

# **Análisis Geostadístico**

## **de datos funcionales**

Proyecto Fin de Máster - Máster en Técnicas Estadísticas

Alumna: María José Ginzo Villamayor  
Tutor: Manuel Febrero Bande



# Resumen

El análisis de datos funcional se ocupa de la modelización estadística de variables aleatorias que toman valores en un espacio de funciones (variables funcionales). Varias técnicas estadísticas estándares tales como la regresión, ANOVA o componentes principales, entre otros, han sido considerados desde el punto de vista funcional. En general, estas metodologías se centran en variables funcionales independientes e idénticamente distribuidas. Sin embargo, en varias disciplinas de las ciencias aplicadas, existe un gran interés en la modelización de datos funcionales espacialmente correlados. En particular, la mayoría de ellos están interesados en el modelado de datos funcionales espacialmente correlacionados. Este es el tema aquí tratado. En concreto, este proyecto trata la predicción de curvas, cuando se dispone de una muestra de las curvas de una región con continuidad espacial.

Se revisan tres métodos para la predicción espacial de los datos funcionales. Inicialmente, se propone un predictor que tiene la misma forma que el predictor kriging clásico, pero teniendo en cuenta las curvas en lugar de datos de una sola dimensión. Los otros predictores surgen de adaptaciones de modelos lineales funcionales con respuesta funcional en el caso de datos funcionales espacialmente correlacionados. Por un lado, se define un predictor que es una combinación de kriging y del modelo funcional lineal point-wise (concurrente). Por otra parte, se utiliza el modelo funcional lineal total para extender dos métodos clásicos geoestadísticos multivariantes para el contexto funcional. El primer predictor se define en términos de parámetros escalares. En el resto de los casos, los predictores implican parámetros funcionales. Se adapta un criterio de optimización, criterio utilizado en predicción espacial multivariante para estimar los parámetros escalares y funcionales que intervienen en los predictores propuestos. En todos los casos se da un enfoque no paramétrico basado en la expansión en términos de bases de funciones que se usa para obtener las curvas a partir de datos discretos.

Las metodologías propuestas se ilustran mediante el análisis de un conjunto de datos real correspondiente a la curva de temperatura que es función del tiempo.

*Palabras clave:* base de funciones; Cokriging; validación cruzada; modelo lineal funcional; Kriging; predicción espacial multivariante.



# Contenidos

<b>Introducción</b>	<b>1</b>
<b>1 Geostatística Univariante</b>	<b>5</b>
1.1 Tipos de datos espaciales . . . . .	6
1.2 Análisis estructural . . . . .	7
1.2.1 Variable regionalizada . . . . .	7
1.2.2 Hipótesis de estacionariedad . . . . .	7
1.3 Estudio del semivariograma . . . . .	10
1.4 Anisotropía . . . . .	12
1.4.1 Anisotropía geométrica . . . . .	13
1.4.2 Anisotropía zonal . . . . .	13
1.5 Condición de positividad . . . . .	13
1.6 Versiones muestrales de algunas medidas de variabilidad espacial . . . . .	14
1.7 Factores a tener en cuenta para la modelización . . . . .	15
1.8 Métodos de estimación de los parámetros del variograma . . . . .	16
1.8.1 Estimación por mínimos cuadrados (MC) . . . . .	16
1.8.2 Estimación máximo verosímil (MV) . . . . .	17
1.9 Kriging: predicción e interpolación . . . . .	18
1.9.1 Kriging Simple (SK) . . . . .	20
1.9.2 Kriging Ordinario (OK) . . . . .	21
1.9.3 Kriging universal(UK) . . . . .	25
1.9.4 Cokriging . . . . .	27
<b>2 Geostatística Multivariante</b>	<b>29</b>
2.1 Análisis estructural . . . . .	29
2.2 Cokriging . . . . .	31
2.2.1 Cokriging simple . . . . .	32
2.2.2 Cokriging ordinario . . . . .	34
2.3 Cokriging universal . . . . .	35
2.4 Condición de insesgadez . . . . .	36

<b>3</b>	<b>Análisis de datos funcionales</b>	<b>37</b>
3.1	Preliminares . . . . .	39
3.2	Técnicas exploratorias para Datos Funcionales . . . . .	40
3.2.1	Media, Varianza, covarianza, correlación, covarianza cruzada y correlación cruzada . . . . .	41
3.2.2	Componentes principales . . . . .	41
3.3	Bases para datos funcionales . . . . .	42
3.3.1	Bases de Fourier . . . . .	42
3.3.2	Bases B-Splines . . . . .	43
3.3.3	Bases de Wavelets . . . . .	43
3.4	Suavización de datos funcionales . . . . .	44
3.5	Elección de la suavización . . . . .	44
3.5.1	Regresión - representación en base . . . . .	45
3.5.2	Regresión Lineal Funcional . . . . .	45
<b>4</b>	<b>Ejemplo real de Kriging con datos reales funcionales</b>	<b>49</b>
4.1	Conceptos básicos de meteorología . . . . .	49
4.2	Kriging ordinario para funciones-valores de datos espaciales . . . . .	53
4.2.1	Predicción y estimación de los parámetros . . . . .	54
4.2.2	Estimación la traza del variograma . . . . .	58
4.2.3	Enfoque no paramétrico . . . . .	58
4.2.4	Resultados de la aplicación . . . . .	60
4.3	Kriging variación de tiempo continua para la predicción espacial de datos funcionales . . . . .	68
4.3.1	Predictor y problema de minimización . . . . .	68
4.3.2	Una solución basada en bases de funciones . . . . .	69
4.3.3	Resultados: predicción espacial de las curvas de temperatura en Galicia	72
4.4	De multivariante a geostadística funcional . . . . .	75
4.4.1	Cokriging basado en datos funcionales . . . . .	76
4.4.2	Kriging funcional: modelo total . . . . .	80
4.4.3	Resultados: Predicción espacial de la temperatura en las curvas Galicia	83
4.5	Comparación de las tres técnicas . . . . .	85
<b>A</b>	<b>Histogramas y boxplots de la Temperatura de Galicia año 2009</b>	<b>99</b>
	<b>Bibliografía</b>	<b>105</b>
	<b>Glosario</b>	<b>107</b>
	<b>Index</b>	<b>108</b>

# Introducción

La mayoría de los fenómenos naturales que se estudian son variables medidas tanto en el espacio como en el tiempo. Considerando una superficie de un suelo, por ejemplo, en ocasiones se puede observar una alta variabilidad en distancias pequeñas. Esta variabilidad es el resultado de los procesos naturales. Si se tienen datos de una de esas variables en  $n$  sitios de una región con continuidad espacial, implícitamente en cada uno de ellos hay una observación funcional y así usando técnicas de suavizado los valores encontrados pueden convertirse en un dato funcional. Por ello, en este proyecto se mostrará cómo las técnicas de permiten identificar tendencia en media y varianza y explorar la estructura de autocorrelación inherente a un conjunto de datos medidos en una región con continuidad espacial mediante el uso de varios ejemplos. Muchos de los datos recogidos en problemas de las ciencias aplicadas son curvas. Para modelizar este tipo de información está el Análisis de Datos Funcionales (FDA), (se puede ver en [26]) se utiliza desde finales de los años noventa. En ciencias como la agronomía, meteorología, ecología y otras el análisis geostadístico, (véase [6]) se utiliza, a menudo, para describir la distribución espacial.

La palabra geoestadística se compone de dos partes, geo y estadística de manera similar a términos como geofísica o geoquímica. Se usa con dos acepciones diferentes:

1. como una colección de todos los métodos estadísticos y probabilísticos aplicados a las geo-ciencias,
2. y como nombre para la teoría de variables regionalizadas.

La geoestadística fue usada primero por la industria minera, debido a los altos costes altos que suponían las perforaciones que se hacían, de ahí que el análisis de los datos fuera de suma importancia. De hecho muchos libros y publicaciones sobre geoestadística están en su mayoría orientados a problemas de minería.

La motivación de este trabajo es ofrecer una solución al problema de predecir curvas en aquellas zonas no muestreadas de una región, basándonos en estas dos ramas estadísticas, el análisis de datos funcionales y la estadística espacial. Los capítulos 1-3 recogen el fundamento teórico más importante, necesario para este proyecto, para ambas ramas.

La modelización de variables medidas en diferentes sitios de una región con continuidad espacial,(véase en [6]) ha tenido desde los años sesenta junto con el desarrollo del análisis geostadístico un gran uso en varias ciencias medioambientales, en minería, geología, ecología.

Por convención, el análisis geostadístico es un procedimiento de dos pasos. El primero, la estructura espacial de la variable es examinada con el análisis del varioograma. Una vez se tiene una estructura espacial y precisado un modelo adecuado (esférico, exponencial o gasussiano, según sea el caso), se utiliza algún procedimiento *kriging* para interpolar la variable no muestreada.

En el contexto geostadístico multivariante consideramos simultáneamente  $p$  procesos aleatorios espaciales, (véase en [6]). Es necesario encontrar un modelo para la covarianza espacial en el que todas las variables estén incluidas en el análisis. Se va a utilizar la información multivariante para predecir cada uno de los puntos no muestreados. Un caso particular de geostadística multivariante es el método *cokriging bivariante*, (en [6]).

Otra rama importante de la geostadística es el análisis de datos espacio-temporales. Los modelos geostadísticos espacio-temporales proporcionan una herramienta probabilística para el análisis de datos y predicciones basadas en la dependencia temporal entre las observaciones. En contraposición a la geostadística multivariante, donde a menudo hay unas cuantas variables en cada ubicación, los conjuntos espacio-temporales de datos pueden ser muy grandes.

La comunidad estadística ha estado interesada en desarrollar modelos para datos funcionales, ya desde trabajos publicados a finales de los años setenta a otros más actuales como el trabajo de Ramsay, ([24]). Nuevas metodologías que combinan los modelos no paramétricos con los datos funcionales se pueden consultar en [8].

De la misma manera que los métodos estadísticos estándares han sido generalizados para ser utilizados en FDA, es posible pensar que los métodos geostadísticos pueden ser adaptados a este tipo de datos.

El proyecto ha sido estructurado del siguiente modo, en el Capítulo 1 se hace una breve descripción de la geostadística univariante, en el Capítulo 2 se generalizan los conceptos anteriores en un contexto multivariante. En dicho contexto,  $m$  procesos aleatorios espaciales se consideran simultáneamente. Es necesario encontrar un modelo de covarianza espacial para todas las variables incluidas en el modelo. Por tanto, se usa la información multivariante para predecir en cada uno de los puntos no muestreados. El Capítulo 3 revisa la teoría fundamental del análisis de datos funcionales. Las técnicas habituales para la modelización de datos funcionales se centran en funciones independientes. Sin embargo, en varias áreas de las ciencias aplicadas, existe un gran interés por la modelización de datos funcionales correlados. De ahí que métodos estadísticos para la modelización de variables correladas, como por ejemplo el análisis geostadístico, hayan sido adaptadas al contexto funcional. Una vez que el lector está entrenado con todos estos conceptos, en el Capítulo 4, se le ofrece una aplicación de los mismos combinando ambas técnicas a unos conjuntos de datos.

En resumen, lo que contiene este proyecto fin de máster es una revisión crítica de los métodos que se han considerado previamente, en estadística espacial con datos funcionales y aplicados a un conjunto de datos real como es el de la temperatura en Galicia.

Este conjunto de datos tiene tanto componente espacial y funcional.

#### **Datos meteorológicos de Galicia**

La predicción espacial en datos meteorológicos es un factor importante para muchos tipos de modelos como los hidrológicos, los de crecimiento y mortalidad de los ecosistemas forestales.

Como caso particular se tiene el modelado de datos de temperatura. A lo largo de la historia geostadística se han desarrollado y utilizado muchos métodos para hacer predicción espacial

de temperatura, aunque, en la mayoría de las ocasiones, no se ha tenido en cuenta su carácter funcional. En este proyecto se usa un conjunto de datos meteorológicos, se dispone de datos de temperatura media del ambiente para Galicia en 66 estaciones distribuidas por Galicia del siguiente modo:

- 13 (en 13 municipios distintos) en la provincia de A Coruña (94 municipios)
- 20 (en 19 municipios distintos) en la provincia de Lugo (67 municipios)
- 15 (en 15 municipios distintos) en la provincia de Orense (92 municipios)
- 18 (en 16 municipios distintos) en la provincia de Pontevedra (62 municipios)

En particular se analiza la información de la temperatura promedio diaria durante el año 2009 (Figura 1).

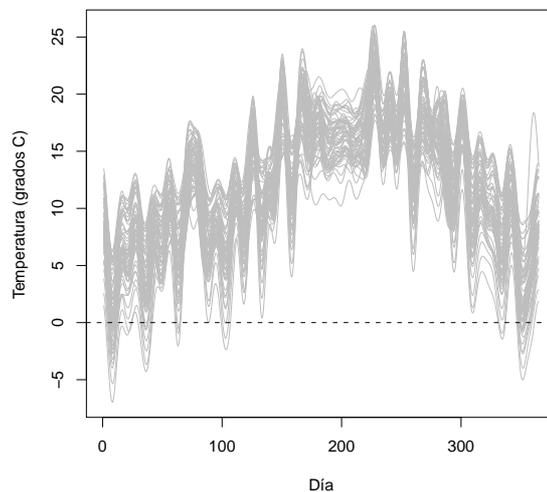


Figura 1: Valores medios de las curvas diarias de la temperatura media observado en 66 estaciones meteorológicas de Galicia.

Los datos de cada estación se obtuvieron la página web de Meteogalicia - Xunta de Galicia (<http://www.meteogalicia.es/web/index.action>). Las coordenadas geográficas de estaciones meteorológicas (Figura 2) también se han obtenido de esa página. Estos datos son analizados en el Capítulo 4.

