gamma(v)]' W(v)^{-1} [\hat{\gamma} - \gamma(v)] \quad (15)$$

Donde $W(v)$ es una matriz diagonal con las varianzas de $\hat{\gamma}$, mientras que los pesos vienen contenidos en w_j , existen diferentes ponderaciones para el ajuste del semivariograma pertenecientes a la librería gstat. La estimación por mínimos cuadrados ponderados es la más utilizada debido a su facilidad de implementación. (Melo, 2015)[8].

5.2. Anamorfosis gaussiana

La transformación de la anamorfosis, es necesaria para llegar a una distribución gaussiana. Gráficamente consiste en deformar el histograma de los datos en un histograma que siga una distribución gaussiana estándar.

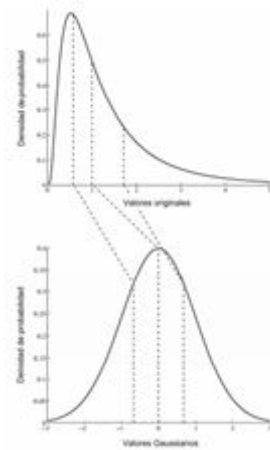


Ilustración 25 . Construcción de la gráfica de la anamorfosis gaussiana
Tomado de (Emery, 2013)[3].

El primer paso de la anamorfosis gaussiana es la acumulación de Z entre una distribución equivalente normal, la variable Y , para un valor de acumulación dado el correspondiente valor transformado es elegido para que el histograma transformado final tenga la característica de la forma de campana de la distribución normal, con media 0 y varianza 1. Sin embargo para obtener la relación uno a uno entre la acumulación y la normal equivalente, en lugar del histograma se trabaja en términos del histograma acumulativo. Para un valor Z_0 de la acumulación, el gaussiano equivalente Y_0 se puede encontrar mediante la ecuación de probabilidades vinculada a la función de anamorfosis ϕ . (Roth & Armstrong)[12].

$$Pr \{Z < z_0\} = Pr \{\phi(Y) < \phi(y_0)\} = Pr \{Y < y_0\} \quad (16)$$

5.3. Anisotropía

Se manifiesta cuando el variograma cambia dependiendo la dirección del espacio; pueden ser identificadas al comparar diferentes variogramas experimentales calculados en diferentes direcciones en el espacio. (Emery, 2013)[3].

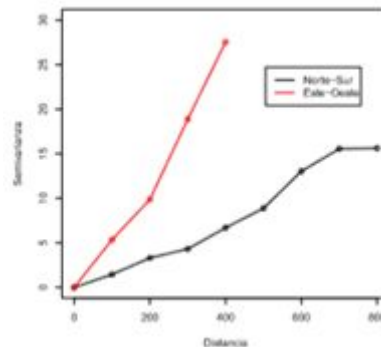


Ilustración 26. Semivariogramas experimentales en dos direcciones fundamentales.
Tomado de (Melo, 2015)

Anisotropía geométrica

Una anisotropía se cataloga como geométrica cuando en el mapa variográfico se observan elipses concéntricas o elipsoides en un caso 3D, los variogramas direccionales tienen una misma forma pero diferentes alcances. En el caso bidimensional la anisotropía geométrica se puede definir completamente por su coeficiente de anisotropía a/b , que hace referencia a la relación entre el semieje mayor a y el semieje menor b de cualquiera de las elipses y el ángulo de anisotropía θ formado por eje semieje mayor y eje de las abscisas. De manera que se puede escribir como

$$\gamma(h) = \gamma_0(|Ah|) \quad (17)$$

Donde γ_0 hace referencia a un variograma isotrópico y A a una matriz de (2x2) o (3x3) que depende de los ángulos y coeficientes de anisotropía. (Emery, 2013)

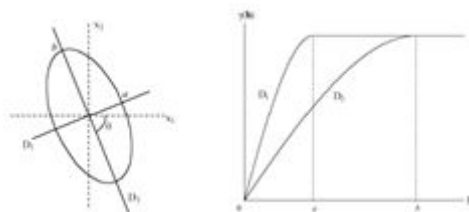


Ilustración 27. Ejemplo bidimensional de anisotropía geométrica.
Tomado de (Emery, 2013)[3].

