



ANÁLISIS GEOESTADÍSTICO DE LA TEMPERATURA MEDIA TERRESTRE EN CROACIA DEL DÍA 27 DE JULIO DE 2008

**Universidad Distrital Francisco José de Caldas
Facultad De Ingeniería
Ingeniería Catastral y Geodesia
Geoestadística**

Presentado a:
Carlos Eduardo Melo Martínez

Presentado por:

Angie Paola Ríos Monroy 20141025001
Lizeth Lorena Flórez- 20142025018

TABLA DE CONTENIDO

1. INTRODUCCIÓN	4
2. JUSTIFICACIÓN.....	4
3. OBJETIVOS.....	5
3.1 OBJETIVO GENERAL	5
3.2 OBJETIVOS ESPECÍFICOS	5
4. ESTADO DEL ARTE.....	5
5. MARCO TEÓRICO	6
6. DESCRIPCIÓN GENERAL DE LOS DATOS	13
7. METODOLOGÍA	14
7.1 ANÁLISIS EXPLORATORIO DE DATOS.....	14
7.2 ANÁLISIS ESTRUCTURAL DE DEPENDENCIA ESPACIAL	16
7.2.1 TENDENCIA.....	16
7.2.2 ANISOTROPÍA	18
7.3 AJUSTE DEL SEMIVARIOGRAMA.....	19
7.4 INTERPOLACIÓN.....	22
7.4.1 MÉTODOS DETERMINÍSTICOS	22
7.4.2 MÉTODOS PROBABILÍSTICOS.....	24
7.5 DISEÑO DE RED.....	26
9. BIBLIOGRAFÍA	28

TABLA DE FIGURAS

Figura 1. Mapa Krigado (a) y errores de estimación (b) de la variable “Área Basal”. Fuente: (Chica-Olmo., 2005).....	6
Figura 2. Partes del Semivariograma. Fuente: (Giraldo, 2003)	8
Figura 3. a) Modelo Esférico b) Modelo Exponencial c) Modelo Gaussiano	10
Figura 4. Puntos geolocalizados de la variable Temperatura el día 27 de julio de 2008	14
Figura 5. Histograma, QQ plot y densidad de la variable temperatura media terrestre. Fuente: elaboración propia.	14
Figura 6. Distribución espacial Tridimensional de la temperatura media terrestre el día 27 de julio de 2008.....	15
Figura 7. Gráficos de Distribución espacial de la variable temperatura, Perfil coordenada en X, Perfil coordenada en Y, histograma de la variable Temperatura.....	16
Figura 8. Dispersogramas de covariables con Temperatura.	16
Figura 9. Modelo lineal clásico (Todas las variables) vs modelo lineal clásico (variables significativas).....	17
Figura 10. Modelo Orden 2	17
Figura 11. Semivariograma clásico multidireccional de temperatura vs residuos del modelo de regresión significativo	18

Figura 12. Modelos teóricos ajustados al semivariograma clásico variable temperatura. Fuente: elaboración propia.	19
Figura 13. Modelos teóricos ajustados al semivariograma Robusto variable temperatura. Fuente: elaboración propia.	20
Figura 14. Modelos teóricos ajustados al semivariograma clásico $ZSi = ZSi - \mu$ para la variable temperatura – μ mediante MVR. Fuente: elaboración propia.	21
Figura 15. Modelos teóricos ajustados al semivariograma robusto de $ZSi = ZSi - \mu$ para la variable temperatura – μ mediante MVR. para la variable temperatura – μ . Fuente: elaboración propia. ..	21
Figura 16. Interpolación IDW usando p óptimo. Fuente: elaboración propia.....	23
Figura 17. Interpolaciones Funciones de Base Radial. Fuente: elaboración propia.....	24
Figura 18. Mapa de interpolación Kriging Simple en la parte izquierda y el mapa de los errores de Ks en la derecha. Fuente: elaboración propia.	24
Figura 19. Mapa de interpolación Kriging Ordinario en la parte izquierda y el mapa de los errores de Ko en la derecha. Fuente: elaboración propia.....	25
Figura 20. Diseños de red con muestreo regular. Fuente: elaboración propia.....	26
TABLA DE TABLAS	
Tabla 1. Formas funcionales de algunas RBFs espacio-temporales. Fuente: (Melo, Melo, & Melo)	11
Tabla 2. Descripción de variables originales. Fuente: elaboración propia.	13
Tabla 3. Estadística Descriptiva de la variable respuesta Temperatura Media Terrestre.....	15
Tabla 4. Tests de normalidad.....	15
Tabla 5. Evaluación de la Anisotropía en la variable Temperatura transformada.	18
Tabla 6. Evaluación de la Anisotropía corregida en la variable Temperatura transformada.	19
Tabla 7. Parámetros de modelos ajustados clásico. Fuente: elaboración propia.	19
Tabla 8. Parámetros de modelos ajustados robusto. Fuente: elaboración propia.	20
Tabla 9. Parámetros de modelos ajustados clásico $ZSi = ZSi - \mu$ para la variable temperatura – μ mediante MVR. Fuente: elaboración propia.	21
Tabla 10. Parámetros de modelos ajustados robusto $ZSi = ZSi - \mu$ para la variable temperatura – μ mediante MVR. Fuente: elaboración propia.	22
Tabla 11. Parámetros optimizados FBR. Fuente: elaboración propia.....	23
Tabla 12. ASEPE para cada diseño de red. Fuente: elaboración propia.	26

1. INTRODUCCIÓN

El análisis espacial de datos experimentales surge por la necesidad de definir métodos que sean capaces de dar respuestas a problemas de estimación o simulación de variables, a partir de esto surge la necesidad de utilizar métodos estadísticos, herramientas que permitan hacer análisis y construcción de mapas, con el fin de hacer predicciones en una región de estudios donde quizá no se pudieron tomar muestras, hacer comparaciones estadísticas y geográficas cuando se tienen en cuenta más de una variable .

A continuación se va a realizar el análisis geoestadístico en base a una muestra de datos de temperatura del 27 de julio de 2008 de Croacia por el método de Geoestadística, con el fin de estimar valores por medio de un análisis de dependencia espacial, presencia o ausencia de normalidad, tendencia y anisotropía, mediante modelos donde se ajustarán los semivariogramas experimental y teórico más adecuado a través del software estadístico R Studio.

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 (Scott, Stone, & Allen, 2004). 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. Dichas técnicas están en capacidad de ser utilizadas por perfiles profesionales como el de Ingenieros Catastrales y Geodestas ya que deben estar en capacidad de realizar análisis espaciales de fenómenos para la toma de decisiones.

3. OBJETIVOS

3.1 OBJETIVO GENERAL

- Modelar la variación espacial de los registros de la temperatura media terrestre diaria de Croacia para obtener predicciones a partir de métodos de interpolación.

3.2 OBJETIVOS ESPECÍFICOS

- Hacer el análisis estadístico de la base de datos tomados el 27 de Julio de 2008 en Croacia.
- Realizar el análisis exploratorio mediante pruebas de normalidad, tendencia y anisotropía de la base de datos de Croacia.
- Analizar cuál es el modelo experimental y teórico más adecuado para ajustar el semivariograma de la base de datos.
- Generar predicciones de temperatura para día mencionado anteriormente.

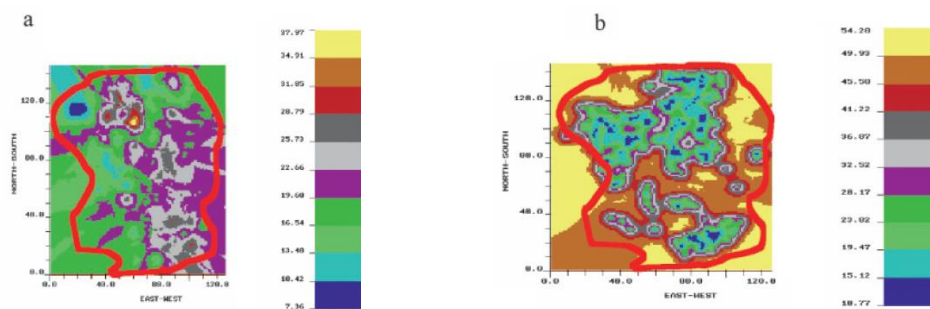
4. ESTADO DEL ARTE

La aplicación de métodos estadísticos en cuanto al análisis de datos espaciales es utilizada para estudiar los problemas en la creación de mapas temáticos, aunque todavía es poco llamativa por la complejidad teórica, en un estudio de recursos forestales se pretende facilitar la comprensión de los aspectos teóricos y reglas para la aplicación de la geoestadística. Mediante una base de datos de parcelas del Inventario Ecológico y Forestal de Cataluña.