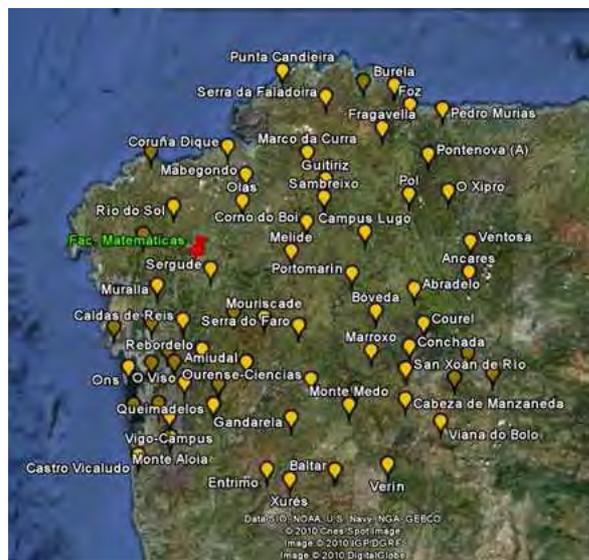


Figura 2: Estaciones metereológicas. El punto marcado en rojo corresponde a la Facultad de Matemáticas, punto no muestreado.

# Capítulo 1

## Geostatística Univariante

La Geostatística es la ciencia que estudia los fenómenos que fluctúan en el espacio y/o tiempo, ofreciendo una colección de herramientas estadísticas para la descripción y modelización de la variabilidad espacial (y temporal).

El término Estadística Espacial se usa para describir una amplia variedad de modelos y métodos adecuados para el análisis de datos referenciados espacialmente. En el libro [6] se puede encontrar una descripción general de éstos.

El análisis espacial comprende el conjunto de conceptos y procedimientos utilizados para abordar el estudio de la estructura y las relaciones territoriales a partir del conocimiento de la posición de las entidades geográficas y las características de las variables seleccionadas para su investigación.

Cuando la distribución espacial de los datos es importante para su estudio e interpretación, la aplicación de técnicas específicas para datos espaciales cobra importancia puesto que puede proporcionar mayor información que las técnicas tradicionales.

Las bases de datos espaciales deben contener observaciones de una (o varias) variables estadísticas de interés y una referencia cartográfica. Estas variables van a ser continuas, la única condición que se le impone es que exista alguna dependencia entre dos variables con distinta referencia cartográfica, es decir, dos observaciones son más similares cuanto más cercanas son sus realizaciones muestrales y a su vez, a medida que la distancia de las localizaciones muestrales aumenta, la correlación entre las variables tiende a anularse. La característica de dependencia es una diferencia importante respecto al análisis estadístico efectuado con datos independientes lo que va a suponer ventajas, ya que las predicciones serán más precisas, como inconvenientes, las estimaciones menos precisas.

No existe ninguna restricción en las referencias cartográficas, que pueden ser una referencia territorial explícita como latitud y longitud (datos geográficos) o una referencia implícita como domicilio o código postal (datos socio-económicos).

En estadística espacial se distinguen tres tipos de datos (véase [6]): datos geoestadísticos o georreferenciados (*geostatistical data*), datos en rejilla o datos en un área (*lattice data*), y datos de procesos puntuales (*point processes data*).

## 1.1 Tipos de datos espaciales

### 1. Datos geoestadísticos o georreferenciados (*geostatistical data*)

Los datos espacialmente continuos son mediciones tomadas en puntos fijos con localizaciones continuas en el espacio. Las mediciones clásicas en salud ambiental, tales como contaminación de suelos y de aire o la radiación natural, suelen pertenecer a esta categoría. La variable medida, sin embargo, puede ser tanto continua como discreta. El objetivo es el de, dados los valores tomados en puntos de muestreo fijos, extender la distribución espacial de los valores de un atributo sobre la región total de estudio. Por lo tanto el análisis de los datos geoestadísticos puede contemplar tanto la modelización del patrón de variabilidad, la determinación de los factores con los que pueda estar relacionado, como la de obtener una buena predicción de la variable en puntos donde no se ha muestreado. Estos métodos son muy utilizados en estudios del área de geociencias (suelo, clima, hidrología, geología minera, etc) por lo que también suelen ser denominados *datos geoestadísticos*.

### 2. Datos en rejilla o datos en un área (*lattice data*)

Los *datos en rejilla* son observaciones procedentes de un proceso aleatorio, observadas sobre una colección contable de regiones espaciales, que pueden estar regular o irregularmente distribuidas, complementados con lo que se denomina estructura de vecindad, es decir, información sobre las regiones vecinas. Matemáticamente una rejilla queda definida como un conjunto de lados y de vértices, es decir, un conjunto de índices de localizaciones con un conjunto asociado de vecinos. Ya que los datos en rejilla están definidos en regiones espaciales, las localizaciones concretas especificadas por el vector suelen referirse al centroide de la región.

### 3. Datos de procesos puntuales (*point processes data*)

Se denominan *datos patrones de puntos o procesos puntuales*, cuando las localizaciones (y no las mediciones) son las variables de interés. Consisten en un número finito de localizaciones observadas en una región determinada. El objetivo de los procesos puntuales es el de conocer la variación de la intensidad de los eventos sobre la región de estudio y el de buscar modelos que ayuden a explicar o comprender el fenómeno. Tan importante como la variable estudiada suele ser el patrón de variación espacial.

Independientemente del tipo de datos, los objetivos principales del estudio de la Estadística Espacial son, como en casi todos los campos de la Estadística, dos:

1. Descripción de los datos. Esto puede incluir no sólo el estudio descriptivo del proceso  $Z(u)$ , tal como se entiende dentro de la estadística clásica, sino también la modelización del tipo de dependencia espacial.
2. Predicción. Sin duda en muchas ocasiones la predicción es el objetivo que motiva el estudio de datos espaciales.

## 1.2 Análisis estructural

En geostatística los datos se observan en  $u$  localizaciones de un conjunto  $D \in \mathbb{R}^d$  donde  $d$  indica la dimensión del espacio. Las localizaciones se expresan en dos o tres coordenadas espaciales, como por ejemplo, longitud, latitud, y/o altitud. Las observaciones se toman en cada localización y se consideran como una realización de un proceso estocástico espacial denotado generalmente por  $Z(u)$ .

La función de distribución acumulada de una variable aleatoria continua dependiente de la localización espacial  $u$ ,  $Z(u)$  viene definida por:

$$F(u; z) = \mathbb{P} \{Z(u) \leq z\}$$

siendo  $\mathbb{P}$  la función de probabilidad asociada. En la práctica se suele trabajar con información de  $n$  datos, por lo que la variable aleatoria se suele denotar por  $Z(u_i) = z(u_i), i = 1, \dots, n$  y en ese caso se trabaja con la función de distribución acumulada condicional dada por:

$$F(u; k|n) = \mathbb{P} \{Z(u) \leq z|n\}$$

En geostatística es importante modelizar el grado de correlación o dependencia espacial entre una cierta variable  $Z(u_i), i = 1, \dots, n$ .

### 1.2.1 Variable regionalizada

Una variable medida en el espacio de forma que presente una estructura de correlación, se dice que es una *variable regionalizada*. De manera más formal se puede definir como un proceso estocástico con dominio contenido en un espacio  $D \in \mathbb{R}^d$ . En términos prácticos  $Z(u)$  puede verse como una medición de una variable aleatoria (por ejemplo concentración de un contaminante) en un punto  $u$  de una región de estudio. Un proceso estocástico es una colección de variables aleatorias indexadas, esto es, para cada  $u$  en el conjunto de índices  $D$ ,  $Z(u)$  es una variable aleatoria. En el caso de que las mediciones sean hechas en una superficie, entonces  $Z(u)$  puede interpretarse como la variable aleatoria asociada a ese punto del plano ( $u$  representa las coordenadas, planas o geográficas, y  $Z$  la variable en cada una de ellas). Estas variables aleatorias pueden representar la magnitud de una variable ambiental medida en un conjunto de coordenadas de la región de estudio.

### 1.2.2 Hipótesis de estacionariedad

Una variable aleatoria  $Z(u), u \in A$  se dice que es *estacionaria* en la región  $A$  si la función de distribución acumulada es invariante bajo cualquier traslación  $C$  efectuada sobre sus localizaciones:

$$F(u_1, \dots, u_k; Z_1, \dots, Z_k) = F(u_1 + C, \dots, u_k + C; Z_1 + C, \dots, Z_k + C)$$

para cualquier vector de traslación  $C$ . Es decir, esta hipótesis establece el grado de homogeneidad espacial del fenómeno.

Como sólo se dispone de una realización discreta de la variable  $Z(u)$ , es inevitable que haya que asumir ciertas hipótesis para poder llevar a cabo algún tipo de estudio estadístico. El tipo de estacionariedad asumido indica qué tipo de inferencia estadística puede realizarse con el modelo probabilístico.

Se considera un proceso estocástico  $\{Z(u), u \in D\}$ , donde  $D$  es un subconjunto de  $\mathbb{R}^d$  (espacio Euclídeo  $d$ -dimensional). El proceso  $Z$  se dice que es *Gaussiano* si, para cualquier natural  $k \geq 1$  y las localizaciones  $u_1, \dots, u_k$ , el vector  $(Z(u_1), Z(u_2), \dots, Z(u_k))$  tiene una distribución normal multivariante.

### Estrictamente Estacionario

Si la correspondiente distribución de  $(Z(u_1), Z(u_2), \dots, Z(u_k))$  es la misma que la del vector  $(Z(u_1 + h), Z(u_2 + h), \dots, Z(u_k + h))$  para cualesquiera  $u_1, u_2, \dots, u_k$  puntos espaciales y cualquier  $h \in \mathbb{R}^d$ , el proceso se dice que es *estrictamente estacionario*.

### Estacionario de segundo orden

El proceso  $Z$  se dice que es *estacionario de segundo orden*, o *estacionariamente débil* si  $\mu(u) \equiv \mu$ , es decir, la media es la misma para todas las  $u$  y  $\text{Cov}\{Z(u_1), Z(u_2)\} = C(u_1 - u_2)$ , para todos  $u_1 \in D, u_2 \in D$ , donde  $C(u)$  es la función de covarianza entre una observación  $u$  y de otra en el 0.  $C(u)$  recibe el nombre de *covariograma*. Como consecuencia de la definición tenemos:

- $\sigma_Z^2 = C(0), \forall x \in D$ .
- $C$  es una función simétrica:  $C(h) = C(-h)$ .
- Aplicando la desigualdad de Cauchy-Schwarz puede demostrarse que  $|C(h)| \leq C(0)$ .
- $C(h)$  es una función definida positiva.
- Una función relacionada es el *correlograma*,  $\rho(h) = C(h)/C(0)$ , verificando que  $|\rho(h)| \leq \rho(0) = 1$

Se puede ver que si todas las varianzas son finitas entonces un proceso estrictamente estacionario es también estacionario de segundo orden. La afirmación inversa es falsa, en general, pero un proceso Gaussiano que es a la vez estacionario de segundo orden será también estrictamente estacionario, ([6]).

### Hipótesis intrínseca

Si se asume que  $\mu(u)$  es una constante, la cual podemos suponer cero sin pérdida de generalidad, se puede definir:

$$2\gamma(u_1 - u_2) = \text{Var}\{Z(u_1) - Z(u_2)\}$$

La ecuación anterior sólo tiene sentido si la parte izquierda depende de  $u_1$  y  $u_2$  sólo a través de su diferencia  $u_1 - u_2$ . Un proceso que satisface esta propiedad se llama *intrínsecamente estacionario*. La función  $2\gamma(\cdot)$  se llama *variograma* y  $\gamma(\cdot)$  *semivariograma*.

1.  $\gamma$  es una función simétrica  $\gamma(h) = \gamma(-h)$ .
2.  $|\gamma(h)| \leq \gamma(0) = 0$ .
3. Variograma estandarizado:  $\gamma_s(h) = \gamma(h)/\sigma_Z^2$ ,  $\gamma_s(0) = 0$ .

La propiedad intrínsecamente estacionario es más débil que la estacionariedad de segundo orden.

Si se supone que el proceso es estacionario de segundo orden se verifica que:

$$\begin{aligned} \text{Var} \{Z(u_1) - Z(u_2)\} &= \text{Var} \{Z(u_1)\} + \text{Var} \{Z(u_2)\} - 2\text{Cov} \{Z(u_1), Z(u_2)\} \\ &= 2C(0) - 2C(u_1 - u_2) \end{aligned}$$

y también si se verifica la condición de estacionario de segundo orden, entonces:

$$\gamma(h) = C(0) - C(h)$$

Como consecuencia, ambas funciones son equivalentes para caracterizar la dependencia espacial de la variable regionalizada  $Z(h)$ .

Además el variograma debe cumplir:

1.  $\gamma(h)$  es una función condicionalmente definida negativa.
2.  $\lim_{|h| \rightarrow +\infty} \frac{\gamma(h)}{|h|^2} = 0$ .

Desde este punto de vista, las distintas formas (más fuertes) de estacionariedad no son necesarias. La estacionariedad tanto intrínseca como la de segundo orden son suposiciones más naturales. Por ello, hay que ser cauto cuando un análisis preliminar de los datos indica que el proceso es intrínsecamente estacionario y no estacionario.

### Isotropía

Un concepto diferente es el de *isotropía*. Si se supone que el proceso es intrínsecamente estacionario con semivariograma  $\gamma(h)$ ,  $h \in \mathbb{R}^d$ . Si  $\gamma(h) = \gamma_0(\|h\|)$ , es decir, si el semivariograma depende del vector  $h$  sólo a través de su longitud  $\|h\|$ , entonces el proceso es *isotrópico*.

Un proceso que es a la vez intrínsecamente estacionario e isotrópico se dice que es *homogéneo*.

### 1.3 Estudio del semivariograma

El estudio del semivariograma, o variograma, es más habitual que el estudio del covariograma. Las razones son dos:

- La hipótesis de intrínsecamente estacionaria incluye la condición de estacionaria de segundo orden.
- Para el cálculo del semivariograma no es necesario conocer la media de la variable.

Por otra parte, si la variable es estacionaria de segundo orden, siempre puede obtenerse la función de covariograma a partir de la función de semivariograma:  $\gamma(h) = \sigma_Z^2 - C(h)$ .

De ahora en adelante, por facilidad de notación, se denotará el variograma isotrópico por  $\gamma(h)$ , donde  $h$  representa la longitud del vector, en lugar del propio vector. El semivariograma representa un índice de cambio que muestra una variable con la distancia. Su forma describe el patrón de variación espacial en términos de su magnitud y forma general.