Anisotropía zonal

La anisotropía zonal, se define como un modelo en el cual la función no se expresa en función de una o varias coordenadas; la función aleatoria tiene un comportamiento constante en el plano horizontal de manera que el variograma $\gamma(h)$ solo depende de la componente vertical de h ; θ hace referencia al

ángulo entre h y la dirección vertical y u hace referencia al vector unitario orientado verticalmente; la anisotropía zonal se puede denotar como

$$\gamma(h) = \gamma(|h| \cos \theta u) \quad (18)$$

Este variograma tiene la misma meseta en todas las direcciones del espacio, a excepción del plano horizontal donde el $\cos \theta = 0$ y $\gamma(h) = 0$.

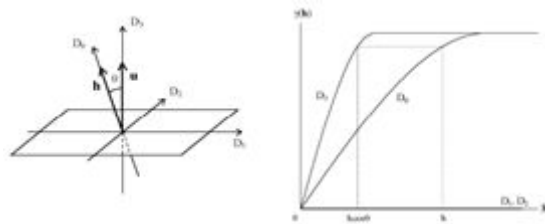


Ilustración 28. Ejemplo tridimensional de anisotropía zonal.
Tomado de (Emery, 2013)[3].

6. Descripción de la base de datos

La base de datos suministrada tiene como zona de estudio Croacia (Europa), y contiene la información de la temperatura media terrestre diaria registrada en 158 estaciones meteorológicas, distribuidas de la siguiente manera

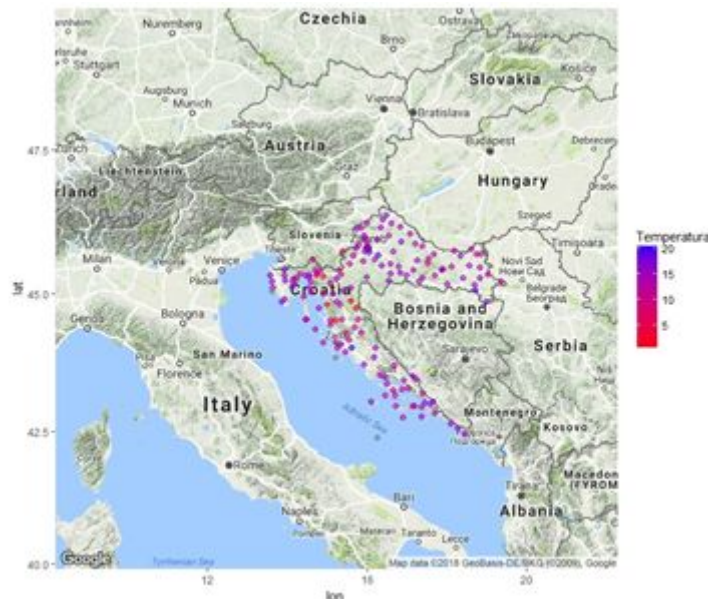


Ilustración 29. Distribución de las estaciones.
Software R.

Las variables a tratar se describen a continuación

Tabla 5. Descripción base de datos

<i>Identificador</i>	<i>Variable</i>	<i>Tipo</i>	<i>Descripción</i>
LST2008_03_13	Temperatura 13 de marzo de 2008	Intervalo	Temperatura media terrestre registrada por las estaciones meteorológicas el día 13 de marzo de 2008.
LAT	Latitud	Razón	Coordenada norte de la estación meteorológica.
LON	Longitud	Razón	Coordenada este de la estación meteorológica.
Hrdem	Modelo de Elevación Digital proyectado en el sistema UTM (zona 33)	Intervalo	Representación de las alturas de cada estación meteorológica.
Hrdsea	Buffer a la costa en km	Intervalo	Distancia entre la costa y cada estación meteorológica.
Hrtwi	Índice de Humedad Topográfica	Intervalo	Humedad en la ubicación de la estación meteorológica.

7. Análisis Exploratorio

Espacialización de los datos

Con el fin de hacer una primera aproximación al comportamiento de la temperatura en Croacia, se realizó un gráfico tridimensional **Ilustración 30**, donde la elevación de cada punto depende de la temperatura registrada.

