
Interpolación espacio-temporal de la temperatura en Reino Unido 2017 con datos faltantes

Spatio-temporal interpolation of temperature in United Kingdom 2017 with missing data

Daniela López Sarmiento^a
daniela.lopez@usantotomas.edu.co

Wilmer Pineda Ríos^b
wilmerpineda@usantotomas.edu.co

Resumen

El análisis de correlación espacial y temporal es útil para conocer la estructura de dependencia en ambas dimensiones, las causas de la variabilidad y realizar la predicción de puntos de la variable de interés, ubicados en sitios cercanos a los observados, o incluso ubicados en el mismo lugar pero en distintos periodos de tiempo. El método de kriging es uno de los métodos más utilizados en interpolación, sin embargo este no admite la presencia de datos incompletos. En este trabajo se realiza una aplicación que permite interpolar datos faltantes en el espacio y el tiempo. En primer lugar se identifica el modelo de covariograma que mejor explica la dependencia espacio-temporal presente en los datos observados descartando los datos faltantes. Después se predicen estos datos con el modelo escogido para completar la información y finalmente se implementa el algoritmo ya conocido de kriging ordinario, como un acercamiento inicial, para interpolar la variable de interés en todo el dominio espacial y temporal. Como aplicación, se utilizan los datos de la temperatura media diaria en Reino Unido durante el año 2017.

—
Palabras clave: Interpolación, modelos espacio-temporales, kriging ordinario, datos faltantes.

Abstract

The spatial and temporal correlation analysis is useful to know the structure of the dependency in both dimensions, the causes of the variability and the prediction of the points of the variable of interest, in places close to those observed, or even located in the same place but in different periods of time. The kriging method is one of the most used methods in interpolation, however this does not admit the presence of incomplete data. In this work an application is proposed that allows interpolating missing data in space and time. First, the covariogram model is identified that best explains the spatio-temporal dependency present in the observed data, discarding the missing data. After this data is predicted with the chosen model to complete the information and finally the algorithm already known for ordinary kriging is implemented, to interpolate the variable of interest throughout the spatial and temporal domain. As an application, the data of the average daily temperature in the United Kingdom is used during the year 2017.

—
Keywords: Interpolation, spatio-temporal models, ordinary kriging, missing data.

^aProf. Estadística. Estudiante Maestría en Estadística Aplicada, U. Santo Tomás, Bogotá

^bPhD(c) en Estadística y Docente de la Facultad de Estadística U. Santo Tomás, Bogotá

Introducción

En los últimos años ha ido aumentando el uso de métodos estadísticos que permiten evaluar la evolución de fenómenos espacio-temporales, debido a los avances tecnológicos y a la necesidad de analizar variables que cambian a través del tiempo y el espacio. La estadística espacio-temporal constituye un campo de estudio en diversas ramas de la ciencia, como el control de la contaminación ambiental, climatología, ciencias sociales, geología, biología, medicina, arqueología o cualquier otra rama relacionada con el estudio de los fenómenos que ocurren en ambos espacios. En términos generales, la geoestadística es útil para encontrar los patrones de dependencia espacial y las causas de variabilidad, con el fin de realizar la interpolación en sitios donde no fue observada la variable de interés.

Sin embargo, en la práctica existe un problema muy frecuente debido a los datos faltantes, ya que generalmente algunos de los puntos observados en el espacio no son medidos en todos los periodos de tiempo. En geoestadística existen diversas aproximaciones para dar solución al problema que presentan dichos vacíos de información durante el proceso de interpolación. Usualmente se eliminan los registros que tienen datos faltantes o simplemente se reemplazan los vacíos con el promedio (Montero et al. 2015). Sin embargo, estas metodologías no son recomendables en la práctica, debido a que en el primer caso se puede estar ignorando información valiosa y en el segundo, se puede estar sesgando el análisis y las conclusiones. Como alternativas a estas limitaciones, se han propuesto varios métodos de interpolación para estimar las observaciones faltantes, que se basan en aplicar los métodos de interpolación espacial a datos espacio-temporales; esto implica que hay cierto grado de pérdida de la información en la dimensión temporal. En este orden de ideas, es preferible implementar métodos de interpolación espacio-temporales, que tengan en cuenta ambas dimensiones. Hoy en día, se utilizan interpolaciones espacio-temporales basadas en métodos espaciales y de regresión. Existen tres métodos principales para estimar datos perdidos en enfoques espaciales y temporales: métodos basados en regresión, ponderación de la distancia inversa y los métodos de kriging (Giraldo et al. 2017).

Actualmente, se encuentra una gran cantidad de aplicaciones de herramientas espaciales y temporales donde se analizan fenómenos climáticos. Por ejemplo, la tesis doctoral de Peña Angulo et al. (2016), donde se estudia la variabilidad espacio-temporal de las temperaturas mensuales en España Peninsular para el período (1951-2010). Para realizar la interpolación de las temperaturas mensuales, se utilizó un método que tiene en cuenta los factores geográficos que influyen en las temperaturas, se estudió la variación temporal de las temperaturas y la espacial a partir de las estimación de la distancia a la que se mantiene un valor de varianza común predeterminado. Otra aplicación fue la de Rodríguez (2005), quien en su tesis doctoral identifica patrones térmicos diferenciados en el territorio utilizando Análisis de Componentes Principales (PCA), método que ha sido ampliamente utilizado para caracterizar la variabilidad de elementos climáticos.

En este trabajo se realiza una aplicación que permite incluir datos incompletos en el proceso de interpolación, utilizando la función “gstat” del paquete “spacetime” en R. Para esto, se utiliza la metodología de kriging ordinario como un primer acercamiento, aplicada a los datos correspondientes a la temperatura media diaria en Reino Unido durante el año 2017.

El trabajo está organizado como sigue: En la primera sección se definen los términos tiempo atmosférico y clima, en la segunda sección se realiza la descripción climatológica de Reino Unido y se define la fuente de información para la aplicación propuesta, en la tercera sección se plantea la teoría sobre la que está basada la estadística espacio-temporal, es decir, en procesos estocásticos espacio-temporales, en la cuarta sección, se muestra el análisis estructural, donde se define la estimación del semivariograma empírico y algunos modelos teóricos de covarianza espacio-temporal. En la quinta sección se muestra el desarrollo teórico de la aplicación, y la forma propuesta para incluir los datos faltantes en el proceso de interpolación. En la sexta sección se muestran los resultados de la aplicación propuesta para la temperatura

de Reino Unido. En la séptima sección se presenta la discusión y en la última sección se muestran las conclusiones.

1. Tiempo atmosférico y clima

Generalmente el tiempo atmosférico y el clima se consideran como términos similares. Sin embargo, el tiempo atmosférico se refiere a la condición de la atmósfera en un lugar determinado, en un lapso de tiempo corto. Este comportamiento de la naturaleza se observa directamente y presenta cambios leves como la variación de la temperatura del aire durante un día.

Por otra parte el clima se define como el estado medio de la atmósfera, o bien como el promedio de los estados del tiempo en un área determinada, calculado sobre observaciones hechas durante muchos años. La Organización Mundial de Meteorología(OMM) establece 30 años. El clima de un lugar está determinado por los elementos y factores climatológicos. Los elementos del clima corresponden a los aspectos físicos que componen el clima: la temperatura del aire, la presión atmosférica, el viento, la humedad del aire y precipitaciones. Los factores del clima son los factores que influyen y/o modifican el comportamiento de cada uno de los elementos del clima. Por ejemplo el relieve, la altitud, las corrientes marinas, la distribución de tierras y océanos, entre otros. (Brenes & Saborío 1995)

Actualmente se mantiene un debate a nivel mundial sobre el cambio climático. Se tienen evidencias del aumento de la concentración de los gases de efecto invernadero (dióxido de carbono, metano, clorofluorocarbonos, entre otros) y el impacto que tiene esto en el aumento de la temperatura media mundial. (Peña Angulo et al. 2016)

2. Clima en Reino Unido

El Reino Unido de Gran Bretaña e Irlanda del Norte es un país de la Unión Europea. Está formado por diversas islas situadas en el océano Atlántico, al oeste de Europa. Sólo tiene fronteras terrestres con Irlanda. Está separado del continente europeo por el canal de la Mancha y el mar del Norte, y de Irlanda por el mar de Irlanda y el canal de San Jorge.