Tomando como base las variables de la base de datos muestran los diversos comportamientos que se pueden presentar, por ejemplo las variables Área Basal y Cobertura Arbustiva, muestran un comportamiento distinto en su distribución espacial, analizando modelos de variograma donde la variable Área Basal tiene un alcance mayor. Por otro lado se encontró Anisotropía se en la variable Volumen del tronco. También realizaron variograma cruzado entre Volumen Tronco y Área Basal donde dedujeron la correlación entre estas en función del vector h de

distancia sugiriendo un procedimiento de estimación de Kriging para la reducción de errores de estimación.

Aplicando Kriging en sus diferentes modalidades para la variable Área Basal junto con el mapa de errores de estimación ver Figura 1, se produjeron estimaciones óptimas de la variable espacial, al minimizar el error de estimación y por la condición de no sesgo el estimador tanto a escala global como local.



*Figura 1. Mapa Krigeado (a) y errores de estimación (b) de la variable “Área Basal”.
Fuente: (Chica-Olmo., 2005).*

El mapa de estimación elaborado debe interpretarse como una imagen más probable de la distribución espacial de la variable estudiada. Su estructura de datos raster hace que sea fácilmente integrable en bases de datos geoespaciales, como cubierta de información SIG. El mapa de errores de estimación es de mucho interés en actividades de planeación medioambiental, a través de la interpretación de los errores relativos de estimación.

En cuanto a la reproducción de estadísticos, de distribución espacial y condicionamiento hacen que las posibilidades de aplicación del método, junto con los kriging, sean muy variadas e interesantes en el estudio de problemas encontrados en la gestión forestal. (Chica-Olmo., 2005)

5. MARCO TEÓRICO

Para realizar un análisis geoestadístico es necesario considerar una serie de pasos: un primer paso consiste en analizar la calidad y la cantidad de datos requeridos, es decir el muestreo espacial. El segundo paso es el análisis exploratorio, el cual se basa en el uso de técnicas estadísticas convencionales y en el análisis estructural de los datos, con el objetivo de

identificar la presencia de anisotropía o isotropía y la tendencia. En el modelado del variograma, se evalúa la relación espacial entre los valores de la variable regionalizada, y se ajusta un modelo de variograma al variograma experimental. Una vez el modelo de variograma se encuentra, los valores de predicción se pueden generar usando la interpolación Kriging para la construcción del mapa de predicción de la variable explicada. Sin embargo, hay métodos deterministas, donde el modelo de interpolación no requiere de un modelo de variograma, como es el caso de las FBRs. Después de esto, para elegir el mejor método de interpolación se utiliza la validación cruzada. El último paso consiste en la generación de los mapas de predicciones de la variable regionalizada y de las desviaciones estándar, junto con su interpretación y análisis. (Melo, Melo, & Melo).

Inicialmente se define una variable regionalizada $\{Z(S_i) : S_i \in D\}$ como aquella que posee información georeferenciada y presenta alguna estructura de dependencia espacial, donde S_i corresponde a cada ubicación de la muestra y D el espacio continuo de 1, 2 o 3 dimensiones.

En este tipo de análisis es fundamental determinar si los datos cumplen la propiedad de la *estacionariedad* que significa la estabilidad, uniformidad u homogeneidad que presentan los mismos en el área de estudio, es decir, propiedades estadísticas similares en el espacio, lo cual es lo ideal. Si ésta existe, se definen los cuatro primeros momentos de dicha variable regionalizada: media μ , varianza σ^2 , covarianza $C(h)$ y semivarianza $\gamma(h)$, donde h es el vector de traslación, cuya longitud es la distancia entre S_i y S_j .

Cuando existe estacionariedad fuerte, sus momentos no varían bajo el efecto de cualquier traslación de h . En particular si cumple con esta invariabilidad en los momentos de orden uno y dos, corresponde a una función estacionaria de segundo orden, es decir, los momentos no dependen de la localización, originando el modelo $Z(S_i) = \mu + \varepsilon(S_i)$. En cambio, si μ y σ^2 varían, existirán problemas de tendencia y heterocedasticidad, casos en los cuales eliminar la tendencia y transformar la variable respuesta suelen ser soluciones iniciales; y por

consiguiendo obteniendo el modelo $Z(S_i) = \mu(S_i) + \varepsilon(S_i)$, donde S_i refleja la tendencia presente en relación con las coordenadas y/o con posibles covariables. Estos dos elementos (media y covarianza) corresponden a los cambios o variabilidad a grande y pequeña escala, respectivamente; el análisis de la variación en ambas escalas es un aspecto esencial que incluye la geoestadística. (Schanbenberger & Gotway, 2005)

Para el *análisis exploratorio* es necesario realizar análisis de relación con las variables mediante un grupo de herramientas estadísticas-gráficas que describen y visualizan las distribuciones espaciales, identificando localizaciones atípicas, descubriendo formas de asociación (autocorrelación espacial) que, a su vez, pueden ser de carácter global o local, y sugiriendo estructuras en el espacio geográfico (heterogeneidad espacial). (Yrigoyen, 2003)

El análisis exploratorio nos permite detectar la presencia de fenómenos mediante el análisis de *semivariogramas*, los cuales son una herramienta usada para cuantificar y analizar la dependencia espacial, está definido por unos parámetros: Pepita (Nugget) C_0 el cual presenta errores de medición y escala, Meseta (Sill) $C_0 + C_1$ y Rango es la distancia entre los pares de puntos ver Figura 2. El semivariograma se calcula mediante la siguiente expresión matemática: (Giraldo, 2003)

$$\gamma(h) = \frac{\sum (Z(S_i) - Z(S_j))^2}{2 |N(h)|}$$

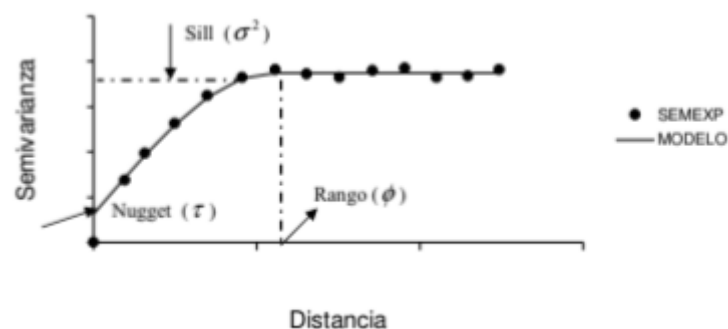


Figura 2. Partes del Semivariograma. Fuente: (Giraldo, 2003)

Mediante los semivariogramas también es posible encontrar *Anisotropía* a partir de la construcción de estos en diferentes direcciones, revelan una dirección privilegiada con distinto patrón de dependencia espacial. (Guedes, Uribe-Opazo, Johann, & Souza, 2015)

De lo contrario si sigue un patrón idéntico en todas las direcciones será dicha variable regionalizada *Isotrópica*.

A Partir del semivariograma se busca un modelo que ajuste el comportamiento de las observaciones mediante *modelos teóricos* de semivariograma, los cuales cumplen con lo descrito anteriormente de Isotropía. Algunos de los modelos teóricos conocidos son: Efecto pepita puro, Esférico, Exponencial, Gaussiano, Estable, Matern y Circular. (Armstrong, 1998)

El primer *efecto pepita puro* corresponde a un modelo puramente aleatorio, donde no existe correlación entre los valores sin importar la cercanía entre los mismos.

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

El modelo más utilizado es el *esférico*, 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

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

Para el 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.

$$\gamma(h) = C(1 - \exp(-|h|/a))$$

El modelo gaussiano Representa un fenómeno extremadamente continuo, donde se ha evidenciado la ocurrencia de inestabilidades numéricas cuando no se hace uso del efecto pepita.

Su rango efectivo es $\sqrt{3a}$.