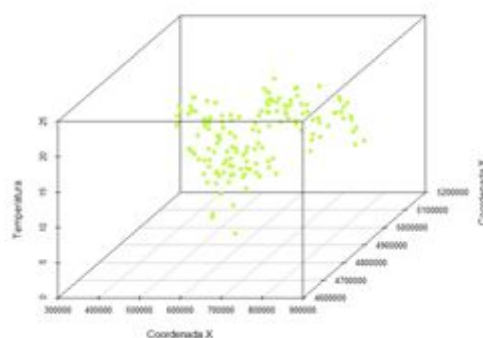


Ilustración 30. Distribución espacial 3D de la temperatura el día 13 de marzo de 2008.
Software R.

Adicionalmente, se presenta una espacialización 2D **Ilustración 29**, donde es posible evidenciar el comportamiento de aumento a disminución de temperatura dependiendo la región de la zona de estudio.

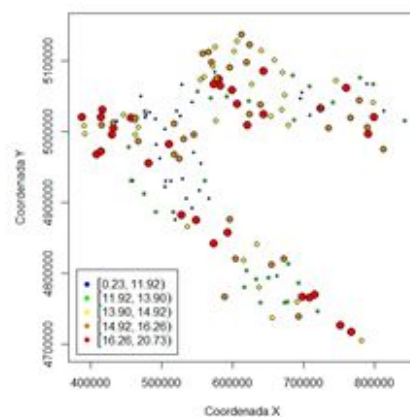


Ilustración 31. Distribución espacial 2D de la temperatura el día 13 de marzo de 2008.
Software R.

Al realizar una comparación de la ilustración anterior con la **Ilustración 31**, se puede inferir que en la zona costera, en la zona central y este de Croacia, es donde se evidencian los registros de mayor temperatura.

Estadísticos de tendencia central

	Hrdem	Hrdsea	Hdrtwi	Latitud	Longitud	Temperatura
Min.	1.00	0.00	12.90	42.39	13.56	0.23
1st Qu	81.25	0.20	16.52	44.30	15.18	12.45
Median	131.50	59.80	18.20	45.22	16.01	14.49
Mean	229.14	91.88	17.75	44.87	16.16	13.87
3rd Qu.	275.50	174.40	19.10	45.54	17.15	15.89
Max.	1453.00	252.30	21.40	46.38	19.36	20.73

Ilustración 32. Estadísticos de tendencia Central

A partir de los estadísticos presentados en la **Ilustración 32.**, es posible evidenciar que la variable a modelar *Temperatura* presenta un intervalo de valores entre 0.23 y 20.73 °C; respecto a sus valores de mediana y media aunque no difieren en un rango considerable, no es posible definir si sigue una distribución normal. En el caso de las variables *Hrdem* y *Hrdsea* sus estadísticos permiten inferir que no siguen una distribución normal y que sus valores de media y mediana pueden estar influenciados por los valores atípicos presentados en las estaciones.

7.1. Análisis de normalidad

En búsqueda de conocer si el comportamiento de la temperatura de la superficie terrestre registrada por las estaciones meteorológicas el día 13 de marzo de 2008 es normal, se utilizaron algunos gráficos y pruebas, como se muestran a continuación:

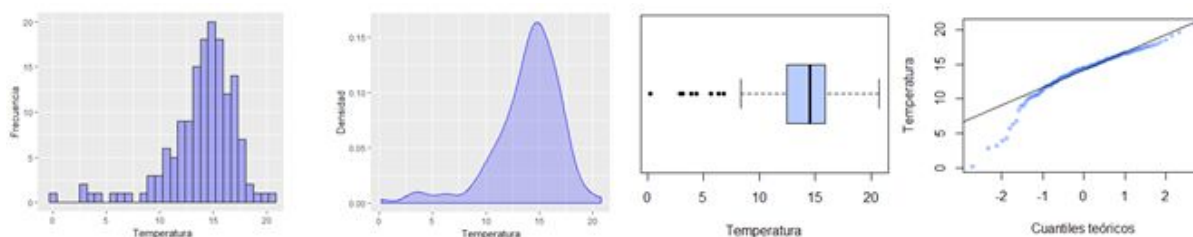


Ilustración 33. Gráficos para la detección de la normalidad en la variable Temperatura de la Superficie Terrestre

El histograma y el diagrama de densidad muestran que se presentan datos asimétricos a la izquierda, es decir, aparentemente no están distribuidos normalmente; una posible causa podría ser la presencia de presuntos datos atípicos en los valores de baja temperatura. El box-plot sugiere que la mediana se encuentra aproximadamente en 14° centígrados y junto con el qq-plot se confirma la presencia de datos atípicos.

