

# **Análisis y Modelamiento Geoestadístico de la temperatura en Croacia para el 3 de julio de 2008**



**UNIVERSIDAD DISTRITAL  
FRANCISCO JOSÉ DE CALDAS**

Juan David Gonzalez Hernandez - 20151025060  
Raúl Camilo Martín Bernal - 20151025909

Geoestadística

Ingeniería Catastral y Geodesia

Universidad Distrital Francisco José de Caldas

*Presentado a* Carlos Eduardo Melo Martínez

Bogotá D.C. agosto de 2020

# Tabla de contenidos

<b>1. Introducción</b>	<b>1</b>
<b>2. Objetivos</b>	<b>2</b>
2.1. Objetivo general . . . . .	2
2.2. Objetivos específicos . . . . .	2
<b>3. Estado del Arte</b>	<b>3</b>
<b>4. Marco teórico</b>	<b>9</b>
4.1. Geoestadística . . . . .	9
4.2. Variable regionalizada . . . . .	9
4.3. Momentos de la variable regionalizada . . . . .	9
4.4. Anamorfosis Gaussiana . . . . .	10
4.5. Funciones de dependencia espacial experimentales . . . . .	10
4.6. Estimación del variograma y covariograma . . . . .	11
4.7. Modelos de variograma . . . . .	12
4.8. Estimación de los parámetros del variograma . . . . .	14
4.9. Anisotropía . . . . .	15
4.10. IDW (Distancia inversa ponderada) . . . . .	16
4.11. Funciones de base radial . . . . .	16
4.12. Kriging . . . . .	17
<b>5. Área de estudio y base de datos</b>	<b>20</b>
<b>6. Análisis exploratorio y estructural</b>	<b>21</b>
6.1. Espacialización de los datos . . . . .	21
6.2. Estadísticas descriptivas . . . . .	23
6.3. Análisis de normalidad . . . . .	24
6.4. Análisis de tendencia . . . . .	26
6.5. Análisis de anisotropía de los residuos . . . . .	28
6.6. Modelamiento del semivariograma . . . . .	29
<b>7. Métodos de interpolación</b>	<b>36</b>
7.1. Metodos determinísticos . . . . .	36
7.1.1. IDW (Distancia inversa ponderada) . . . . .	36
7.1.2. Funciones de base radial . . . . .	38

7.2. Métodos geoestadísticos . . . . . 39

**8. Diseño de red 42**

**9. Conclusiones 44**

# Índice de figuras

6.1. Mapa de la temperatura media terrestre registrada para el día 3 de julio de 2020 . . . . .	21
6.2. Distribución espacial variable de temperatura para el día 3 de julio de 2008 . . . . .	22
6.3. Mapa de altura de las estaciones . . . . .	23
6.4. Histograma, Q-Q plot y Box-Plot de la variable de temperatura . . . . .	25
6.5. Superficie de tendencia de la variable temperatura . . . . .	27
6.6. Semivariogramas direccionales. . . . .	28
6.7. Semivariogramas experimentales KO . . . . .	30
6.8. Semivariogramas teóricos KO . . . . .	31
6.9. Estimación parámetros semivariograma exponencial KO . . . . .	32
6.10. Semivariogramas experimentales KS . . . . .	33
6.11. Semivariogramas teóricos KS . . . . .	34
6.12. Estimación parámetros semivariograma exponencial KS . . . . .	35
7.1. Optimización parámetro p . . . . .	37
7.2. Resultados interpolación IDW . . . . .	37
7.3. Resultados interpolación función de base radial, spline completamente regularizado . . . . .	39
7.4. Resultados interpolación predicción temperatura Kriging Simple y Kriging Ordinario . . . . .	40
7.5. Superficie varianzas Kriging Simple y Kriging Ordinario . . . . .	41

# Índice de cuadros

4.1. Funciones de base radial . . . . .	17
5.1. Descripción base de datos . . . . .	20
5.2. Estadística temperatura 3 de julio de 2008 de Croacia . . . . .	20
6.1. Estadísticas descriptivas todas las variables de la base de datos . . . . .	24
6.2. Estadísticos de normalidad de la variable temperatura . . . . .	25
6.3. Estadísticos de normalidad de la variable temperatura transformada . . . . .	26
6.4. Modelos de análisis de tendencia . . . . .	26
6.5. Estadísticos normalidad de los residuos del modelo 2 . . . . .	27
6.6. Resultados estimación de anisotropía . . . . .	29
6.7. Parámetros semivariogramas exponenciales obtenidos en los diferentes métodos KO . . . . .	33
6.8. Parámetros semivariogramas exponenciales obtenidos en los diferentes métodos KS . . . . .	35
7.1. Valores de RMSPE para cada valor p . . . . .	36
7.2. Estadísticas resultados predicción IDW . . . . .	38
7.3. Resultado optimización funciones de base radial sin tendencia; 500 iteraciones; 15 vecinos . . . . .	38
7.4. Estadísticas resultados función de base radial, spline completamente regularizado . . . . .	39
7.5. Parámetros semivariogramas exponenciales para realización de las interpolaciones Kriging . . . . .	39
7.6. Estadísticas de las predicciones de temperatura de los krigeados . . . . .	40
7.7. RMSPE Krigeados . . . . .	41
8.1. Resultados network.design . . . . .	42
8.2. Costo de las redes de estaciones meteorológicas . . . . .	43

# 1. Introducción

La geoestadística es una rama de la estadística que se especializa en el análisis y modelación de la variabilidad de fenómenos regionalizados a través de una o varias funciones y que ha sido ampliamente aplicada en diversas ramas de las ciencias aplicadas, en ingenierías en áreas como: minería, pesca, geofísica, medio ambiente, hidrogeología, estudios forestales, cartografía, meteorología, etc. En palabras de quien es considerado el padre de la geoestadística es definida como la aplicación de la teoría de las variables regionalizadas a la estimación de procesos o fenómenos geológicos en el espacio Matheron (1963). Una variable regionalizada  $Z(x)$  se puede definir como una variable aleatoria distribuida en el espacio de tal manera que presenta una estructura espacial de correlación esto indica intuitivamente que mientras más cercanos estén situados dos puntos están más correlacionados y mientras más separados hay menos relación entre ellos. Para el presente trabajo la variable a analizar es la temperatura comprendida por el país de Croacia el día 3 de Julio del año 2008 y que grosso modo ha sido utilizada en 3 etapas; 1. Análisis exploratorio de los datos que nos permite conocer un conjunto de información sobre la muestra bajo estudio para realizar un correcto análisis estadístico; 2. Análisis estructural en donde caracterizamos la estructura espacial de la variable temperatura en la región, este proceso es importante porque es aquí donde obtenemos un modelo de una función que refleja la correlación espacial de la variable, esta función es conocida como semivariograma o función semivarianza; 3. La aplicación de interpolaciones para la predicción de valores en puntos no muestrales Viera y González (2002).

El análisis de la temperatura media terrestre es de gran interés para aquellos que estudian el cambio climático debido a que conocer las condiciones climáticas de un país a una escala nacional permitiría crear escenarios para el futuro en términos de poder mitigar los impactos y gestionar de forma eficiente aquellos recursos que se ven afectados por las condiciones climáticas.

## **2. Objetivos**

### **2.1. Objetivo general**

Aplicar los métodos de interpolación determinísticos y geoestadísticos sobre los datos de temperatura terrestre en Croacia previamente tratados estadísticamente para predecir la temperatura media terrestre para el día 3 de julio de 2008.

### **2.2. Objetivos específicos**

1. Analizar y si es el caso solucionar los problemas de anormalidad, tendencia y anisotropía.
2. Determinar el semivariograma teórico que mejor se ajuste a los datos de temperatura.
3. Hacer uso de los métodos de interpolación para generar los mapas correspondientes de predicción y errores .

### 3. Estado del Arte

En este capítulo se realizara una breve descripción de algunos casos de estudio en relación con los diferentes métodos geoestadísticos para interpolación de variables medio ambientales en algunas regiones del Mundo.

#### 1. Comparación de métodos kriging multivariante para la estimación espacial de la temperatura del aireVarela (2001)

Una de las principales preocupaciones de cualquier servicio meteorológico es contar con información depurada y representada espacialmente, teniendo en cuenta las peculiaridades morfológicas de la región en la que incide. En este sentido, el carácter puntual de las observaciones obtenidas mediante la red de estaciones constituyen una limitación para la correcta vigilancia y asistencia en las diversas tareas de planificación (hidrológica, agrícola, etc.). Para acercarse la realidad meteorológica en puntos no muestreados es necesario aplicar técnicas de estimación espacial. El contexto donde se desarrolló este trabajo corresponde con la creación de una herramienta que permita visualizar los daos meteorológicos en tiempo real y cuya estructura metodológica principal es la geoestadística, concretamente los algoritmos de kriging. Las siguientes características de la geoestadística fueron determinantes en estudio: 1) Las medidas de variabilidad espacial como la nube del semivariograma son de utilidad en la validación de datos meteorológicos; 2) No ignora los patrones de dependencia espacial que normalmente se observan en las variables meteorológicas; 3) Admite la incorporación de información secundaria correlacionada; 4) puede emplearse en la construcción de modelos de incertidumbre y en simulación estocástica; 5) se adapta bien al diseño de redes bajo observación. En este artículo se desarrollaron tres métodos geoestadísticos multivariantes: Kriging simple con diversas medias locales, Kriging con deriva externa y cokriging colocado cuyos resultados se compararon mediante validación cruzada.

**Datos empleados:** Datos registrados por la red hidrometeorológica automática del servicio Vasco de meteorología (S.V.M). Con una resolución temporal de diez minutos (144 tomas diarias), cada estación envía en tiempo real a la sede central del (S.V.M). Para el ensayo de la aplicación se han seleccionado temperaturas registradas durante el día 24 de abril de 1999 y que cumplen con el requisito de observarse una tendencia con la elevación del terreno, dicho parámetro geoclimático es uno de los más determinantes en



la distribución espacial de la temperatura. Se han extraído nueve registros diezminutarios con coeficientes de correlación próximos a (-0.6, -0.7, -0.8, -0.9, -0.95) afectados también de manera desigual por otra deriva en dirección norte-sur.

**Modelo de semivariograma:** Los semivariogramas presentados capturan el grado de disimilitud media entre observaciones correspondientes a una realidad, la temperatura del aire, que es cambiante en el tiempo. Pero no puede aplicarse la misma hipótesis sobre su variabilidad espacial a lo largo del día. Para las primeras tomas el semivariograma no presenta meseta, indicativo de que la variabilidad espacial tiene lugar con una tendencia en general, en efecto la temperatura disminuye en forma progresiva de Norte a Sur para lo que se plantea la hipótesis que el valor esperado de la función aleatoria depende de la observación y sus incrementos en primer orden no son estacionarios. A medida que la tendencia con dicho eje se debilita, los semivariogramas van presentando mesetas y alcances mejor definidos. El contraste entre costa e interior es uno de los factores de diferenciación climática más perceptible. En cuanto al ajuste del semivariograma teórico el modelo correspondió al exponencial.

**Predicción espacial:** Se aplicara Kriging en un vecindario  $W(u)$  para todas las realizaciones seleccionadas de  $Z(u)$ , bajo el supuesto de que tal función aleatoria es localmente intrínseca (cuasi-estacionaria). Se determinó que el vecindario  $W(u)$  este formado por las 8 estaciones más próximas a estimar  $u$ , para ello se empleó la distancia del variograma de tal forma que la selección se realizó en la dirección de máxima continuidad. Para ahorrar tiempos de cómputo se seleccionó ese vecindario tan pequeño y la introducción de los parámetros en los sistemas de kriging se realizó en términos de covarianzas y no de semivarianzas. Los sistemas kriging usados en este estudio fueron el kriging simple, kriging ordinario, Kriging simple con diversas media locales en el cual también añadieron información del terreno mediante modelos digitales de elevación, Kriging con una deriva externa también introduciendo la elevación. El hecho de evaluar localmente la relación entre la elevación y la temperatura permite tener en cuenta posibles cambios en el gradiente térmico altitudinal a lo largo del área de estudio.

**Resultados de la estimación y conclusiones:** El funcionamiento de todas las predicciones Kriging se realizó mediante validación cruzada, procedimiento que consiste en considerar temporalmente cada una de las estaciones como un punto no muestreado y estimar su valor a partir del resto. El criterio de comparación es el error medio cuadrático (EMC). El algoritmo que ignora la elevación KO, ofrece a primera vista resultados globales aceptables, no obstante suaviza en exceso los detalles locales de la variación espacial de la temperatura. Los valores más bajos se sobreestiman, mientras que los valores más altos se subestiman, la estimación fue especialmente pobre en las zonas más abruptas insuficientemente muestreadas por la red, como ejemplo la toma diezminutaria 45, parece claro asumir que el comportamiento lineal se extiende por encima de los 1200m, altitud máxima a la que se dispone información, sin embargo las estimaciones no

superaron el valor termométrico alcanzado en dicha cota. Como era de esperar en este estudio la ventaja de emplear algoritmos multivariantes aumenta a medida que la relación entre la temperatura del aire y la elevación del terreno es más fuerte. Con un coeficiente de correlación superior a 0.9, KSML, KDE, CKC superaron en todo momento a KO. De entre estos métodos la mejor alternativa fue KSML que sorprende porque en este método se interpolan los residuos de la regresión global entre ambas variables y se añadió al ajuste, le siguió el método DKE que tiene en cuenta los posibles cambios en el gradiente altitudinal dentro de cada vecindario. Finalmente se concluye que la aplicación de métodos geoestadísticos multivariantes puede proporcionar mapas más efectivos, incluso con las relaciones entre temperatura y elevación no demasiado fuertes, además el estudio se podría robustecer aún más con la inclusión de variables explicativas como la orientación de la ladera.