El Reino Unido se encuentra entre las latitudes medias y altas del hemisferio norte, y al oeste del continente euroasiático. Le corresponde un clima frío y húmedo, sin embargo la corriente del Golfo equilibra los climas de esta región, facilitando que se den veranos cálidos, inviernos frescos y abundantes precipitaciones durante todo el año. La estructura del relieve modifica el clima entre la costa y el interior, y las zonas bajas y las altas.

Para la aplicación de este trabajo, se utilizaron los datos públicos obtenidos a partir del sitio web del Proyecto Europeo de Evaluación del Clima y Bases de Datos (ECA & D), que presenta los datos de observaciones diarias de estaciones meteorológicas en toda Europa y se está extendiendo gradualmente a los países de Oriente Medio y África del Norte. El ECA & D se considera el Centro Regional de Clima con datos observados de alta resolución en la Región VI de la Organización Meteorológica Mundial (Europa y Medio Oriente)].

El objetivo de ECA & D es monitorear y analizar el clima y sus cambios. Los datos se encuentran disponibles públicamente para descargar. En la base de datos se incluye una colección de series diarias obtenidas de divisiones climatológicas de los Servicios Meteorológicos e Hidrológicos Nacionales (SMHN),

observatorios y centros de investigación en toda Europa y el Mediterráneo.

El proyecto ECA & D es iniciado por la Red Europea de Apoyo al Clima (ECSN) y está coordinado en el Real Instituto Meteorológico de los Países Bajos, que ahora financia el proyecto después de que fue financiado por EUMETNET, la red de Servicios meteorológicos nacionales europeos.

Para efectos de este estudio, se utilizan los datos diarios de la temperatura media en Reino Unido para el año 2017, medida en $\times 0.1^\circ C$ y observada en diferentes estaciones georreferenciadas. La forma de recolección de la información se basa en un software desarrollado para ECA & D, que permite monitorear y evaluar el clima en diferentes estaciones ubicadas en distintos puntos de Reino Unido. El sistema de coordenadas para los puntos observados corresponde al WGS84 (World Geodetic System 1984) en la zona UTM 30.

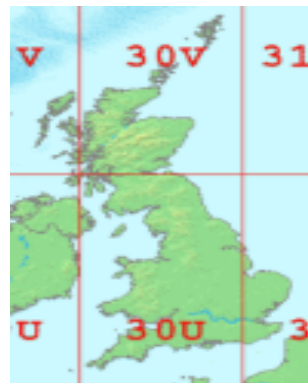


Figura 1: Zonas UTM de Reino Unido. Tomado de: https://es.wikipedia.org/wiki/Sistema_de_coordenadas_universal_transversal_de_Mercator.

A continuación, se presentan las variables que se tuvieron en cuenta:

- **STOID:** Identificador de la estación en la que se mide la temperatura.
- **STANAME:** Nombre de la estación.
- **LAT:** Latitud en coordenadas decimales (+: Norte, -: Sur)
- **LON:** Longitud en coordenadas decimales (+: Este, -: Oeste)
- **DATE:** Fecha.
- **TG:** Temperatura media diaria en $\times 0.1$ grados Celsius

3. Procesos estocásticos espacio-temporales

La geoestadística es una rama que permite analizar mediciones observadas en puntos fijos, definidos en cualquier lugar del espacio, por lo que sus localizaciones definen una superficie espacialmente continua. La teoría detrás de la geoestadística, está basada en procesos estocásticos en los que el espacio de parámetros $D \in \mathbb{R}^d$ es continuo. Esto es una característica esencial de muchos fenómenos observados en

la naturaleza, como por ejemplo la contaminación del aire, la temperatura, la precipitación, la concentración de minerales, entre otros.

Cuando se incluye el factor tiempo, se habla de continuidad espacio-temporal que es una propiedad que caracteriza la relación entre las observaciones ubicadas en diferentes puntos de un dominio espacio-temporal, es decir, si se tienen dos observaciones $Z(s_i, t_i)$ y $Z(s_j, t_j)$, su relación depende del vector diferencia $(s_i - s_j, t_i - t_j)$.

Para la modelización espacio-temporal, se hace una formulación general, basada en el caso espacial, propuesta por Cressie (1993) donde $(s_i, t_i), i = 1, 2, \dots, n$ es un conjunto de ubicaciones del espacio euclídeo d -dimensional $\mathbb{R}^d$ y de la dimensión del tiempo en $\mathbb{R}$. El fenómeno de interés toma un valor aleatorio $Z(s, t)$ en cada localización s y tiempo t . Extendiendo la geoestadística al caso espacio-temporal, se tiene el siguiente proceso aleatorio:

$$\{Z(s, t) : s \in D, t \in T\}, \quad (3.1)$$

con $D \subset \mathbb{R}^2$ y $T \subset \mathbb{R}$ dentro del dominio espacio-temporal (Mas 2016).

3.1. Definiciones básicas

A lo largo de esta sección, se supondrá que para cada conjunto $(s, t) \in \mathbb{R}^d \times \mathbb{R}$, existe la media del proceso que se define como:

$$E(Z(s, t)) = \mu(s, t) < \infty, \quad (3.2)$$

y la función de covarianza como:

$$C((s_i, t_i), (s_j, t_j)) = C(Z(s_i, t_i), Z(s_j, t_j)) < \infty. \quad (3.3)$$

Para dos pares determinados de ubicaciones espacio-temporales (s_i, t_i) y (s_j, t_j) en $\mathbb{R}^d \times \mathbb{R}$.

Definición 1.1.1 Un proceso $\{Z(s, t)\}$ se denomina Gaussiano si, para cada $n \in \mathbb{N}$ y cualquier conjunto de ubicaciones $\{(s_1, t_1), \dots, (s_n, t_n)\}$, el vector aleatorio $\mathbf{Z} = (Z(s_1, t_1), \dots, Z(s_n, t_n))'$ sigue una distribución normal multivariante.

Definición 1.1.2 Un proceso $\{Z(s, t)\}$ es estrictamente estacionario si, para cada $n \in \mathbb{N}$ y para cualquier conjunto de localizaciones $(s_1, t_1), (s_2, t_2), \dots, (s_n, t_n)$, la función de distribución conjunta del vector $(Z(s_1, t_1), Z(s_2, t_2), \dots, Z(s_n, t_n))'$ es idéntica a la del vector

$$(Z(s_1 + h, t_1 + u), Z(s_2 + h, t_2 + u), \dots, Z(s_n + h, t_n + u))', \quad (3.4)$$

para cualquier h y u .

Esta condición es muy difícil de satisfacer en la práctica. Sin embargo, existen otros tipos de estacionariedad menos restrictivos, que son la estacionariedad de segundo orden y la intrínseca (Montero et al. 2015).

Definición 1.1.3 Un proceso espacio-temporal $\{Z(s, t)\}$ es estacionario de segundo orden si:

1. La función media existe y no depende de la localización ni del tiempo, esto es, $E(Z(s, t)) = \mu(s, t) = \mu$ para todo $(s, t) \in \mathbb{R}^d \times \mathbb{R}$.

2. La función de covarianza existe y sólo depende de la diferencia entre las localizaciones involucradas, esto es, $C(Z(s_i, t_i), Z(s_j, t_j)) = C(h, u)$ para todo $(s, t) \in \mathbb{R}^d \times \mathbb{R}$ siendo $h = s_i - s_j$ el vector diferencia entre dichas localizaciones y $u = t_i - t_j$ el vector diferencia entre tiempos. La función $C(\cdot, \cdot)$ recibe el nombre de covariograma.

Definición 1.1.4 Un proceso espacio-temporal $\{Z(s, t)\}$ es intrínsecamente estacionario, si el incremento $Z(s+h, t+u) - Z(s, t)$ es estacionario de segundo orden en el espacio y el tiempo para cualquier (h, u) fijo.