Test de Normalidad Shapiro Wilk		Test de Normalidad Kolmogorov-Smirnov	
Variable: temp		Variable: temp	
w	0.89959	D	0.12158
p-value	9.708e-09	p-value	0.02172
		Hipótesis Alternativa: dos-colas "two-sided"	

Ilustración 34. Tests de Normalidad aplicados a la variable Temperatura de la Superficie Terrestre

Las pruebas de normalidad realizadas, indican que la variable Temperatura de la Superficie Terrestre no sigue una distribución normal, ya que el $p - value < \alpha$ ($\alpha = 0.05$), por ello se rechaza la hipótesis nula (*La variable sigue una distribución normal*).

Para solucionar el problema de no normalidad, se transformó la variable mediante el método de Anamorfosis Gaussiana. Como se puede ver en la **Ilustración 35.**, se asocia un valor gaussiano a cada valor crudo de la variable original, en búsqueda de que el histograma de la variable se deforme en un histograma Gaussiano estándar. El comportamiento de la variable mejoró bastante, como es notable en el histograma y diagrama de densidad realizado después de la transformación.

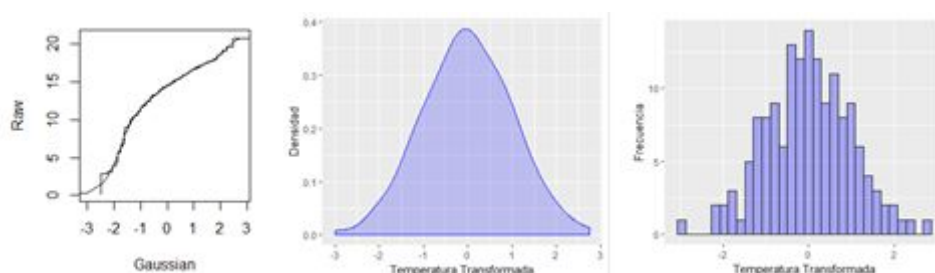


Ilustración 35. Gráficos para la detección de la normalidad en la variable Temperatura de la Superficie Terrestre Transformada.

Comprobando lo evidenciado mediante las gráficas anteriormente mencionadas, se realizaron tres pruebas de normalidad, en las cuales el $p - value > \alpha$, por lo tanto a un nivel de confianza del 95% no se rechaza la hipótesis nula, es decir la variable Temperatura de la Superficie Terrestre transformada sigue una distribución normal.

Test de Normalidad Shapiro Wilk Variable: temp.trans		Test de Normalidad Anderson-Darling Variable: temp.trans	
w	0.99864	A	0.075104
p-value	1	p-value	0.9992

Test de Normalidad Kolmogorov-Smirnov Variable: temp.trans	
D	0.026037
p-value	0.9999
Hipótesis Alternativa: dos-colas "two-sided"	

Ilustración 36. Tests de Normalidad aplicados a la variable Temperatura de la Superficie Terrestre Transformada

7.2. Análisis de tendencia

Se realizó un primer modelo para conocer cuáles de las variables $HRdem$, $HRdsea$, $HRdtwi$, LON , LAT de la base de datos, en donde LON y LAT son las variables proyectadas en UTM Este y Norte, explican de manera adecuada la variable Temperatura de la Superficie Terrestre Transformada. En la salida que se muestra en la *ilustración*, se observa que a un nivel de significancia del 5% las variables $HRdem$, $HRdsea$, LON , LAT son significativas estadísticamente, y según el valor del R^2 ajustado un 24% de la variabilidad de la Temperatura de la Superficie Terrestre, es explicada por las variables dependientes mencionadas anteriormente.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-29.4205	12.5026	-2.35	0.0199
HRdem	-0.0016	0.0003	-4.54	0.0000
HRdsea	-0.0075	0.0034	-2.23	0.0275
HRdtwi	0.0560	0.0451	1.24	0.2166
LON	0.00004132	0.00002087	1.98	0.0496
LAT	0.00005440	0.00002309	2.36	0.0198

Observations	153
R ²	0.268
Adjusted R ²	0.243
Residual Std. Error	0.873 (df = 147)
F Statistic	10.766*** (df = 5; 147)

Ilustración 37. Primera Regresión Lineal realizada. Salida software R.