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

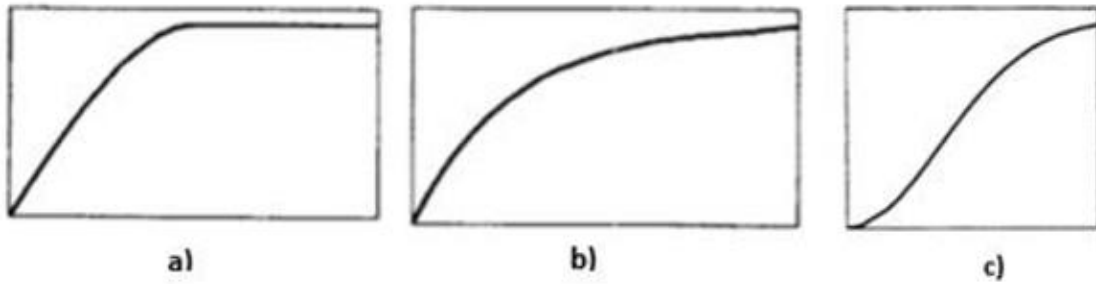


Figura 3. a) Modelo Esférico b) Modelo Exponencial c) Modelo Gaussiano

Una vez el modelo de variograma se encuentra, los valores de predicción se pueden generar usando la interpolación para llevar a cabo el análisis espacial donde se busca convertir algún fenómeno cuya distribución espacial puede ser irregular o cuyo muestreo no ha sido el más apropiado, en una representación de tipo raster, es decir, transformarlo en una estructura geográfica diferente que represente predicciones confiables mediante los *Métodos de Interpolación Espacial*. (Goodchild, Longley, & Maguire, 1999) Es fundamental la localización de los n vecinos más cercanos ya que el vecindario, a través de cierto radio de influencia, define la predicción en determinado punto en los interpoladores locales. Mediante métodos determinísticos y probabilísticos.

Dentro de los *métodos determinísticos* destacan: Distancia Inversa Ponderada (*IDW*) y las Funciones de Base Radial (*FBR*). El primero considera que el valor de un punto determinado, se puede definir como la suma ponderada de los valores tomados de la variable regionalizada en la ubicación de sus vecinos más cercanos, su ecuación general está dada por

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

Donde; $\hat{Z}(S_0)$ hace referencia al valor que se busca predecir en la posición (S_0) , es el número de observaciones que rodean al punto de predicción, λ_i son los pesos de cada una

de las mediciones, estos inversamente proporcionales a la distancia y $Z(S_i)$ que es el valor de variable en la posición. Los pesos se determinan haciendo uso de

$$\lambda_i = \frac{d_{io}^{-p}}{\sum_{i=0}^n d_{io}^{-p}}$$

Donde $\sum_{i=0}^n \lambda_i = 1$, el peso por su parte está controlado por un factor con cada incremento de las distancia, de tal manera que, si las distancia entre las posiciones medidas y la posición predecir. El valor óptimo para se determina minimizando la raíz cuadrada media del error de predicción. (Melo C. , 2015)

Las FBR a diferencia de IDW pueden predecir valores superiores e inferiores al rango de los datos de muestra, y son útiles al calcular superficies suavizadas que no presentan variaciones abruptas. Algunas de las funciones son la observadas en la Tabla 1 EXP (Exponencial), MQ (Multicuadrática), GAU (Gaussiano), IMQ (Multicuadrática Inversa), TPS (Spline Capa Delgada), CRS (Completamente Regularizada Spline) y ST (Spline con Tensión).

RBF	Forma funcional	RBF	Forma funcional
EXP	$\phi(\delta_{st}) = e^{-\eta \delta_{st}}, \quad \eta \neq 0$	GAU	$\phi(\delta_{st}) = e^{-\eta \delta_{st}^2}, \quad \eta \neq 0$
MQ	$\phi(\delta_{st}) = \sqrt{\eta^2 + \delta_{st}^2}, \quad \eta \neq 0$	IMQ	$\phi(\delta_{st}) = 1 / \sqrt{\eta^2 + \delta_{st}^2}, \quad \eta \neq 0$
RBF	Forma funcional		
TPS	$\phi(\delta_{st}) = \begin{cases} (\eta \cdot \delta_{st})^2 \log(\eta \cdot \delta_{st}) & \text{si } \delta_{st} \neq 0, \eta > 0 \\ 0 & \text{si } \delta_{st} = 0 \end{cases}$		
CRS	$\phi(\delta_{st}) = \begin{cases} \ln(\eta \cdot \delta_{st}/2)^2 + E_1(\eta \cdot \delta_{st}/2)^2 + C_E & \text{si } \delta_{st} \neq 0, \eta > 0 \\ 0 & \text{si } \delta_{st} = 0 \end{cases}$ <p>donde \ln es el logaritmo natural, $E_1(x)$ es la función integral exponencial y C_E es la constante de Euler.</p>		
ST	$\phi(\delta_{st}) = \begin{cases} \ln(\eta \cdot \delta_{st}/2) + K_0(\eta \cdot \delta_{st}) + C_E & \text{si } \delta_{st} \neq 0 \\ 0 & \text{si } \delta_{st} = 0 \end{cases}$ <p>donde $K_0(x)$ es la función modificada de Bessel.</p>		

Tabla 1. Formas funcionales de algunas RBFs espacio-temporales. Fuente: (Melo, Melo, & Melo)

El parámetro de suavizamiento óptimo η , el cual es de libre elección, se encuentra al minimizar la raíz del cuadrado medio del error de predicción (RMSPE, por sus siglas en inglés) haciendo uso de la validación cruzada.

En los *métodos probabilísticos* el *Kriging* es uno de los métodos para obtener predicciones mediante una combinación ideal de ponderaciones λ , a partir de un subconjunto de localizaciones vecinas, estos interpoladores requieren de un análisis de la estructura de dependencia espacial establecida en una función de semivarianza o covarianza,. Para evaluar la calidad de las predicciones se realiza el proceso de validación cruzada mediante estadísticos como Raíz Media del Cuadrado de los Errores (RMPSE), el estandarizado y el Error Estándar Promedio de los Errores de Predicción (ASEPE).

Existen varios métodos de Kriging dentro de los cuales se encuentran: Simple, Ordinario, Universal, Indicador.

Kriging Simple considera la 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 una aproximación de esta media, la cual se puede considerar $Z(s) = \mu(s) + \varepsilon(s)$

El método de Kriging ordinario también asume 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, está definida por

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

La existencia de tendencia es asumida por el método Kriging universal, dado por una predicción $\hat{Z}(S_0) = \hat{Z}(S_0) = \lambda' Z$ ajustando el modelo de semivariograma para los residuos.

6. DESCRIPCIÓN GENERAL DE LOS DATOS

Para el análisis de calidad y cantidad de datos se hizo de acuerdo a la base de datos correspondiente LST2008_07_27 la cual contiene 158 registros de los cuales cuatro no contienen información los cuales se puede apreciar en la figura 4 de una tonalidad verde, por tanto los análisis realizados son desarrollados a partir de información proporcionada por 154 estaciones meteorológicas croatas de la fecha 27 de julio del 2008. En la Figura 4 se puede observar la distribución de las estaciones meteorológicas donde fue registrada la información, se aprecia que no tienen una distribución uniforme, que algunas zonas del país no cuentan con estaciones y en otras áreas está muy densificada su presencia.

Variable	Descripción	Rango
HRdem	Modelo digital de elevación (m)	1-1453
HRdsea	Distancia a línea costera (km)	0-252.3
HRtwi	Índice de humedad topográfico	12.9-21.4
LST2008_07_27	Temperatura Media Terrestre en Croacia el 27 de julio de 2008 (°C)	21.21 -39.37
N	Coordenada y de cada observación	4694138-5137646
E	Coordenada x de cada observación	387331-842443

Tabla 2. Descripción de variables originales. Fuente: elaboración propia.