La pendiente del semivariograma indica la intensidad de cambio del atributo (variable) analizado con la distancia al mismo tiempo que el porcentaje de disminución en la dependencia espacial. El máximo valor que alcanza un semivariograma se llama **meseta (sill)** o varianza a priori, e indica la escala bajo la cual los datos definen un proceso estacionario de segundo orden. La meseta puede ser o no finita. Los semivariogramas que tienen meseta finita cumplen con la hipótesis de estacionariedad fuerte; mientras que cuando ocurre lo contrario, el semivariograma define un fenómeno natural que cumple sólo con la hipótesis intrínseca.

El *lag* o *distancia* para la que el *sill* es alcanzado se llama **rango** o **alcance** y define el límite de la dependencia espacial. El rango se interpreta como la zona de influencia. Existen algunos modelos de semivariograma en los que no existe una distancia finita para la cual dos observaciones sean independientes; por ello se llama rango efectivo a la distancia para la cual el semivariograma alcanza el 95% de la meseta. Cuanto más pequeño sea el rango, más cerca se está del modelo de independencia espacial. El rango no siempre aparece de manera explícita en la fórmula del semivariograma.

Finalmente un semivariograma con término independiente define la varianza llamada **Efecto Pepita (nugget)**, la cual define la variabilidad intrínseca en los datos y que no ha sido captada por el rango de distancia analizadas así como cualquier variación puramente aleatoria. Representa una discontinuidad puntual del semivariograma en el origen. Puede ser debido a errores de medición en la variable o a la escala de la misma. En algunas ocasiones puede ser indicativo de que parte de la estructura espacial se concentra a distancias inferiores a las observadas.

Denotando por  $|h| = \|h\|$ , algunos ejemplos de semivariogramas isotrópicos son:

1. Modelo *Nugget effect*.

Es la estructura más básica del semivariograma indicando falta de estructura espacial. Viene dado por:

$$\gamma(|h|) = \begin{cases} 0 & \text{si } |h| = 0, \\ 1 & \text{en otro caso.} \end{cases}$$

## 2. Modelo lineal.

Define un modelo no acotado en función de dos constantes positivas. La función tiende a infinito para distancias grandes lo que hace que este modelo no se corresponda con un proceso estacionario,

$$\gamma(|h|) = \begin{cases} 0 & \text{si } |h| = 0 \\ c_0 + c_1 h & \text{si } |h| > 0. \end{cases}$$

## 3. Modelo esférico.

Definido por un rango actual  $a$ , una varianza a priori (*sill*)  $c_1$  y un efecto *nugget*  $c_0$ ,

$$\gamma(|h|) = \begin{cases} c_0 + c_1 \cdot \left[ 1.5 \frac{|h|}{a} - 0.5 \left( \frac{|h|}{a} \right)^3 \right] & \text{si } |h| \leq a \\ c_0 + c_1 & \text{si } |h| \geq a. \end{cases}$$

## 4. Modelo Exponencial.

Definido por un rango efectivo  $a$  (rango integral  $a/3$ ), una varianza a priori (*sill*)  $c_1$  y un efecto *nugget*  $c_0$ ,

$$\gamma(|h|) = c_0 + c_1 \cdot \left[ 1 - \exp\left(-\frac{3|h|}{a}\right) \right].$$

## 5. Modelo Gaussiano.

Definido por un rango efectivo  $a$ , una varianza a priori  $c_1$  y un efecto *nugget*  $c_0$ ,

$$\gamma(|h|) = c_0 + c_1 \cdot \left[ 1 - \exp\left(-\frac{(3|h|)^2}{a^2}\right) \right].$$

## 6. Modelo potencial (*Power*).

Definido por un factor  $0 < \omega < 2$ , una pendiente positiva  $c_1$  y un efecto *nugget*  $c_0$ ,

$$\gamma(|h|) = c_0 + c_1 \cdot |h|^\omega.$$

## 7. Modelo del efecto *Hole* o sinusoidal.

Se utiliza para definir componentes cíclicas subyacentes. Con una varianza a priori de  $c_1$  y un efecto *nugget* de  $c_0$ , se define como

$$\gamma(|h|) = c_0 + c_1 \cdot \left[ 1.0 - \cos\left(\frac{|h|}{a} \cdot \pi\right) \right].$$

Para ser un modelo de variograma correctamente definido, este modelo de efecto *hole* debe ser sólo aplicado en una dirección.

8. Familia *Matérn*.

Esta clase se define mejor en términos de la función de covarianza definida por

$$C_0(|h|) = \frac{1}{2^{\theta_2-1} \Gamma(\theta_2)} \left( \frac{2\sqrt{\theta_2}|h|}{\theta_1} \right)^{\theta_2} k_{\theta_2} \left( \frac{2\sqrt{\theta_2}|h|}{\theta_1} \right).$$

En este caso,  $\theta_1 > 0$  define un parámetro de dependencia espacial y  $\theta_2 > 0$  es un parámetro de forma. La función  $\Gamma(\cdot)$  representa la función gamma mientras que  $k_{\theta_2}$  es la función modificada de Bessel de tercera clase de orden  $\theta_2$ . Como casos especiales se tiene que  $\theta_2 \rightarrow \frac{1}{2}$  tiende a la forma exponencial y el límite de  $\theta_2 \rightarrow \infty$  daría lugar a la forma Gaussiana.

Todos los modelos anteriores (excepto el lineal y el potencial) son acotados, lo que significa que el *sill* (umbral) se alcanza realmente en el límite a una cierta distancia marcada por el rango. Además se tienen las siguientes observaciones:

- Para el modelo de efecto *nugget*, el *sill* es alcanzado tan pronto la distancia se hace positiva.
- El modelo esférico alcanza realmente el *sill* a una distancia equivalente a su rango.
- Los modelos exponencial y Gaussiano alcanzan asintóticamente el *sill*, definiéndose un rango práctico como aquella distancia para la que el modelo se encuentra al 95% del *sill*.

Los modelos acotados también se suelen llamar *modelos de transición*. Por otra parte, los modelos potencial y lineal no tienen *sill*, y por tanto, carece de su correspondiente función de covarianza.

Alrededor del origen podemos distinguir tres tipos de comportamiento:

1. Comportamiento parabólico: (modelo Gaussiano) tal comportamiento es característico de fenómenos altamente regulares.
2. Comportamiento lineal: (modelos esférico o exponencial) tal comportamiento es característico de aquellos fenómenos que muestran dependencias espaciales a cortas distancias aumentándolas linealmente con las mismas.
3. Comportamiento discontinuo: (modelo de efecto *nugget*) evidencia de gran variabilidad intrínseca a los datos que enmascara cualquier posible dependencia espacial.

El comportamiento cerca del origen del modelo potencial depende de los valores del parámetro  $\omega$ , siendo lineal para  $\omega = 1$  y parabólico para valores de  $\omega$  cercanos a 2.

## 1.4 Anisotropía

Hay varias formas de trabajar con procesos anisótropicos considerando éstos como generalizaciones más o menos directas de procesos isotrópicos.

### 1.4.1 Anisotropía geométrica

La forma más simple es considerar la *anisotropía geométrica*. Esta se refiere a un semivariograma de la forma

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

donde  $\gamma_0(\|Ah\|)$  es un variograma isotrópico y  $A$  una matriz de dimensión  $d \times d$  representando una transformación lineal de  $\mathbb{R}^d$ . Lógicamente si  $A$  es la identidad, esto se reduce al caso isotrópico. La idea subyacente es esta situación es que el proceso no es isotrópico en el espacio original, pero sí en algún espacio transformado linealmente, el cual puede, por ejemplo, corresponder a una transformación de las coordenadas. En el caso más lógico y usual en el que  $A$  es una matriz definida positiva, los contornos de igual covarianza se corresponden con elipses inscritas en círculos.

### 1.4.2 Anisotropía zonal

Una posible generalización de anisotropía surge de la simple observación de que si  $Z_1, \dots, Z_p$  son procesos independientes intrínsecamente estacionarios, entonces  $Z = Z_1 + \dots + Z_p$ , es también intrínsecamente estacionario, con semivariograma dado por  $\gamma(h) = \gamma_1(h) + \dots + \gamma_p(h)$ , denotando  $\gamma_1, \dots, \gamma_p$  los semivariogramas de  $Z_1, \dots, Z_p$  respectivamente. Así,

$$\gamma(h) = \sum_{i=1}^p \gamma_0(A_i h),$$

siendo  $\gamma_0$  un semivariograma isotrópico y  $A_1, \dots, A_p$  matrices, es un semivariograma válido que generaliza la anisotropía geométrica. Esta anisotropía se llama *anisotropía zonal*.

Una idea más complicada es asumir que, para alguna función no lineal  $g(u)$ , el proceso  $Z(g(u))$ , en lugar del original  $Z(u)$ , es un proceso isotrópico estacionario. Esta idea puede, de hecho analizarse, tanto la no estacionariedad como la no isotropía.

## 1.5 Condición de positividad

Una restricción importante en geostatística es que no se puede definir una covarianza espacial o una función de semivariograma de forma arbitraria. Necesariamente la primera de ellas debe cumplir con la condición de positividad. En el caso más general en el que  $\text{Cov}\{Z(u_1), Z(u_2)\} = C(u_1, u_2)$ , el cual no supone ninguna condición de estacionariedad, la condición de positividad significa que la relación

$$\sum_i \sum_j a_i a_j C(u_i, u_j) \geq 0$$

se cumple para cualquier conjunto finito de puntos  $u_1, \dots, u_n$  y coeficientes reales arbitrarios  $a_1, \dots, a_n$ . Es necesario que la parte izquierda de la ecuación es la varianza de  $\sum_i a_i Z(u_i)$ . También se tienen estas mismas condiciones en la versión de variogramas. Supongamos que

$\gamma(\cdot)$  es el semivariograma de un proceso estacionario de segundo orden, entonces si  $a_1, \dots, a_n$  son constantes con  $\sum a_i = 0$ , tenemos

$$\sum_i \sum_j a_i a_j \gamma(u_i - u_j) \leq 0$$

Esta es la condición de *no positividad condicional*, (explicada en [6]). La anterior condición es una condición necesaria para que  $\gamma(\cdot)$  sea un semivariograma válido en el caso general.

## 1.6 Versiones muestrales de algunas medidas de variabilidad espacial

Se considera ahora el problema de la estimación del variograma y de otras medidas de variabilidad espacial. En general disponemos de un proceso  $\{Z(u), u \in D\}$  observado en un número finito de localizaciones  $u_1, \dots, u_N$ .

El estimador más simple viene dado por el *método de los momentos*, el cual, suponiendo que los puntos de muestreo  $u_1, \dots, u_N$  estén definidos en un retículo regular, viene definido por

$$2\hat{\gamma}(h) = \frac{1}{\#N(h)} \sum_{(u_i, u_j) \in N(h)} \{Z(u_i) - Z(u_j)\}^2.$$

$N(h)$  denota todos aquellos pares  $(u_i, u_j)$  para los que  $u_i - u_j = h$  y  $\#N(h)$  denota el cardinal de  $N(h)$ . En el caso, por otra parte más común, de que los puntos de muestreo no estén en un retículo regular, se aplica la fórmula anterior pero con la nueva definición de  $N(h)$ ,

$$N(h) = \{(u_i, u_j) : u_i - u_j \in T(h)\},$$

siendo  $T(h)$  alguna vecindad o región de tolerancia sobre  $h$ .

Si disponemos de dos atributos medidos sobre las mismas localizaciones espaciales, puede ser de interés evaluar la variabilidad espacial (cruzada) entre ambos atributos  $Z, Y$ . La medida adecuada viene de la mano del estimador del *variograma cruzado* dado por:

$$2\hat{\gamma}_{ZY}(h) = \frac{1}{\#N(h)} \sum_{(u_i, u_j) \in N(h)} (Z(u_i) - Z(u_j))(Y(u_i) - Y(u_j)).$$

Una posible objeción al método de los momentos es que no es robusto frente a valores extremos de  $Z$ .

Otra objeción surge del hecho del sesgo de la distribución: si suponemos que el proceso es Gaussiano, para valores concretos de  $u$  y  $h$ , la distribución de  $\{Z(u+h) - Z(u)\}^2$  es de la forma  $2\gamma(h)\chi_1^2$ , y la distribución de  $\chi_1^2$  está sesgada. Sin embargo, si  $X \sim \chi_1^2$  entonces  $X^{1/4}$  tiene una distribución casi simétrica y por tanto las medias muestrales de  $|Z(u_1) - Z(u_2)|^{1/2}$  se comportarán mejor que las de  $\{Z(u_1) - Z(u_2)\}^2$ .

La función de *covarianza muestral* puede ser también obtenida mediante el método de los momentos se la siguiente forma:

$$\hat{C}(h) = \frac{1}{\#N(h)} \sum_{(u_i, u_j) \in N(h)} \{Z(u_i)Z(u_j) - m^2\}$$

donde  $m = \frac{1}{\#N(h)} \sum_i Z(u_i)$ . Si además definimos  $\sigma^2 = \frac{1}{\#N(h)} \sum_i Z(u_i)^2 - m^2$ , podemos estandarizar la covarianza para definir el *correlograma*

$$\rho(h) = \frac{C(h)}{\sigma^2}$$

Por otra parte, el *madograma* es una medida similar al variograma del método de los momentos en el que el cuadrado de las diferencias entre  $Z(u_i)$  y  $Z(u_j)$  es sustituido por la diferencia absoluta, dando lugar a la siguiente expresión

$$2\hat{\gamma}_M(h) = \frac{1}{\#N(h)} \sum_{(u_i, u_j) \in N(h)} |Z(u_i) - Z(u_j)|$$

Algunas de las anteriores medidas de variación espacial pueden ser usadas cualitativamente para conseguir estructuras de continuidad espacial. Más de un tipo de medidas puede ser obtenida al mismo tiempo. En general, las características observadas a través del eje de las abscisas (distancias) son comunes para todas las medidas de variabilidad / continuidad. Sin embargo los valores del eje de ordenadas que definen el variograma son específicos del tipo de variograma elegido.