4. Análisis estructural espacio-temporal

Dado que es necesario conocer la estructura de dependencia espacio-temporal o correlación presente en los datos observados, inicialmente se construye un semivariograma o una función de covarianza empírica. Sin embargo, lo usual es que no cumpla los supuestos necesarios para que la función sea un modelo válido. Por lo que se deben ajustar varios modelos válidos y compararlos con la parte empírica. A continuación se presentan algunas definiciones necesarias para el desarrollo de la propuesta:

1. El variograma espacio-temporal está definido por la siguiente función:

$$2\gamma((s_i, t_i), (s_j, t_j)) = V(Z(s_i, t_i) - Z(s_j, t_j)), \quad (4.1)$$

mientras que el semivariograma se define por la mitad de esta cantidad. (Montero et al. 2015)

2. Si el proceso $Z(s, t)$ tiene un semivariograma intrínsecamente estacionario en el espacio y el tiempo, entonces se dice que tiene semivariograma intrínsecamente estacionario, expresado de la siguiente manera:

$$\gamma((s_i, t_i), (s_j, t_j)) = \gamma(h, u), \quad (4.2)$$

donde $h = s_i - s_j$ y $u = t_i - t_j$ representan la distancia en el espacio y el tiempo respectivamente.

4.1. Estimación del semivariograma empírico

Existen dos alternativas principales para estimar el semivariograma (y su función de covarianza si el proceso también es estacionario de segundo orden). Una es el estimador clásico propuesto por Matheron (1989) y el estimador robusto propuesto por Cressie & Hawkins (1980), donde $Z(\cdot, \cdot)$ es un proceso intrínsecamente estacionario observado en un conjunto de n ubicaciones espacio-temporales $\{(s_1, t_1) \dots (s_n, t_n)\}$.

■ Estimador clásico

Su forma general está dada por:

$$\gamma(h(l), u(k)) = \frac{1}{2\#N(h(l), u(k))} \sum_{(s_i, t_i), (s_j, t_j) \in N(h(l), u(k))} (Z(s_i, t_i) - Z(s_j, t_j))^2, \quad (4.3)$$

donde $N(h(l), u(k)) = \{(s_i, t_i), (s_j, t_j) : s_i - s_j \in G(h(l)), t_i - t_j \in G(u(k))\}$, $G(h(l))$ es la región de tolerancia sobre $\mathbb{R}^d$ alrededor de $h(l)$ y $G(u(k))$ una región de tolerancia sobre $\mathbb{R}$ alrededor de $u(k)$, y $\#N(h(l), u(k))$ el número de elementos diferentes en $N(h(l), u(k))$, con $l = 1, \dots, L$ y $k = 1, \dots, K$.

En cuanto a la función covarianza, el estimador por momentos toma la siguiente forma:

$$\hat{C}(h(l), u(k)) = \frac{1}{\#N(h(l), u(k))} \sum_{(s_i, t_i), (s_j, t_j) \in N(h(l), u(k))} (Z(s_i, t_i) - \bar{Z})(Z(s_j, t_j) - \bar{Z}), \quad (4.4)$$

donde $\bar{Z} = \frac{1}{n} \sum_i Z(s_i, t_i)$ es el estimador de la media μ del campo aleatorio, y $N(h(l), u(k))$ es definido previamente. La ventaja del estimador clásico es su facilidad de cálculo. Sin embargo, tiene una gran desventaja y es que no es robusto, por lo tanto es muy sensible a valores extremos.

■ Estimador Robusto

Este estimador es propuesto por Cressie & Hawkins (1980), con el fin de mitigar el efecto de los valores atípicos y se define de la siguiente forma:

$$\hat{\gamma}(h(l), u(k)) = \left(\frac{1}{2\#N(h(l), u(k))} \sum_{(s_i, t_i), (s_j, t_j) \in N(h(l), u(k))} |Z(s_i, t_i) - Z(s_j, t_j)|^{\frac{1}{2}} \right)^4 \times \left(0.457 + \frac{0.494}{\#N(h(l), u(k))} \right)^{-1}. \quad (4.5)$$

En general, los estimadores de la función de covarianza empírica no cumplen las condiciones suficientes que permiten realizar la interpolación. Por lo que, en la práctica, es necesario seleccionar un modelo de covarianza o semivariograma paramétrico, es decir, válido, donde se estiman los parámetros del modelo que mejor se ajustan a los valores del estimador empírico.

4.2. Modelos de covarianza espacio-temporal

Dado que el semivariograma empírico usualmente no satisface los supuestos para ser válido, no puede utilizarse para la predicción espacio-temporal de kriging. Por esto, es necesario ajustar una serie de modelos teóricos válidos y escoger el que mejor describa el tipo de continuidad asumida para la variable regionalizada.

Bevilacqua (2008), propone un estimador basado en la verosimilitud compuesta ponderada que mejora la eficiencia y reduce la carga de cálculo:

$$WCL(\theta, \delta) = \frac{1}{W_{N, \delta}} \sum_{i=1}^N \sum_{j>i}^N \ell(U_{ij}; \theta) w_{ij}(\delta), \quad (4.6)$$

Donde n es el número de puntos observados s_i , $\delta = (\delta, \delta_t)'$ es un vector compuesto por dos vectores de diferencias, una espacial y la otra temporal, $w_{ij}(\delta)$ son los pesos que dependen de las diferencias $(s_i - s_j)$ y $(t_i - t_j)$ y $W_{N, \delta} = \sum_{i=1}^N \sum_{j>i}^N w_{ij}(\delta)$.

4.2.1. Modelo métrico

También conocido como modelo de distancia combinada, se basa en la hipótesis de que la covarianza espacio-temporal tiene una estructura uniforme a lo largo de todo el dominio, lo que implica que la

función de covarianza espacial marginal sigue el mismo tipo de modelo que la función de covarianza temporal marginal y, además, ambos tienen la misma meseta y no necesariamente el mismo rango. Su función de covarianza está dada por:

$$C(h, u) = C(|h| + c|u|), (h, u) \in \mathbb{R}^d \times \mathbb{R}, c > 0, \quad (4.7)$$

Donde $|h| + c|u|$ es la distancia en $\mathbb{R}^d \times \mathbb{R}$ y c es una constante positiva.

4.2.2. Modelo suma

El modelo suma, propuesto por Rouhani & Hall (1989), consiste en descomponer la covarianza espacio-temporal en la suma de la función de covarianza netamente espacial y la función de covarianza temporal. Este modelo también es conocido como modelo zonal o modelo lineal y está definido de la siguiente manera:

$$C((s_i, t_i)(s_j, t_j)) = C_s(s_i, s_j) + C_t(t_i, t_j), \forall s_i, s_j \in D \subset \mathbb{R}^d, \forall t_i, t_j \in T \subset \mathbb{R},$$

donde C_s es la función de covarianza espacial definida en $\mathbb{R}^d$ y C_t la función de covarianza temporal definida en $\mathbb{R}$.

Este modelo fue ampliamente utilizado en un principio, por su simplicidad. Sin embargo tiene una gran desventaja debido a que la suma de los dos modelos de covarianza, el espacial y el temporal, generalmente es definida positiva semi-positiva, lo cual implica que la matriz de covarianzas en las ecuaciones de kriging puede no ser invertible. Una de las soluciones propuestas para este inconveniente, es combinar el modelo suma con otros modelos.

4.2.3. Modelo producto

Este es uno de los modelos de covariograma separables más utilizados en la literatura (Rodríguez-Iturbe & Mejía 1974). De hecho, se ha utilizado en muchas ocasiones, debido a su simplicidad, y no tanto por lo bien que se ajusta a los datos de estudio. Generalmente se define, en términos del covariograma como:

$$C((s_1, t_1)(s_2, t_2)) = C_s(s_1, s_2)C_t(t_1, t_2), \forall s_i, s_j \in D \subset \mathbb{D}^d, \forall t_i, t_j \in T \subset \mathbb{R}, \quad (4.8)$$

siendo C_s y C_t funciones definidas positivas.

En este modelo producto cabe resaltar que la definición de separabilidad implica que la estructura de covarianza temporal es la misma para cualquier separación espacial. De igual forma, la estructura de dependencia espacial mantiene la misma forma para los diferentes rezagos de tiempo.

4.2.4. Modelo producto-suma

Los modelos producto y suma presentados por separado, pueden combinarse con el fin de generar estructuras de covarianza espacio-temporales (Montero et al. 2015). El modelo estacionario de suma de producto está dado por:

$$C(h, u) = k_1 C_s(h) C_t(u) + k_2 C_s(h) + k_3 C_t(u), \quad (4.9)$$

donde C_s y C_t son funciones de covarianza y $k_1 > 0$, $k_2 \geq 0$, $k_3 \geq 0$ son constantes que garantizan la validez del modelo (Montero et al. 2015).