Con el objetivo de obtener un modelo que ajuste mejor la realidad, se utilizó el método de Stepwise, con los criterios Backward- Forward y AIC, dando como resultado la regresión lineal expuesta en la *ilustración*, que según el estadístico F, a un nivel de confiabilidad del 99% ajusta los datos, y con variables significativas estadísticamente a un nivel de significancia del 5%, a saber $HRdem$, $HRdsea$ y LAT .

Con este modelo se identifica la presencia de tendencia de orden 1, y por ello es necesario llevar a cabo el modelado del variograma a los residuos de esta regresión, debido a que estos están libres de tendencia.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-25.8706	12.1931	-2.12	0.0355
HRdem	-0.0018	0.0003	-6.15	0.0000
HRdsea	-0.0068	0.0033	-2.03	0.0446
LON	0.000003694	0.000002061	1.79	0.0751
LAT	0.000004973	0.000002283	2.18	0.0309
Observations	153			
R ²	0.260			
Adjusted R ²	0.240			
Residual Std. Error	0.874 (df = 148)			
F Statistic	13.025*** (df = 4; 148)			

Ilustración 38. Regresión Lineal seleccionada. Salida software R.

Test de Normalidad Shapiro Wilk		Test de Normalidad Anderson-Darling	
Variable: residuales.mejormodelo		Variable: residuales.mejormodelo	
w	0.98368	A	0.72472
p-value	0.0676	p-value	0.05758

Test de Normalidad Kolmogorov-Smirnov	
Variable: residuales.mejormodelo	
D	0.076788
p-value	0.3277
Hipótesis Alternativa: dos-colas "two-sided"	

Ilustración 39. Tests de Normalidad aplicados a los residuos del modelo seleccionado.

Se evalúa la normalidad de los residuos del modelo mediante las pruebas expuestas en la *ilustración*, los valores *p-value* arrojados son mayores al nivel de significancia $\alpha = 5\%$, por lo tanto, no rechazamos la hipótesis nula y aceptamos que los residuos siguen una distribución normal.

7.3. Análisis de anisotropía a los residuos

Mediante la función *estimateAnisotropy(object, depVar, formulaString)* de la librería *intamap* diseñada para R, se evaluó la existencia de anisotropía en los residuos. El radio anisotrópico es 1 y el ángulo de orientación de la elipse es 0, lo que indica que no hay anisotropía en los residuos.

Estimación de la Anisotropía	
R	1
Theta (Grados)	0
Q	[0, 0, 0]
doRotation	FALSE

Ilustración 40. Evaluación de la Anisotropía en los residuos Salida Software R.

8. Modelamiento del semivariograma

Inicialmente se realiza el cálculo del semivariograma experimental, la distancia utilizada se tomó como la distancia máxima entre las observaciones dividida entre dos, como se puede observar a continuación:

Coordenadas		
	X	Y
Min	387331.2	5020958
Max	842442.8	5016272
Distancia Max	455135.7	
Distancia a usar	227567.9	

Ilustración 41. Cálculo de la Distancia a utilizar para el modelado del semivariograma.

Los semivariogramas experimentales obtenidos son los siguientes:

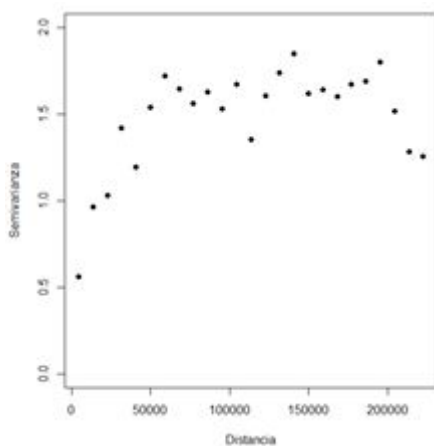


Ilustración 42. Semivariograma experimental robusto.
Salida software R.

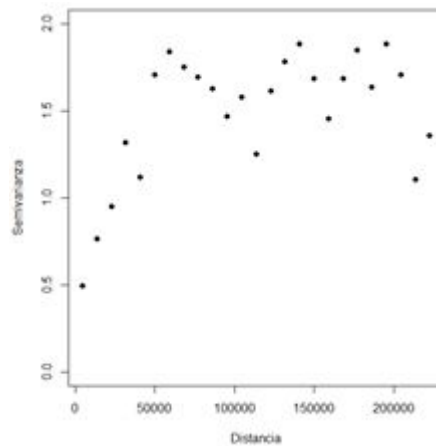


Ilustración 43. Semivariograma experimental mediana.
Salida software R.