Se tienen dos reglas prácticas que deberían ser tenidas en cuenta al estimar un variograma:

1. el variograma empírico sólo debe ser considerado para distancias para las que el número de pares es superior a 30, (ver [? ]);
2. la distancia de fiabilidad para un variograma experimental es  $h < D/2$  siendo  $D$  la distancia máxima que presentan las localizaciones muestreadas.

## 1.7 Factores a tener en cuenta para la modelización

En la práctica de la modelización tres son los puntos claves del problema:

1. Determinación del variograma o covarianza experimental.
2. Análisis de los posibles variogramas permisibles.
3. Utilización de información auxiliar, como el conocimiento físico del área y fenómeno bajo estudio o medidas robustas tales como el madrograma.

La modelización consiste en la conjunción de estas diferentes fuentes de información para construir un modelo adecuado que retenga la mayoría de las características de los atributos bajo estudio.

Algunas decisiones importantes en una modelización se centran en:

1. Cómo obtener un modelo isotrópico o anisotrópico;
2. Cuál es el número y tipo de estructuras básicas que constituyan el variograma;
3. Elección adecuada de los parámetros asociados a los variogramas (*sill*, rango y *nugget*).

Respecto a la decisión sobre modelo isotrópico o anisotrópico, en la práctica se suele comparar diferentes variogramas experimentales calculados en distintas direcciones (para decidir si existe anisotropía geométrica, por lo menos, tres direcciones deben ser consideradas). Una opción que nos puede ayudar para la detección de anisotropía direccional, es un *scatterplot* del mapa del variograma en el sistema de coordenadas usual. El centro del mapa corresponde con el origen del variograma  $\gamma(0) = 0$ . Cuando la variación es isotrópica, el incremento es similar en cada dirección y por contra, la anisotropía geométrica aparece como líneas de contorno elípticas indicando la dirección de máxima variabilidad espacial. Por esto, la evaluación de un mapa de semivariogramas requiere considerar algunas direcciones e intervalos de distancias (*lags*).

El último peldaño en el proceso de la modelización consiste en la determinación de los parámetros asociados a los modelos seleccionados.

## 1.8 Métodos de estimación de los parámetros del variograma

Asumamos que muestreamos a partir de un proceso espacial homogéneo para el que el variograma ha sido estimado por alguno de los métodos anteriores.

Los semivariogramas  $\hat{\gamma}(h)$  y  $\bar{\gamma}(h)$  carecen de la propiedad de no positividad condicional. Como consecuencia, es posible que algunas predicciones espaciales derivadas a partir de tales estimadores presenten varianzas negativas. La forma más clara y común de evitar esta dificultad es reemplazando el semivariograma empírico  $\hat{\gamma}(h)$  por algún modelo paramétrico, que se sabe que cumple con la condición de semidefinido negativo. En general, no es necesario restringirse a modelos isotrópicos, aunque suelen ser los primeros a ser considerados. Se consideran tres métodos:

1. Estimación por mínimos cuadrados (*least squares estimation*).
2. Estimación por máxima verosimilitud (*maximum likelihood*).

### 1.8.1 Estimación por mínimos cuadrados (MC)

Supongamos que tenemos estimado el semivariograma  $\gamma(h)$  en un conjunto finito de valores de  $h$ , y queremos ajustar un modelo especificado por una función paramétrica  $\gamma(h; \theta)$  en términos de un vector finito de parámetros  $\theta$ . Este vector suele contener tres parámetros, efecto *nugget*, *sill* y rango. Supóngase que se ha utilizado el estimador de los momentos  $\hat{\gamma}$  y sea  $\hat{\gamma}$  el vector que contiene los valores estimados y  $\gamma(\theta)$  el vector de los valores derivados por el modelo sobre los mismos valores de  $h$ .

Se tienen tres posibilidades para los métodos de **estimación mínimo cuadrática no lineal**:

- Mínimos cuadrados ordinarios (MCO), en los que se toma  $\theta$  como aquel valor que minimiza  $\{\hat{\gamma} - \gamma(\theta)\}' \{\hat{\gamma} - \gamma(\theta)\}$ . (donde  $'$  indica la transposición de matrices).

- Mínimos cuadrados ponderados (MCP), en los que se toma  $\theta$  como aquel valor que minimiza  $\{\hat{\gamma} - \gamma(\theta)\}' W(\theta)^{-1} \{\hat{\gamma} - \gamma(\theta)\}$ . En este caso  $W(\theta)$  es una matriz diagonal cuyos elementos de la diagonal son las varianzas de  $\hat{\gamma}$ , aunque no covarianzas como si lo hace el MCG. Por tanto MCP admite varianzas de  $\hat{\gamma}$ , aunque no covarianzas aunque si lo hace el MCG.
- Mínimos cuadrados generalizados (MCG), en los que se toma  $\theta$  como aquel valor que minimiza  $\{\hat{\gamma} - \gamma(\theta)\}' V(\theta)^{-1} \{\hat{\gamma} - \gamma(\theta)\}$ . Aquí  $V(\theta)$  denota la matriz de covarianzas de  $\hat{\gamma}$ , la cual depende de  $\theta$ .

En general, los tres estimadores MCO, MCP y MCG, aparecen en orden creciente de eficiencia pero decreciente en simplicidad. Notar que MCO es fácilmente implementable por algún procedimiento de mínimos cuadrados no lineales, mientras que MCP y MCG requieren la especificación de las matrices  $W(\theta)$  y  $V(\theta)$ .

### 1.8.2 Estimación máximo verosímil (MV)

Si se asume que se muestrea a partir de un proceso Gaussiano, entonces es bastante sencillo obtener la forma exacta de la verosimilitud y maximizarla numéricamente. Considérese el proceso espacial  $Z \sim N(X\beta, \Sigma)$ , con  $Z$  un vector de observaciones  $n$ -dimensional,  $X$  una matriz  $n \times q$  de covariables ( $q < n$ ;  $X$  de rango completo),  $\beta$  un vector de dimensión  $q$  de parámetros desconocidos y  $\Sigma$  la matriz de covarianzas de las observaciones. En la práctica se puede asumir que  $\Sigma = \alpha V(\theta)$  siendo  $\alpha$  un parámetro de escala conocido y  $V(\theta)$  es una matriz de covarianzas estandarizadas determinadas por el parámetro desconocido  $\theta$ . Con  $Z$ ,  $Z \sim N(X\beta, \Sigma)$ , su función de densidad es de la forma

$$(2\pi)^{-n/2} \det(\Sigma)^{-1/2} \exp \left\{ -\frac{1}{2} (Z - X\beta)' \Sigma^{-1} (Z - X\beta) \right\}.$$

Y por tanto, la log-verosimilitud negativa será:

$$l(\beta, \alpha, \theta) = \frac{n}{2} \log(2\pi) + \frac{n}{2} \log(\alpha) + \frac{1}{2} \log \det(V(\theta)) + \frac{1}{2\alpha} (Z - X\beta)' V(\theta)^{-1} (Z - X\beta).$$

Aunque este método es computacionalmente factible, su mayor dificultad frente a por ejemplo el método de MCP lo hace menos usado.

Supóngase  $\{Y_1, \dots, Y_n\}$  son variables aleatorias independientes y normales  $N(\mu, \sigma^2)$  con parámetros desconocidos  $\mu$  y  $\sigma^2$ . Los estimadores máximo verosímiles de  $\mu$  y  $\sigma^2$  son  $\hat{\mu} = \bar{Y} = \frac{1}{n} \sum_i Y_i$  y  $\hat{\sigma}^2 = \frac{1}{n} \sum_i (Y_i - \bar{Y})^2$ . Pero este estimador resulta sesgado y se suele usar el estimador insesgado de  $\sigma^2$ ,  $\frac{1}{n-1} \sum_i (Y_i - \bar{Y})^2$ . Supóngase ahora que en lugar de trabajar con el vector  $Y_1, \dots, Y_n$  lo hacemos con la densidad conjunta de  $(Y_1 - \bar{Y}, \dots, Y_n - \bar{Y})$ , cuya distribución no depende de  $\mu$ . Ahora el estimador verosímil de  $\sigma^2$  es directamente  $\frac{1}{n-1} \sum_i (Y_i - \bar{Y})^2$ .

Esta idea puede ser extendida al modelo general,  $Z \sim N(X\beta, \Sigma)$ . Si se define  $W = A'Z$  un vector de  $n - q$  contrastes linealmente independientes, es decir, las  $n - q$  columnas de  $A$

son linealmente independientes y  $A'X = 0$ , tenemos que  $W \sim N(0, A'\sigma A)$ , y el logaritmo negativo de la función de verosimilitud en  $W$  será de la forma,

$$l_W(\alpha, \theta) = \frac{n-q}{2} \log(2\pi) + \frac{n-q}{2} \log(\alpha) + \frac{1}{2} \log |A'V(\theta)A| + \frac{1}{2\alpha} W' (A'V(\theta)A) W.$$

Es posible elegir  $A$  que satisfaga  $A'A = I - X(X'X)^{-1}X'$ ,  $A'A = I$ . En este caso la expresión anterior se simplifica a:

$$\begin{aligned} l_W(\alpha, \theta) &= \frac{n-q}{2} \log(2\pi) + \frac{n-q}{2} \log(\alpha) + \frac{1}{2} \log |X'X| + \\ &+ \frac{1}{2} \log |X'V(\theta)^{-1}X| + \frac{1}{2} \log \det(V(\theta)) + \frac{1}{2\alpha} G^2(\theta), \end{aligned}$$

donde  $G^2(\theta)$  denota la suma de cuadrados generalizados de los residuos

$$G^2 = (Z - X\hat{\beta})' V^{-1} (Z - X\hat{\beta})$$

y  $\hat{\beta} = (Z - X\hat{\beta}) X'V^{-1}Z$  es el estimador MCG de  $\beta$  basado en la matriz de covarianzas  $V$ .

La estimación MVR proporciona mejores estimaciones que el MV pues, en general, da lugar a estimadores con menores desviaciones para muestras con pocos datos. La estimación MVR es ampliamente utilizada en modelización geostadística. Sin embargo, es más sensible que el estimador MV de máxima verosimilitud bajo incorrecta especificación del vector de medias  $\beta$ .

En general, como los variogramas estimados por los métodos de verosimilitud (MV, MVR) no están basados en los variogramas empíricos, hay claras diferencias entre éstos y los estimadores MCO o MCP.

## 1.9 Kriging: predicción e interpolación

Una vez conocida la varianza específica el objetivo es predecir e interpolar procesos espaciales.

El problema tiene los siguientes fundamentos dado un conjunto de observaciones de un atributo espacial  $Z(u_1), Z(u_2), \dots, Z(u_n)$ , el objetivo es predecir el valor de  $Z(u_0)$ , para algún  $u_0 \notin \{u_1, \dots, u_n\}$ .

*Kriging* es un nombre genérico adoptado en geostadística para dar nombre a una metodología de interpolación basada en una familia de algoritmos de regresión generalizados por mínimos cuadrados.

Todas las clases de estimaciones de *krigings*, no son más que variantes de las estimaciones de regresión lineal básicas, las cuales para predecir el valor del atributo  $Z$  en la localización  $u_0$ , denotado por  $Z^*(u_0)$ , vienen definidas por

$$Z^*(u_0) - m(u_0) = \sum_{\alpha=1}^{n(u_0)} \lambda_{\alpha}(u_0) [Z(u_{\alpha}) - m(u_{\alpha})]$$

donde  $\lambda_{\alpha}$  define la ponderación, peso asignado a los datos que intervienen en el sumatorio,  $m(u_0)$  y  $m(u_{\alpha})$  son los correspondientes valores esperados de  $Z(u_0)$  y  $Z(u_{\alpha})$  respectivamente

y  $n(u_0) + 1$  elementos:  $n(u_0)$  atributos. A partir de ahora y sin pérdida de generalidad para que la notación resulte más sencilla se usará  $n$  en lugar de  $n(u_0)$ . Obsérvese que sólo actúan aquellas localizaciones  $u_\alpha$  vecinas a la localización de predicción de  $u_0$ .

Cualquier base de *kriging* tiene como objetivo la minimización de la varianza del error  $\sigma_{\mathbb{E}}^2(u)$ , la cual en su formato general viene dado por

$$\sigma_{\mathbb{E}}^2(u) = \text{Var} [Z^*(u) - Z(u)]$$

donde el superíndice \* indica el valor estimado para esa localización.

Además  $\sigma_{\mathbb{E}}^2(u) = \text{Var} [Z^*(u) - Z(u)]$  se minimiza bajo la restricción de insesgadez, es decir,  $\mathbb{E} [Z^*(u) - Z(u)] = 0$ . Normalmente, la variable aleatoria que define el atributo en estudio se descompone en una componente residual  $R(u)$  y otra determinista que define la tendencia  $m(u)$ ,

$$Z(u) = R(u) + m(u)$$

La componente residual se modeliza como una variable aleatoria estacionaria de media cero y covarianza  $C_r(h)$ , siendo

$$\begin{aligned} \mathbb{E} \{R(u)\} &= 0 \\ \text{Cov} \{R(u), R(u+h)\} &= \mathbb{E} \{R(u) \cdot R(u+h)\} = C_R(h) \end{aligned}$$

De esta forma, el valor esperado de la variable aleatoria  $Z$  en una cierta localización  $u$  viene dado por el valor de la componente tendencia en esa localización  $\mathbb{E} \{Z(u)\} = m(u)$ .

De acuerdo con el modelo considerado para la tendencia, podemos considerar las siguientes variantes *kriging* lineales: simple (SK), ordinario (OK), con modelo de tendencia o universal (UT), en bloques y factorial. En cuanto a los no lineales se pueden mencionar: lognormal, multi-Gaussiano, de rango, indicatriz y disyuntivo.

En la Tabla 1.1 se resumen los principales tipos de kriging lineal que detallaremos a continuación.

<b>Kriging para <math>Z(u)</math></b>	<b>Media <math>m(u)</math></b>	<b>Análisis estructural <math>Y(u)</math></b>	<b>Propiedades</b>
Simple	Constante conocida	Covariograma	Son óptimos si hay normalidad multivariada. Independiente de la distribución son los mejores predictores linealmente insesgados.
Ordinario	Constante desconocida	Semivariograma	
Universal	No constante y desconocida	Semivariograma	

Tabla 1.1: Principales tipos de kriging lineal y propiedades

Antes de comenzar el estudio de cada uno de ellos, se presentan dos particularidades importantes:

1. La condición de estacionariedad, necesaria para el análisis estructural, no es ahora imprescindible para la predicción kriging. Por esta razón, en muchas de las expresiones que se utilizarán se podemos encontrar  $\gamma(u_\alpha - u_\beta)$  o  $C(u_\alpha - u_\beta)$  en vez de  $\gamma(h)$  o  $C(h)$  utilizadas anteriormente.