5. Predicción espacio-temporal incluyendo datos faltantes

El objetivo del método de kriging es predecir un valor de la variable en un punto desconocido $Z(s_0, t_0)$ en un punto no observado (s_0, t_0) . Para esto, toda la información disponible de la variable regionalizada es utilizada. Sea $\{Z(s, t) : s \in D, t \in T\}$ con $D \subset \mathbb{R}^d$ y $T \subset \mathbb{R}$ un campo aleatorio espacio-temporal, observado en un conjunto de n ubicaciones espacio-temporales $\{Z(s_1, t_1), \dots, Z(s_n, t_n)\}$.

De todos los posibles predictores lineales, la técnica de kriging selecciona aquel predictor lineal que sea insesgado y además sea óptimo en el sentido que minimice el error cuadrático medio de la predicción (Melo 2013). Por este motivo el predictor resultante recibe el nombre de BLUP (Best Linear Unbiased Predictor).

Para realizar la predicción lineal, se utiliza el siguiente predictor lineal :

$$Z^*(s_0, t_0) = \sum_{i=1}^n \lambda_i Z(s_i, t_i) = \boldsymbol{\lambda}' \mathbf{Z}, \quad (5.1)$$

A continuación se describe la aplicación llevada a cabo para incluir los datos faltantes en el proceso de predicción. Una vez que ya se tiene escogido el modelo de covariograma que mejor se ajusta a los datos. Asumiendo que $\{Z(s, t)\}$ es un proceso intrínsecamente estacionario, con función de semivarianza conocida $\gamma(h, u)$. Con el fin de encontrar los valores de $\boldsymbol{\lambda}$ que minimicen el Error Cuadrático Medio (ECM), se impone la restricción (5.2) con el fin de obtener insesgamiento.

$$\boldsymbol{\lambda}' \mathbf{1} = \sum_{i=1}^n \lambda_i = 1 \quad (5.2)$$

Por lo tanto el sistema de ecuaciones asociado al método kriging ordinario, en términos de la función de semivariograma es:

$$\sum_{j=1}^n \lambda_j \gamma(s_i - s_j, t_i - t_j) + \alpha = \gamma(s_i - s_0, t_i - t_0), \forall i = 1, \dots, n \quad (5.3)$$

Donde α es el multiplicador de Lagrange asociado a la condición de insesgamiento y γ es la estimación de los valores de la semivarianza basados en un modelo de covariograma dado. Posteriormente, se construye la matriz simétrica de semivariograma estimada $\boldsymbol{\Gamma} = (\hat{\gamma}(s_i - s_j, t_i - t_j))_{i,j}$ con la estimación de los sitios tiempos observados.

El procedimiento desarrollado para encontrar los pesos $\boldsymbol{\lambda}$ se muestra a continuación:

$$\boldsymbol{\lambda}' \boldsymbol{\Gamma} + \alpha \mathbf{1}_n = \boldsymbol{\gamma} \quad (5.4)$$

$$\lambda' \mathbf{1}_n = 1. \quad (5.5)$$

Expresado en forma matricial,

$$\begin{bmatrix} \Gamma & \mathbf{1} \\ \mathbf{1}' & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \alpha \end{bmatrix} = \begin{bmatrix} \gamma \\ 1 \end{bmatrix} \quad (5.6)$$

$$\begin{bmatrix} \lambda \\ \alpha \end{bmatrix} = \begin{bmatrix} \Gamma & \mathbf{1} \\ \mathbf{1}' & 0 \end{bmatrix}^{-1} \begin{bmatrix} \gamma \\ 1 \end{bmatrix} \quad (5.7)$$

Finalmente se obtiene la siguiente ecuación:

$$\lambda = \left[\gamma + \mathbf{1} \left(\frac{1 - \mathbf{1}' \Gamma^{-1} \mathbf{1}}{\mathbf{1}' \Gamma^{-1} \mathbf{1}} \right) \right]' \Gamma^{-1} \quad (5.8)$$

6. Resultados

Posterior al planteamiento teórico en el que se fundamenta la estimación de la función de covarianza y la predicción espacio-temporal de kriging, se presentan los resultados de la aplicación propuesta.

El primer paso para abordar esta propuesta consiste en un análisis descriptivo de las variables observadas en el espacio y en el tiempo, donde se grafica el comportamiento espacial y temporal de la temperatura media diaria en 34 estaciones diferentes de Reino Unido. Después, se construye un semivariograma empírico que permite identificar la estructura de dependencia espacio-temporal de los datos, y de acuerdo a su comportamiento, se ajustan una serie de modelos teóricos con el fin de compararlos y escoger el que tenga un mejor ajuste, es decir un menor error cuadrático medio. Algunos de los modelos teóricos más conocidos son el exponencial, esférico, gaussiano, Mátern, periódico, ondulado, entre otros. Una vez que se tiene el ajuste del semivariograma teórico escogido, se realiza la predicción espacio-temporal de la temperatura en los sitios y tiempos no observados, incluyendo los datos faltantes. Para esto existen diversos métodos, como la familia kriging ordinario, simple, universal e indicador, cokriging, así como los polígonos de Thiessen y ponderación en función inversa de la distancia, entre otros (Abad Encalada 2014). Para efectos de este estudio, se utilizará el método de kriging ordinario, como un acercamiento inicial, que se basa en un predictor lineal que se caracteriza por ser insesgado y además por ser óptimo, en el sentido que minimiza el error cuadrático medio de la predicción (Melo 2013).

Para el manejo de los datos espacio-temporales se utiliza el software R, mediante el uso del paquete “spacetime” y para la estimación de la covarianza espacio-temporal, modelos de variograma e interpolación espacio-temporal se utiliza “gstat” (Pebesma 2001). La implementación de la propuesta fue programada en R, en la que se predicen los valores de la temperatura en los sitios tiempos no observados, y se rellenan los datos faltantes con dichas predicciones. Posteriormente, se implementa el algoritmo ya conocido de kriging, que se encuentra definido a profundidad en Montero et al. (2015) para predecir la temperatura media en todo el dominio de interés espacial y temporal.

La información que se tiene disponible, corresponde a la temperatura diaria en 2017 (365 días) en 34 estaciones de Reino Unido, es decir 12410 sitios tiempos, con el 98.61%(12238) de observaciones y el 1.4% con datos faltantes. En la tabla 1, se muestra el resumen de la temperatura media durante todo el año 2017, donde se evidencia que la temperatura media mínima observada fue de $-54 \times 0.1^\circ C$ y la máxima de $260 \times 0.1^\circ C$. Además, la media y la mediana se encuentran muy cercanas, por lo que se afirma que la temperatura media en Reino Unido sigue una distribución simétrica.

Tabla 1: Resumen de la Temperatura media diaria en Reino Unido para 2017

Resumen Temperatura Media	
Min	$-54 \times 0.1^\circ C$
Q1	$72 \times 0.1^\circ C$
Mediana	$108 \times 0.1^\circ C$
Media	$106,4 \times 0.1^\circ C$
Q3	$144 \times 0.1^\circ C$
Max	$260 \times 0.1^\circ C$
Datos incompletos	172

Fuente: Elaboración propia

Posteriormente, se muestran las series temporales en cada una de las 34 estaciones, en las que se observa el comportamiento de la temperatura durante el año 2017 (Figura 2), con algunos espacios debido a los datos faltantes en determinados puntos del tiempo. En general, se observa que las series no son estacionarias en media ni en varianza, y además los datos presentan variaciones regulares y previsibles que se repiten cada año (estacionalidad), donde el mes de julio presenta los picos más altos, con entre $150 \times 0.1^\circ C$ y $200 \times 0.1^\circ C$, mientras que en los primeros y los últimos meses del año presentan las temperaturas más bajas.