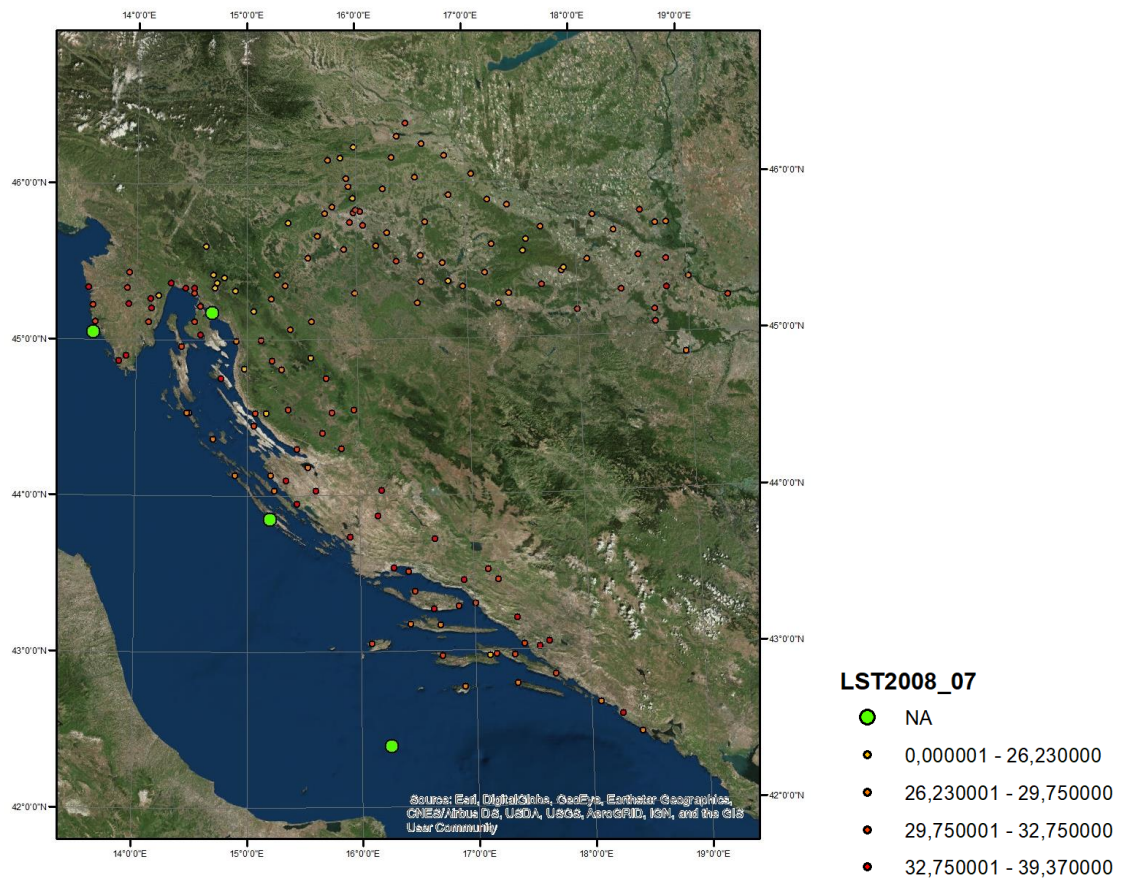


Figura 4. Puntos geolocalizados de la variable Temperatura el día 27 de julio de 2008

7. METODOLOGÍA

7.1 ANÁLISIS EXPLORATORIO DE DATOS

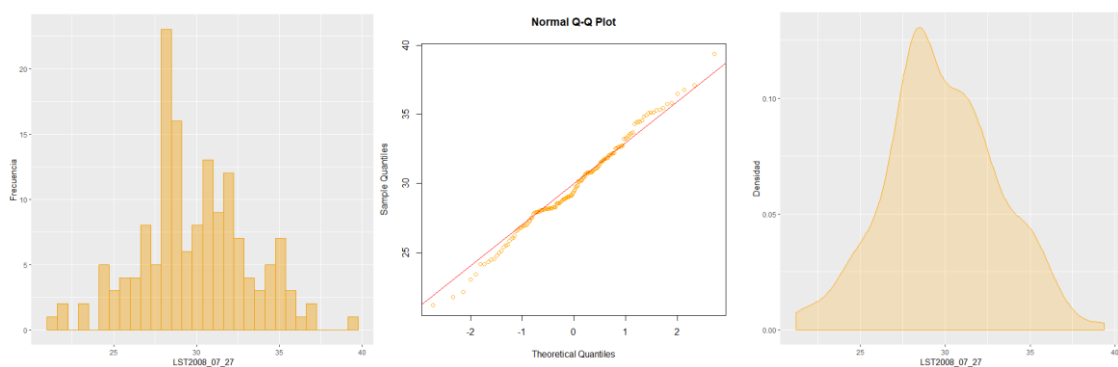


Figura 5. Histograma, QQ plot y densidad de la variable temperatura media terrestre. Fuente: elaboración propia.

N° Obse	Media	Mediana	Desv. Est	Varianza
154	29.83	29.41	3.31	10.98

Tabla 3. Estadística Descriptiva de la variable respuesta Temperatura Media Terrestre.

A partir de las estadísticas convenciones se puede observar que la variable Temperatura Media Terrestre tiene un comportamiento en su mayor parte cumpliendo con una distribución normal, sin embargo presenta valores de temperatura, menores a 25 °C y mayores a 35 °C, que afectan la muestra; la media y la mediana presentan valores similares, en cuanto a la variable estandarizada nos hace suponer de que no será necesario transformar la variable. El cual podemos confirmar con los siguientes test de normalidad que se mostrará en la tabla 4, donde observamos que el $p - value > \alpha$ por lo tanto no rechazamos la hipótesis nula, es decir la variable temperatura sigue una distribución normal.

Test	Statistic	P value
Shapiro-Wilk	0.9929	0.6449
Kolmogorov-Smirnov	0.0685	0.4654
Anderson-Darling	0.4927	0.2145

Tabla 4. Tests de normalidad.

También se realiza un análisis de la variable Temperatura Media Terrestre por medio de un diagrama tridimensional en el cual la elevación de cada punto nos representa la temperatura registrada para el día 27 de julio de 2008.

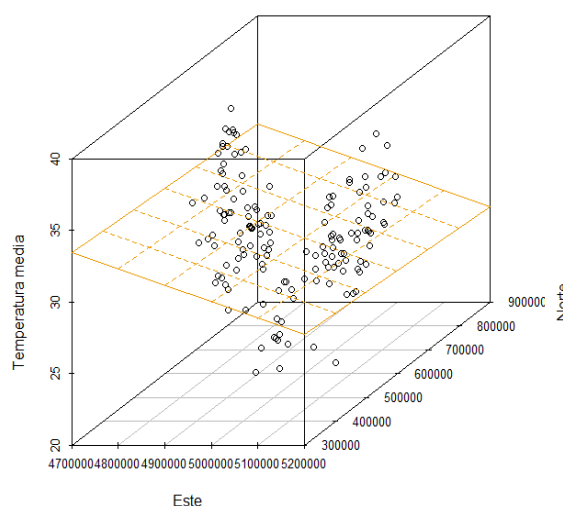


Figura 6. Distribución espacial Tridimensional de la temperatura media terrestre el día 27 de julio de 2008.

7.2 ANÁLISIS ESTRUCTURAL DE DEPENDENCIA ESPACIAL

7.2.1 TENDENCIA

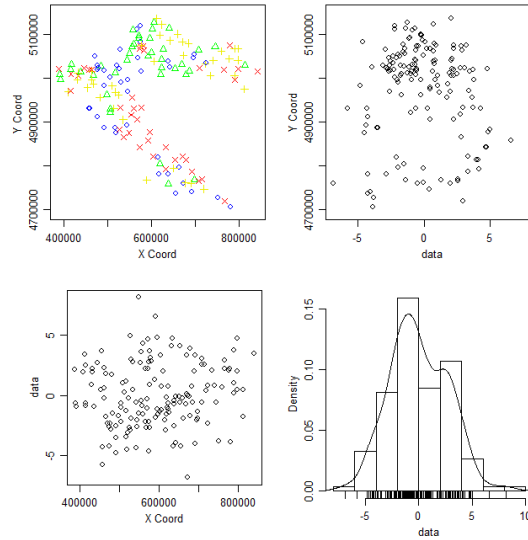


Figura 7. Gráficos de Distribución espacial de la variable temperatura, Perfil coordenada en X, Perfil coordenada en Y, histograma de la variable Temperatura.

En la figura 7 se observa una baja tendencia para confirmarlo realizamos regresiones lineales de la variable Temperatura Media Terrestre como variable dependiente, y como variables independientes *HRdem*, *HRsea*, *HRtwi*, N y E, para ello realizamos los dispersogramas como se muestra en la Figura 8 de la variable dependiente con respecto a las independiente, donde la variable *HRdem* posee una correlación más alta en comparación a las otras dos variables, esto nos permite determinar las variables que se incluirán para realizar un regresión adecuada.