2. En la etapa de predicción la dependencia espacial se supone totalmente modelizada por las funciones  $\gamma(h)$  y/o  $C(h)$  obtenidas en la etapa de análisis estructural.

### 1.9.1 Kriging Simple (SK)

Supóngase que hay una variable regionalizada estacionaria con media  $m$  y covarianza conocidas. De manera análoga a como se define en modelos lineales el modelo establecido en este caso es igual a la media más un error aleatorio con media cero. La diferencia es que en este caso los errores no son independientes. Sea  $Z(u)$  la variable de interés medida en el sitio  $u$ .

$$\begin{aligned}\mathbb{E}[Z(u)] &= m \\ Z(u) &= m + \epsilon(u), \text{ con } \mathbb{E}[\epsilon(u)] = 0.\end{aligned}$$

El predictor de la variable de interés en un sitio  $u_0$  donde no se tiene información se define como:

$$Z^*(u_0) = m + \epsilon^*(u_0),$$

con  $\epsilon^*(u_0)$  que corresponde a la predicción del error aleatorio en el sitio  $u_0$ . Despejando de la ecuación anterior  $\epsilon^*(u_0) = Z^*(u_0) - m$ . El predictor del error aleatorio se define por:

$$\epsilon^*(u_0) = \sum_{\alpha=1}^n \lambda_{\alpha} \epsilon(u_{\alpha}) = \sum_{\alpha=1}^n \lambda_{\alpha} (Z(u_{\alpha}) - m)$$

de donde el predictor de la variable de estudio es:

$$Z^*(u_0) = m + \left[ \sum_{\alpha=1}^n \lambda_{\alpha} (Z(u_{\alpha}) - m) \right] = m + \sum_{\alpha=1}^n \lambda_{\alpha} \epsilon(u_{\alpha})$$

El predictor es insesgado si:  $\mathbb{E}^*(Z(u_0)) = \mathbb{E}(Z(u_0)) = m$ . Luego el predictor será insesgado cuando  $\mathbb{E}(\epsilon^*(u_0)) = 0$ .

$$\mathbb{E}(\epsilon^*(u_0)) = \sum_{\alpha=1}^n \lambda_{\alpha} \mathbb{E}(\epsilon(u_{\alpha})) = \sum_{\alpha=1}^n \lambda_{\alpha} (0) = 0.$$

Por tanto en este caso no existen restricciones para las ponderaciones tendientes al cumplimiento de la condición de insesgades. La estimación de los pesos del método kriging simple se obtiene de tal forma que se minimice  $V(\epsilon^*(u_0) - \epsilon(u_0))$ .

$$\begin{aligned}V(\epsilon^*(u_0) - \epsilon(u_0)) &= \mathbb{E}(\epsilon^*(u_0) - \epsilon(u_0))^2 \\ &= \mathbb{E}\left(\left(\sum_{\alpha=1}^n \lambda_{\alpha} \epsilon(u_{\alpha})\right) - \epsilon(u_0)\right)^2 \\ &= \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \mathbb{E}(\epsilon(u_{\alpha}) \epsilon(u_{\beta})) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} \mathbb{E}(\epsilon(u_{\alpha}) \epsilon(u_0)) + \mathbb{E}(\epsilon(u_0))^2\end{aligned}$$

usando:

1.  $\mathbb{E}[\epsilon(u_0)] = 0$
2.  $\mathbb{E}(\epsilon(u_\lambda)\epsilon(u_\beta)) = \text{Cov}(\epsilon(u_\lambda), \epsilon(u_\beta)) = C_{\lambda\beta}$
3.  $\mathbb{E}(\epsilon(u_0))^2 = \sigma^2$

$$V(\epsilon^*(u_0) - \epsilon(u_0)) = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_\alpha \lambda_\beta \mathbb{E}(\epsilon(u_\alpha)\epsilon(u_\beta)) - 2 \sum_{\alpha=1}^n \lambda_\alpha \lambda_\beta C_{\lambda\beta} + \sigma^2$$

derivando respecto a  $\lambda_1$  se tiene:

$$\frac{\partial V(\epsilon^*(u_0) - \epsilon(u_0))}{\partial \lambda_1} = 2 \sum_{\alpha=1}^n \lambda_\alpha C_{1\alpha} - 2C_{10}$$

igualando a cero  $\sum_{\alpha=1}^n \lambda_\alpha C_{1\alpha} = C_{10}$ . En general para cualquier  $\lambda$ ,  $\lambda = 1, 2, \dots, n$ , se obtiene:

$$\frac{\partial}{\partial \lambda_\alpha} = \sum_{\beta=1}^n \lambda_\beta C_{\alpha\beta} = C_{\alpha 0}$$

Con las  $n$  ecuaciones resultantes se construye el siguiente sistema de ecuaciones:

$$\begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} C_{10} \\ C_{20} \\ \vdots \\ C_{n0} \end{bmatrix}$$

La varianza de predicción Kriging Simple será,

$$\sigma_{SK}^2 = \sigma^2 - \sum_{\alpha=1}^n \lambda_\alpha C_{\alpha 0}.$$

Finalmente, el peso asignado a la media  $m$  vendrá dado por,

$$\lambda(u_0) = 1 - \sum_{\alpha=1}^n \lambda_\alpha(u_0).$$

## 1.9.2 Kriging Ordinario (OK)

Supóngase que se hacen mediciones de la variable de interés  $Z$  en los puntos  $u_\alpha$ ,  $\alpha = 1, \dots, n$ , de la región de estudio, es decir, se tienen realizaciones de las variables  $Z(u_1), \dots, Z(u_n)$ , y se desea predecir  $Z(u_0)$ , en el punto  $u_0$  donde no hubo medición. El *Kriging Ordinario* (OK) tiene en cuenta las posibles fluctuaciones locales de la tendencia o media, limitando el dominio de estacionariedad de la media a la vecindad local  $W(u) : m(u^*)$  es una constante (desconocida) para todo  $u^* \in W(u)$ .

El estimador *Kriging Ordinario* puede predecirse como una combinación lineal de las  $n$  variables aleatorias así:

$$\begin{aligned} Z^*(u_0) &= \lambda_1 Z(u_1) + \lambda_2 Z(u_2) + \dots + \lambda_n Z(u_n) \\ &= \sum_{\alpha=1}^n \lambda_\alpha Z(u_\alpha) \end{aligned}$$

en donde los  $\lambda_\alpha$  representan los pesos o ponderaciones de los valores originales. Dichos pesos se calculan en función de la distancia entre los puntos muestreados y el punto donde se va a hacer la correspondiente predicción. La suma de los pesos debe ser igual a uno para que la esperanza del predictor sea igual a la esperanza de la variable. Esto último se conoce como el *requisito de insesgadez*.

Estadísticamente la propiedad de insesgadez se expresa a través de  $\mathbb{E}(Z^*(u_0)) = \mathbb{E}(Z(u_0))$ . Asumiendo que el proceso es estacionario de media  $m$  (desconocida) y utilizando las propiedades del valor esperado, se demuestra que la suma de las ponderaciones debe ser igual a uno, es decir,  $\sum_{\alpha=1}^n \lambda_\alpha = 1$ . Se dice que  $Z^*(u_0)$  es el mejor predictor, lineal en este caso, porque los pesos se obtienen de tal manera que minimicen la varianza del error de predicción, es decir que minimicen la expresión  $V(Z^*(u_0) - Z(u_0))$ .

Esta última es la característica distintiva de los métodos kriging, ya que existen otros métodos de interpolación como el de distancias inversas o el poligonal, que no garantizan varianza mínima de predicción. La estimación de los pesos se obtiene minimizando  $V(Z^*(u_0) - Z(u_0))$  sujeto a  $\sum_{\alpha=1}^n \lambda_\alpha = 1$ .

Se tiene que  $V(Z^*(u_0) - Z(u_0)) = V(Z^*(u_0)) - 2\text{Cov}[Z^*(u_0), Z(u_0)] + V[Z(u_0)]$ .

Desagregando las componentes de la ecuación anterior se obtiene lo siguiente:

$$\begin{aligned} V[Z^*(u_0)] &= V\left[\sum_{\alpha=1}^n \lambda_\alpha Z(u_\alpha)\right] \\ &= \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_\alpha \lambda_\beta \text{Cov}[Z(u_\alpha), Z(u_\beta)] \end{aligned}$$

En adelante se usará la siguiente notación:  $\text{Cov}[Z(u_\alpha), Z(u_\beta)] = C_{\alpha\beta}$  y  $V[Z(u_0)] = \sigma^2$ . De lo anterior

$$\begin{aligned} \text{Cov}[Z^*(u_0), Z(u_0)] &= \text{Cov}\left[\sum_{\alpha=1}^n \lambda_\alpha Z(u_\alpha), Z(u_0)\right] \\ &= \sum_{\alpha=1}^n \text{Cov}[\lambda_\alpha Z(u_\alpha), Z(u_0)] \\ &= \sum_{\alpha=1}^n \lambda_\alpha C_{\alpha 0} \end{aligned}$$

Entonces reemplazando, se tiene que:

$$V(Z^*(u_0) - Z(u_0)) = \sum_{\alpha}^n \sum_{\beta}^n \lambda_\alpha \lambda_\beta C_{\alpha\beta} - 2 \sum_{\alpha}^n \lambda_\alpha C_{\alpha 0} + \sigma^2(0)$$

Luego se debe minimizar la función anterior sujeta a la restricción  $\sum_{\alpha=1}^n \lambda_{\alpha} = 1$ . Este problema de minimización con restricciones se resuelve mediante el método de los Multiplicadores de Lagrange.

$$\sigma_k^2 = \sum_{\alpha}^n \sum_{\beta}^n \lambda_{\alpha} \lambda_{\beta} C_{\alpha\beta} - 2 \sum_{\alpha}^n \lambda_{\alpha} C_{\alpha 0} + 2\mu \left( \sum_{\alpha=1}^n \lambda_{\alpha} - 1 \right)$$

Siendo  $\mu$  los Multiplicadores de Lagrange. Siguiendo el procedimiento acostumbrado para obtener valores extremos de una función, se deriva e iguala a cero, en este caso con respecto a  $\lambda_{\alpha}$  y a  $\mu$ .

$$\frac{\partial (\sigma_k^2)}{\partial \lambda_1} = 2 \sum_{\beta=1}^n \lambda_{\beta} C_{1\beta} - 2C_{10} + 2\mu$$

$$\text{Por tanto } \sum_{\beta=1}^n \lambda_{\beta} C_{1\beta} + \mu = C_{10}$$

De manera análoga se determinan las derivadas con respecto a  $\lambda_2, \dots, \lambda_n$ :

$$\sum_{\beta=1}^n \lambda_{\beta} C_{2\beta} + \mu = C_{20}$$

$$\vdots$$

$$\sum_{\beta=1}^n \lambda_{\beta} C_{n\beta} + \mu = C_{n0}$$

por último derivamos con respecto a  $\mu$ :

$$\frac{\partial (\sigma_k^2)}{\partial \mu} = 2 \sum_{\alpha=1}^n \lambda_{\alpha} - 2$$

$$\text{Por tanto } \sum_{\alpha=1}^n \lambda_{\alpha} = 1$$

De las ecuaciones anteriores resulta un sistema de  $(n + 1)$  ecuaciones con  $(n + 1)$  incógnitas, que matricialmente puede ser escrito como:

$$\begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} & 1 \\ C_{21} & C_{22} & \dots & C_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} C_{10} \\ C_{20} \\ \vdots \\ C_{n0} \\ 1 \end{bmatrix}$$

El sistema anterior se puede plantear:

$$C_{\alpha\beta} \lambda = C_{\alpha 0}$$

por lo cual los pesos que minimizan el error de predicción se determinan mediante la función de covariograma a través de

$$\lambda = (C_{\alpha\beta})^{-1}C_{\alpha 0}.$$

Encontrando los pesos se calcula la predicción en el punto  $u_0$ . De forma análoga se procede para cada punto donde se quiera hacer predicción. La varianza de predicción del Kriging Ordinario será por tanto:

$$\sigma_{OK}^2 = \sigma^2 - \sum_{\alpha=1}^n \lambda_{\alpha} C_{\alpha 0} - \mu$$

### Validación del kriging

Una modelización estadística general exige una validación a posteriori de sus resultados, y de forma particular la modelización geostadística requiere de dicha validación y la basa en una reestimación de los valores conocidos bajo las condiciones de implementación de los modelos construidos. Estas implementaciones incluyen los modelos de variogramas, el tipo de *kriging* y la elección de la estrategia general de modelización.

Existen diferentes métodos para evaluar la bondad de ajuste del modelo de semivariograma elegido con respecto a los datos muestrales y por ende de las predicciones hechas con kriging. La técnica más empleada es la de Validación Cruzada (*cross-validation*, CV) sirve para comparar valores estimados por los modelos con los reales. La idea consiste en un proceso iterativo en el que cada vez se excluye un dato de la muestra y se estima con el resto de los datos el modelo de semivariograma escogido, predecir vía kriging el valor de la variable en estudio en la ubicación del punto que se excluyó. Cada uno de estos valores se compara, por ejemplo mediante regresión lineal, con el valor real. Buenos coeficientes de correlación / determinación serán indicativos de una correcta modelización. Si el modelo de semivarianza elegido describe bien la estructura de autocorrelación espacial, entonces la diferencia entre el valor observado y el valor predicho debe ser pequeño. Este procedimiento se realiza de forma secuencial con cada uno de los puntos muestrales y así se obtiene un conjunto de *n errores de predicción*. Lo usual es calcular medidas que involucren a estos errores de predicción para diferentes modelos de semivarianza y seleccionar aquél que optimice algún criterio como por ejemplo el del mínimo error cuadrático medio (MECM). Una forma descriptiva de hacer la validación cruzada es mediante un gráfico de dispersión de los valores observados contra los valores predichos. En la medida en que la nube de puntos se ajuste más a una línea recta que pase por el origen, mejor será el modelo de semivariograma utilizado para realizar el kriging.