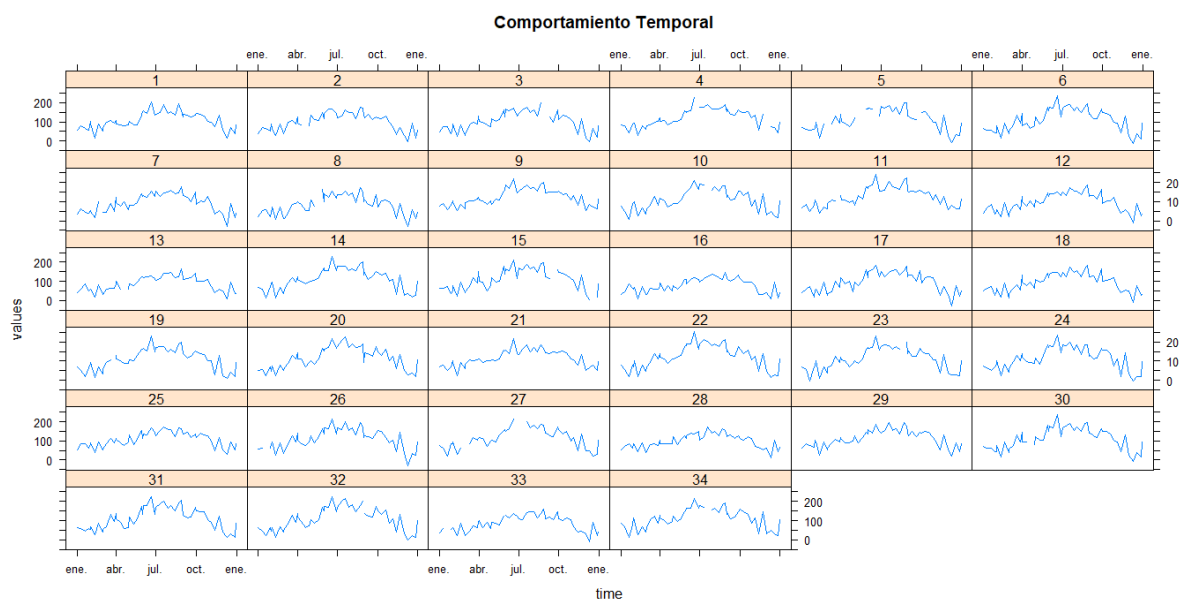


Figura 2: Series de tiempo de la temperatura media en Reino Unido en cada una de las 34 estaciones observadas.

En la figura 3, se observa el comportamiento espacial de la temperatura media durante algunos días del año 2017 en cada una de las 34 estaciones. Se evidencia que los primeros días del año, entre enero y marzo se presentan las temperaturas más bajas, entre -31 y $83 \times 0.1^\circ C$ junto con los últimos meses de noviembre y diciembre, mientras que los días de junio y julio son las que presentan temperaturas más altas, entre 160 y $256 \times 0.1^\circ C$.

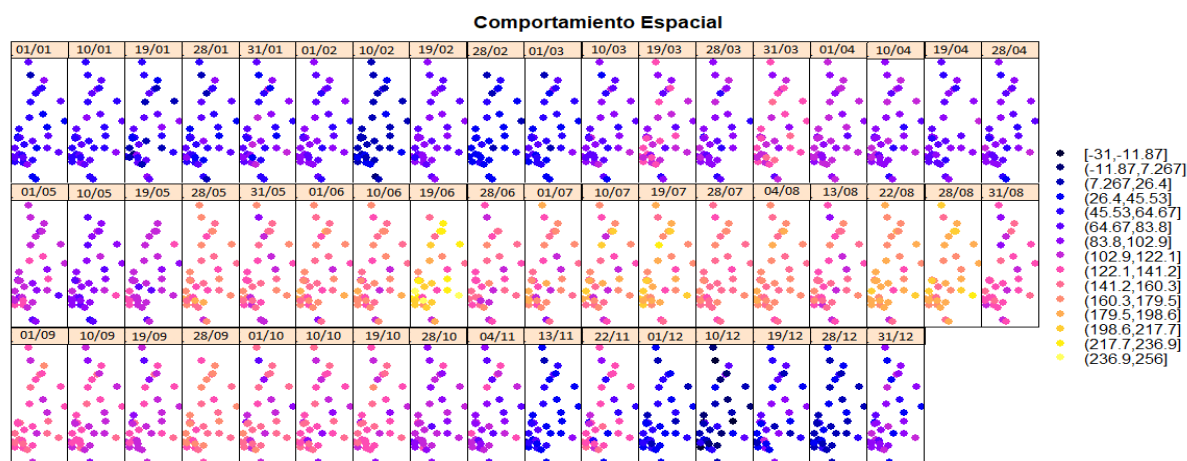


Figura 3: Comportamiento espacial de la temperatura media en Reino Unido en algunos días del año 2017.

6.1. Funciones de semivarianza

Antes de definir un proceso óptimo de predicción, es fundamental expresar, en una función de semivarianza, la estructura de dependencia espacial o correlación presente en las observaciones. Sin embargo, dichas funciones deben cumplir ciertos requisitos, mencionados en el marco teórico que en la práctica usualmente no se satisfacen, debido a que solo se cuenta con una realización observada del campo aleatorio bajo estudio, es decir, un conjunto de datos georreferenciados en un dominio dado. Por esta razón, es necesario ajustar una serie de modelos teóricos, conocidos también como modelos válidos.

Con el fin de definir el modelo válido que mejor se ajuste a los datos observados, se utilizó el modelo separable, el modelo métrico y el modelo producto-suma. En la figura 4, se muestra el semivariograma empírico espacio-temporal para la temperatura media en Reino Unido obtenido a partir de la función “variogram” de la librería “gstat” en R que utiliza el método de verosimilitud compuesta ponderada (Bevilacqua 2008) para encontrar la estimación de los parámetros del modelo. En la derecha, se presenta en dos dimensiones y a la izquierda en tres dimensiones. En cuanto a la parte espacial, se observa en ambos gráficos que el variograma tiene un comportamiento oscilante y a medida que pasa el tiempo, se va extinguiendo dicho comportamiento, además disminuye la correlación espacial de las observaciones. A partir de este gráfico, se obtienen las estimaciones de los parámetros para el efecto pepita, la meseta y el rango de la dimensión espacial y temporal.

Ajuste del Modelo Separable Posteriormente, se construye el variograma espacio-temporal separable, y se ajustan diferentes combinaciones de modelos con el fin de comparar y escoger el que menor error cuadrático medio produzca. En la tabla 2, se observa que el modelo que tiene menor ECM es el ondulado en la parte espacial y exponencial en la parte temporal.

Ajuste del Modelo Métrico En la tabla 3 se observa que en general el ECM en los variogramas métricos disminuye considerablemente respecto al de los separables. Como se mencionó en la sección 2, este tipo de variograma supone que la covarianza espacio temporal tiene una estructura uniforme a lo largo de todo el dominio, lo que implica que la función de covarianza espacial marginal sigue el mismo tipo de modelo que la función de covarianza temporal marginal. En este caso, se ajusta un tipo de modelo conjunto y se muestra que el que tiene menor ECM es el exponencial.

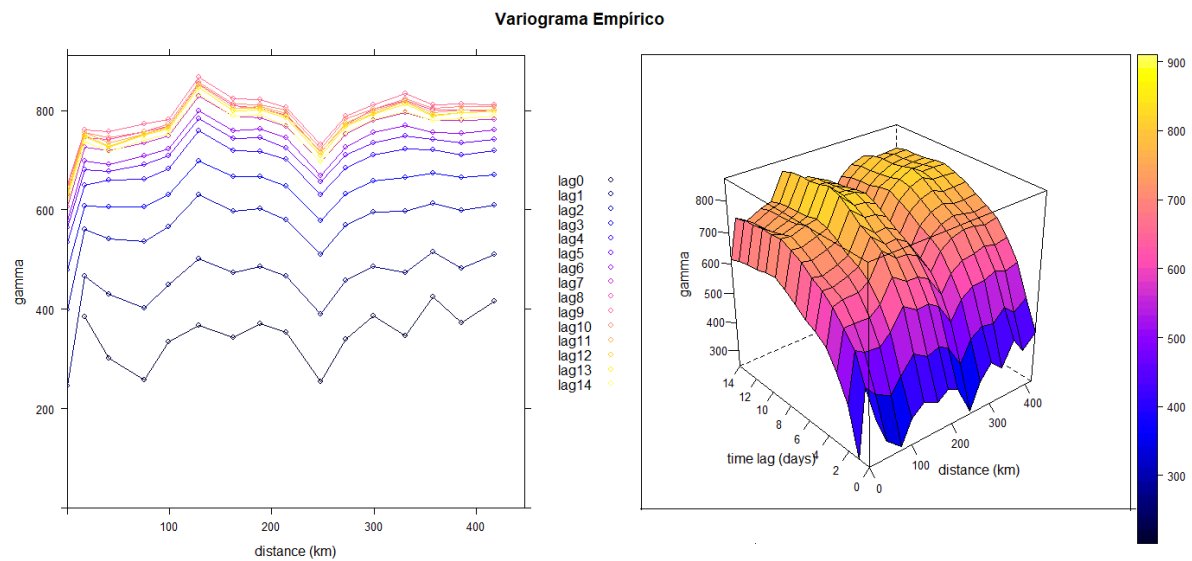


Figura 4: Variograma empírico espacio temporal.