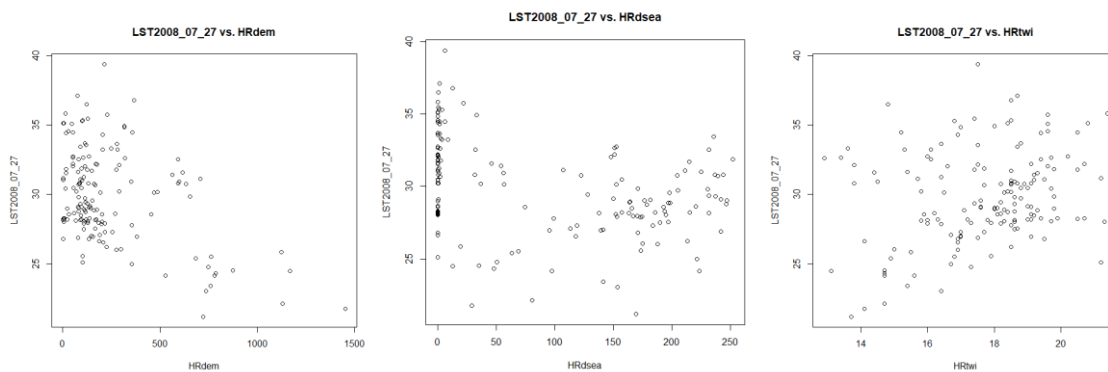


Figura 8. Dispersogramas de covariables con Temperatura.

Se realizó la regresión lineal simple con las variables indicadas anteriormente donde nos indica que algunas variables no son significativas para el modelo, el R cuadrado ajustado es de 0.35, para obtener una mejor información se realizó una regresión lineal por la metodología de Stepwise con los correspondientes parámetros de Backward- Forward y AIC obteniendo un modelo con variables significativamente para el modelo, el R cuadrado ajustado es de 0.343. Indicando una baja tendencia.

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-22.104765751	37.397092127	-0.591	0.5554
HRdem	-0.004558093	0.001050026	-4.341	0.0000262 ***
HRdsea	-0.030554715	0.010126699	-3.017	0.0030 **
HRTw1	0.179080667	0.137195525	1.305	0.1938
N	0.000009126	0.000006897	1.323	0.1878
E	0.000012359	0.000006257	1.975	0.0501 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 2.658 on 148 degrees of freedom				
Multiple R-squared: 0.3781, Adjusted R-squared: 0.3571				
F-statistic: 17.99 on 5 and 148 DF, p-value: 0.00000000000006433				
Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	32.5751790	0.3756043	86.727	< 2e-16 ***
HRdem	-0.0059661	0.0008391	-7.110	0.00000000000432 ***
HRdsea	-0.0142595	0.0023895	-5.968	0.0000000164775 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 2.687 on 151 degrees of freedom				
Multiple R-squared: 0.3516, Adjusted R-squared: 0.343				
F-statistic: 40.93 on 2 and 151 DF, p-value: 6.256e-15				

Figura 9. Modelo lineal clásico (Todas las variables) vs modelo lineal clásico (variables significativas).

Para estar más seguros de la no presencia de tendencia se realizó un modelo de orden dos el cuyos resultados se muestran en la figura 10.

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1209.79958796782989	491.25048211434779	-2.463	0.0149 *
HRdem	-0.00619821959259	0.00085141083650	-7.280	0.00000000000178 ***
HRdsea	-0.00860647228774	0.00348351898538	-2.471	0.0146 *
N	0.00050950298415	0.00020007664687	2.547	0.0119 *
NN	-0.00000000005221	0.00000000002037	-2.564	0.0113 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
Residual standard error: 2.634 on 149 degrees of freedom				
Multiple R-squared: 0.385, Adjusted R-squared: 0.3684				
F-statistic: 23.31 on 4 and 149 DF, p-value: 0.00000000000005575				

Figura 10. Modelo Orden 2.

En esta regresión se realizó el mejor modelo con el proceso mencionado anteriormente, y se obtuvo un R cuadrado ajustado es de 0.36 el cual nos sigue indicando una baja presencia de tendencia. por esta razón no se realizará un kriging universal en los métodos de interpolación..

7.2.2 ANISOTROPÍA

Al realizar una estimación de la anisotropía la figura 11 sugiere que si existe anisotropía ya que en las direcciones 0° y 90° alcanzan una meseta similar de $\sigma^2 \approx 13$ en rangos a entre 150.000 y 200.000, en dirección 45° una meseta de $\sigma^2 \approx 7$ mientras que a 135° $\sigma^2 \approx 4$ al igual cuando se incluyen los residuos presenta valores similares en todas las direcciones, sugiriendo una evaluación y corrección de anisotropía. Lo anterior es comprobado mediante la estimación de este supuesto arrojando según la tabla 5 un valor de radio superior a 1, una dirección diferente de 0 y el ángulo de rotación es True por lo que afirma presencia de anisotropía con rotación y dirección. por esta razón se realizó la corrección de anisotropía, arrojando los resultados expuestos en la tabla 6.

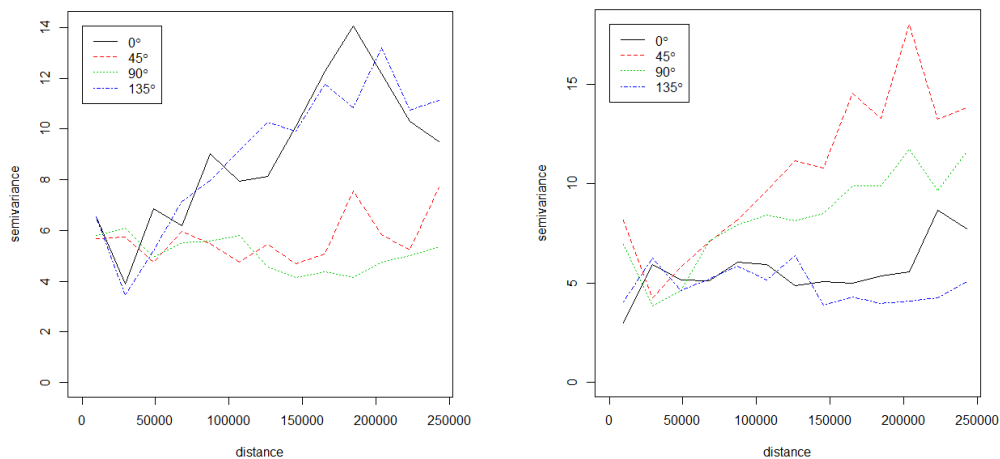


Figura 11. Semivariograma clásico multidireccional de temperatura vs residuos del modelo de regresión significativo

Estimación de la anisotropía	
Radio	1,1700
Dirección	-48,0656
Ángulo de rotación	True

Tabla 5. Evaluación de la Anisotropía en la variable Temperatura transformada.

Estimación de la anisotropía Corregida	
Radio	1
Dirección	0
Ángulo de rotación	False

Tabla 6. Evaluación de la Anisotropía corregida en la variable Temperatura transformada.

7.3 AJUSTE DEL SEMIVARIOGRAMA

Se ajustaron dos semivariogramas uno para la variable Temperatura Media Terrestre y el otro para Temperatura Media Terrestre - mu. Para ello se tuvo en cuenta tomar la distancia máxima entre dos observaciones y dividirla entre dos, la cual es 252609 con esta se realizaron los semivariogramas.

PARA LA VARIABLE TEMPERATURA MEDIA TERRESTRE

Clásico

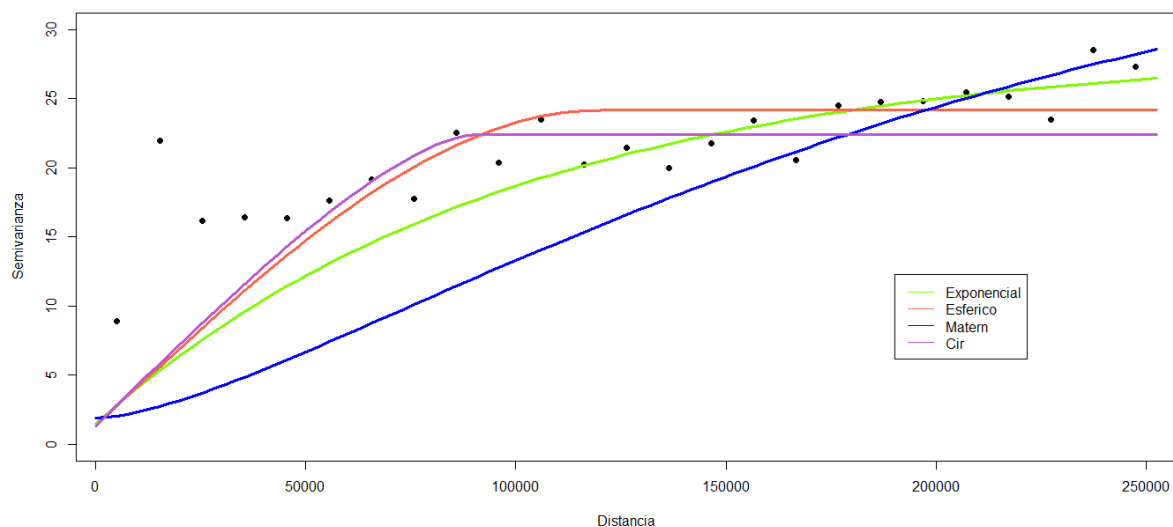


Figura 12. Modelos teóricos ajustados al semivariograma clásico variable temperatura.
Fuente: elaboración propia.