## **2. Uso combinado de técnicas geoestadísticas y SIG para la realización de mapas robustos de variables climáticas en Asturias Martínez (2007)**

El estudio de variables climáticas es muy importante desde diferentes puntos de vista: hidrogeológicos, aprovechamiento agrícola y forestal, establecimiento de riesgos, etc. Los mapas derivados de estos estudios como mapas de precipitación y temperaturas medias terminan siendo documentos básicos y esenciales para estudios como la recarga de acuíferos en estudios hidrogeológicos necesarios para el establecimiento del potencial hidrogeológico los acuíferos en una zona tan densamente poblada como el área central metropolitana asturiana. Las variables climáticas son por definición procesos espacio-temporales, pero dada la escasez de datos de medida, suelen ser tratados separadamente bien como variables regionalizadas en un lapso temporal prefijado (caso de las precipitaciones y temperaturas medias) o bien como series temporales (caso del estudio de la evolución de la precipitación o de la temperatura en una estación bien informada). En la realización de mapas cartográficos de variables climáticas para la región sobresalen los estudios realizados por Felicísimo Pérez (1980) y "Marquinez, Lastra y García (2003)" donde destacan la importancia de factores orográficos como la altitud, distancia al oeste, distancia al mar. La región de Asturias en el paralelo 43° N, motiva la interacción de masas de aire de origen polar y subtropical, lo cual genera borrascas que se desplazan de Oeste a Este y son el origen de precipitaciones intensas principalmente en periodos estivales, en cuanto a la temperatura se destaca el papel del régimen de vientos, de componente meridional en invierno y septentrional en verano, y que contribuyen por tanto a la suavidad de temperaturas. C. Melo (2015)

**Área de estudio:** La Región de Asturias en el norte de España, con una superficie de 10.604 km<sup>2</sup>. Su límite norte es el mar Cantábrico, y el límite natural del sur es la Cordillera Cantábrica. Esta región se caracteriza por un relieve abrupto, con altitudes que van desde el nivel del mar hasta 2648 m en solo 40 km, y con un 80 % de las tierras que superan las pendientes del 20 % y el 34,5 % del territorio más del 50 % de pendiente.

**Datos empleados:** Puntos de muestreo de la red meteorológica regional y las variables climáticas consideradas serán los valores medios o máximos mensuales de precipitaciones y/o temperaturas. Dicha red posee amplias zonas de montaña sin muestrear, dado que no existen datos por encima de 1500 metros de altitud.

**Modelo de varianza espacial:** Se procede a inferir el variograma experimental a partir de los variables-dato, a realizar su ajuste con modelos teóricos. Entre éstos, los modelos esférico y exponencial. Para el mismo rango el modelo esférico traduce una mayor continuidad del fenómeno. Para dar cuenta de las microvariabilidades y/o del efecto en los errores de medida, se utiliza el efecto pepita puro, que combinado con el variograma (o variogramas) de la estructura espacial a la escala del muestreo, sirve para explicar la variabilidad espacial. Dicho modelo conceptual se conoce bajo el nombre de superposición de estructuras anidadas. El variograma para la precipitación media en 24 horas. Se trata de un modelo exponencial con alcance 20 km y en el que el efecto pepita supone un 30 % de la variabilidad total. El mismo análisis para las precipitaciones máximas en 24 horas el alcance es menor (alrededor de 15-16 kilómetros) y el efecto pepita mayor (50 % de la variabilidad total, que es 10 veces la de la lluvia media), lo cual coincide con lo que debía esperarse: las precipitaciones máximas poseen un carácter repentino menos regular que la lluvia media. Las temperaturas muestran un modelo espacial de varianza similar al de las precipitaciones, anisótropo, de regularidad media, con un mayor alcance en dirección E-W (en torno a 40 km). El régimen de temperatura la región está sumamente influenciado por la distancia al mar, que actúa como colchón térmico), por la altitud, existiendo una correlación negativa en distancias del orden de 30 km, y por el régimen predominante de vientos.

Debe entenderse que dichos modelos dependen del grado de muestreo del fenómeno analizado. Si se poseyese una red de muestreo más fina, incorporando datos provenientes de zonas montañosas, es previsible que dicho modelo se altere, apareciendo variabilidades a más pequeña escala propias de los microclimas existentes en la región.

**Predicción espacial y conclusiones:** Los métodos de predicción espacial fueron basados en Kriging ordinario en donde se determinaron 4 mapas para la variable precipitación, lluvia media en año seco, lluvia media en año húmedo, lluvia máxima año seco y lluvia máxima año húmedo. Y para la temperatura media se muestra una evolución mes a mes en donde se puede observar un mapa estratificado Norte-Sur en los meses de otoño invierno, hasta llegar a una inversión térmica en los meses de verano donde las menores temperaturas se alcanzan en el litoral y las mayores temperaturas en las zonas centrales. Además del análisis que se realizó a los correspondientes mapas obtenidos concluyeron que la suavidad de temperaturas en Asturias se debe entre otros factores al régimen de vientos y a la proximidad al mar, que los mapas de incertidumbre proporcionados por la técnica de krigeado de variables indicadoras pueden ser utilizados para diseñar una red óptima de muestreo para la monitorización y análisis en tiempo real de dichas variables,

ante las sospechas de cambio climático.

### 3. Kriging-based approach to predict missing air temperature data Shtiliyana y col. (2017)

Los registros de datos meteorológicos son necesarios para evaluar escenarios en estudios de recursos naturales. En muchos de estos estudios, las observaciones del clima en la superficie son los datos fundamentales para los modelos de simulación, para los cuales la ausencia de datos de una estación en particular en un momento dado puede introducir sesgos en las respuestas o generar tendencias falsas. Los datos faltantes son un problema común en los conjuntos de datos de observación meteorológica. Los registros de historial de estaciones a menudo están incompletos porque los instrumentos pueden romperse o estar en mal funcionamiento, o la transmisión de datos puede ser interrumpida. Algunos fenómenos meteorológicos como la precipitación, pueden provocar fallas en el registro de los datos.

**Entrada de datos faltantes basada en Kriging:** Aquí, utilizaron el Kriging ordinario (KO), que es de computo práctico y más fácil de implementar que otras variantes de Kriging. La semivarianza experimental se calcula con base en el vector de retraso que representa la separación entre dos ubicaciones espaciales, el vector de coordenadas de muestra espacial y el número de pares de muestras separados por el retraso. La fórmula se aplica sobre los pares de puntos. Usando esta fórmula, se produce un semivariograma para describir la autocorrelación espacial de la variable. En comparación con otros métodos, las ventajas clave de los semivariogramas son: el método se basa en datos sin procesar y no requiere indicadores precalculados; Se pueden detectar cambios tanto lineales como no lineales; el cambio identificado se relaciona con la dinámica esperada, que representa toda la serie considerada, no solo el comienzo y el final de una serie dada; y se puede analizar la dinámica de la serie, Los cambios en los parámetros del semivariograma se relacionan con cambios en diferentes aspectos de la serie. En general, el semi-variograma se ajusta utilizando modelos de variogramas autorizados como lineales, esféricos, exponenciales, gaussianos, etc. Un semivariograma se describe por su umbral, rango práctico y el efecto de pepita. En este estudio, las interpolaciones se basan en los gradientes entre los puntos de los vecinos más cercanos en la serie de tiempo. Los valores que faltan se completan en función de todas las características inherentes al conjunto histórico. Se utilizan dentro de un método de interpolación geoestadística para calcular el mejor predictor lineal no sesgado en el contexto geoestadístico al acomodar los registros mediante un proceso basado en la memoria. Usando el concepto basado en la memoria, cualquier valor de temperatura en una serie de tiempo se ve afectado por sus vecinos anteriores y, a su vez, afecta sus valores del próximo vecino.

**Evaluación del método:** Para ilustrar el proceso de generación de brechas, se calculó cuántas veces se eligió una hora dada para ser una brecha con brechas de una longitud dada. Esta prueba, en la que se pueden superponer las brechas, sirvió para verificar si se manejaron los casos de brechas anidadas o encadenadas y cómo se colocaron las brechas. El rendimiento del método de imputación basado en Kriging se evaluó en función de los conjuntos de datos de temperatura de cinco sitios en Europa y un sitio situado en el Océano Índico. Para cuatro de los sitios europeos, los datos de temperatura por hora estaban disponibles, mientras que para Rothamsted (Reino Unido) el conjunto de datos de varios años incluía temperaturas máximas y mínimas diarias. Los valores de temperatura máxima y mínima diarios se extrajeron de todos los conjuntos de datos por hora.

**Análisis de semivariogramas:** El modelo de variograma esférico se utilizó para calcular los parámetros de Kriging porque la relación con la distancia en la escala de tiempo varió demasiado con otros modelos. La tendencia de aumento monótono de los semivariogramas empíricos no se manifiesta antes de que se alcance el valor del umbral, lo que de hecho indicaría la presencia de diferentes fenómenos y la necesidad de calcular otros semivariogramas. Por lo tanto, tiene una consecuencia limitada para la predicción local de valores. Esto es conveniente desde el punto de vista de la predicción, ya que las predicciones típicas de Kriging dependen principalmente de los valores de semivariogramas en distancias cortas. Debido a este hecho, de acuerdo con estudios anteriores, no se intentó desvirtuar los datos utilizando una función apropiada.

**Variaciones del conjunto de datos completados:** Después de obtener los nuevos valores para las ranuras faltantes, comparamos los conjuntos de datos originales y los rellenados. Se obtuvo un valor promedio predicho de la temperatura gracias a 200 simulaciones. Todas las imputaciones de los datos faltantes por hora se hicieron con el Kriging basado en ocho vecinos que rodean el valor faltante. Se presentan algunos resultados ilustrativos, que muestran que la tendencia general de las distribuciones se respetó en todos los sitios. Los datos reemplazados introdujeron algunos extremos: cuando el extremo se sitúa al comienzo del conjunto de datos, esto podría considerarse como un problema de frontera, debido al número limitado de vecinos para el Kriging. Cuando el extremo está en un lugar diferente, pueden surgir problemas si el primer vecino más cercano del valor faltante se coloca lejos. Si el vecino es un extremo en sí mismo, esto podría ser un problema para la generación del nuevo valor porque el método no tiene en cuenta esta situación y el valor predicho también podría ser extremo. Se obtuvo una buena adaptación con brechas de cuatro horas en Oensingen en julio-octubre con solo pequeñas diferencias entre los valores predichos y los reales.

## 4. Marco teórico

### 4.1. Geoestadística

El termino fue acuñado por Matheron para describir su trabajo en el cual realizo con respecto a problemas de predicción espacial sobre datos geoestadísticos. Estas ideas se desarrollaron extensamente e independientemente de la corriente principal de la estadística espacial. Cressie (2019) considera que la geoestadística es una de las tres principales ramas de estadística espacial, sin embargo se ve como las herramientas que se usan en la practica de la geoestadística se evidencian su origen independiente de la estadística espacial Gil (2013).

### 4.2. Variable regionalizada

Es toda aquella variable la cual como su nombre lo indica se encuentra en cierta región del espacio por tal motivo esta georreferenciada, matemáticamente se define como una variable aleatoria en un punto del espacio. Matheron (1963) acuño el termino a principios de la década de 1960. Una de sus principales características se basa en que esta toma un determinado valor en cada punto espacial de un dominio definido.

### 4.3. Momentos de la variable regionalizada

1. Valor esperado de una variable regionalizada

$$E(Z(s_i)) = \mu \quad (4.1)$$

2. Varianza de una variable regionalizada

$$\begin{aligned} V(Z(s_i)) &= E(Z(s_i) - E(Z(s_i)))^2 \\ V(Z(s_i)) &= E(Z(s_i) - \mu)^2 \\ V(Z(s_i)) &= \sigma^2 \end{aligned} \quad (4.2)$$

### 3. Covarianza de una variable regionalizada

$$\begin{aligned}C(Z(s_i), Z(s_j)) &= E((Z(s_i) - E(Z(s_i)))(Z(s_j) - E(Z(s_j)))) \\C(Z(s_i), Z(s_j)) &= E((Z(s_i) - \mu)(Z(s_j) - \mu)) \\C(Z(s_i), Z(s_j)) &= C(h) = C_{ij}^h\end{aligned}\tag{4.3}$$

Donde  $C_{ij}^h$  es definido como covariograma

## 4.4. Anamorfosis Gaussiana

Para estimar los variogramas es necesario que la variable que se estudia siga una distribución normal en caso tal de que esto no se cumpla puede ser necesaria una transformación llamada anamorfosis, en la cual gráficamente consiste en deformar el histograma para obtener que se vea una campana de Gauss en su histograma, lo cual consiste en asociar a cada valor de de la variable original el valor gaussiano y tener la misma frecuencia acumulada Emery (2013).

## 4.5. Funciones de dependencia espacial experimentales

### Covariograma y correlograma

La función de covarianza muestral entre parejas de observaciones a una distancia  $h$  se calcula empleando la formula clásica de la covarianza muestral por:

$$\begin{aligned}C(h) &= COV(Z(x+h), Z(x)) = \sum_{i=1}^n \frac{(Z(x+h) - m)(Z(x) - m)}{n} \\C(h) &= \frac{\sum_{i=1}^n (Z(x+h))(Z(x))}{n} - m^2\end{aligned}\tag{4.4}$$

Donde  $m$  es el valor promedio en todo punto de la región de estudio y  $n$  es el numero de parejas de puntos a una distancia.