Sin embargo, la utilización de la validación cruzada para seleccionar modelos de semivariogramas, tiene algunas restricciones:

1. Un remuestreo del modelo de semivariograma no influye en los pesos y *krigings*. Así, los valores de *sill* total no pueden ser obtenidos por validación cruzada de valores restimados.
2. El *sill* relativo y el comportamiento del semivariograma en el origen *nugget*, no pueden ser usados simultáneamente con la validación cruzada.

3. Si el modelo es inadecuado, entonces no está claro qué parámetros deben ser cambiados.

### Representación de las predicciones

Una vez se ha hecho la predicción en un conjunto de puntos diferentes de los muestrales vía kriging, se debe elaborar un mapa que de una representación global del comportamiento de la variable de interés en la zona estudiada. Los más empleados son los mapas de contornos, los mapas de residuos y los gráficos tridimensionales. En el caso de los mapas de contornos, en primer lugar se divide el área de estudio en un enmallado y se hace la predicción en cada uno de los nodos de este mismo. Posteriormente se unen los valores predichos con igual valor, generando así las líneas de contorno (isolíneas de distribución). Este gráfico permite identificar la magnitud de la variable en toda el área de estudio. Es conveniente acompañar el mapa de interpolaciones de la variable con los correspondientes mapas de isolíneas de los errores y de las varianzas de predicción (posiblemente estimados a través de métodos matemáticos), con el propósito de identificar zonas de mayor incertidumbre respecto a las predicciones.

### Intervalos de Confianza

Assumiendo que los errores de predicción siguen una distribución normal estándar y que son independientes, un intervalo de confianza del  $100(1 - \alpha)\%$ ,  $0 < \alpha < 1$ , para  $Z(u)$  es:  $\left[ z^*(u) - z_{1-\frac{\alpha}{2}} \sigma_k, z^*(u) + z_{1-\frac{\alpha}{2}} \sigma_k \right]$ , con  $z^*$  el valor calculado de la predicción y  $z_{1-\frac{\alpha}{2}}$  el percentil de una normal estándar.

### 1.9.3 Kriging universal(UK)

Este tipo de *kriging* (UK) considera que la media local no es conocida y varía suavemente en cada vecindad local  $W(u)$ . La componente de tendencia se modeliza como una combinación lineal de funciones  $f_l(u)$  de las coordenadas. Para tratar este tipo de variables es frecuente descomponer la variable  $Z(u)$  como la suma de la tendencia, tratada como una función determinística, más una componente estocástica estacionaria de media cero. Assúmase que  $Z(u) = m(u) + \epsilon(u)$ , con  $\mathbb{E}(\epsilon(u)) = 0$ ,  $V(\epsilon(u)) = \sigma^2$  y por consiguiente  $\mathbb{E}(Z(u)) = m(u)$ . La tendencia puede expresarse mediante  $m(u) = \sum_{l=1}^p a_l f_l(u)$  donde las funciones  $f_l(u)$  son conocidas y  $p$  es el número de términos empleados para ajustar  $m(u)$ . El predictor Kriging Universal se define como:

$$Z^*(u_0) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(u_{\alpha})$$

este será insesgado si:

$$\begin{aligned}
 \mathbb{E}(Z^*(u_0)) &= m(u_0) \\
 \mathbb{E}\left(\sum_{\alpha=1}^n \lambda_{\alpha} Z(u_{\alpha})\right) &= m(u_0) \\
 \mathbb{E}\left(\sum_{\alpha=1}^n \lambda_{\alpha} m(u_{\alpha})\right) &= m(u_0) \\
 \sum_{\alpha=1}^n \lambda_{\alpha} \left(\sum_{l=1}^p a_l f_l(u_{\alpha})\right) &= \sum_{l=1}^p a_l f_l(u_0) \\
 \sum_{l=1}^p a_l \left(\sum_{\alpha=1}^n \lambda_{\alpha} f_l(u_{\alpha})\right) &= \sum_{l=1}^p a_l f_l(u_0) \\
 \sum_{\alpha=1}^n \lambda_{\alpha} f_l(u_{\alpha}) &= \sum_{l=1}^p f_l(u_0)
 \end{aligned}$$

La obtención de los pesos en el Kriging Universal, análogo a los otros métodos kriging, se hace de tal forma que la varianza del error de predicción sea mínima.

$$\begin{aligned}
 V(Z^*(u_0) - Z(u_0)) &= \mathbb{E}(Z^*(u_0) - Z(u_0))^2 \\
 &= \mathbb{E}\left(\left(\sum_{\alpha=1}^n \lambda_{\alpha} (m(u_{\alpha}) - \epsilon(u_{\alpha}))\right) - (m(u_0) - \epsilon(u_0))\right)^2 \\
 &= \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} \mathbb{E}(\epsilon(u_{\alpha}) \epsilon(u_{\beta})) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} \mathbb{E}(\epsilon(u_{\alpha}) \epsilon(u_0)) + \mathbb{E}(\epsilon(u_0))^2
 \end{aligned}$$

usando:

1.  $C_{\alpha\beta} = \text{Cov}(\epsilon(u_{\alpha}), \epsilon(u_{\beta}))$
2.  $\sigma^2 = \mathbb{E}(\epsilon(u_0))^2$

se tiene

$$V(Z^*(u_0) - Z(u_0)) = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C_{\alpha\beta} - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C_{\alpha 0} + \sigma^2$$

Luego incluyendo la restricción dada por la condición de insesgadez, se debe minimizar:

$$\sigma^2 = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C_{\alpha\beta} + 2 \sum_{\alpha=1}^n \lambda_{\alpha} C_{\alpha 0} + \sigma^2 + \sum_{l=1}^p \mu_l \left[ \sum_{\alpha=1}^n \lambda_{\alpha} f_l(u_{\alpha}) - f_l(u_0) \right]$$

o en términos de la función de semivarianza

$$\sigma^2 = - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} + 2 \sum_{\alpha=1}^n \lambda_{\alpha} + \sigma^2 + \sum_{l=1}^p \mu_l \left[ \sum_{\alpha=1}^n \lambda_{\alpha} f_l(u_{\alpha}) - f_l(u_0) \right]$$

derivando la expresión anterior respecto a  $\lambda_1, \lambda_2, \dots, \lambda_n, \mu_1, \mu_2, \dots, \mu_p$  e igualando a cero las correspondientes derivadas se obtienen las siguientes ecuaciones:

$$\begin{aligned} \sum_{\beta=1}^n \lambda_{\beta} \gamma_{\alpha\beta} + \sum_{l=1}^p \mu_l f_l(u_{\alpha}) &= \gamma_{\alpha 0}, \alpha = 1, 2, \dots, n \\ \sum_{\beta=1}^n \lambda_{\beta} f_l(u_{\beta}) &= f_l(u_0), \beta = 1, 2, \dots, p \end{aligned}$$

en términos matriciales

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1n} & f_{11} & \dots & f_{1n} \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2n} & f_{21} & \dots & f_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \dots & \gamma_{nn} & f_{n1} & \dots & f_{pn} \\ f_{11} & f_{12} & \dots & f_{1n} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_{p1} & f_{p2} & \dots & f_{pn} & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_n \\ \mu_1 \\ \dots \\ \mu_n \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \\ \dots \\ \gamma_{n0} \\ f_{10} \\ \dots \\ f_{p0} \end{bmatrix}$$

donde  $f_{l\beta} = f_l(u_{\beta})$  es la  $l$ -ésima función en el punto  $j$ -ésimo. La varianza de predicción del Kriging Universal esta dada por:

$$\sigma_{UK}^2 = \sum_{\alpha=1}^n \lambda_{\alpha} \gamma_{\alpha 0} + \sum_{l=1}^p \mu_l f_l(u_0)$$

Nótese que si  $p = 1$  y  $f_l(u) = 1$ , el sistema de ecuaciones del Kriging Universal y la varianza de predicción coinciden con las del Kriging Ordinario. En este orden de ideas puede decirse que el Kriging Ordinario es un caso particular del Kriging Universal.

#### 1.9.4 Cokriging

El término *cokriging* se utiliza para los métodos de regresión en los que intervienen varios atributos. Supongamos, pues, que disponemos de dos variables regionalizadas  $Z, Y$  definidas en las mismas localizaciones. La ecuación para la estimación del valor de la variable principal  $Z$  en la localización  $u_0$  viene dada por

$$Z^*(u_0) = \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1}(u_0) Z(u_{\alpha_1}) + \sum_{\alpha_2=1}^{n_2} \lambda_{\alpha_2}^*(u_0) Y(u_{\alpha_2}^*)$$

Este tipo de *kriging* requiere un modelo para la matriz de funciones de covarianza, incluida la covarianza de  $Z$ ,  $C_Z(h)$ , la covarianza de  $Y$ ,  $C_Y(h)$ , la covarianza cruzada de  $Z - Y$ ,  $C_{ZY}(h) = \text{Cov}\{Z(u), Y(u+h)\}$ , y la covarianza cruzada de  $Y - Z$ ,  $C_{YZ}(h)$ .



## Capítulo 2

# Geostatística Multivariante

Hasta ahora se ha estudiado como estimar una propiedad utilizando los valores conocidos de dicha propiedad obtenidos en puntos vecinos o cercanos o bien como hacer uso de una función de tendencia para guiar la estimación de la propiedad.

Se consideran entonces procesos espaciales multivariantes:

$\{Z(U) = (Z_1(u), Z_2(u), \dots, Z_p(u))' / u \in D\}$ ,  $D \subset \mathbb{R}^d$  De forma que:

- Estos  $p$  procesos espaciales univariantes se suponen intercorrelacionados.
- Cada  $Z_i(u)$  se observa en un conjunto  $S_i = \{u_{i,1}, u_{i,2}, \dots, u_{i,n_i}\}$  de  $n_i > 0$  localizaciones. Dos conjuntos  $S_i$  y  $S_j$  son, en general, diferentes para  $i \neq j$ .
- $Z_i(u) = m_i(u) + Y_i(u)$ ,  $i = 1, 2, \dots, p$  donde  $m_i(u)$  es la componente determinista y representa los cambios o evolución a gran escala e  $Y_i(u)$  es la componente aleatoria (errática) y representa el comportamiento local o evolución a pequeña escala.

Sin pérdida de generalidad, suele suponerse que la variable de interés (variable a predecir) es  $Z_1$  mientras que las restantes se denominan variables secundarias.

### 2.1 Análisis estructural

Al igual que en el caso univariante, sólo se dispone de una realización del proceso multivariante y por eso es necesario asumir ciertas hipótesis de estacionariedad sobre  $Z(u)$ .

#### Proceso multivariante de segundo orden

- Existe  $\mathbb{E}[Z_i(u)] = m_i, \forall i = 1, 2, \dots, p, \forall u \in D$ .
- Existe  $\text{Cov}(Z_i(u), Z_j(u+h)) = C_{i,j}(h); \forall i, j = 1, 2, \dots, p, \forall u \in D$  y  $\forall h \in \mathbb{R}^d$ .

Las funciones  $C_{i,j}(h)$  reciben el nombre de *covariogramas cruzados*.

1. En general los covariogramas cruzados, al contrario que en el caso unidimensional, no son simétricos:  

$$C_{i,j}(h) = \text{Cov}(Z_i(u), Z_j(u+h)) \neq \text{Cov}(Z_j(u), Z_i(u+h)) = C_{i,j}(-h).$$
2. Se verifica que  $C_{i,j}(h) = C_{j,i}(-h)$
3. Por la desigualdad de Cauchy-Schwarz  

$$|C_{i,j}(h)|^2 \leq |C_{i,j}(h)| |C_{j,j}(h)| \leq |C_{i,i}(0)| |C_{j,j}(0)|,$$
 por lo que los covariogramas cruzados están siempre acotados superiormente.

### Proceso multivariante intrínsecamente estacionario

- Existe  $\mathbb{E}[Z_i(u) - Z_i(u+h)] = m_i(h), \forall i = 1, 2, \dots, p, \forall u \in D.$
- Existe  

$$\frac{1}{2} \text{Cov}((Z_i(u) - Z_i(u+h)), (Z_j(u) - Z_j(u+h))) = \gamma_{i,j}(h);$$

$$\forall i, j = 1, 2, \dots, p, \forall u \in D \text{ y } \forall h \in \mathbb{R}^d.$$

Las funciones  $\gamma_{i,j}(h)$  reciben el nombre de *semivariogramas cruzados*.

1. Los semivariogramas cruzados son simétricos:  $\gamma_{i,j}(h) = \gamma_{i,j}(-h).$
2. Los semivariogramas cruzados son nulos en el origen:  $\gamma_{i,j}(0) = 0.$
3. En el caso de variables estacionarias de segundo orden, existe una relación entre semivariograma y covariograma cruzados:

$$\gamma_{i,j}(h) = C_{i,j}(0) - \frac{1}{2} [C_{i,j}(h) + C_{i,j}(-h)]$$

4. El semivariograma cruzado es estimable sólo si  $S_i = S_j.$

Es importante señalar que ésta no es la única definición de semivariograma cruzado que puede encontrarse en la bibliografía ya que existen diversas formas de generalizar los semivariogramas unidimensionales al caso multidimensional. Así, por ejemplo, una segunda definición muy extendida es:

$$\frac{1}{2} \text{Var}(Z_i(u) - Z_j(u+h)) = \gamma_{i,j}(h); \forall i, j = 1, 2, \dots, p; \forall u \in D \text{ y } \forall h \in \mathbb{R}^d.$$

A esta segunda expresión se le suele denominar **pseudosemivariograma cruzado**.

1. Por lo general, los pseudosemivariograma cruzados no son simétricos:  

$$\gamma_{i,j}(h) = \gamma_{j,i}(-h) \neq \gamma_{i,j}(-h)$$
2. Los pseudosemivariograma cruzados pueden ser no nulos en el origen:  $\gamma_{i,j}(0) \geq 0.$
3. En el caso de variables estacionarias de segundo orden, no existe una relación entre pseudosemivariograma y covariograma cruzados.