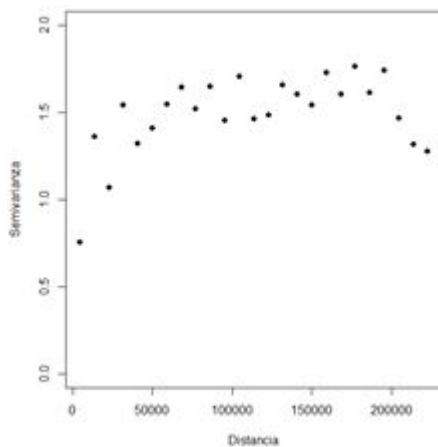


Ilustración 44. Semivariograma experimental clásico.

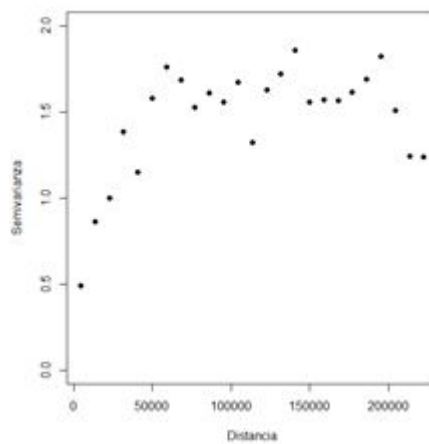


Ilustración 45. Semivariograma experimental media recortada.

Según lo observado, los semivariogramas clásico y robusto presentan características similares, de forma paralela, los semivariogramas media recortada y mediana, son semejantes; sin embargo se

ajustaron varios semivariogramas teóricos sobre los experimentales para poder elegir un modelo teórico adecuado. Estos se muestran a continuación:

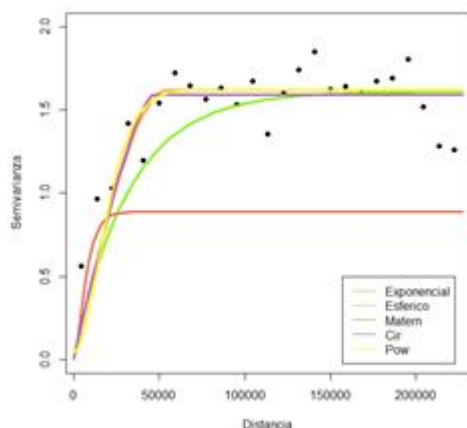


Ilustración 45. Modelo LST experimental robusto.
Salida software R.

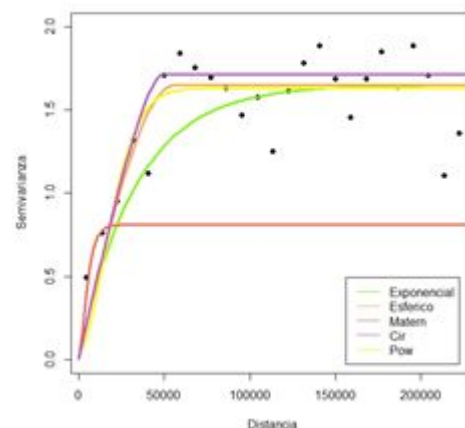


Ilustración 47. Modelo LST experimental mediana.
Salida software R.

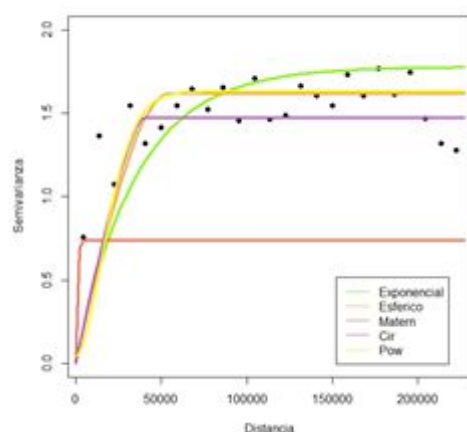


Ilustración 48. Modelo LST experimental clásico.
Salida software R.