Asumiendo que el fenómeno es estacionario y estimando la varianza de la variable regionalizada a través de la varianza muestral se tiene que el correlograma muestral esta dado por: Henao (2002)

$$r(h) = \frac{COV(Z(x+h), Z(x))}{S_{x+h} - S_x} = \frac{C(h)}{S_X^2} = \frac{C(h)}{c(0)}\tag{4.5}$$

## Variograma y semivariograma

Según Matheron (1963) el variograma es una curva que representa el grado de continuidad de un fenómeno. La definición de esta es dada por medio de un proceso intrínsecamente estacionario en forma que:

$$2\gamma(h) = \text{Var}(Z(S_i) - Z(S_j)) \quad (4.6)$$

Donde  $S_i$  y  $S_j \in D$  y  $h = S_i - S_j$  es el vector de distancias. El variograma además, debe cumplir con dos condiciones, la primera es que debe ser una función condicionalmente definida negativa y tener un ritmo de crecimiento inferior a  $h^2$ . C. E. Melo (2012) La mitad del variograma se conoce como la función de semivarianza y caracteriza las propiedades de dependencia espacial del proceso. Dada una realización del fenómeno, la función de semivarianza es estimada, por el método de momentos, a través del semivariograma experimental Henao (2002) que se explicara a continuación.

$$\bar{\gamma}(h) = \frac{\sum (Z(x+h) - Z(x))^2}{2n} \quad (4.7)$$

C. Melo (2015)

### 4.6. Estimación del variograma y covariograma

#### Estimador clásico

La estimación del variograma mas sencillo es el obtenido mediante el método de estimador de momentos si los puntos de muestreo estuviesen localizados sobre una rejilla regular, el estimador vendrá definido por :

$$2\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{(s_i, s_j) \in N(h)} (Z(s_i) - Z(s_j))^2 \quad (4.8)$$

Para la covarianza, el estimador obtenido por el método de los momentos sería.

$$\hat{C}(h) = |N(h)| \sum_{(s_i, s_j) \in N(h)} (Z(s_i) - \hat{Z})(Z(s_j) - \hat{Z}) \quad (4.9)$$

C. Melo (2015)



## Estimador robusto

Aunque el método de estimación clásico presenta la ventaja de su facilidad matemática, presenta algunos inconvenientes prácticos como que no es robusto frente a valores extremos, Cressie (2019) presentan una variación de mayor robustez como estimador insesgado del variograma que se define de la siguiente forma C. Melo (2015).

$$2\hat{\gamma}(h) = \frac{1}{|N(h)|(0,457 + 0,494/|N(h)|)} \sum_{(s_i, s_j) \in N(h)} |Z(s_i) - Z(s_j)|^{1/2} \quad (4.10)$$

e

## 4.7. Modelos de variograma

A continuación se presentan algunos modelos de variogramas con sus respectivas formas funcionales.

### Modelo pepita puro

Se trata de un modelo aleatorio, es decir, ruido blanco, donde no hay presencia de correlación entre los valores.

$$\gamma(h) = \begin{cases} 0 & \text{si } h = 0 \\ C_0 & \text{si } h > 0 \end{cases} \quad (4.11)$$

Bajo el supuesto de estacionariedad cualquiera de las tres funciones de dependencia espacial mencionadas, es decir semivariograma, covariograma o correlograma, puede ser usada en la determinación de la relación espacial entre los datos. Sin embargo como se puede observar en las fórmulas, la única que no requiere hacer estimación de parámetros es la función de semivarianza. Por esta razón es que en la practica se emplea el semivariograma y no las otras dos funciones Henao (2002).

## Modelo esférico

Es un modelo en el cual lo que se observa en general se observa una forma en la cual presenta un crecimiento hasta cierta distancia y después una estabilización donde se aplanan la curva en un determinado punto cuya abscisa es de  $2a\sqrt{3}$

$$\gamma(h) = \begin{cases} 0 & \text{si } h = 0 \\ C_0 + c_s\left(\frac{3}{2}\left(\frac{h}{a_s}\right) - \frac{1}{2}\left(\frac{h}{a_s}\right)^3\right) & \text{si } 0 \leq h \leq a_s \\ C_0 + C_s & \text{si } h > a_s \end{cases} \quad (4.12)$$

## Modelo exponencial

Se observa un modelo exponencial en el cual se tiene un crecimiento acelerado inicial disminuyendo y estabilizándose en cierto punto, su rango efectivo se encuentra en  $3a$  que es donde alcanza el 95 % de su límite.

$$\gamma(h) = \begin{cases} C_0 + C_s(1 - \exp(-3h/a_e)) & \text{si } h > 0 \\ 0 & \text{si } h = 0 \end{cases} \quad (4.13)$$

## Modelo Gaussiano

Este representa un fenómeno continuo donde se evidencia la ocurrencia de inestabilidades numéricas en el caso donde no se tiene en cuenta o no se hace uso del efecto pepita Armstrong (1950) su rango efectivo es de  $\sqrt{3a}$

$$\gamma(h) = \begin{cases} C_0 + C_s(1 - \exp(-3h^2/a_g^2)) & \text{si } h > 0 \\ 0 & \text{si } h = 0 \end{cases} \quad (4.14)$$

### Modelo lineal con meseta

$$\gamma(h) = \begin{cases} 0 & \text{si } h = 0 \\ C_0 + c_S(h/a_l) & \text{si } 0 \leq h \leq a_l \\ C_0 + C_s & \text{si } h > a_l \end{cases} \quad (4.15)$$

### Modelo efecto agujero

$$\gamma(h) = \begin{cases} C_0 + C_s(1 - \frac{\sin(h)}{h}) & \text{si } h > 0 \\ 0 & \text{si } h = 0 \end{cases} \quad (4.16)$$

### Modelo Circular

$$\gamma(h) = \begin{cases} 0 & \text{si } h = 0 \\ C_0 + C_S(1 - \frac{2}{\pi} \cos^{-1}(\frac{h}{a_c}) + \frac{2h}{\pi a_c} \sqrt{(1 - h/a_c)^2}) & \text{si } 0 \leq h \leq a_c \\ C_0 + C_s & \text{si } h > a_c \end{cases} \quad (4.17)$$

## 4.8. Estimación de los parámetros del variograma

En ocasiones las predicciones generadas a través de semivariogramas estimados tienen varianzas negativas, esto es debido a que estos carecen de ser definidos semipositivos, una de las formas más comunes es reemplazar el semivariograma empírico con un modelo parámetro definido como  $\gamma(h, v)$  Este se debe aproximar a la dependencia espacial encontrada en el mismo variograma empírico y debe cumplir con la condición de ser semidefinido positivo. El objetivo de este procedimiento es elegir un mejor modelo acorde a las observaciones para luego utilizarlo en el proceso de predicciónC. Melo (2015). Los principales métodos de estimación son

- La estimación por mínimos cuadrados ordinarios consiste en minimizar el valor de  $\hat{v}$

$$\sum_{j=1}^n (\hat{\gamma}(h_j) - \gamma(h_j, v))^2 = [\hat{\gamma} - \gamma(v)]' [\hat{\gamma} - \gamma(v)] \quad (4.18)$$

Este presenta problemas como que en este caso las estimaciones estén correlacionadas y tienen varianzas diferentes.

- Mínimos cuadrados generalizados en el cual consiste en minimizar

$$|\hat{\gamma} - \gamma(v)|' V(v)^{-1} |\hat{\gamma} - \gamma(v)| \quad (4.19)$$

Donde  $V(v)$  es la matriz de varianzas-covarianzas de  $\hat{\gamma}$  que depende del valor de  $v$  desconocido y cuyos elementos pueden ser difíciles de estimar

- Otra estimación es la de mínimos cuadrados ponderados la que minimiza

$$\sum_{j=1}^n w_j [\hat{\gamma}(h_j) - \gamma(h_j, v)]^2 = [\hat{\gamma} - \gamma(v)]' W(v)^{-1} [\hat{\gamma} - \gamma(v)] \quad (4.20)$$

Donde  $W(v)$  es la matriz diagonal de varianzas de  $\hat{\gamma}$  y  $w_j$  son los pesos, esta ultima estimación tiene mayor uso ya que presenta mayor facilidad en su implementación C. Melo (2015).

## 4.9. Anisotropía

La anisotropía se manifiesta cuando los datos no solo dependen de la ubicación con respecto a los demás sino también las direcciones en las que se encuentren para ello según Emery (2013) Las anisotropías se presentan cuando el variograma difiere según las direcciones del espacio. Según Armstrong (1950) Se pueden distinguir dos tipos diferentes de anisotropía: anisotropía geométrica y zonal

### Anisotropía geométrica

Si se encuentran variogramas que tienen el mismo umbral en todas las direcciones, aunque sus rangos sean diferente, hay presencia de anisotropía geométrica. Se puede realizar un diagrama en que muestre el rango o la pendiente en función de la dirección, si la curva es una elipse, entonces se dice que hay anisotropía geométrica, Para este caso se transforma el sistema de coordenadas del elipse en un círculo para eliminar la anisotropía Emery (2013).

## Anisotropía zonal

Existen variogramas en los cuales no se puede corregir con una transformación de coordenadas, en especial cuando la meseta es variable según la dirección del espacio. En términos matemáticos es un modelo en el cual la función aleatoria no depende de una o varias coordenadas. Una forma equivalente de definir la anisotropía zonal consiste en verla como una anisotropía geométrica donde el eje mayor de la elipse (elipsoide) de alcances se vuelve infinitamente grande Emery (2013).

### 4.10. IDW (Distancia inversa ponderada)

Es un método de interpolación el cual se basa en la distancia inversa y calcula un promedio ponderado:

$$\hat{Z}(S_0) = \sum_{i=1}^n \lambda_i Z(S_i) \quad (4.21)$$

Donde  $\hat{Z}(S_0)$  es el valor de predicción en la posición  $S_0$ ,  $n$  el número de observaciones que rodean el punto,  $\lambda_i$  los pesos de cada una de las mediciones realizadas y  $Z(S_i)$  el valor de la variable en la posición determinada. C. Melo (2015).

Dichos pesos se determinan mediante:

$$\lambda_i = \frac{d_{i0}^{-p}}{\sum_{i=1}^n d_{i0}^{-p}} \quad (4.22)$$

donde  $\sum_{i=1}^n \lambda_i = 1$  el peso se encuentra controlado por un factor  $p$  con cada incremento de distancia determinada por la posición de predicción y cada una de las posiciones medidas, dicho parámetro  $p$  es influyente en el peso de los valores medidos en el punto a predecir, por lo cual es necesario determinar un valor óptimo para este en el cual se determina minimizando la raíz cuadrática media del error de predicción (RMSPE).

### 4.11. Funciones de base radial

Las funciones de base radial son interpoladores que no necesitan de muchos cálculos y hacen parte de los métodos determinísticos al igual que la IDW. La superficie debe pasar por cada valor de muestra medido y por esto se consideran exactas sus predicciones.

Las RBFs se utilizan para calcular superficies lisas a partir de una gran cantidad de puntos de datos. Las

funciones producen buenos resultados para superficies que varían suavemente, como la elevación. Las técnicas son inapropiadas cuando hay grandes cambios en el valor de la superficie dentro de una corta distancia horizontal y cuando se sospecha que los datos de la muestra son propensos a errores o incertidumbre Johnston (2001).

En la siguiente tabla se presentan algunas formas funcionales de las funciones de base radial

Función	Forma funcional
Multicuadrática	$\Phi_{ij} = \sqrt{\eta^2 + \delta_{ij}^2}, \eta \neq 0(4.23)$
Inversa Multicuadrática	$\Phi_{ij} = \frac{1}{\sqrt{\eta^2 + \delta_{ij}^2}}, \eta \neq 0(4.24)$
Spline con tensión	$\Phi_{ij} = \begin{cases} \ln(\eta * \frac{\delta}{2}) + K_0(\eta\delta) + C_E & si \quad \delta > 0 \\ 0 & si \quad \delta = 0 \end{cases} \quad (4.25)$
Spline completamente regularizada	$\Phi_{ij} = \begin{cases} \ln(\eta * \frac{\delta}{2})^2 + E_1(\eta * \delta/2)^2 + C_E & si \quad \delta > 0 \\ 0 & si \quad \delta = 0 \end{cases} \quad (4.26)$
Spline capa delgada	$\Phi_{ij} = \begin{cases} \ln(\eta * \frac{\delta}{2}) + K_0(\eta\delta) + C_E & si \quad \delta > 0, \eta > 0 \\ 0 & si \quad \delta = 0 \end{cases} \quad (4.27)$
Spline completamente regularizada	$\Phi_{ij} = \begin{cases} (\eta\delta)^2 \log(\eta\delta) & si \quad \delta > 0, \eta > 0 \\ 0 & si \quad \delta = 0 \end{cases} \quad (4.28)$
Exponencial	$\Phi_{ij} = e^{\eta\delta}(4.29)$
Gaussiana	$\Phi_{ij} = e^{\eta\delta^2}(4.30)$

Cuadro 4.1: Funciones de base radial

## 4.12. Kriging

Kriging ha sido utilizado durante muchas décadas como sinónimo de interpolación geoestadística. En Matheron (1963) se describe una aplicación de gran importancia que es proporcionada por un procedimiento geoestadístico llamado "Kriging". Consiste en estimar el grado de un panel calculando el promedio ponderado de las muestras disponibles, algunas de las cuales se encuentran ubicadas sobre un espacio S