Tabla 2: ECM del ajuste del modelo separable.

Modelo	Space	Time	ECM
Sep_1	Períodico	Mátern	65807,56
Sep_2	Ondulado	Exponencial	65604,02
Sep_3	Períodico	Gaussiano	65956,92
Sep_4	Ondulado	Esférico	65604,03

Tabla 3: ECM del ajuste del modelo métrico.

Modelo	Joint	ECM
Met_1	Períodico	19346,69
Met_2	Ondulado	3615,42
Met_3	Mátern	2604,05
Met_4	Exponencial	2603,09

Ajuste del Modelo Producto-Suma En la tabla 4 se muestran los tipos de modelos usados en el modelo producto suma, y se observa que los que tienen el modelo ondulado en la dimensión espacial tienen un menor ECM. En este caso, el modelo escogido es el ondulado en la parte espacial y el exponencial en la temporal.

Tabla 4: ECM del ajuste del modelo producto-suma.

Modelo	Space	Time	ECM
Prodsum_1	Periodico	Exponencial	78649,62
Prodsum_2	Ondulado	Esférico	54512,58
Prodsum_3	Periodico	Mátern	78649,62
Prodsum_4	Ondulado	Exponencial	35866,31

Finalmente, el modelo que se escogió fue el métrico debido a que presenta un menor ECM en comparación con los otros modelos. En la figura 5, se muestra la comparación de los modelos ajustados y el variograma empírico. Gráficamente se confirma que el modelo métrico es el que tiene el comportamiento más similar al empírico. El modelo producto-suma muestra el comportamiento de ondas parecidas al empírico, sin embargo no ajusta la curva que se observa en el empírico. En cuanto al separable, es el que menos se ajusta a la estructura de dependencia espacio-temporal de los datos.

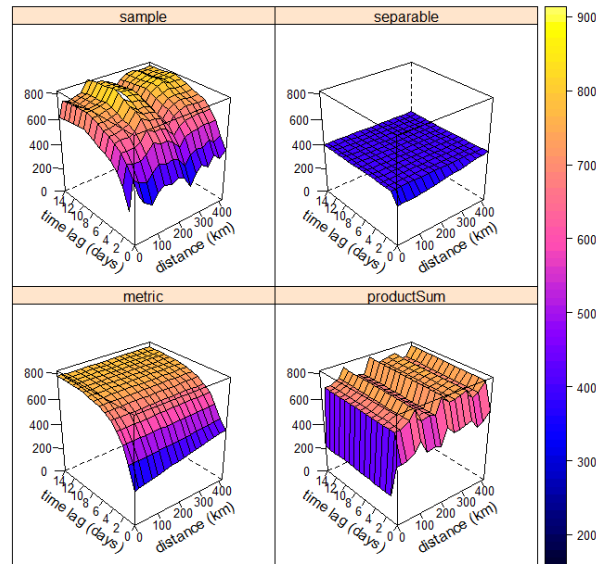


Figura 5: Comparación del covariograma empírico y los modelos teóricos.

6.2. Interpolación

Después de identificar el modelo de covariograma que mejor explica la dependencia espacio-temporal presente en los datos observados, se implementa la propuesta de predicción de kriging incluyendo los datos perdidos, para lo que fue necesario realizar la programación en R, en la que se predicen los valores de la temperatura en los 172 sitios tiempos no observados, y se rellenan los datos faltantes con dichas predicciones. Posteriormente se implementa el algoritmo ya conocido de kriging, que se encuentra definido en la sección 3 o por Montero et al. (2015) para predecir la temperatura media en todo el dominio de interés espacial y temporal.

Para aplicar el predictor lineal definido en la ecuación (5.1), se necesitan los pesos λ_i que se obtienen de la ecuación (5.8). Primero, se calcula la matriz cuadrada $\mathbf{\Gamma}^{-1}$ con ceros en la diagonal, para la cual, solo se toman los sitios tiempos observados. Para este caso, quedaron 12238 registros y a partir de las distancias euclidianas de sus coordenadas y de sus tiempos, se obtienen las estimaciones de los gammas $\mathbf{\Gamma}(\hat{\gamma}(s_i - s_j, t_i - t_j))$. Para dicha estimación, se utiliza la función “variogramSurface” de “gstat” en R.

Luego, se calcula el vector $\gamma(s_i - s_0, t_i - t_0)$ para los sitios tiempos no observados, es decir los 172 datos faltantes y se calculan los pesos para cada uno. Después de hacer la predicción, se reemplazan dichos valores por los vacíos y se anexan a los datos observados. Para poder realizar la predicción espacio-temporal de kriging con el algoritmo usual de la función “krigeST” del paquete “gstat”.

Por facilidad visual, sólo se presenta la interpolación para los 4 meses donde se presentan los cambios de estaciones del año, como se muestra en la tabla 5.

Tabla 5: Estaciones del año en el hemisferio norte. Fuente: <https://www.astromia.com/tierraluna/estaciones.htm>.

Inicio	H Norte	Días duración
20-21 Marzo	Primavera	92,9
21-22 Junio	Verano	93,7
22-24 Septiembre	Otoño	89,6
21-22 Diciembre	Invierno	89

En la figura 6 se presenta la interpolación de la temperatura media diaria para el mes de marzo de 2017 en Reino Unido, en donde se pasa de invierno a primavera. Se observa que los primeros 7 días son los más fríos, con temperaturas en la zona centro de 60 a $80 \times 0.1^\circ C$, mientras que las temperaturas más altas se presentan los últimos 3 días de marzo, alrededor de $120 \times 0.1^\circ C$.

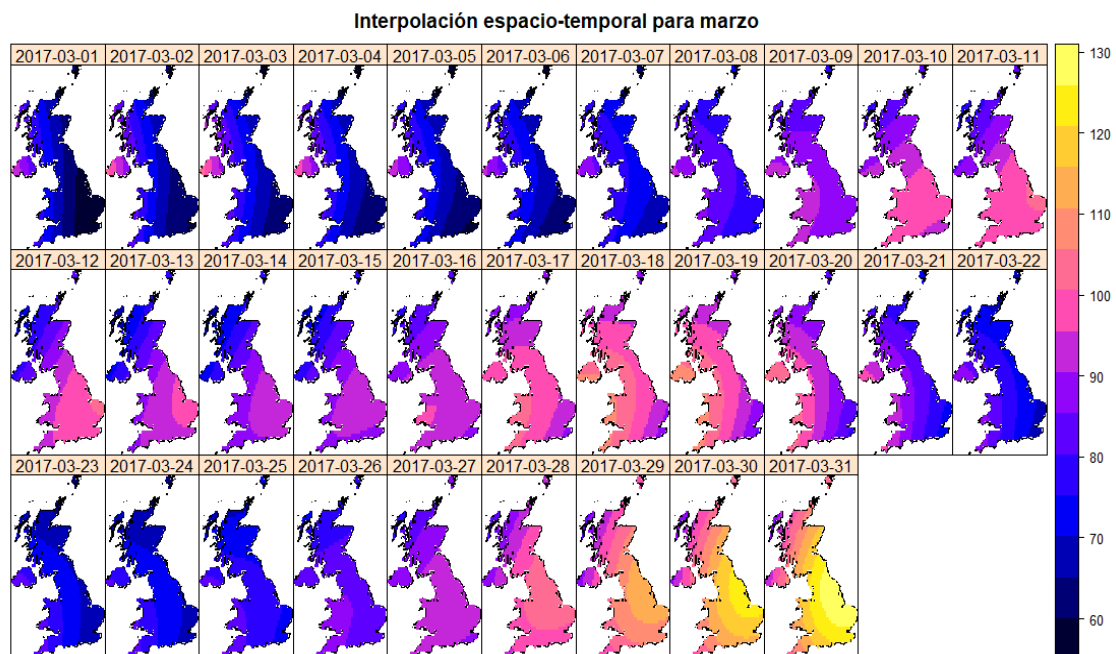


Figura 6: Interpolación espacio-temporal para marzo.