	Co	C1	α	Kappa	AIC
Exponencial	1.5	27.17	99999.98	-	789.34
Esférico	1.3	22.87	119999.99	-	794.19
Circular	1.3	21.13	90000.00	-	793.68
Matern	1.9	38.82	144999.95	0.9	755.68

Tabla 7. Parámetros de modelos ajustados clásico. Fuente: elaboración propia.

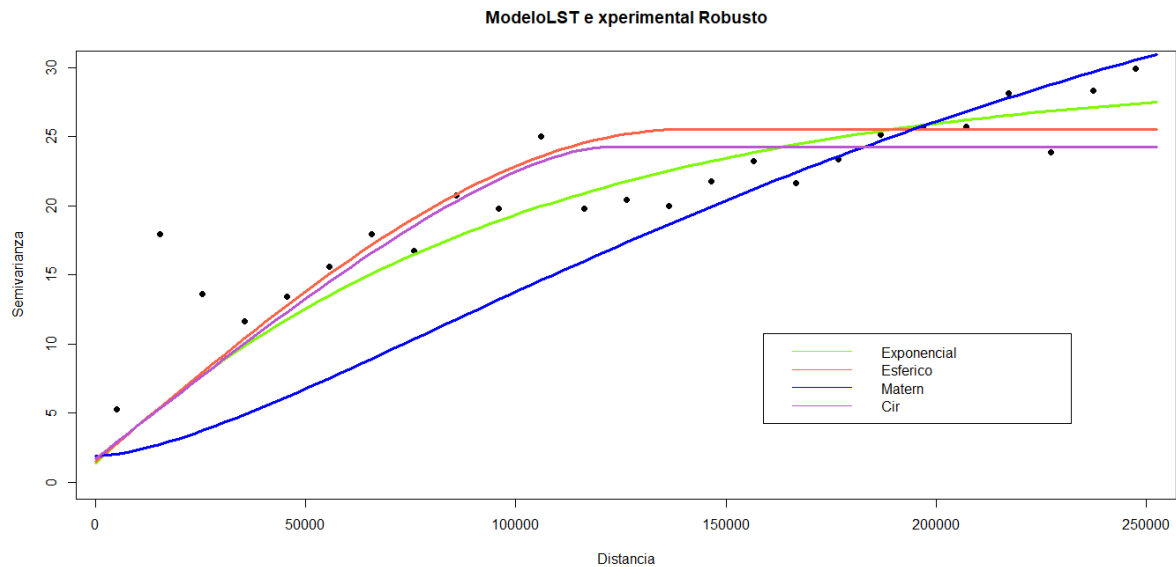


Figura 13. Modelos teóricos ajustados al semivariograma Robusto variable temperatura.
Fuente: elaboración propia.

	Co	C1	α	Kappa	AIC
Exponencial	1.4	28.40	99999.98	-	790.66
Esférico	1.5	24.04	139999.99	-	791.47
Circular	1.7	23.55	120000.00	-	789.53
Matern	1.9	45.29	159999.93	0.9	756.65

Tabla 8. Parámetros de modelos ajustados robusto. Fuente: elaboración propia.

Analizando visualmente los semivariogramas experimentales, se determinó que el semivariograma clásico con el semivariograma teórico Matern, por medio del criterio del AIC el cual nos aproxima a una realidad más cercana.

PARA LA VARIABLE TEMPERATURA MEDIA TERRESTRE - MU

Clásico

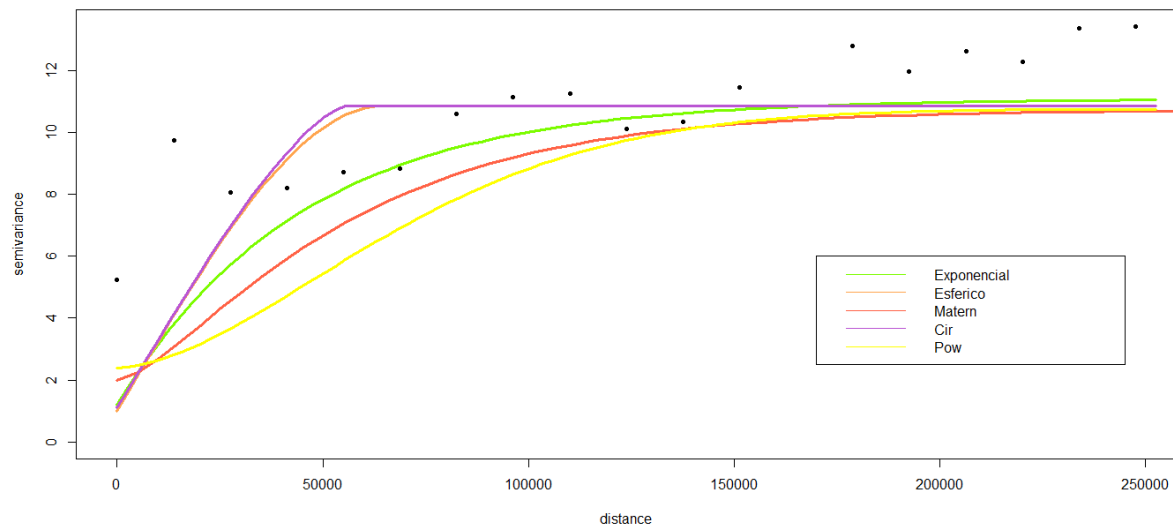


Figura 14. Modelos teóricos ajustados al semivariograma clásico $Z(Si) = Z(Si) - \mu$ para la variable temperatura – μ mediante MVR. Fuente: elaboración propia.

	Co	C1	α	Kappa	AIC
Exponencial	1.2	9.87	44999.99	-	741.24
Esférico	1	9.85	65000.00	-	743.33
Circular	2	8.72	39999.98	0.9	738.06
Matern	1.1	9.75	56000.00	-	745.59

Tabla 9. Parámetros de modelos ajustados clásico $Z(Si) = Z(Si) - \mu$ para la variable temperatura – μ mediante MVR. Fuente: elaboración propia.

Robusto

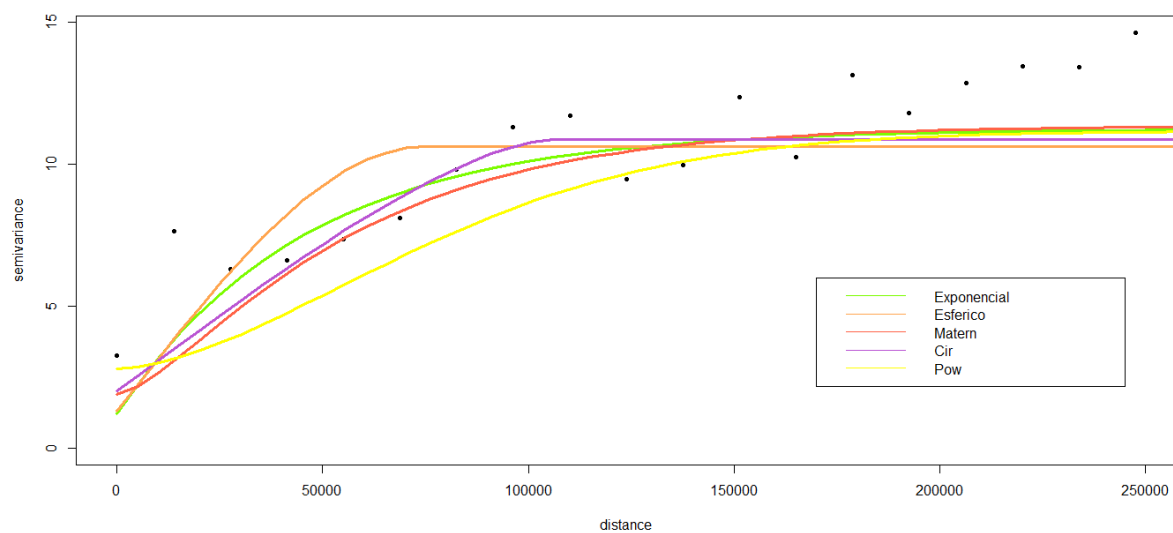


Figura 15. Modelos teóricos ajustados al semivariograma robusto de $Z(Si) = Z(Si) - \mu$ para la variable temperatura – μ mediante MVR. para la variable temperatura – μ . Fuente: elaboración propia.