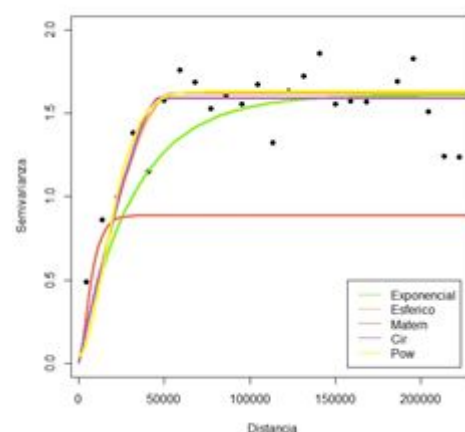


Ilustración 49. Modelo LST experimental media recortada.
Salida software R.

Modelo	Valor AIC			
	Robusto	Clásico	Mediana	Media Recortada
Exponencial	424.1941	424.5216	421.5338	420.7006
Esférico	429.2582	429.7524	430.3206	429.7524
Matern	397.7078	393.6756	401.4363	410.2446
Circular	432.2621	431.2509	432.9086	429.9656
Power Exponential	469.428	469.8946	469.8946	469.8946

Ilustración 50. Valores del AIC para los semivariogramas teóricos.

En base al análisis realizado visualmente sobre los semivariogramas experimentales, se seleccionó el semivariograma experimental clásico, debido al buen ajuste logrado mediante los modelos teóricos y a la forma que presenta. Además de la detallada visualización de los semivariogramas teóricos ajustados, para la elección de éste se utilizó el criterio AIC, en el cual como se observa en la **Ilustración 50**, el menor valor es el del semivariograma Matérn, sin embargo, en la práctica no fue posible modelarlo adecuadamente mediante el ajuste a sentimiento, por lo tanto, no lo se tuvo en cuenta y el modelo seleccionado fue el Exponencial.

Posteriormente se estimaron cuatro semivariogramas exponenciales, cada uno, con un método diferente, a saber, Mínimos Cuadrados Ordinarios (OLS), Mínimos Cuadrados Ponderados (WLS), Máxima Verosimilitud (ML) y Máxima Verosimilitud Restringida (RML). Como es visible en la **Ilustración 51**, los semivariogramas estimados por Máxima Verosimilitud y Máxima Verosimilitud Restringida, presentan los mejores ajustes, estos son bastante similares, y por ello, para la elección del más apto, se utilizó nuevamente el criterio de Akaike AIC.

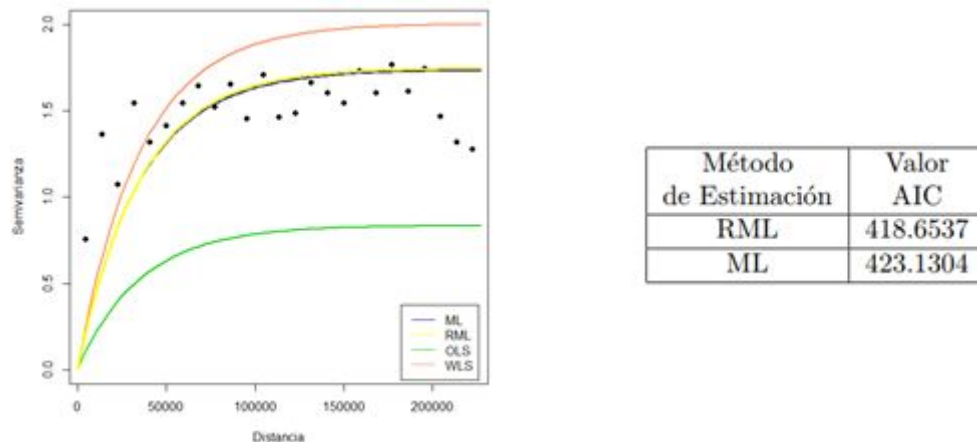


Ilustración 51. Semivariogramas Teóricos Exponenciales ajustados por OLS, WLS, ML, RML (izquierda). Valores AIC para semivariogramas Teóricos Exponenciales (derecha).

Finalmente, para modelar los residuos de la Temperatura de la Superficie Terrestre, se seleccionó el semivariograma Clásico (semivariograma experimental) y el semivariograma Exponencial estimado mediante RML (semivariograma teórico) (**Ilustración 52**). El último mencionado con parámetros $pepita = 0.0288$, $meseta\ parcial = 1.744$ y $rango = 35\ 540$.