## Kriging Simple

En este proceso se considera una media del proceso  $Z(s)$  conocida y constante, pero existen casos donde no se conoce y de acuerdo a la distribución espacial en las muestras se hace necesario realizar un proceso de declustering para llegar a la aproximación de esta media, la cual se puede considerar:

$$Z(s) = \mu(s) + \varepsilon(s) \quad (4.31)$$

## Kriging Ordinario

En esta se considera la media del proceso  $Z(s)$  constante pero desconocida por lo que usa un promedio local para el vecindario definido a través de un modelo de semivariograma sin considerar tendencia, la predicción para un punto de coordenadas esta definida por:

$$\hat{Z}(S_0) = \sum_{i=1}^n \lambda_i Z(S_i) \quad (4.32)$$

## Kriging Universal

Este kriging corresponde a un proceso o estacionario, por lo que se asume la existencia de tendencia, asociada en general con coordenadas espaciales, no obstante, se presentan algunos casos en donde la tendencia está en función de variables complementarias o covariables. C. Melo (2015)

Sea  $Z(s) = \mu(s) + \epsilon(s)$  un proceso aleatorio Gaussiano donde  $\mu(s)$  describe la tendencia en el espacio de la variable regionalizada Z. Para un punto ubicado en  $s_i$  se tiene

$$Z(s) = \mu(s) + \epsilon(s) \quad (4.33)$$

Aquí

$$E(\varepsilon(s_i)) = 0 \quad y \quad V(\varepsilon(s_i)) = \sigma^2 \quad (4.34)$$

Dado que hay tendencia  $Z(s) = \mu(s) = x(s)' \beta$  con  $x(s)' = (1, x_i, y_i)$  y  $\beta = (\beta_0, \beta_1, \beta_2)$ , para una tendencia de primer orden que solo este en función de las coordenadas espaciales x e y. Para generar la predicción en un punto coordenado  $S_0$  se hace uso de la siguiente expresión

$$\hat{Z}_0 = \hat{Z}(S_0) = \lambda' Z \quad (4.35)$$

Donde el predictor debe cumplir con  $\lambda'1 = 1$  para que este sea insesgado C. Melo (2015)



## 5. Área de estudio y base de datos

La base de datos corresponde a datos de Temperatura, altura, Humedad y distancia de las estaciones meteorológicas ubicadas en Croacia con sus respectivas coordenadas Norte y Este para el año 2008, Croacia es un país que hace parte de la Unión Europea y se encuentra entre Europa central, Europa meridional y el Mar Mediterráneo. La base de datos contiene lo siguiente:

Variable	Descripción
Tmt	Temperatura media terrestre registrada por las estaciones meteorológicas para el día 26 de febrero de 2008
N	Coordenada Norte estación meteorológica
E	Coordenada Este estación meteorológica
HRdem	Altura de la estación el modelo digital de elevación
HRtwi	Humedad en la ubicación de la estación
HRdsea	Distancia entre la costa y la estación meteorológica

Cuadro 5.1: Descripción base de datos

Se trabajaran los datos del día 3 de julio de 2008 en la cual se tiene registrada la temperatura media terrestre de 154 de 158 estaciones, ya que 4 presentan ausencia de este dato para la fecha. La temperatura media terrestre varía entre  $20^{\circ}C$  y  $37^{\circ}C$  estas altas temperaturas corresponden a que la fecha es del verano del respectivo año. Las siguientes son unos de los estadísticos de la variable de temperatura media terrestre:

Mínimo	Media	Máximo	Vacíos
20.51	29.2487	37.09	4

Cuadro 5.2: Estadística temperatura 3 de julio de 2008 de Croacia

## 6. Análisis exploratorio y estructural

### 6.1. Espacialización de los datos

Se realiza la espacialización de los datos graficando sus coordenadas Este y Norte y simbolizándolo por la temperatura, en ella se puede observar que las grandes temperaturas se concentran mas hacia la costa sur con el mar mediterráneo pero también se ven puntos con temperaturas altas en la zona Nororiental y Norte del país

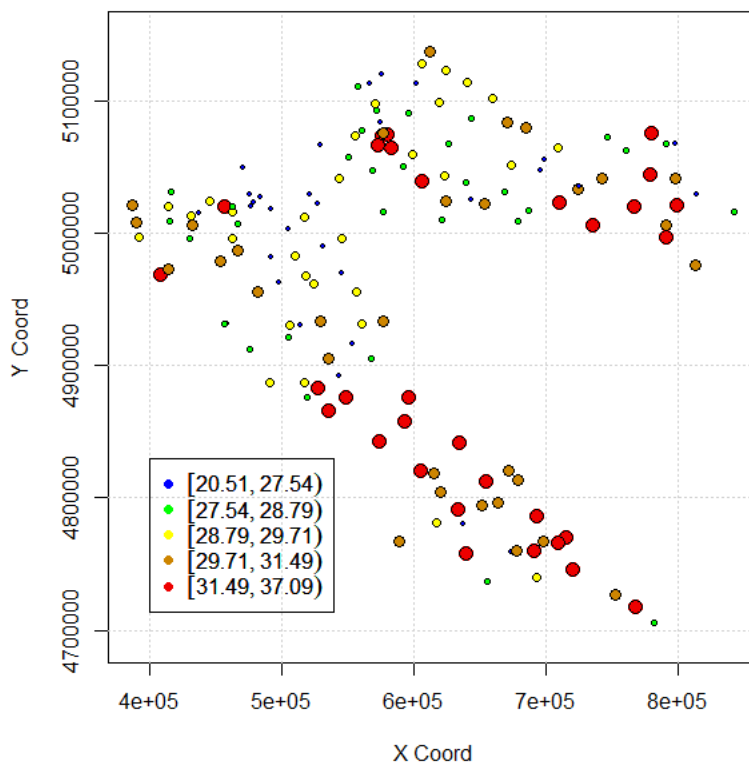


Figura 6.1: Mapa de la temperatura media terrestre registrada para el día 3 de julio de 2020

Adicionalmente se realiza un gráfico en 3 dimensiones donde en el eje z se utiliza para representar la Temperatura registrada para este punto

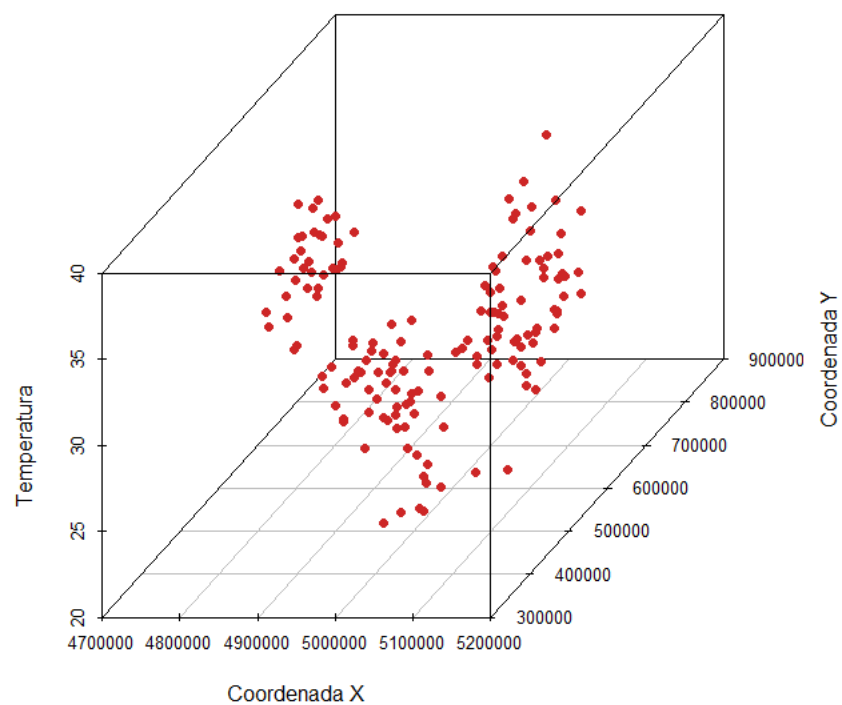


Figura 6.2: Distribución espacial variable de temperatura para el día 3 de julio de 2008

De las anteriores gráficas se podría observar cierto grado de tendencia hacia la costa por lo que podríamos evidenciar al analizarla cierta tendencia pero probablemente no muy significativa.

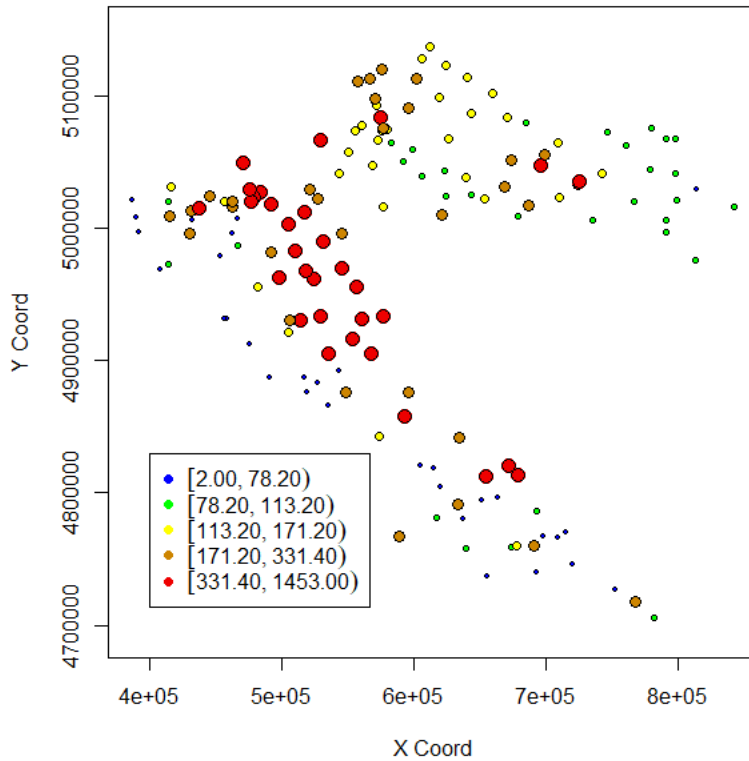


Figura 6.3: Mapa de altura de las estaciones

Observando un gráfico con las alturas de las estaciones meteorológicas podemos observar la clara relación donde las mayores temperaturas se encuentran en las zonas de menor altura.

## 6.2. Estadísticas descriptivas

A continuación se presenta una tabla con las estadísticas descriptivas de cada una de las variables de la base de datos

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Hrdem	154	234.532	259.248	2	84.8	278.2	1,453
Hrdsea	154	94.262	91.041	0.000	0.400	174.925	252.300
Hrtwi	154	17.691	1.875	12.900	16.425	19.000	21.400
Tmt	154	29.249	2.990	20.510	27.910	31.025	37.090
N	154	4,971,725.000	109,266.100	4,705,463.000	4,907,309.000	5,047,359.000	5,137,646.000
E	154	594,304.100	108,316.200	387,331.200	517,075.200	673,193.800	842,442.800

Cuadro 6.1: Estadísticas descriptivas todas las variables de la base de datos

Las variables de altura, humedad y distancia entre la estación y la costa y sus estadísticas permiten inferir en que no siguen una distribución normal y que sus valores de media y mediana pueden ser influenciado por valores extremos los cuales se ven reflejados por ejemplo en que los puntos del lado nororiental son mas elevados que los demás, igual que la altura y humedad que pueden variar hasta valores atípicos a lo largo del territorio.

### 6.3. Análisis de normalidad

Para mirar si la temperatura media terrestre registrada para el 3 de julio es normal se realiza un análisis por medios gráficos por los cuales se observa que no son del todo los normales y al realizar los análisis de normalidad con la prueba de Shapiro-Wilk y de Kolmogorov-Smirnov se evidencia que los datos no siguen una distribución normal, ya que en el caso de la prueba de Shapiro-Wilk se rechaza la hipótesis nula de que la variable sigue una distribución normal.