En la figura 7 se presenta la interpolación de la temperatura media diaria para el mes de junio de 2017 en Reino Unido, donde se pasa de primavera a verano. Se observa que los días más calientes fueron del 18 al 21 en la zona centro de Reino Unido con temperaturas mayores a $190 \times 0.1^\circ C$, mientras que las temperaturas más bajas se presentan los días del 4 al 7 de junio, alrededor de $120 \times 0.1^\circ C$ en la zona sur oriente y $140 \times 0.1^\circ C$ en la zona centro.

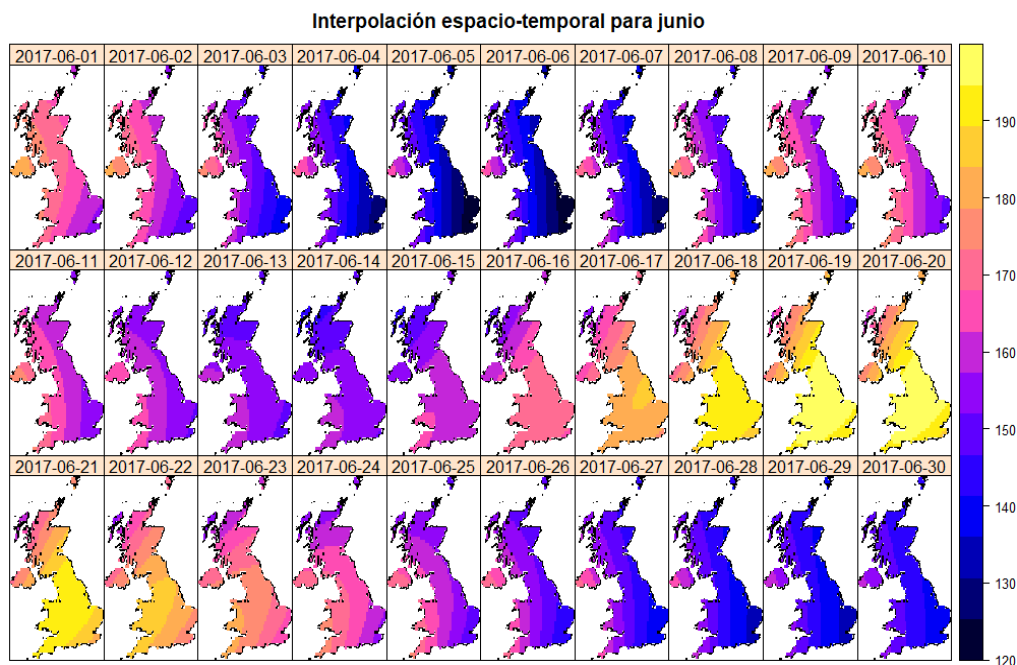


Figura 7: Interpolación espacio-temporal para junio.

En la figura 8 se presenta la interpolación de la temperatura media diaria para el mes de septiembre de 2017 en Reino Unido en donde se pasa de verano a otoño. Se observa que los días del 3 al 7 presentan temperaturas alrededor de $150 \times 0.1^\circ C$, los días del 15 al 19 presentan temperaturas bajas, alrededor de $115 \times 0.1^\circ C$ en la zona sur oriente y $125 \times 0.1^\circ C$ en la zona occidental, alrededor de $120 \times 0.1^\circ C$, y los últimos días del mes vuelve a subir la temperatura.

En la figura 9 se presenta la interpolación de la temperatura media diaria para el mes de diciembre de 2017 en Reino Unido en donde se pasa de otoño a invierno. Se observa que los primeros 7 días se presentan las temperaturas más altas del mes, alrededor de $80 \times 0.1^\circ C$ en la zona occidental, y de $90 \times 0.1^\circ C$ en la oriental, mientras que los días posteriores la temperatura disminuye considerablemente, en especial en la zona oriental, con temperaturas por debajo de $20 \times 0.1^\circ C$.

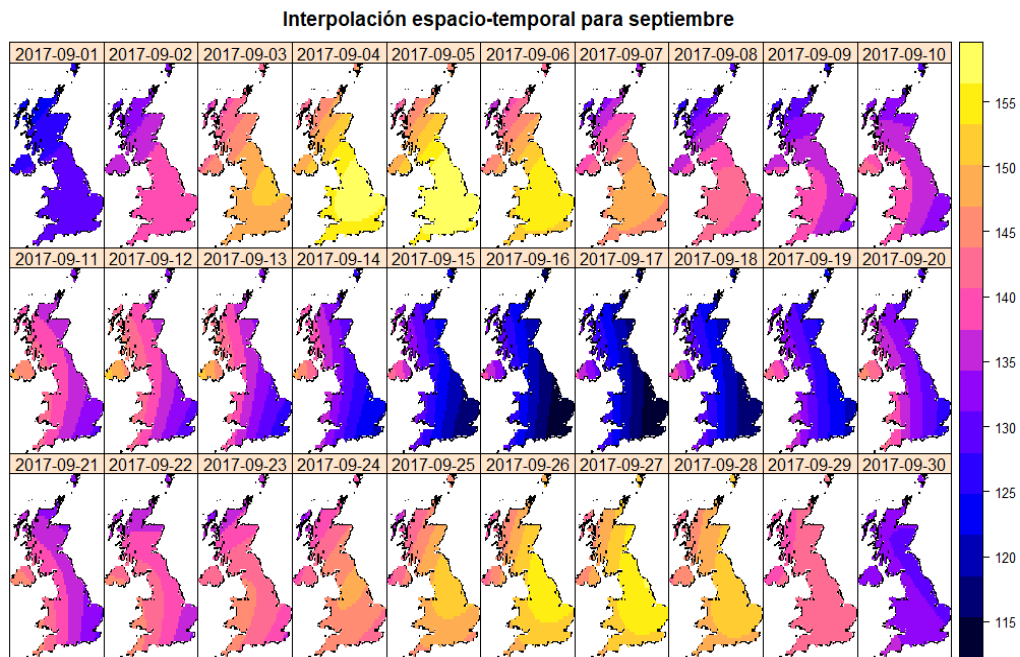


Figura 8: Interpolación espacio-temporal para septiembre.

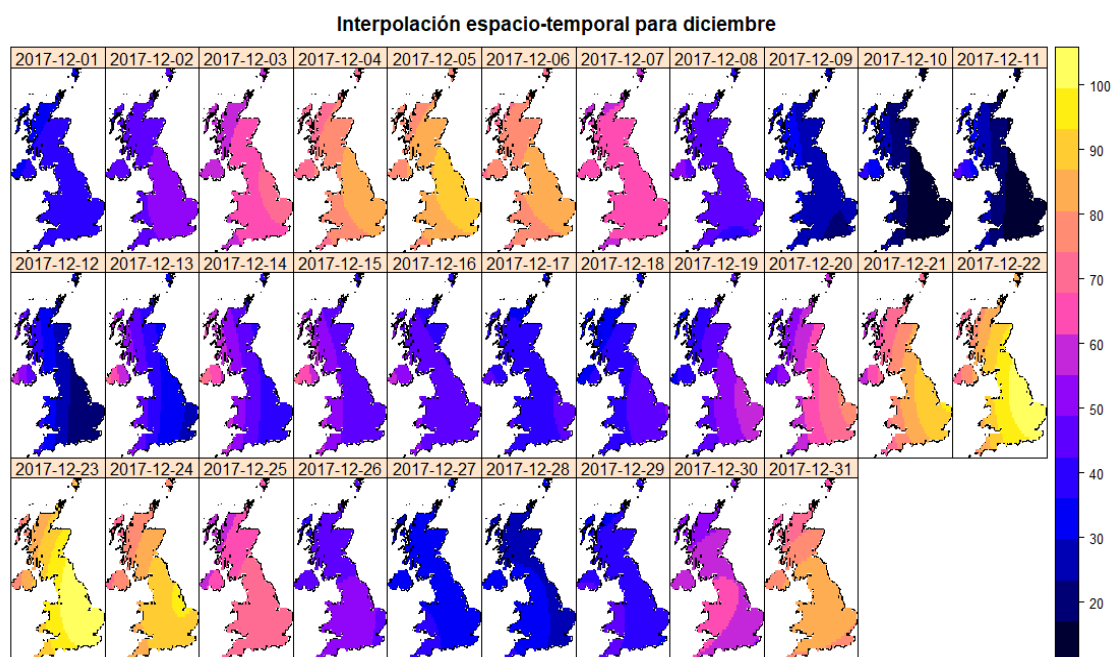


Figura 9: Interpolación espacio-temporal para diciembre.