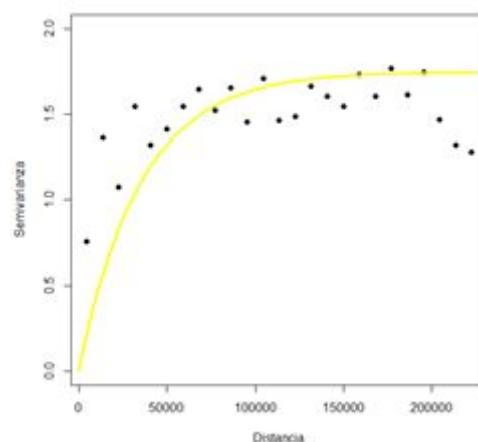


Ilustración 52. Semivariograma Teórico Exponencial estimado mediante RML

9. Referencias

- [1] Armstrong, M. (1950). *Basic linear geostatistics*. Publicado originalmente por Springer -Verlag Berlín Heidelberg Nueva York en 1998. ISBN 978-3-540-61845-4.
- [2] Cressie, N., 1993: *Statistics for Spatial Data*. John Wiley and Sons, 900 pp.
- [3] Emery, X. (2013). *Geoestadística*. Chile: Universidad de Chile. Ingeniería de minas.
- [4] Goovaerts, P., 1997: *Geostatistics for Natural Resources Evaluation*. Oxford University Press, 467 pp, 2000: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.*, 228, 113–129.
- [5] Henao, R. G. (s.f.). *Introducción a la geoestadística*. (2015). Bogotá: Universidad Nacional de Colombia, Departamento de estadística.
- [6] Jones R, M Noguer, D Hassell, D Hudson, S Wilson, G Jenkins & J Mitchell (2004) Generating high resolution climate change scenarios using PRECIS. Hadley Center for Climate Prediction and Research, UK.
- [7] Melo, C. E. (2012). *Análisis geoestadístico espacio tiempo basado en distancias y splines con aplicaciones*. Barcelona: Universidad de Barcelona. Programa de doctorado en Estadística.
- [8] Melo, C. E. (2015). *Estadística espacial teoría y aplicaciones*. Bogotá: Universidad Distrital Francisco José de Caldas.
- [9] Richard D. Hunter and Ross K. Meentemeyer. Climatologically Aided Mapping of Daily Precipitation and Temperature (2005); Department of Geography, Sonoma State University, Rohnert Park, California. Doi: <https://doi.org/10.1175/JAM2295.1>. Recuperado de: <https://journals.ametsoc.org/doi/pdf/10.1175/JAM2295.1>
- [10] Samper, F.J. & J. Carrera 1990. *Geoestadística. Aplicaciones a la Hidrogeología Subterránea*. Centro Internacional de Métodos Numéricos en Ingeniería. Universitat Politècnica de Catalunya. Barcelona.
- [11] Stott P, Stone D A, Allen, M.R. (2004) Human contribution to the European heatwave of 2003, *Nature* 432, 610-614.
- [12] Roth, C., & Armstrong, M. (s.f.). *Confidence intervals for local estimation: Application to the witwatersrand basin*. Francia: Center de Géostatistique.
- [13] Socha, Liliana. *Análisis tendencial de la variación climática “temperatura y precipitación” espacio-temporal del departamento de Boyacá (Colombia)*, (2014). Universidad Militar Nueva Granada. Trabajo de Grado. Recuperado de:

<http://unimilitar-dspace.metabiblioteca.org/bitstream/10654/11999/1/ANALISIS%20TENDENCIAL%20DE%20LA%20VARIACION%20CLIMATICA.pdf>

[14] Vargas, A., Santos, A., Cárdenas, E., & Obregón, N. (2011). *Análisis de la Distribución e Interpolación espacial de las lluvias en Bogotá, Colombia*. Pontificia Universidad Javeriana. Recuperado de: <https://revistas.unal.edu.co/index.php/dyna/article/download/25783/26201>.

[15] Warrick, A. W., D. E. Myers & D. R. Nielsen. 1986. Geostatistical Methods Applied to Soil Science. Methods of Soil Analysis. Part 1. Physical and Mineralogical Methods- Agronomy Monograph 9: 53 - 81.

[16] WORLD METEOROLOGICAL ORGANIZATION (WMO). Guide to Hydrological Practices. Volume I. Hydrology - From Measurement to Hydrological Information, WMO-No. 168, Sixth edition, 2008.