	Co	C1	α	Kappa	AIC
Exponencial	1.2	10.03	45999.99	-	731.47
Esférico	1.3	9.32	75000.00	-	738.28
Circular	1.9	9.44	39999.99	0.9	740.62
Matern	2	8.86	105000.00	-	729.18

Tabla 10. Parámetros de modelos ajustados robusto $Z(Si) = Z(Si) - \mu$ para la variable temperatura – μ mediante MVR. Fuente: elaboración propia.

Analizando visualmente los semivariogramas experimentales, se determinó que el semivariograma robusto con el semivariograma teórico Circular, por medio del criterio del AIC el cual nos aproxima a una realidad más cercana.

7.4 INTERPOLACIÓN

7.4.1 MÉTODOS DETERMINÍSTICOS

Estos métodos son realizados para obtener una interpolación espacial cuando no se puede realizar un ajuste del semivariograma.

IDW

Para la elaboración de la interpolación mediante el método IDW, se debe determinar el valor de p óptimo para la interpolación; en este caso se obtuvo que el valor de este parámetro es 1.2154, la elección de este valor se hizo usando el criterio del menor RMSPE. A continuación se muestra la interpolación con este método.

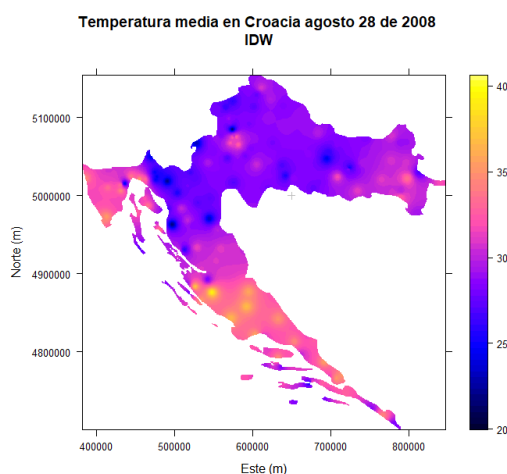
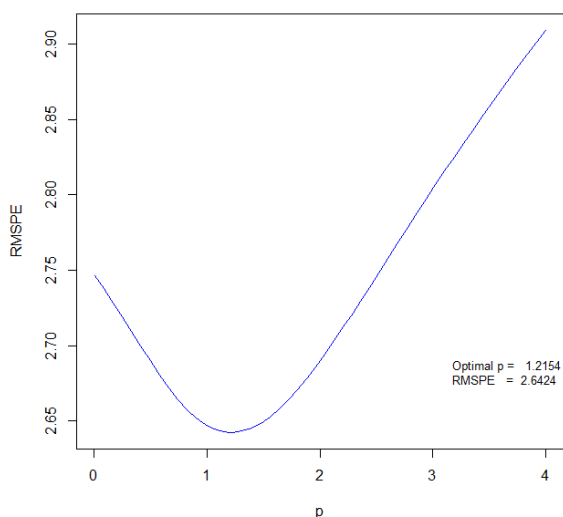


Figura 16. Interpolación IDW usando p óptimo. Fuente: elaboración propia.

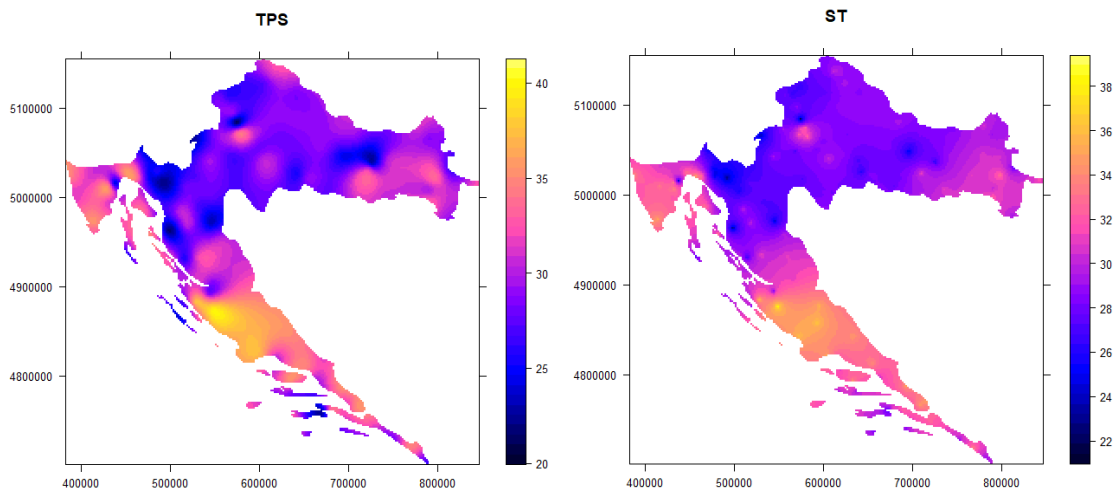
El resultado de las anteriores interpolaciones es consistente con los datos observados; las temperaturas más altas tienden a localizarse en la zona costera del país, mientras que hacia el nor-occidente se ubican los valores más bajos.

FBR

Para llevar a cabo la interpolación con estos métodos, se optimizaron los parámetros y en promedio se realizaron 100 iteraciones evaluadas en 40 puntos usando la función `graph.rbf` de la librería `geospt`, obteniendo los siguientes resultados:

FBR	η	ρ	RMSE
Multicuadrática Inversa	1	0	3.620189
Spline con Tensión	0.003457209	0.0964551	2.320347
Spline Capa Delgada	0.6250498	1.997422	3.186084
CRS	0.004177314	0.006956263	2.320327

Tabla 11. Parámetros optimizados FBR. Fuente: elaboración propia.



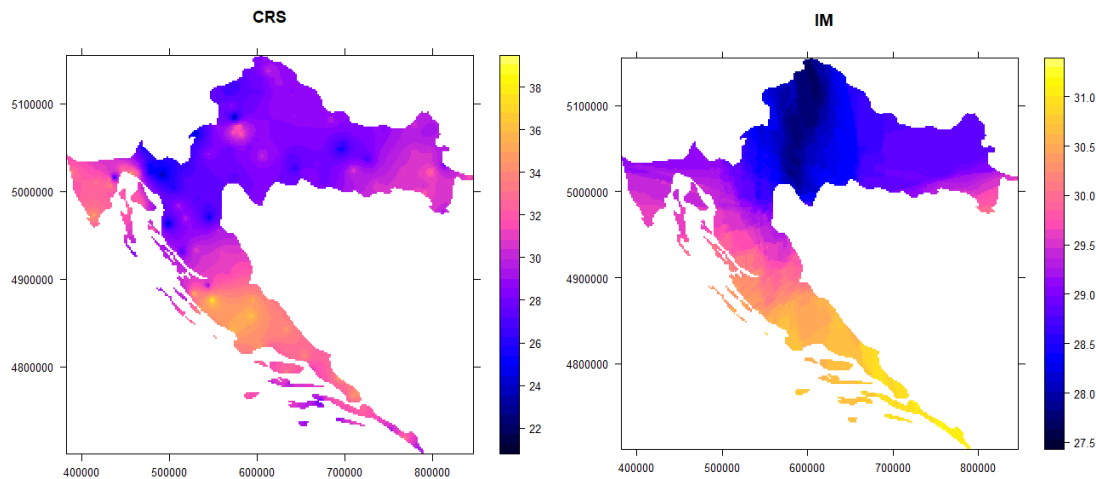


Figura 17. Interpolaciones Funciones de Base Radial. Fuente: elaboración propia.

7.4.2 MÈTODOS PROBABILÍSTICOS

KRIGING SIMPLE

Se toma el semivariograma para la variable Temperatura Media Terrestre - mu para obtener el mejor ajuste de los datos, se procedió a realizar el krigeado para la elaboración del mapa de predicciones con su respectivo mapa de errores.