4. El pseudosemivariograma cruzado siempre es estimable, aún cuando  $S_i \neq S_j$ .

En todo caso, si los covariogramas, o los semivariogramas, o los pseudosemivariograma cruzados son sólo función de la distancia  $\|h\|$  y no de la dirección del vector, se denominan isotrópicos. En caso contrario se habla de anisotropía.

A continuación estudiaremos algunas técnicas geoestadísticas propuestas para obtener estimaciones de la propiedad de interés cuando se dispone de observaciones de otras variables secundarias relacionadas con la variable en estudio.

Entre este tipo de técnicas se encuentran:

- Cokriging Simple y Ordinario
- Cokriging colocado (collocated cokriging)

Al igual que en el caso de geoestadística univariada, lo fundamental es contar con una herramienta que mida la correlación espacial de las variables involucradas y su interrelación.

La correlación espacial de cada una de las variables involucradas se obtiene como antes a través de la función de covarianza o del variograma.

La correlación espacial conjunta o la interrelación se obtiene a través de la función de covarianza cruzada que se estudiará a continuación

## 2.2 Cokriging

Planteamiento básico de la estimación por Cokriging:

Considerar la estimación de  $Z_1(u)$  como una combinación lineal de las observaciones disponibles de  $Z_1$  más combinaciones lineales de las observaciones de las variables relacionadas.

Ejemplo:

- $Z$ , propiedad o variable principal, por ejemplo porosidad.
- $Y$ , información o variable secundaria, por ejemplo impedancia acústica.

$$Z_{cok}^*(u) = \sum_{\alpha=1}^N \lambda_{\alpha}(u)Z(u_{\alpha}) + \sum_{\alpha=1}^M \beta_{\alpha}(u)Y(x_{\alpha})$$

El primer sumando es una combinación lineal de la variable principal y el segundo de la variable secundaria.

En el caso general lo único que se complica es la notación:

- $Z$ , propiedad o variable principal, por ejemplo porosidad.
- $Y_i$ , variables secundarias, por ejemplo atributos sísmicos.

$$Z_{cok}^*(u) = \sum_{\alpha=1}^N \lambda_{\alpha}(u)Z(u_{\alpha}) + \sum_{\alpha_1=1}^{N_1} \beta_{\alpha_1}(u)Y_1(x_{\alpha_1}) + \dots + \sum_{\alpha_k=1}^{N_k} \beta_{\alpha_k}(u)Y_k(x_{\alpha_k})$$

El primer sumando es una combinación lineal de la variable principal y los siguientes de las variables secundarias.

$$Z_{cok}^*(u) = \sum_{\alpha=1}^N \lambda_{\alpha}(u)Z(u_{\alpha}) + \sum_{j=1}^K \sum_{\alpha_j=1}^{N_j} \beta_{\alpha_j}(u)Y_j(x_{\alpha_j})$$

### 2.2.1 Cokriging simple

El caso más simple se denomina **cokriging simple** y la hipótesis básica es la estacionaridad de todas las variables junto con el hecho de que se asume que las medias de todas las variables son conocidas. Esto es,

$$\begin{aligned} \mathbb{E}(Z(u)) &= m \text{ conocida} \\ \mathbb{E}(Y_j(u)) &= m_j \text{ conocida } \forall j. \end{aligned}$$

A continuación se obtienen las ecuaciones de cokriging simple en el caso en que se considera sólo una variable secundaria. En este caso el estimador propuesto es,

$$Z_{cok}^*(u) = m + \sum_{\alpha=1}^N \lambda_{\alpha}(u)(Z(u_{\alpha}) - m) + \sum_{\alpha_1=1}^{N_1} \beta_{\alpha_1}(u)(Y_1(x_{\alpha_1}) - m_1)$$

Al igual que en el caso anterior, las condiciones de optimalidad son:

1. Estimador insesgado,  $\mathbb{E}(Z_{cok}^*(u)) = \mathbb{E}(Z(u))$ .
2.  $\text{Var}[Z(u) - Z_{cok}^*(u)]$  mínima.

La primera condición se obtiene automáticamente al utilizar que:

$$\begin{aligned} \mathbb{E}(Z(u_{\alpha}) - m) &= 0 \\ \mathbb{E}(Y_1(x_{\alpha_1}) - m_1) &= 0 \end{aligned}$$

Con lo cual,

$$\mathbb{E}(Z_{cok}^*(u)) = m = \mathbb{E}(Z(u))$$

La condición de varianza mínima se obtiene derivando respecto a los parámetros  $\alpha$  y  $\beta$  e igualando a cero cada una de las derivadas obtenidas.

$$\begin{aligned} \frac{\partial \text{Var}[Z(u) - Z_{cok}^*(u)]}{\partial \lambda_j} &= 0, j = 1, 2, \dots, N \\ \frac{\partial \text{Var}[Z(u) - Z_{cok}^*(u)]}{\partial \beta_j} &= 0, j = 1, 2, \dots, N_1 \end{aligned}$$

Para calcular explícitamente la expresión de la varianza hay que proceder con cautela debido a que aparecen nuevos términos a considerar.

$$\text{Var} [Z(u) - Z_{cok}^*(u)] = \text{Var} [Z(u)] + \text{Var} [Z_{cok}^*(u)] - 2\text{Cov} (Z(u), Z_{cok}^*(u))$$

$$\begin{aligned} T_1 &= \text{Var} [Z(u)] = \sigma^2 \\ T_2 &= \text{Var} [Z_{cok}^*(u)] \\ &= \text{Var} \left( \sum \lambda_i (Z(u_i - m)) \right) + \text{Var} \left( \sum \beta_j (Y(x_j - m_j)) \right) \\ &+ 2\text{Cov} \left( \sum \lambda_i (Z(u_i - m)), \sum \beta_j (Y(x_j - m_j)) \right) \\ &= \sum \sum \lambda_i \lambda_j C_Z(u_i - u_j) + \sum \sum \beta_i \beta_j C_Y(x_i - x_j) \\ &+ 2 \sum \sum \lambda_i \beta_j C_{ZY}(u_i - x_j) \\ T_3 &= 2\text{Cov} (Z(u), Z_{cok}^*(u)) \\ &= 2 \sum \lambda_i C_Z(u - u_i) + 2 \sum \beta_j C_{ZY}(u - x_j) \end{aligned}$$

Al calcular las derivadas respectivas se obtiene que

$$\begin{aligned} \frac{\partial \text{Var} [Z(u) - Z_{cok}^*(u)]}{\partial \lambda_i} &= 2 \sum_{j=1}^N \lambda_j C_Z(u_i - u_j) + 2 \sum_{j=1}^N \beta_j C_{ZY}(u_i - x_j) \\ &- 2C_Z(u - u_i), \quad i = 1, 2, \dots, N \\ \frac{\partial \text{Var} [Z(u) - Z_{cok}^*(u)]}{\partial \beta_i} &= 2 \sum_{j=1}^N \beta_j C_Y(x_i - x_j) + 2 \sum_{j=1}^N \lambda_j C_{ZY}(u_i - x_j) \\ &- 2C_{ZY}(u - x_i), \quad i = 1, 2, \dots, N_1 \end{aligned}$$

Ahora la expresión detallada del sistema de ecuaciones es

$$\begin{bmatrix} C_Z & C_{ZY} \\ C_{YZ} & C_Y \end{bmatrix} \begin{bmatrix} \lambda \\ \beta \end{bmatrix} = \begin{bmatrix} C_{ZU} \\ C_{YU} \end{bmatrix}$$

Siendo cada una de las submatrices las siguientes:

$$\begin{bmatrix} C_Z \\ C_{YZ} \end{bmatrix} = \begin{bmatrix} C_Z(0) & C_Z(u_1 - u_2) & \dots & C_Z(u_1 - u_N) \\ C_Z(u_2 - u_1) & C_Z(0) & \dots & C_Z(u_2 - u_N) \\ \vdots & \vdots & \dots & \vdots \\ C_Z(u_N - u_1) & C_Z(u_N - u_2) & \dots & C_Z(0) \\ C_{YZ}(x_1 - u_1) & C_{YZ}(x_2 - u_1) & \dots & C_{YZ}(x_1 - u_N) \\ C_{YZ}(x_2 - u_1) & C_{YZ}(x_2 - u_2) & \dots & C_{YZ}(x_2 - u_N) \\ \vdots & \vdots & \dots & \vdots \\ C_{YZ}(x_1 - u_N) & C_{YZ}(x_N - u_2) & \dots & C_{YZ}(x_N - u_N) \end{bmatrix}$$

$$\begin{bmatrix} C_{ZY} \\ C_Y \end{bmatrix} = \begin{bmatrix} C_{YZ}(u_1 - x_1) & C_{YZ}(u_1 - x_2) & \dots & C_{YZ}(u_1 - x_N) \\ C_{YZ}(u_2 - x_1) & C_{YZ}(u_2 - x_2) & \dots & C_{YZ}(u_2 - x_N) \\ \vdots & \vdots & \dots & \vdots \\ C_{YZ}(u_N - x_1) & C_{YZ}(u_N - x_2) & \dots & C_{YZ}(u_N - x_N) \\ C_Y(0) & C_Y(x_1 - x_2) & \dots & C_Y(x_1 - x_N) \\ C_Z(x_2 - x_1) & C_Y(0) & \dots & C_Y(x_2 - x_N) \\ \vdots & \vdots & \dots & \vdots \\ C_Y(x_N - x_1) & C_Y(x_N - x_2) & \dots & C_Y(0) \end{bmatrix}$$

$$\begin{bmatrix} \lambda \\ \beta \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} \quad \begin{bmatrix} C_{YU} \\ C_{YU} \end{bmatrix} = \begin{bmatrix} C_Z(u - u_1) \\ C_Z(u - u_2) \\ \vdots \\ C_Z(u - u_N) \\ C_Y(u - x_1) \\ C_Y(u - x_2) \\ \vdots \\ C_Y(u - x_N) \end{bmatrix}$$

### 2.2.2 Cokriging ordinario

Al igual que en el caso de kriging ordinario, se asume que las medias de las variables son desconocidas y se imponen condiciones para filtrarlas.

El estimador propuesto es:

$$Z_{cok}^*(u) = \sum_{\alpha=1}^N \lambda_{\alpha}(u)Z(u_{\alpha}) + \sum_{j=1}^K \sum_{\alpha_j=1}^{N_j} \beta_{\alpha_j}(u)Y_j(x_{\alpha_j})$$

Con lo cual,

$$\mathbb{E}(Z_{cok}^*(u)) = m \sum \lambda_{\alpha} + \sum_{j=1}^K m_j \sum_{\alpha_j=1}^{N_j} \beta_{\alpha_j}$$

Y se obtienen las condiciones  $\sum \lambda_{\alpha} = 1$ ,  $\sum_{\alpha_j=1}^{N_j} \beta_{\alpha_j} = 0$ ,  $j = 1, 2, \dots, K$ .

Ahora se procede nuevamente como en el kriging ordinario pero con  $K + 1$  parámetros de Lagrange. Cuando se tiene tan solo una variable secundaria, el sistema de ecuaciones del cokriging ordinario es,

$$\begin{bmatrix} C_Z & C_{ZY} & 1 & 0 \\ C_{YZ} & C_Y & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \beta \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} C_{ZU} \\ C_{YU} \\ 1 \\ 0 \end{bmatrix}$$

Obsevaciones:

1. Con sólo 2 variables se requieren 4 funciones de covarianza. En general, con  $N$  variables secundarias se requieren  $2^{N+1}$  funciones de covarianza.
2. Debe existir una correlación lineal entre las variable principal y las variables secundarias.
3. Las variables secundarias deben poseer un número mucho mayor de observaciones que la variable principal.
4. Imposible estimar las covarianzas cruzadas con datos NO coincidentes.
5. Resultados satisfactorios se obtienen con datos parcialmente coincidentes.
6. Con datos totalmente coincidentes. Conveniente para estimar de manera consistente el tope y la base de un yacimiento. No se obtiene una mejora sustancial sobre los métodos de kriging cuando la variable secundaria es la información sísmica.

Cuando las variables están *intrínsecamente* relacionadas, es decir, cuando ocurre que los modelos de variograma o covarianza de todas las variables son proporcionales a un mismo modelo de variograma o covarianza, entonces el kriging y el cokriging con datos totalmente coincidentes son iguales.

## 2.3 Cokriging universal

Representamos por  $\{Z_i(u_{i,j}), i = 1, \dots, p; j = 1, \dots, n_1\}$  la muestra a partir de la cual pretendemos predecir  $Z_1(u_0)$ ,  $u_0 \in D$ . Al igual que en el caso univariante, dependiendo de las suposiciones sobre las funciones de tendencia  $m_1(u), \dots, m_p(u)$  se distinguen tres tipos de predicción lineal óptima multivariante:

1. Cokriging simple. Si las funciones de tendencia son conocidas.
2. Cokriging ordinario. Si las funciones de tendencia son desconocidas pero constantes.
3. Cokriging universal. Si las funciones de tendencia son desconocidas.