7. Discusión

Como se muestra en la sección anterior, a partir de la aplicación propuesta se obtuvo la interpolación de la temperatura media diaria en Reino Unido durante 2017, sin excluir los datos incompletos. Si bien, la metodología de interpolación de kriging ordinario ya ha sido ampliamente utilizada en contextos espaciales, temporales o espacio-temporales, esta no admite la presencia de datos incompletos, lo que resulta ser el aporte de este trabajo.

Cabe resaltar que existe una gran cantidad de variables auxiliares que pueden afectar la temperatura diaria en cada sitio y que no se tuvieron en cuenta, como la humedad, la presión, los vientos, las precipitaciones, las corrientes oceánicas, la proximidad al mar, entre otros. En este caso, se contempló únicamente el comportamiento de la variable de la temperatura.

8. Conclusiones y trabajos futuros

- Los resultados de la interpolación espacio-temporal de la temperatura media en Reino Unido se pueden obtener incluyendo los datos faltantes en el proceso de predicción. De acuerdo al modelo métrico exponencial que presentó el menor ECM, se concluye que la predicción en los sitios no observados es óptima y está alineada a las temperaturas observadas en los 34 sitios de Reino Unido durante el 2017.
- El código utilizado para incluir datos faltantes en el proceso de interpolación fue creación propia en el que se rellenaron los vacíos de información con las predicciones de kriging. Después de rellenar la información se hizo uso del método tradicional de kriging, explicado en la sección 4, que también se encuentra en el libro de Montero et al. (2015).
- Este primer acercamiento solo contempla el comportamiento de la variable en sí misma y no se consideran otros factores climatológicos que pueden mejorar la precisión del modelo, además no considera el efecto temporal individual de las series de tiempo consideradas. Adicionalmente, el kriging ordinario está planteado para encontrar la predicción en sitios no observados y no pretende reemplazar las metodologías existentes en modelos para series temporales.
- Existen pruebas paramétricas y no paramétricas para probar la isotropía para el caso espacial. Varios autores han propuesto pruebas de isotropía para el caso espacio-temporal, por ejemplo Ma (2004), Porcu et al. (2007), Melo Martínez (2012), entre otros. Sin embargo, esta aplicación no era el objetivo principal de este trabajo. Para futuros trabajos se recomienda ahondar en este tema y adaptar dichas pruebas al espacio y el tiempo.
- Como trabajo futuro, también se recomienda realizar la interpolación espacio-temporal, incluyendo datos faltantes en el contexto de datos de conteo.

Agradecimientos

Primeramente a Dios por darme la oportunidad de crecer profesionalmente a través de este proceso. A Wilmer Pineda quien desde el principio me brindó su apoyo y dedicó gran parte de su tiempo para guiarme en este trabajo y a Jeison Rodríguez quien me ayudó ejecutando algunos procesos que demandaban una alta carga computacional.

Referencias

- Abad Encalada, I. O. (2014), Estimación de precipitación espacial mediante correlación con variables secundarias y la ayuda de tecnologías de información geográficas, Master's thesis.
- Bevilacqua, M. (2008), 'Composite likelihood inference for space-time covariance models'.
- Brenes, Á. & Saborío, V. F. (1995), *Elementos de climatología: su aplicación didáctica a Costa Rica*, number 1, EUNED.
- Carrera-Hernández, J. & Gaskin, S. (2007), 'Spatio temporal analysis of daily precipitation and temperature in the basin of mexico', *Journal of Hydrology* **336**(3-4), 231–249.
- Cressie, N. A. (1993), 'Statistics for spatial data: Wiley series in probability and mathematical statistics', *Find this article online*.
- Cressie, N. & Hawkins, D. (1980), 'Robust estimation of the variogram', *Journal of the International Association for the Mathematical Geology*, **2**(12), 1330–40.
- Giraldo, R., Pacheco, Ó. & Orozco, A. (2017), 'Geoestadística aplicada a series de tiempo autorregresivas: un estudio de simulación', *Integración: Temas de matemáticas* **35**(1), 83–102.
- Kondrashov, D. & Ghil, M. (2006), 'Spatio-temporal filling of missing points in geophysical data sets', *Nonlinear Processes in Geophysics* **13**(2), 151–159.
- Le, N. D., Sun, W. & Zidek, J. V. (1997), 'Bayesian multivariate spatial interpolation with data missing by design', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **59**(2), 501–510.
- Li, L. & Revesz, P. (2002), A comparison of spatio-temporal interpolation methods, in 'International Conference on Geographic Information Science', Springer, pp. 145–160.
- Li, L. & Revesz, P. (2004), 'Interpolation methods for spatio-temporal geographic data', *Computers, Environment and Urban Systems* **28**(3), 201–227.
- Ma, C. (2004), 'Spatial autoregression and related spatio-temporal models', *Journal of multivariate analysis* **88**(1), 152–162.
- Mas, J. F. (2016), 'An introduction to r for spatial analysis & mapping', *GeoFocus. Revista Internacional de Ciencia y Tecnología de la Información Geográfica* (17), 185–186.
- Matheron, G. (1989), 'Estimating and choosing. an essay on probability in practice. translated from the french and with a preface by am hasofer'.
- Melo Martínez, C. E. (2012), 'Análisis geoestadístico espacio tiempo basado en distancias y splines con aplicaciones'.
- Melo, O. O. (2013), *Estadística Espacial*, Pro-Offset Editorial S.A.
- Millano, J., Paredes, F. & Vivas, I. (2017), 'Efecto de la oscilación meridional (enso) y la temperatura superficial del océano atlántico sobre la distribución espacio-temporal de las lluvias en el estado cojedes', *Revista Agrollania de Ciencia y Tecnología* **4**, 103–116.
- Min, W. & Wynter, L. (2011), 'Real-time road traffic prediction with spatio-temporal correlations', *Transportation Research Part C: Emerging Technologies* **19**(4), 606–616.
- Montero, J.-M., Mateu, J. et al. (2015), *Spatial and spatio-temporal geostatistical modeling and kriging*, Vol. 998, John Wiley & Sons.

- Pebesma, E. J. (2001), ‘Gstat user’s manual’, *Dept. of Physical Geography, Utrecht University, Utrecht, The Netherlands*.
- Peña Angulo, D., Gonzalez Hidalgo, J. C. & Brunetti, M. (2016), ‘Variabilidad espacio temporal de las temperaturas mensuales en la España Peninsular’. Presentado: 12 12 2016.
- Porcu, E., Mateu, J., Zini, A. & Pini, R. (2007), ‘Modelling spatio-temporal data: A new variogram and covariance structure proposal’, *Statistics & probability letters* **77**(1), 83–89.
- Rodriguez-Iturbe, I. & Mejía, J. M. (1974), ‘The design of rainfall networks in time and space’, *Water Resources Research* **10**(4), 713–728.
- Rodríguez, J. S. (2005), Variabilidad espacio-temporal de la temperatura del aire en Cataluña, PhD thesis, Universitat Rovira i Virgili.
- Rouhani, S. & Hall, T. J. (1989), Space-time kriging of groundwater data, *in* ‘Geostatistics’, Springer, pp. 639–650.
- Schabenberger, O. & Gotway, C. A. (2017), *Statistical methods for spatial data analysis*, CRC press.
- Sherman, M. (2011), *Spatial statistics and spatio-temporal data: covariance functions and directional properties*, John Wiley & Sons.
- Sibanda, S., Grab, S. & Ahmed, F. (2018), ‘Spatio-temporal temperature trends and extreme hydro-climatic events in southern zimbabwe’, *South African Geographical Journal* **100**(2), 210–232.
- Stoffer, D. S. (1986), ‘Estimation and identification of space-time armax models in the presence of missing data’, *Journal of the American Statistical Association* **81**(395), 762–772.
- Subba Rao, T. (2013), ‘Spatial statistics and spatio-temporal data’, *Journal of Time Series Analysis* **34**(2), 280–280.
- Teegavarapu, R. S. & Chandramouli, V. (2005), ‘Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records’, *Journal of Hydrology* **312**(1-4), 191–206.
- Torres, H. & Castaño, O. (1991), ‘Las series de tiempo en la investigación espacio-temporal sobre descargas eléctricas atmosféricas en Colombia’, *Revista Colombiana de Estadística* **12**(23-24).