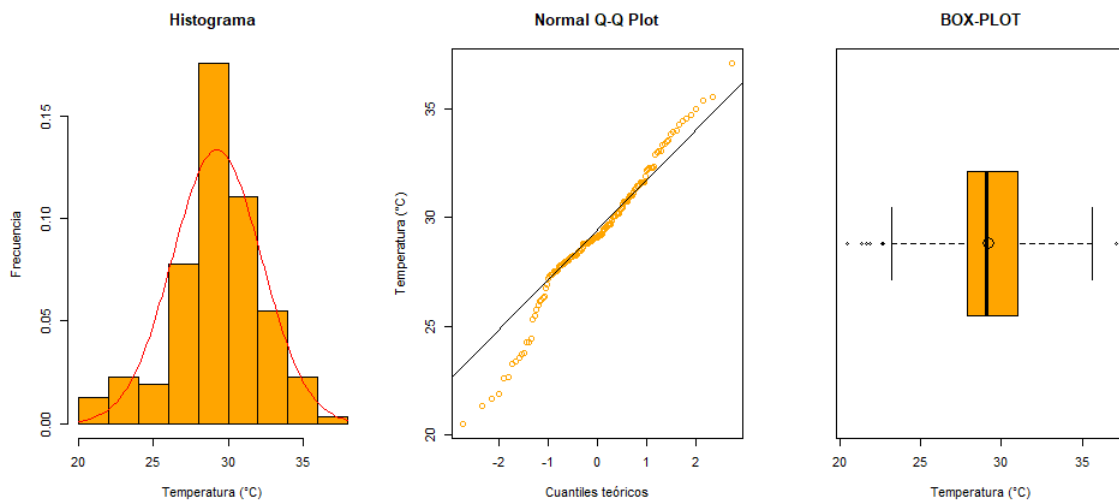
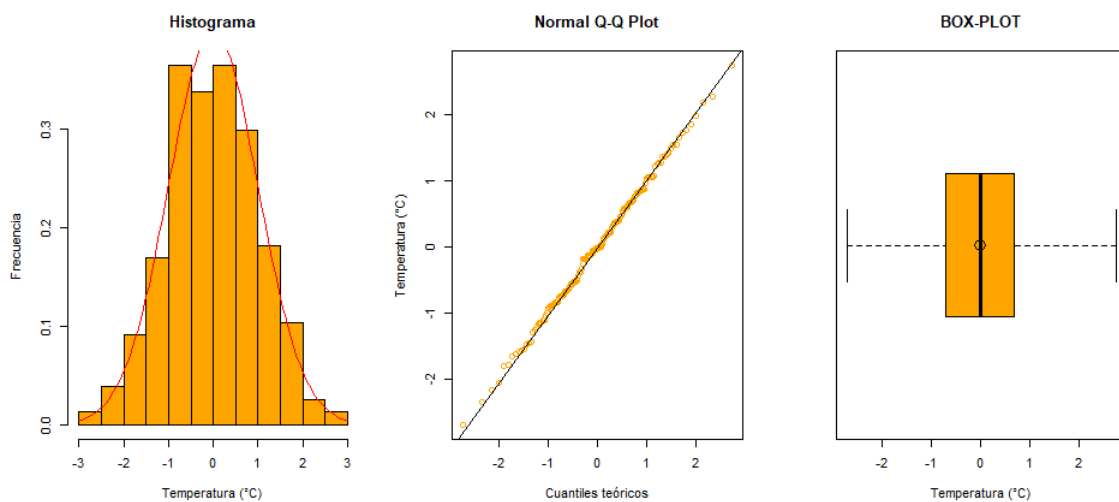


Figura 6.4: Histograma, Q-Q plot y Box-Plot de la variable de temperatura

Test	Estadístico	p-value
Shapiro-Wilk	0.97504	0.00673
Kolmogorov-Smirnov	0.096803	0.1116

Cuadro 6.2: Estadísticos de normalidad de la variable temperatura

Para corregir este problema se transforma la variable con Anamorfosis Gaussiana explicada en capítulos anteriores y a continuación se observan los resultados en donde ya la variable transformada sigue una distribución gráfica y esta se ve en los métodos gráficos y que en las dos pruebas estadísticas se acepta la hipótesis de que se sigue una distribución normal.



captionHistograma, Q-Q plot y Box-Plot de la variable de temperatura transformada

Test	Estadístico	p-value
Shapiro-Wilk	0.99833	0.9999
Kolmogorov-Smirnov	0.044157	0.9248

Cuadro 6.3: Estadísticos de normalidad de la variable temperatura transformada

## 6.4. Análisis de tendencia

Para analizar la tendencia se genera un modelo lineal para poder establecer cual de las variables HRdem, HRdsea, HRTwi y sus coordenadas Estes E y Nortes N de las estaciones explican la variable de temperatura.

	Modelo 1	Modelo 2
Hrdem	−0.001***	−0.002***
Hrdsea	−0.001	
Hrtwi	0.074*	0.071*
N	−0.00000	−0.00000***
E	0.00000	
Constant	6.123	11.809***
Observations	154	154
R <sup>2</sup>	0.347	0.343
Adjusted R <sup>2</sup>	0.325	0.330
Residual Std. Error	0.821 (df = 148)	0.818 (df = 150)
F Statistic	15.732*** (df = 5; 148)	26.138*** (df = 3; 150)

Significancia: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Cuadro 6.4: Modelos de análisis de tendencia

En el Modelo 1 se corre un modelo de todas las variables en donde podemos ver que solo la variable de HRTwi resulta ser significativa y un  $R^2$  ajustado de 0.325, en el Modelo 2 se tiene en cuenta el criterio de AIC con el método de Stepwise en donde podemos conseguir un mejor modelo, con las variables de la coordenada N significativa y negativa por lo que podemos inferir que hacia el sur se encuentran valores mayores de temperatura algo ya observado en la especialización de los datos, pero sin embargo con un  $R^2$  ajustado bajo de solo 0.33 por lo tanto no hay tendencia.

Test	Estadístico	p-value
Shapiro-Wilk	0.99388	0.7638
Kolmogorov-Smirnov	0.045933	0.9013

Cuadro 6.5: Estadísticos normalidad de los residuos del modelo 2

Se puede comprobar a partir de estadísticos que los residuos del mejor modelo (Modelo 2) siguen una distribución normal, también se realiza un gráfico de la superficie de tendencia en el cual se observa cierto grado de tendencia.

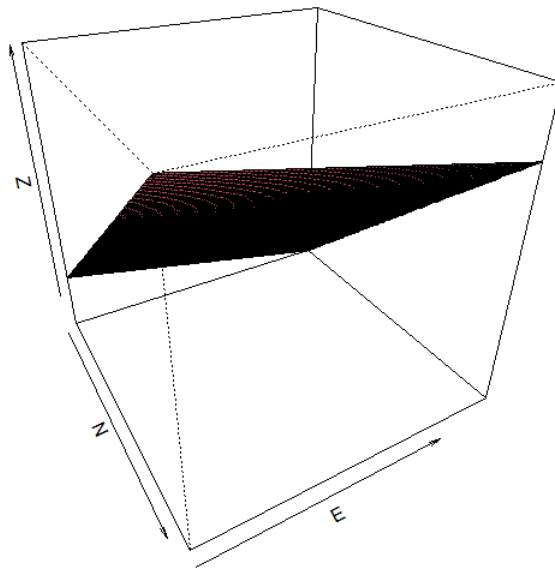


Figura 6.5: Superficie de tendencia de la variable temperatura

Al obtenerse para el mejor modelo un  $R^2$  ajustado bajo de solo 0.33 se decidió que no hay tendencia por lo tanto se realizan los modelos de Kriging Ordinario y Kriging Simple, KO y KS en la sección de hallar el modelo del semivariograma y no realizar Kriging Universal en los métodos geoestadísticos.



## 6.5. Análisis de anisotropía de los residuos

La anisotropía implica que la dependencia espacial es una función de solo la distancia y la dirección de la separación espacial entre las ubicaciones, se puede evaluar esta por métodos gráficos, en el siguiente gráfico se observa como al parecer dependiendo de la dirección los semivariogramas parece que siguen el mismo comportamiento, sin embargo este criterio mediante gráficos es difíciles de evaluar por ende también se evalúa esta mediante la función `estimateAnisotropy` de la librería `intamap`.

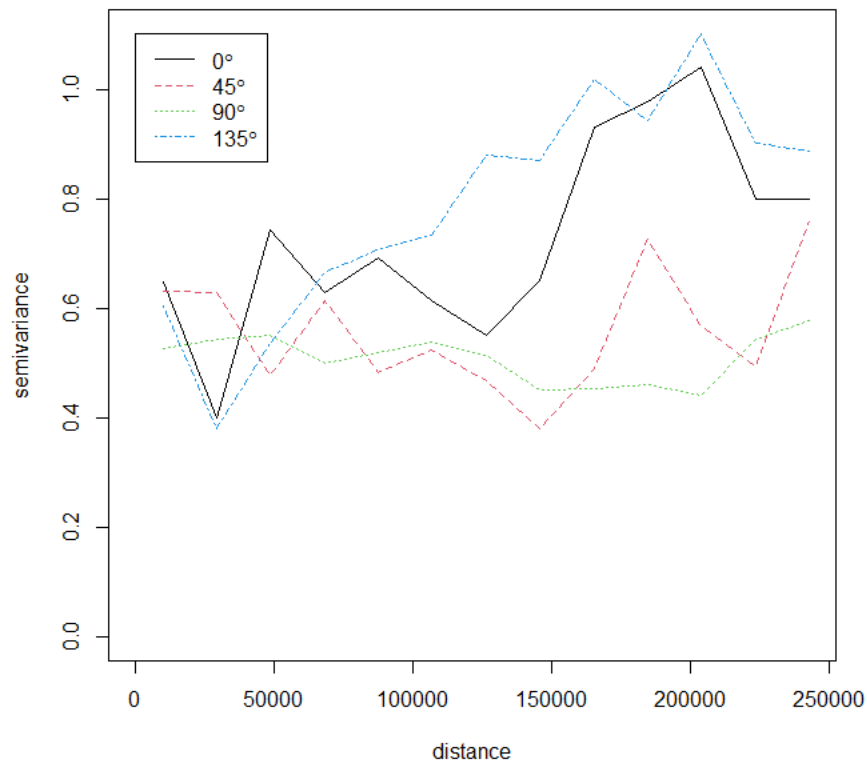


Figura 6.6: Semivariogramas direccionales.

En los resultados de la estimación de la anisotropía se observan los diferentes valores de dirección y escala de la anisotropía y el valor de `'doRotation'` en la cual mediante un valor booleano si la anisotropía estimada es estadísticamente significativa, para ello realiza el test de isotropía usando una aproximación del 95 % de confiabilidad por ende al obtener como resultado `'FALSE'` no es necesario realizar ninguna acción Skoien (2018) es decir que la anisotropía no es significativa y no es necesario rotar ni escalar los datos para poder realizar la selección de los semivariogramas y las interpolaciones.

Estimación Anisotropía	
R	1.053742
Theta	79.24073
Q	[4.82e-10,4.38e-10,-8.83e-12]
doRotation	FALSE

Cuadro 6.6: Resultados estimación de anisotropía

## 6.6. Modelamiento del semivariograma

Se realiza primero el calculo de los semivariogramas experimentales y para esto tomamos la distancia para interpolar como la mitad de la distancia maxima, que para el caso el resultado fue de 252609.5 metros.

$$Distancia = \frac{DistanciaMaxima}{2} = 252609,5m \quad (6.1)$$

Se debe encontrar el mejor modelo para utilizar en cada método de interpolación para ello se realiza lo siguiente para cada metodo.

## Semivariograma para Kriging Ordinario(KO)

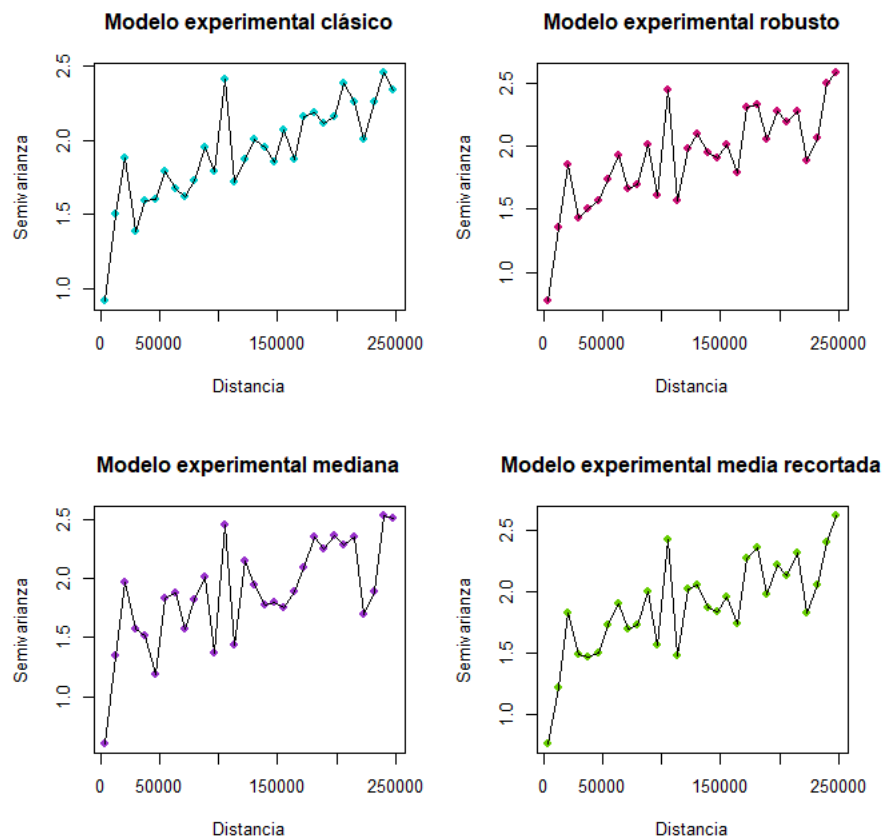


Figura 6.7: Semivariogramas experimentales KO

De acuerdo a las anteriores gráficas podemos ver que con los estimadores Clásico y Robusto se tiene el menor ruido aunque son similares se observa menos ruido en el Clásico. Se ajustan varios semivariogramas teóricos sobre los experimentales para elegir el modelo teórico adecuado.