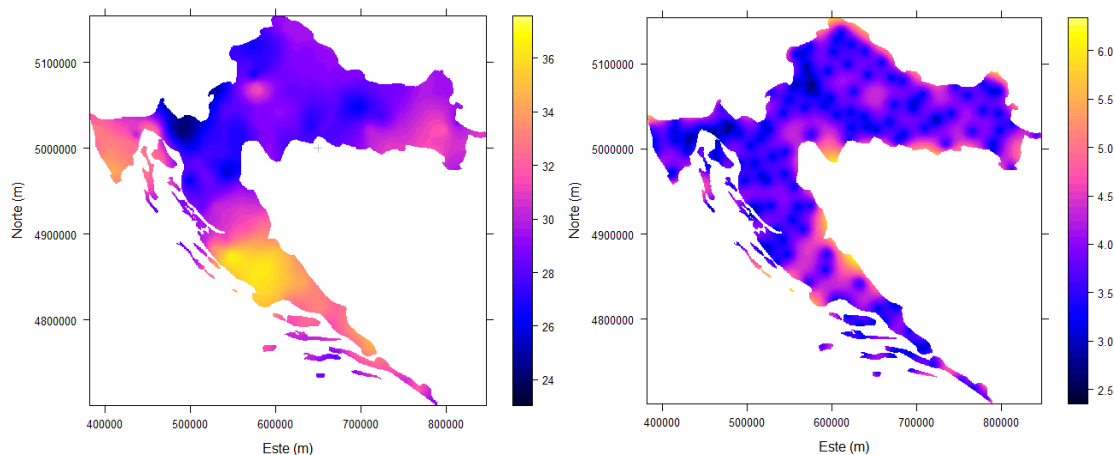


Figura 18. Mapa de interpolación Kriging Simple en la parte izquierda y el mapa de los errores de Ks en la derecha. Fuente: elaboración propia.

Después de realizar la interpolación obtenemos un rango aproximado de temperatura media para el día 27 de julio de 2008, se observan valores de 24 y 36 °C como se observa en la figura

18 , se observan los errores de esta interpolación la cual va en un rango de 2.5 a 6°C donde la mayoría de la zona presenta errores de 2.5 a 4 °C.

KRIGING ORDINARIO

Se toma el semivariograma para la variable Temperatura Media Terrestre para obtener el mejor ajuste de los datos, se procedió a realizar el krigeado para la elaboración del mapa de predicciones con su respectivo mapa de errores.

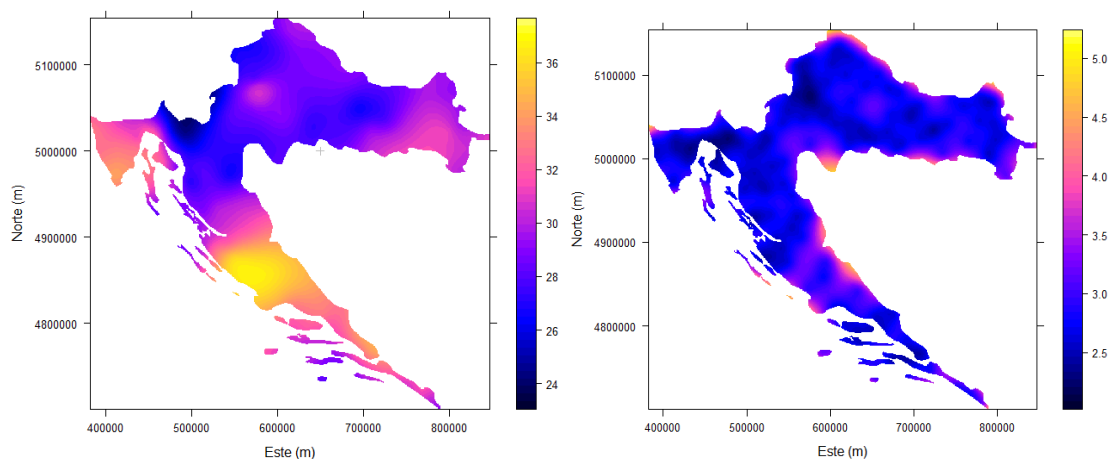


Figura 19. Mapa de interpolación Kriging Ordinario en la parte izquierda y el mapa de los errores de Ko en la derecha. Fuente: elaboración propia.

Después de realizar la interpolación obtenemos un rango aproximado de temperatura media para el día 27 de julio de 2008, se observan valores de 24 y 36 °C como se observa en la figura 18 , se observan los errores de esta interpolación la cual va en un rango de 2.5 a 5°C donde la mayoría de la zona presenta errores de 2.5 a 3°C.

Obteniendo menores errores que el krigeado simple, se puede comprobar que para el día 27 de julio de 2008 es mejor realizar un kriging Ordinario.

7.5 DISEÑO DE RED

A partir de la interpolación espacial obtenida anteriormente, se diseña la red. El uso de un diseño de red muestral permite establecer la relación entre varianzas de predicción máximas, es decir, aquellas que se obtienen en el centro de cada cuadrado, además de los costos asociados a ellas. (Giraldo, 2003). Mediante la función diseñada para R `network.design()`, se diseñaron puntos de 20, 40 y 100 respectivamente, utilizando el variograma seleccionado anteriormente, Matern con parámetros de *pepita* = 1.9, *meseta* = 38.82, *rango* = 145000 y *kappa* = 0.9. el cual fue utilizado para realizar el kriging ordinario.

Los siguientes valores de ASEPE:

	ASEPE
NDP20	2,7397
NDP40	2,2750
NDP100	1,9187

Tabla 12. ASEPE para cada diseño de red. Fuente: elaboración propia.

Con base en los valores anteriores, se eligió la mejor red.



Figura 20. Diseños de red con muestreo regular. Fuente: elaboración propia.

8. CONCLUSIONES

En el análisis exploratorio realizado, se observó que el comportamiento de la temperatura media Terrestre de Croacia, a lo largo de las mediciones capturadas en las 154 estaciones tomadas el 27 de Julio del 2008, seguía una distribución normal, por lo tanto no fue necesaria la transformación de la variable, sin embargo se corrigió la rotación por anisotropía.

En el ajuste de los semivariogramas para la variable temperatura, se obtuvo el semivariograma teórico Matern más adecuado para el modelo experimental robusto, con valores de $pepita = 1.9$, $meseta = 38.82$, $rango = 145000$ y $kappa = 0.9$. Ya que presentaron un mejor ajuste bajo dos estructuras diferentes, por lo que se recomienda el uso en futuras aplicaciones.

Obteniendo menores errores que el krigeado simple, se puede comprobar que para el día 27 de julio de 2008 es mejor realizar un kriging Ordinario ya que presenta menos errores.

La calidad de las predicciones obtenidas depende en gran medida de las características que posea el conjunto de datos usado, ya que según se use un método u otro, se obtendrán ventajas y desventajas específicas.

9. BIBLIOGRAFÍA

- Armstrong, M. (1998). *Basic linear geostatistics*. Berlín & Nueva York.
- Chica-Olmo., M. (2005). *La geoestadística como herramienta de análisis espacial de datos de inventario forestal*. Granada España: Universidad de Granada. Departamento de Geodinámica.
- Giraldo, R. (2003). *Introducción a la geoestadística*. Bogotá: Universidad Nacional de Colombia, Departamento de estadística.
- Goodchild, M., Longley, P., & Maguire, D. (1999). *Geographical information systems*. John Wiley & sons.
- Guedes, L. P., Uribe-Opazo, Johann, J., & Souza, E. G. (2015). *Anisotropía no estudo da variabilidade espacial de algumas variáveis químicas do solo*. Universidade Estadual do Oeste do Paraná.
- Melo, C. (2015). *Estadística espacial teoría y aplicaciones*. Bogotá: Universidad Distrital Francisco José de Caldas.
- Melo, C. E., Melo, O. O., & Melo, S. E. (s.f.). Nuevas funciones geoestadísticas y funciones de base radial en el programa R. *Journal of Statistics Software*, MMMMMM YYYY, Volume VV, Issue II.
- Schanbenberger, O., & Gotway, C. (2005). *Statistical methods for spatial data analysis*. Chapman & Hall/CRC.
- Scott, P., Stone, D. A., & Allen, M. (2004). *Nature Human contribution to the European heatwave of 2003*. 610-614: 432.
- Yrigoyen, C. C. (2003). *Econometría espacial aplicada a la predicción - extrapolación de datos microterritoriales*. Madrid.