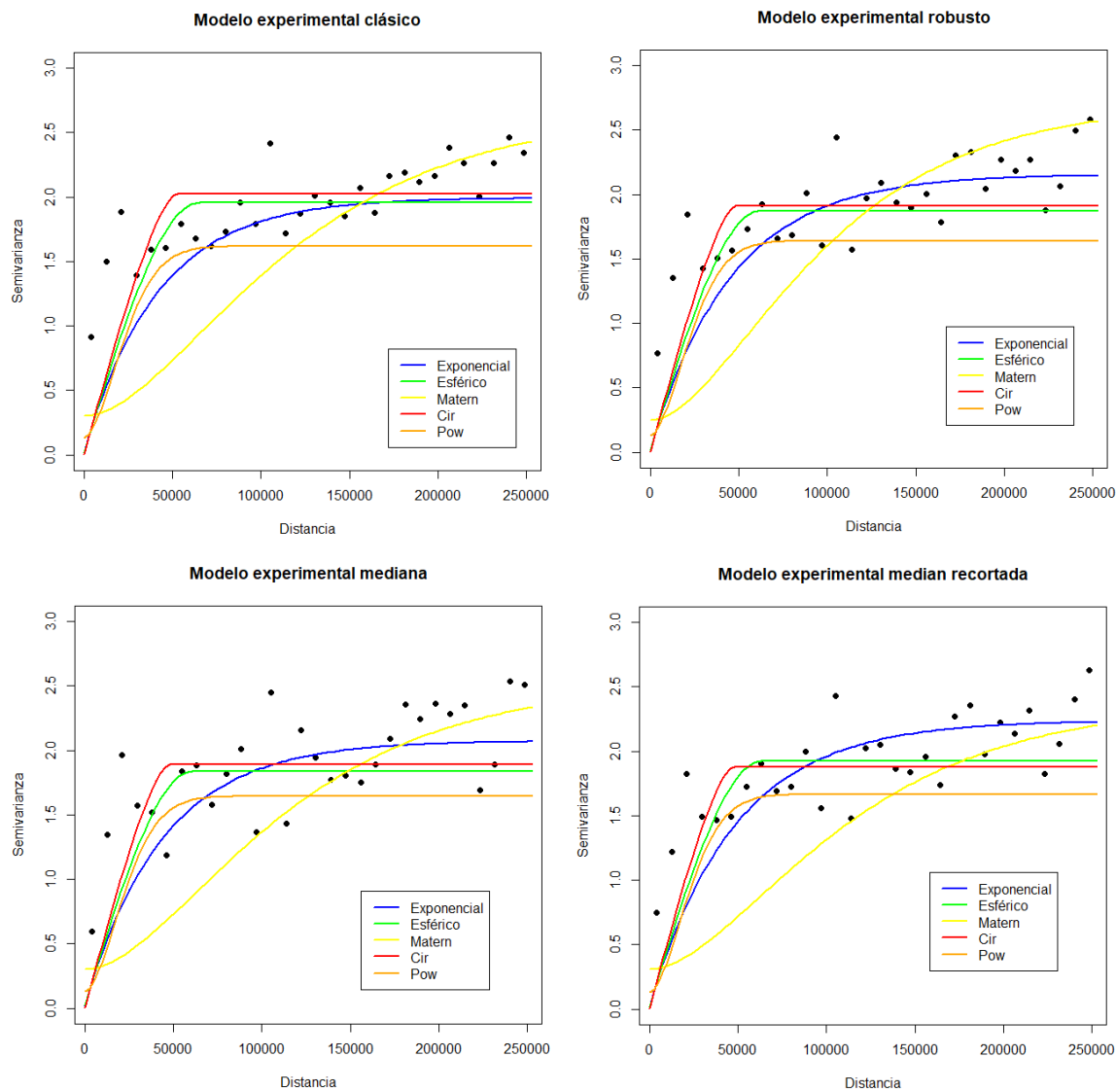


Figura 6.8: Semivariogramas teóricos KO

De acuerdo a los gráficos anteriores de los semivariogramas teóricos y la siguiente tabla donde se evalúa el AIC de todos los modelos para seleccionar el mejor modelo, se evidencia que para el caso los mejores fueron los de los semivariogramas clásico y de la Mediana, se selecciona sin con el menor AIC el el modelo exponencial ajustado sobre el semivariograma clásico.

	Clásico	Robusto	Mediana	Media recortada
Exponencial	<b>432.89</b>	434.67	433.80	435.48
Esferico	444.31	443.70	443.35	444.10
Matern	444.69	461.64	444.15	440.72
Circular	449.64	451.25	451.34	451.53
Power	463.50	464.34	464.49	465.27

Se procede a la estimación de los parámetros de este mediante máxima verosimilitud (ML), máxima verosimilitud restringida (RML) y mínimos cuadrados ponderados (WLS).

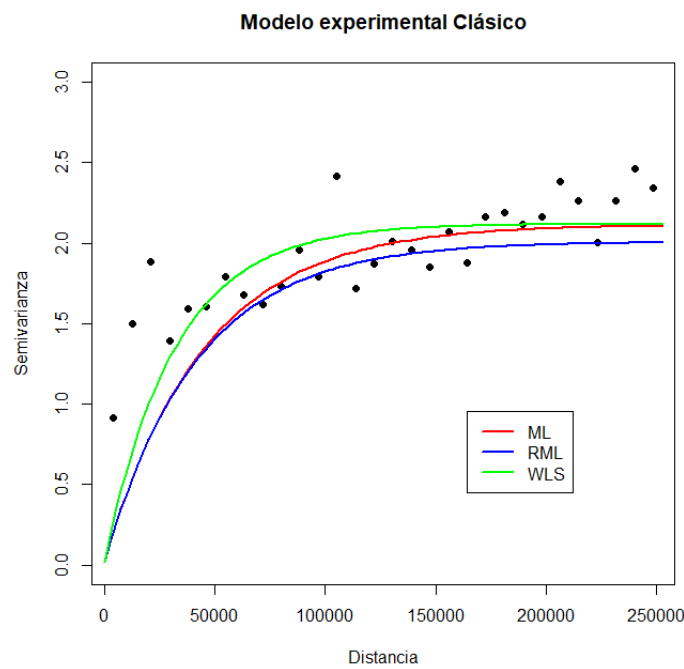


Figura 6.9: Estimación parámetros semivariograma exponencial KO

Se selecciona como el mejor el estimado mediante mínimos cuadrados ponderados ya que también presenta un ajuste en las distancias entre la pepita y la meseta y también en general a lo largo del semivariograma Clásico. En la tabla anterior se presentan los parámetros obtenidos por cada método de ajuste.

Estimación	Pepita( $C_0$ )	Meseta ( $C_1$ )	Rango (a)
ML	0.02	1.6	45000
RML	0.02	1.6	42000
WML	0.02	2.1	32000

Cuadro 6.7: Parámetros semivariogramas exponenciales obtenidos en los diferentes métodos KO

### Semivariograma para Kriging Simple (KS)

Se realiza el mismo procedimiento para hallar el modelo de semivariograma para Kriging simple pero para ella tomando la variable transformada estandarizada por la media es decir restando la media a cada una de los valores.

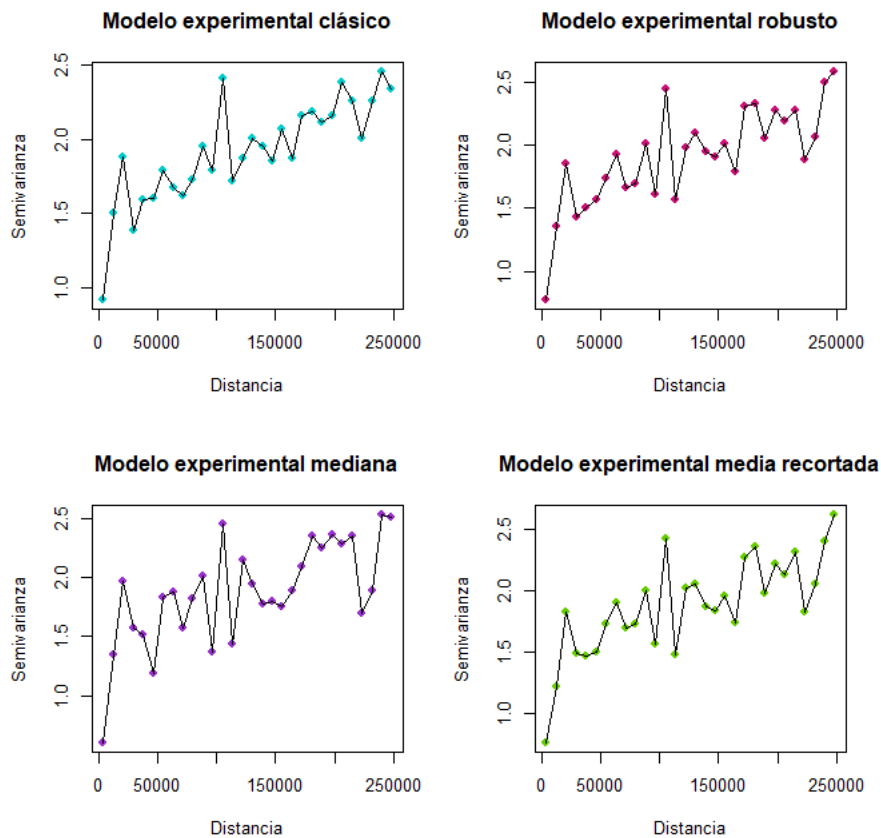


Figura 6.10: Semivariogramas experimentales KS

Al igual que para el calculo del semivariograma para Kriging Ordinario las anteriores gráficas podemos ver que con los estimadores Clásico y Robusto se tiene el menor ruido aunque son similares se observa menos ruido en el Clásico.

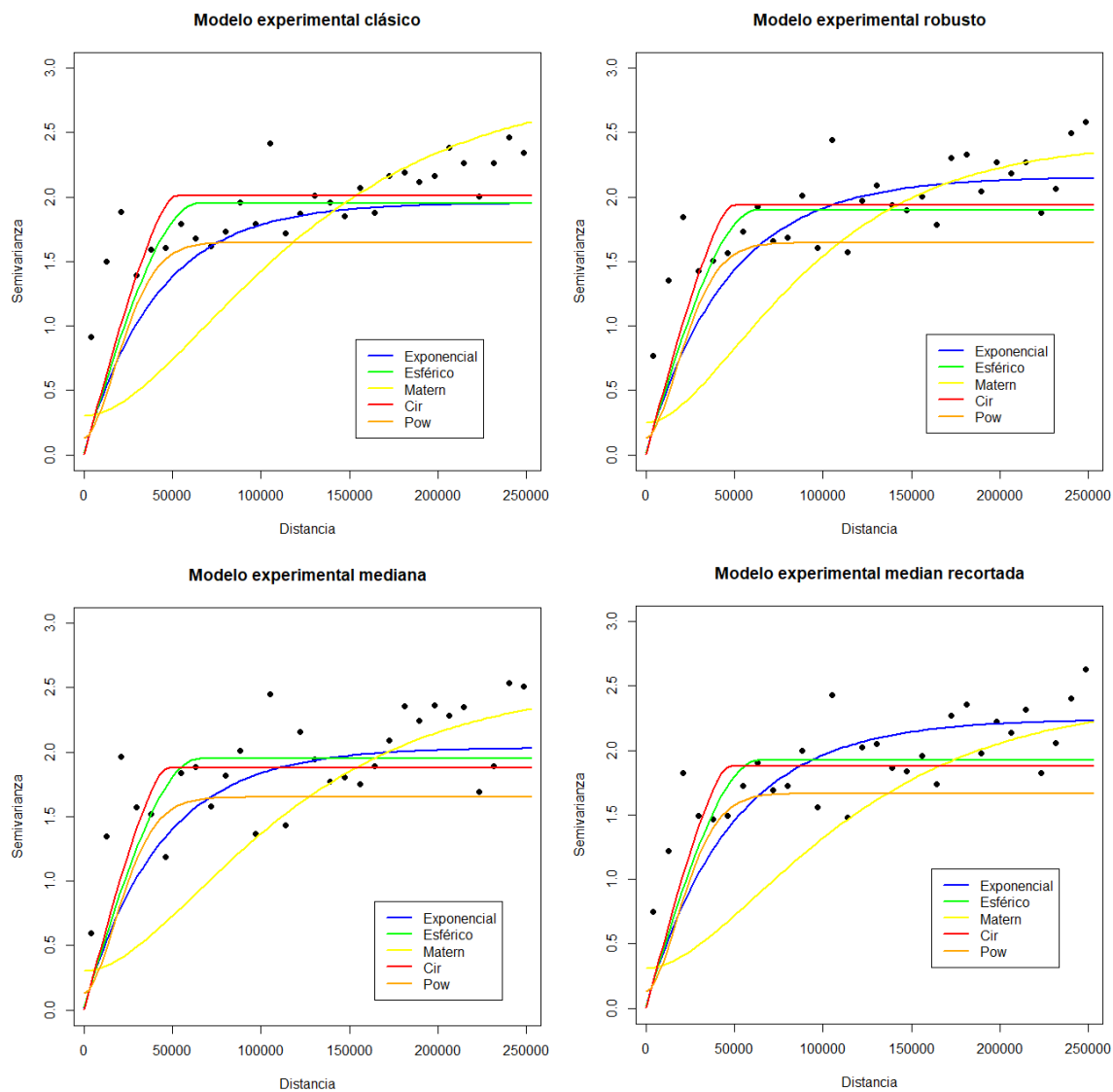


Figura 6.11: Semivariogramas teóricos KS

De acuerdo a los gráficos anteriores de los semivariogramas teóricos y la siguiente tabla donde se evalúa el AIC de todos los modelos para seleccionar el mejor modelo, se selecciona al igual que para el Kriging Ordinario el semivariograma por menor AIC el modelo exponencial ajustado sobre el semivariograma clásico.

	Clásico	Robusto	Mediana	Media recortada
Exponencial	<b>432.41</b>	434.66	433.35	435.52
Esferico	444.28	443.89	444.17	444.07
Matern	445.46	459.85	444.14	440.85
Circular	449.88	450.99	451.50	451.49
Power	464.64	464.49	464.69	465.25

Se procede a la estimación de los parámetros de este mediante máxima verosimilitud (ML), máxima verosimilitud restringida (RML) y mínimos cuadrados ponderados (WLS).

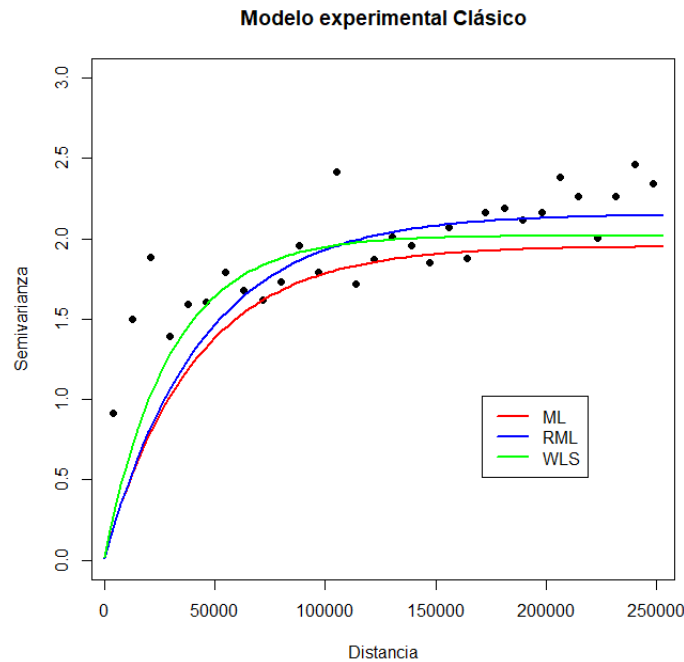


Figura 6.12: Estimación parámetros semivariograma exponencial KS

Se selecciona como el mejor el estimado mediante máxima verosimilitud restringida ya que también presenta un ajuste en los datos incluyendo los valores mayores para gamma también en general a lo largo del semivariograma Clásico. En la tabla anterior se presentan los parámetros obtenidos por cada método de ajuste.

Estimación	Pepita( $C_0$ )	Meseta ( $C_1$ )	Rango ( $a$ )
ML	0.02	2	41000
RML	0.01	2.1	44000
WML	0.02	2	30000

Cuadro 6.8: Parámetros semivariogramas exponenciales obtenidos en los diferentes métodos KS



## 7. Métodos de interpolación

Teniendo en cuenta el modelo exponencial para el semivariograma estimado por mínimos cuadrados y al considerar una tendencia no muy explicativa, se realiza el kriging con las técnicas de kriging simple y ordinario para los métodos geoestadísticos, para los métodos deterministas se tendrán la distancia inversamente ponderada y las funciones de base radial con sus respectivos parámetros optimizados.

### 7.1. Métodos determinísticos

#### 7.1.1. IDW (Distancia inversa ponderada)

Para calcular el promedio ponderado del inverso de la distancia es necesario primero hallar un  $p$  óptimo, por defecto esta toma el valor de 2 en la siguiente tabla se observan los valores de la raíz media del error de la predicción (RMSPE) para cada valor de  $p$ .

	P1=1	P2=P-OPTIMO	P3=2	P4=4	P5=5	P6=6
RMSPE	2.52	2.5	2.51	2.55	2.62	2.69

Cuadro 7.1: Valores de RMSPE para cada valor  $p$

Como se observa el  $p$  óptimo y el  $p = 2$  tienen valores muy cercanos, mediante métodos gráficos se observa que el  $p$  óptimo es menor a 2 así para el IDW, y mediante la función de optimización del  $p$  podemos llegar que para la estimación se utilizara un  $p$ -óptimo de 1.5775117.

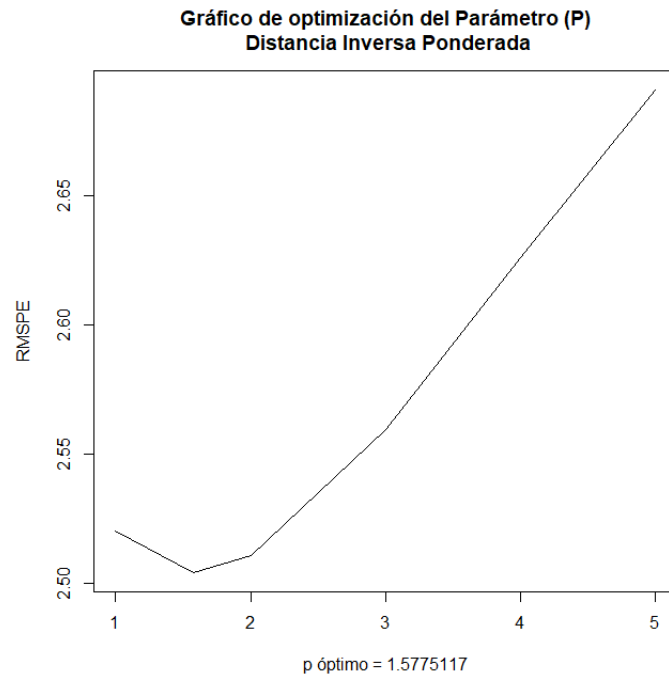


Figura 7.1: Optimización parámetro p

El resultado de las interpolaciones es consistente con los datos observados ver tabla 6.2; las temperaturas más altas tienden a localizarse en la zona costera y hacia la zona central y nororiental mientras que en la zona occidental se registran temperaturas mas bajas.

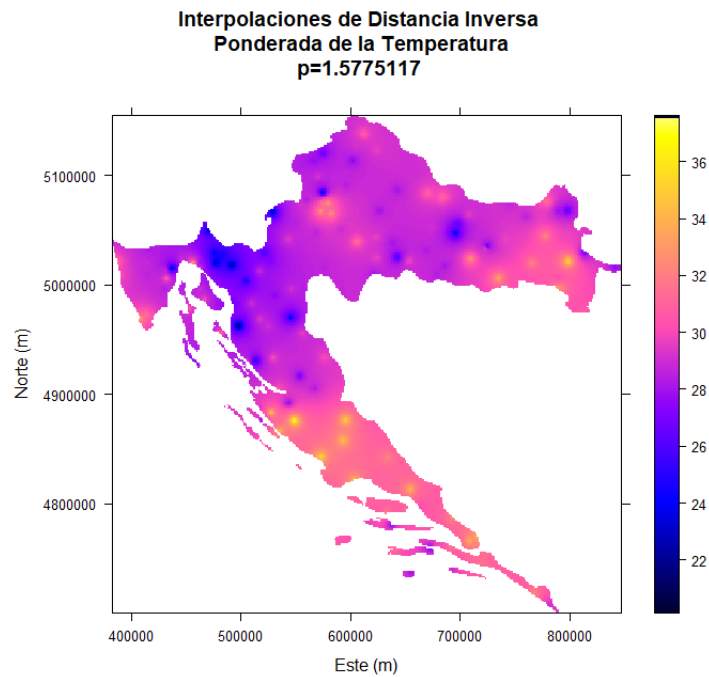


Figura 7.2: Resultados interpolación IDW

Mínimo	Media	Máximo
21.2182	29.28603	36.53699

Cuadro 7.2: Estadísticas resultados predicción IDW

### 7.1.2. Funciones de base radial

Para realizar la interpolación con funciones de base radial primero se realizó la optimización de los parámetros mediante la función `graph.rbf` de la librería `geospt` con 500 iteraciones, 15 vecinos y sin tener en cuenta tendencia por resultados anteriores, se obtuvieron los siguientes valores de los cuales se selecciona el de menor RMSPE que para el caso fue la función de Spline completamente regularizada y se realiza la interpolación con esta.

Función de base radial	$\eta$ óptimo	$\rho$ óptimo	RMSPE
Multicuadrática	$1e^{-5}$	0	2.498
Inversa Multicuadrática	2	0	2.59
Spline con Tensión	0.00027482429	0.1054797	2.38
<b>Spline completamente regularizado</b>	<b>0.000406909</b>	<b>0.1075205</b>	<b>2.33</b>
Spline capa delgada	2	1.999798	2.82
Exponencial	0.1	0.1	2.59
Gaussiana	0.1	0.1	2.59

Cuadro 7.3: Resultado optimización funciones de base radial sin tendencia; 500 iteraciones; 15 vecinos

Los resultados de la interpolación realizada por este método no es tan consistente con los datos observados como en los demás métodos en esta se presenta un mínimo menor y un máximo mayor con grandes diferencias de 6 grados en el mínimo y en el máximo al que teníamos de los datos ver tabla 6.2 sin embargo sigue siendo consistente en que las temperaturas mas elevadas y las menores se encuentran en los mismos lugares, mas sin embargo no se tiene como una muy buena predicción dado a que son grandes diferencias con los valores observados.

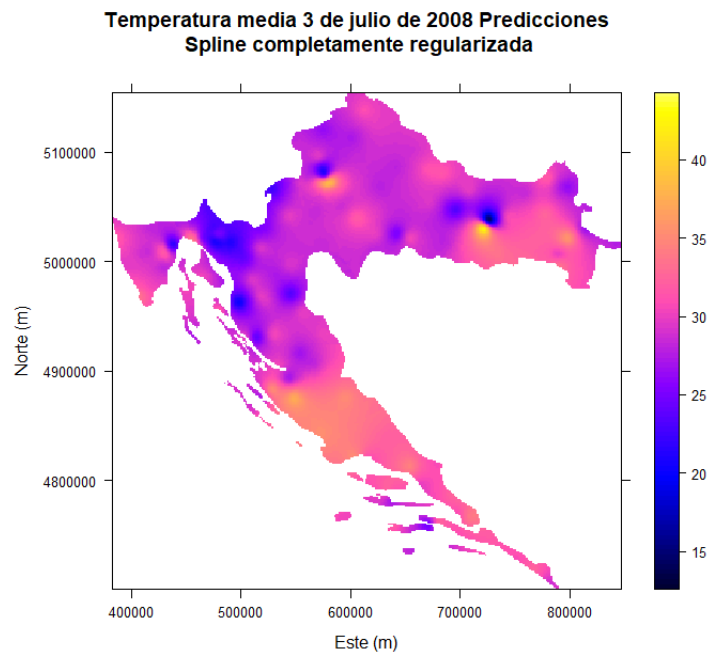


Figura 7.3: Resultados interpolación función de base radial, spline completamente regularizado

Mínimo	Media	Máximo
14.5252	29.4996	42.3413

Cuadro 7.4: Estadísticas resultados función de base radial, spline completamente regularizado

## 7.2. Métodos geoestadísticos

### Kriging simple y ordinario

Teniendo en cuenta los semivariogramas seleccionados, y el valor de la media de la variable de temperatura transformada se realizan los dos métodos de kriging sobre toda la grilla de Croacia y se realiza la antitransformación de la anamorfosis gaussiana para obtener valores de temperatura nuevamente, en los dos se obtienen resultados similares a continuación se presentan los mapas de predicción de cada uno de los métodos.

Kriging	Estimación	Pepita( $C_0$ )	Meseta ( $C_1$ )	Rango (a)
Simple	RML	0.01	2.1	44000
Ordinario	WML	0.02	2.1	32000

Cuadro 7.5: Parámetros semivariogramas exponenciales para realización de las interpolaciones Kriging

## Predicción

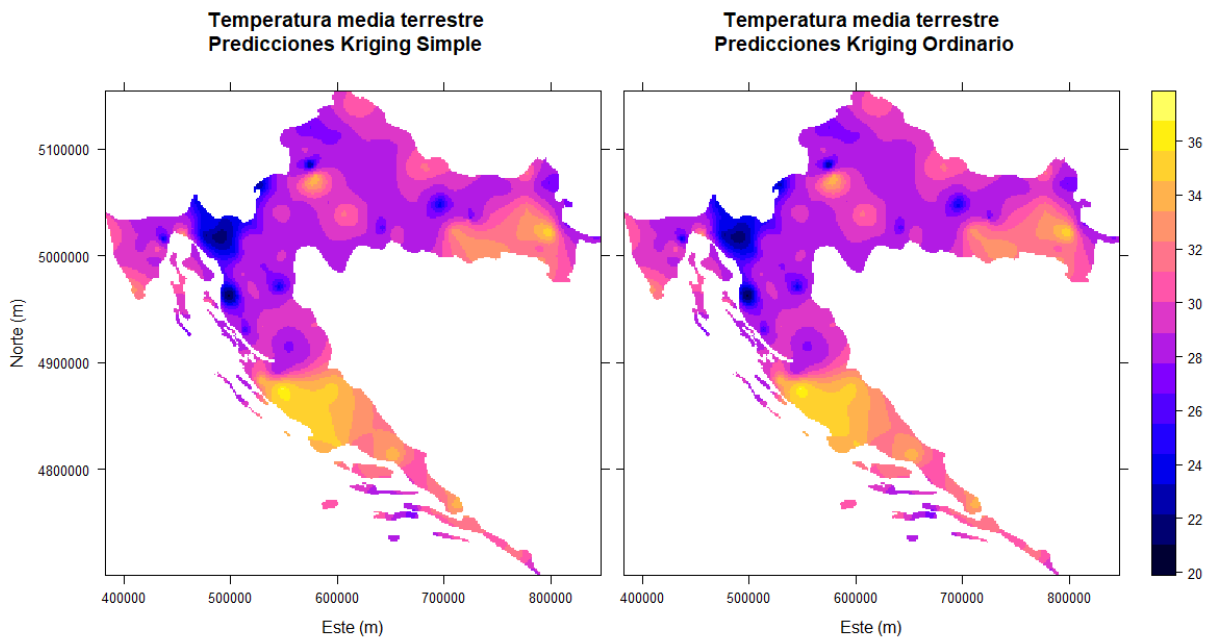


Figura 7.4: Resultados interpolación predicción temperatura Kriging Simple y Kriging Ordinario

Se pueden observar mediante las estadísticas de estos que son valores son valores son resultados muy similares a parte de tener varianzas muy cercanas, resultados muy similares de las observaciones, ver tabla 6.2.

	Mínimo	Media	Máximo	Var.Minima	Var Maxima
Kriging Simple	20.93788	29.59083	36.82473	0.0271	1.3876
Kriging Ordinario	20.98765	29.58121	36.76404	0.0495	1.6974

Cuadro 7.6: Estadísticas de las predicciones de temperatura de los krigeados

A continuación se presentan las superficies de varianzas de los dos krigeados.

## Varianza

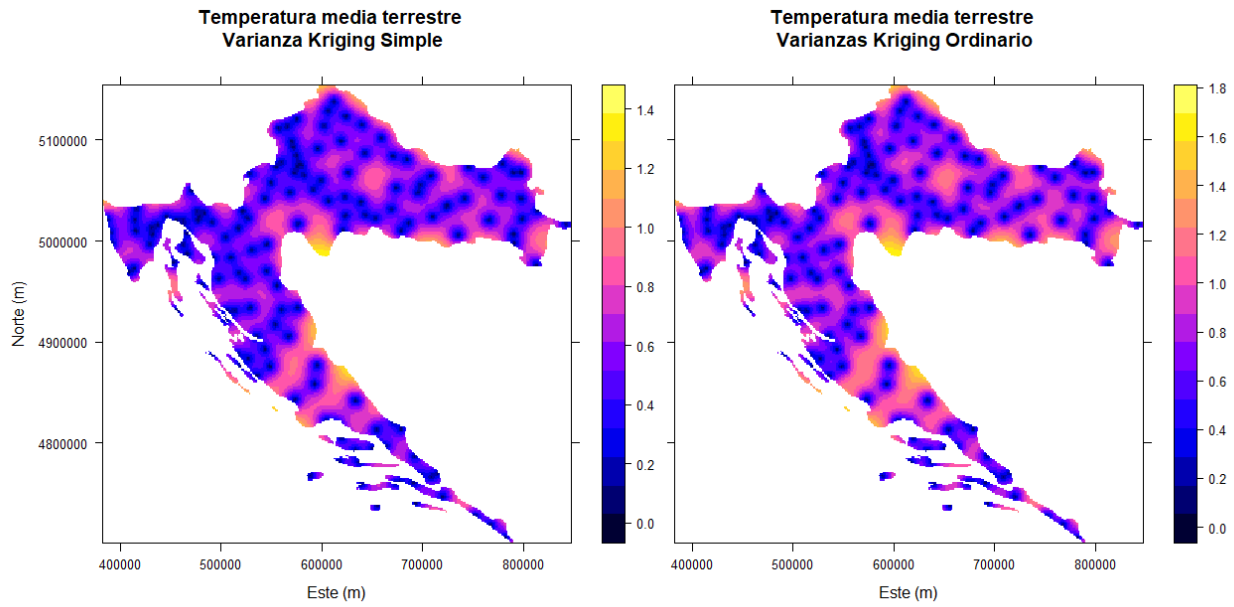


Figura 7.5: Superficie varianzas Kriging Simple y Kriging Ordinario

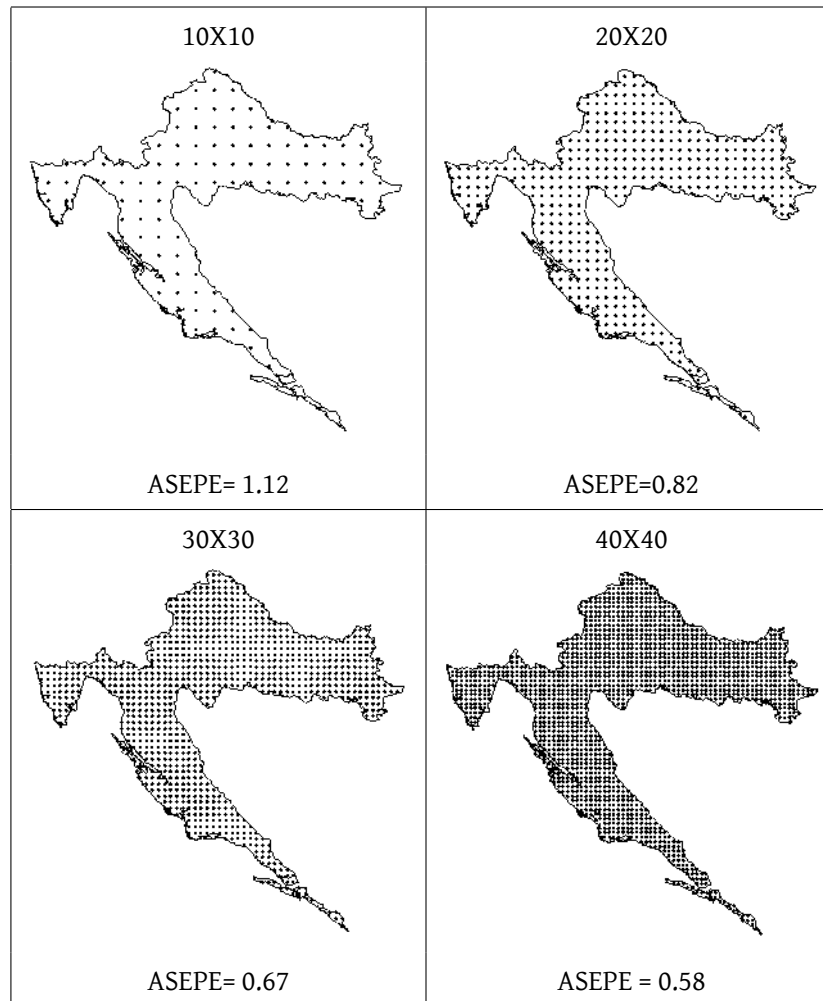
Se evidencia en que las variaciones son mínimas entre  $0,0271^{\circ}C$  y  $1,38^{\circ}C$ , para el caso del Krigin simple y entre  $0,049^{\circ}C$  y  $1,69^{\circ}C$  para Kriging Ordinario, así que la estimación del Krigeado es buena para la variable de temperatura, se decide elegir entre los dos el Kriging Ordinario ya que al realizar la validación cruzada se obtiene un menor RMSPE ligeramente mas bajo que el del Kriging Simple.

	RMSPE
Kriging Simple	0.8493
Kriging Ordinario	0.8480

Cuadro 7.7: RMSPE Krigeados

## 8. Diseño de red

Podemos ahora luego de seleccionar el modelo crear una red en la cual se minimize la varianza para ello utilizamos la función de `network.design` de la librería `geospt` de R en la cual utilizamos el Kriging Ordinario con modelo exponencial con valores  $C_0 = 0,02$ ,  $C_1 = 2,1$  y  $a = 32000$  y la función crea una red y devuelve el valor promedio del error estándar de las predicciones para el krigeado (ASEPE) Santacruz (2015). Se evaluaron 4 mallas de 10x10, 20x20 y 40x40 y se obtuvieron las siguientes mallas y valores de ASEPE, como era de esperarse entre mayor es el numero de puntos menor va a ser el valor del error.



Cuadro 8.1: Resultados `network.design`

Una estación meteorológica en la zona de estudio en Europa según una tabla de precios de DAVIS weather

instruments DAVIS (s.f.), una estación básica pero que cuente con la medición de las variables de la base de datos como temperatura y humedad (las demás variables se pueden considerar constantes por lo tanto al ubicarla ya tendríamos las variables de altura, coordenadas y distancia a la costa) tiene un costo de 85 Euros por lo tanto tendríamos los siguientes costos para cada red.

Grilla	Costo en Euros
10X10	8.500
20x20	34.000
30x30	76.500
40x40	136.000

Cuadro 8.2: Costo de las redes de estaciones meteorológicas

Al observar los costos y los errores (ASEPE) que obtenemos con las diferentes rejillas se llega a la conclusión que probablemente no sea tan conveniente utilizar las rejillas con mas densidad de puntos dado a su elevado precio por ende y dado que no se ve mucha diferencia en los errores generados se selecciona que la selección ideal para el diseño de red es la rejilla de 20X20 con un costo de 34.000 euros, 154 millones de pesos colombianos al cambio.



## 9. Conclusiones

- En el análisis realizado, se observó como las temperaturas obtenidas por las estaciones el día 3 de julio de 2008 presentaban anomalía por lo que tuvo que realizarse una transformación en las cuales se transforman los datos para obtener la normalidad, aunque después de realizar las predicciones es necesario usar la antitransformada para obtener valores en términos de temperatura.
- Al analizar la tendencia observamos que no se obtuvo un coeficiente de determinación bajo por tal motivo no fue necesario modelar esta tendencia por tal motivo se siguió trabajando con la variable transformada. Al analizar anisotropía observamos que era necesario analizarla por métodos gráficos y con la función de la librería *intamap* para poder verificar que los datos no presentaran anisotropía.
- En el ajuste de semivariograma para la variable temperatura se seleccionó como mejor modelo en modelo teórico exponencial para el modelo experimental clásico ya que era en el cual se observaba menor ruido y se eligió el ajuste realizado por mínimos cuadrados ponderados con los valores de  $pepita = 0,02$ ,  $meseta = 2,1$  y  $rango = 32000$ .
- Para los métodos determinísticos en el IDW se obtuvieron resultados más consistentes con las observaciones que en las funciones de base radial en donde se seleccionó el Spline completamente regularizado más sin embargo hubo más variaciones en esta con respecto a los demás métodos.
- Con el semivariograma elegido se calcularon los krigeados en donde se obtuvo resultados entre el Kriging Simple y Ordinario muy similares pero presentando ligeramente menos error en el Simple por tal motivo se seleccionó este como mejor y fue el mejor de todas las estimaciones ya que presenta un menor RMSPE, con este se realizó un diseño de red en donde se observa que entre más puntos se obtiene un menor error.

# Referencias

- Armstrong, M. (1950). «Basic linear geostatistics.» En: (vid. págs. 13, 15).
- Cressie, Noel A.C (2019). *Statistics for Spatial Data*. Vol. 53. 9, págs. 1689-1699. ISBN: 9788578110796. DOI: 10.1017/CB09781107415324.004. arXiv: arXiv:1011.1669v3 (vid. págs. 9, 12).
- DAVIS (s.f.). «DAVIS wheater instruments price list». En: (). URL: <https://www.e-wetter.eu/downloads/davispricelisteng.pdf> (vid. pág. 43).
- Emery, Xavier (2013). «Geoestadística». En: *Universidad de Chile* (vid. págs. 10, 15, 16).
- Felícísimo Pérez, AM (1980). *Introducción al clima de Asturias. Régimen pluviotérmico* (vid. pág. 5).
- Gil, Adriana (2013). «Modelació De La Variació Espacial De Variables Edáficas Y Su Aplicació Ee». En: (vid. pág. 9).
- Henao, Ramón Giraldo (2002). *INTRODUCCION A LA GEOESTADISTICA*. Inf. téc. Universidad Nacional de Colombia (vid. págs. 10-12).
- Johnston, Kevin (2001). «Using arcgis geostatistical analyst». En: (vid. pág. 17).
- Marquinez, Jorge, Javier Lastra y Pilar Garcia (2003). «Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis». En: *Journal of hydrology* 270(1-2), págs. 1-11 (vid. pág. 5).
- Martinez, Juan Luis Fernández (2007). «Uso combinado de técnicas geoestadísticas y SIG para la realización de mapas robustos de variables climáticas en Asturias.» En: *XII Congreso Internacional de Energía y Recursos Minerales* (vid. pág. 5).
- Matheron, Georges (1963). «Principles of geostatistics». En: *Economic geology* 58(8), págs. 1246-1266 (vid. págs. 1, 9, 11, 17).
- Melo, Carlos (2015). *Estadística Espacial Teoría y Aplicaciones* (vid. págs. 5, 11, 12, 14-16, 18, 19).
- Melo, Carlos Eduardo (2012). «Análisis geoestadístico espacio tiempo basado en distancias y splines con aplicaciones». En: (vid. pág. 11).
- Santacruz, Al (2015). *Geostatistical Analysis and Design of Optimal Spatial Sampling Networks geospt*. R package version 1.0. URL: <https://www.rdocumentation.org/packages/geospt/versions/1.0-2> (vid. pág. 42).
- Shtiliyanova, Anastasiya y col. (2017). «Kriging-based approach to predict missing air temperature data». En: *Computers and Electronics in Agriculture* 142, págs. 440-449 (vid. pág. 7).
- Skoien, Jon (2018). *Procedures for Automated Interpolation intamap*. R package version 1.4. URL: <https://www.rdocumentation.org/packages/intamap/> (vid. pág. 28).
- Varela, Roberto Hernández (2001). «Comparación de métodos de kriging multivariante para la estimación espacial de la temperatura del aire». En: *Estudios geográficos* 62(243), págs. 285-308 (vid. pág. 3).

Viera, Martín A Díaz y Ricardo Casar González (2002). «Geoestadística aplicada». En: *Instituto de Geofísica, Universidad Nacional Autónoma de México, Ministerio de Ciencia, Tecnología y Medio Ambiente de Cuba: México DF, Mexico*, págs. 31-57 (vid. pág. 1).