Evidentemente, el último caso es el más general y, por ello, el más utilizado. En el caso de cokriging universal se asume que cada tendencia  $m_i(u)$  puede expresarse como combinación lineal de funciones regresoras conocidas:

$$m_i(u) = \sum_{l=0}^{L_i} a_{i,l} f_i^l(u); \quad i = 1, 2, \dots, p.$$

El predictor cokriging se define como el mejor predictor lineal insesgado (BLUP) calculado con todas las variables observadas:

$$Z_1^*(u_0) = \sum_{i=1}^p \sum_j^{n_i} \lambda_{i,j} Z_i(u_{i,j}).$$

## 2.4 Condición de insesgadez

Para que el predictor sea insesgado son condiciones necesarias:

$$\begin{aligned} \sum_j^{n_1} \lambda_{1,j} f_1^l(u_{1,j}) &= f_1^l(u_0); \quad l = 0, 1, \dots, L_1; u_{1,j} \in S_1. \\ \sum_j^{n_i} \lambda_{i,j} f_i^l(u_{i,j}) &= 0; \quad l = 0, 1, \dots, L_i; u_{i,j} \in S_i, i = 2, \dots, p. \end{aligned}$$

Si ambas se verifican, entonces:

$$\mathbb{E}[Z_1^*(u_0)] = \sum_{l=0}^{L_1} a_{1,l} f_1^l(u) = \mathbb{E}[Z_1(u_0)].$$

### Condición de varianza mínima

Al igual que en el caso unidimensional, se trata de un problema de minimización de varianza de predicción sujeto a las restricciones necesarias para asegurar la insesgadez del predictor. La función a minimizar se plantea utilizando  $L_1 + \dots + L_p + p$  Multiplicadores de Lagange:

$$\begin{aligned} \text{Var}(Z_1^*(u_0) - Z_1(u_0)) &+ 2 \sum_{l=0}^{L_1} \mu_{1,l} \left( \sum_{j=1}^{n_1} \lambda_{1,j} f_1^l(u_{1,j}) - f_1^l(u_0) \right) \\ &+ 2 \sum_{i=2}^p \sum_{l=0}^{L_i} \mu_{i,l} \left( \sum_{j=1}^{n_i} \lambda_{i,j} f_i^l(u_{i,j}) \right) \end{aligned}$$

Mediante derivadas parciales de la expresión anterior se obtiene la expresión matricial de la cual se deduce el vector de pesos óptimo y la varianza de predicción. A continuación se muestra el aspecto del sistema matricial a resolver en el caso de  $p = 2$  variables estacionarias de segundo orden y con tendencias  $m_1$  y  $m_2$  desconocidas pero constantes (Cokriging Ordinario).

## Capítulo 3

# Análisis de datos funcionales

En muchas áreas se ha empezado a trabajar con grandes bases de datos, que cada vez con más frecuencia, corresponden a observaciones de una variable aleatoria tomadas a lo largo de un intervalo continuo (o en discretizaciones cada vez más extensas de este intervalo continuo).

En campos como la espectrometría, el resultado de la medición es una curva que representa a la muestra concreta que al menos se ha evaluado en una centena de puntos. Este tipo de datos, que llamaremos datos funcionales, surgen de manera natural en muchas disciplinas, y como este ejemplo podríamos citar muchos otros en diversos campos como la economía, ingeniería, medio ambiente, . . . . Ante estos nuevos retos surge como respuesta la estadística de datos funcionales que originalmente identificaba dato funcional con función en un intervalo continuo. Básicamente, los problemas a los que se debe enfrentar la estadística con datos funcionales responde a las mismas necesidades que la estadística clásica. Estos se podrían categorizar de la siguiente manera:

1. Explorar y describir el conjunto de datos funcionales resaltando sus características más importantes.
2. Explicar y modelar la relación entre una variable dependiente y una independiente (modelos de regresión).
3. Métodos de Clasificación Supervisada o no Supervisada de un conjunto de datos respecto a alguna característica.
4. Contraste, validación y predicción.

En ambos casos todas las técnicas incluidas están restringidas al espacio de funciones  $L^2$ , es un espacio con características específicas que lo hacen especialmente tratable. Una variable aleatoria  $X$  se dice que es una **variable funcional** si toma valores en un espacio funcional  $\mathbb{E}$  (Espacio normado o seminormado completo), como se puede ver en [8].

Un conjunto de datos funcionales  $\{X_1, \dots, X_n\}$  es la observación de  $n$  variables funcionales  $X_1, \dots, X_n$  idénticamente distribuidas.

La primera dificultad que siempre tendremos al analizar datos funcionales, es encontrar una representación adecuada para los datos.

Estas definiciones se pueden aplicar a muchos tipos de espacios. En particular,  $\mathbb{R}^p$  con las métricas usuales es un espacio funcional y por tanto puede deducirse que toda técnica que se desarrolle para datos funcionales puede ser aplicada con ciertas garantías en el entorno multivariante. El espacio más comúnmente usado cuando se habla de datos funcionales es el espacio  $L^2[S]$ , esto es, las funciones de cuadrado integrable en el intervalo  $S = [a, b] \subset \mathbb{R}$ . Desde un punto de vista más general podemos tener datos funcionales en la familia:  $L^p[S, \mu] = \{f : S \rightarrow \mathbb{R} \text{ tal que } \int |f(t)|^p d\mu\}$ , donde  $(S, \mu)$  es un espacio de medida y  $1 < p < \infty$ . Estos espacios son semi-normados salvo el caso  $p = 2$  que es el único de esta familia que es un espacio de Hilbert separable. Cuando se desarrolla una nueva técnica para datos funcionales la primera preocupación es siempre determinar en que espacio funcional vamos a trabajar. Esto determinará decisivamente el conjunto de herramientas que podremos usar. Una preocupación similar la tendremos al aplicar una técnica de datos funcionales a un conjunto de datos. La métrica del espacio funcional que se elija para encuadrar estos datos debe ser coherente con la interpretación física del fenómeno que describan.

En general, la representación de un dato funcional en una base ortonormal proporcionará ventajas tanto desde el punto de vista teórico como práctico sirviendo de puente entre la inevitable discretización del dato funcional y su verdadera forma funcional.

Una base es un conjunto de funciones conocidas e independientes  $\{\phi_k\}_{k \in \mathbb{N}}$  tales que cualquier función puede ser aproximada, tan bien como se quiera, mediante una combinación lineal de  $K$  de ellas con  $K$  suficientemente grande. De esta forma, la observación funcional puede aproximarse como  $x(t) \approx \sum_{k=1}^K c_k \phi_k(t)$ .

Si los elementos de la base son fácilmente diferenciables hasta orden  $q$  tenemos  $x^{(q)}(t) \approx \sum_{k=1}^K c_k \phi_k^{(q)}(t)$ .

Básicamente, la idea clave cuando se pueden usar bases ortonormales es representar cada dato funcional en la base usando aquellas coordenadas que son más significativas. Debido a la alta dimensión de los datos funcionales, se elige en general un número  $K$  para representar los datos en el subespacio, convirtiendo el problema de dimensión infinita en un problema multidimensional. La elección del parámetro  $K$  y de la base más adecuada para los datos observados se antoja crucial y, en principio, no hay ninguna regla que permita hacer una selección óptima de forma universal. El parámetro  $K$  es, en cierto modo, un parámetro de suavización de los datos funcionales. Si  $K$  es bajo tendremos un modelo muy manejable pero posiblemente habremos perdido información relevante. Si  $K$  es alto representaremos muy bien los datos pero el problema de la dimensión cobra importancia. Si atendemos a la elección de la base, para datos periódicos se suele emplear la base de Fourier y para datos no periódicos la base B-spline o la Wavelet. Una base muy popular está basada en la expansión de Karhunen-Loève que no es más que la extensión del análisis de componentes principales multivariante a procesos estocásticos y por añadidura a datos funcionales. Calculando a partir del operador momento de segundo orden muestral las correspondientes autofunciones y autovalores es posible construir específicamente una base ortonormal adaptada para cada conjunto de datos. Esta técnica se denomina Componentes Principales Funcionales (FPCA) y ha dado lugar a muchas técnicas interesantes para datos funcionales. Sin embargo, esta técnica puede ser muy sensible a la aparición de datos atípicos y la representación del dato funcional puede no ser relevante para el objetivo del estudio como podría ser la relación con otra variable funcional o no. La decisión sobre qué base

elegir debe tomarse en función del objetivo del estudio y los datos y aprovechando las ventajas e inconvenientes que presenta cada tipo de base. Si se trunca cualquiera de estas bases en un número determinado de elementos obtendremos una semimétrica que también podremos usar para manejar los datos funcionales. En este caso, cualquier métrica o semi-métrica en el espacio no es más que una forma de determinar qué elementos del espacio están cercanos y cuáles lejanos.

La estadística con datos funcionales tiene frontera con otros campos relevantes de la estadística como el análisis multivariante, el análisis de datos longitudinales o las series temporales. Como se comentó anteriormente, una técnica de datos funcionales puede aplicarse con ciertas garantías a datos multivariantes. Al revés, en general, no es cierto. Para la mayoría de las técnicas multivariantes que basan mucho de su trabajo en propiedades del álgebra matricial puede ser un problema casi insalvable tratar datos funcionales de alta frecuencia con seguridad, muy fuerte colinealidad. Según aumenta el grado de resolución con el que somos capaces de ver una curva, más difícil resulta para las técnicas multivariantes obtener un resultado convirtiendo el aumento de resolución en una dificultad más que en una oportunidad de obtener mejor información. Algo similar podría decirse del análisis de datos longitudinales. En este campo se obtienen medidas repetidas a lo largo del tiempo para el mismo sujeto, pero en general, éste es un número pequeño y las técnicas multivariantes pueden adaptarse para trabajar con ellas. La principal dificultad para tratar datos longitudinales como datos funcionales suele ser precisamente la baja calidad de representación de las curvas. La relación con el campo de las series temporales es totalmente diferente. Así, ejemplos clásicos de datos funcionales se han construido a base de cortar una serie temporal en ciclos homogéneos. Por ejemplo, en [25] se usan los datos de un índice bursátil estadounidense troceados por años (como unidad funcional) para deducir a partir de la forma de cada curva anual la tipología de los distintos años (de expansión, de crisis, ...). Considerados los datos como una serie temporal, el objetivo es predecir alguno de los periodos del próximo año. Como conjunto de datos funcionales, el objetivo es resumir la información y el resultado será siempre un dato funcional, esto es, un ciclo anual completo. Por tanto, la relación entre estos dos campos es peculiar. Muchas veces trabajan sobre la misma información pero desde ópticas completamente diferentes.

### 3.1 Preliminares

Se presentan unas definiciones preliminares sobre medidas clásicas y distancias entre vectores en  $\mathbb{R}^p$  en contexto funcional. Se asume que  $X(t), Y(t), Z(t), t \in T$  son funciones definidas en algún espacio de funciones.

- *Producto interior*

$$\langle X(t), Y(t) \rangle = \int_T X(t)Y(t)dt.$$

Propiedades:

1. Simetría:  $\langle X(t), Y(t) \rangle = \langle Y(t), X(t) \rangle$ .
2. Positividad:  $\langle X(t), X(t) \rangle \geq 0$ .

3. Bilinealidad:  $\langle aX(t) + bY(t), Z(t) \rangle = a \langle X(t), Z(t) \rangle + b \langle Y(t), Z(t) \rangle$ , para cualesquiera números reales  $a$  y  $b$ .

- Norma

$$\|X(t)\| = \sqrt{\|X(t)\|^2}, \text{ donde}$$

$$\|X(t)\|^2 = \langle X(t), X(t) \rangle = \int_T X(t)X(t)dt.$$

Propiedades:

1.  $\|X(t)\| \geq 0$ .
2.  $\|aX(t)\| = |a| \|X(t)\|$ , para cualquier número real  $a$ .
3.  $\|X(t) + Y(t)\| \leq \|X(t)\| + \|Y(t)\|$ .
4.  $|\langle X(t), Y(t) \rangle| \leq \|X(t)\| \|Y(t)\| = \sqrt{|\langle X(t), X(t) \rangle| |\langle Y(t), Y(t) \rangle|}$ .
5.  $-1 \leq \frac{|\langle X(t), Y(t) \rangle|}{(\|X(t)\| \|Y(t)\|)} \leq 1$

Un dato funcional  $\chi_i(t) t \in T$ , se representa generalmente como un conjunto finito de pares  $(t_j, y_{ij}), t_j \in T, j = 1, \dots, M$  e  $y_{ij} = \chi_j(t_j)$  (si no hay ruido blanco) o  $y_{ij} = \chi_j(t_j) + \epsilon_j$  (si hay ruido blanco),  $\epsilon_j$  tiene media cero. El conjunto de puntos  $\{t_j\}_{j=1}^M \subset T$  puede ser considerado el mismo para todas las funciones en un conjunto de datos funcionales. Métodos de interpolación (si no hay ruido blanco) o métodos no paramétricos de suavización (en caso contrario) son comunmente usados para representar los conjuntos discretos  $(t_j, y_{ij}), j = 1, \dots, M$ , como una función real  $\chi_i$ . En este sentido se puede decir que el hereda la metodología de la estimación no paramétrica funcional.

### 3.2 Técnicas exploratorias para Datos Funcionales

En [25] se recogen como herramientas para resumir los datos: la media funcional, la varianza funcional y la función de covarianza. En un capítulo posterior se emplean las componentes principales funcionales como herramientas del análisis descriptivo. Básicamente esto era todo el análisis descriptivo de un conjunto de datos funcionales. Sin embargo, el análisis descriptivo se revela decisivo para el tratamiento de datos funcionales. La cuestión se complica si pensamos que nuestros datos pueden estar sujetos a métricas no usuales y por tanto, las representaciones usuales engañarían nuestra mirada. En este campo se echan en falta herramientas descriptivas que en otros ámbitos como el multivariante se han desarrollado expresamente. Manejando diferentes conceptos sobre profundidad estadística también se han definido extensiones de medidas robustas para datos funcionales, incluyendo incluso el bootstrap para datos funcionales como herramienta para analizar la variabilidad de los distintos estimadores. Durante todo este trabajo estamos en  $L^2(T)$ .

### 3.2.1 Media, Varianza, covarianza, correlación, covarianza cruzada y correlación cruzada

- Media:  $\bar{X}_t = \frac{1}{n} \sum_{i=1}^n X_i(t)$ .
- Varianza:  $\text{Var}(X) = \frac{1}{n-1} \sum_{i=1}^n (X_i(t) - \bar{X}(t)) (X_i(t) - \bar{X}(t))'$ .
- Covarianza:  $\text{Cov}(X(t_1), X(t_1)) = \frac{1}{n-1} \sum_{i=1}^n (X_i(t_1) - \bar{x}(t_1)) (X_i(t_2) - \bar{x}(t_2))$ .
- Correlación:  $\text{Corr}(X(t_1), X(t_1)) = \frac{\text{Cov}(X(t_1), X(t_1))}{\sqrt{\text{Var}(X(t_1))\text{Var}(X(t_2))}}$ .
- Covarianza cruzada:  $\text{Cov}(X(t_1), Y(t_2)) = \frac{1}{n-1} \sum_{i=1}^n (X_i(t_1) - \bar{X}(t_1)) (Y_i(t_2) - \bar{Y}(t_2))$ .
- Correlación cruzada:  $\text{Corr}(X(t_1), Y(t_2)) = \frac{\text{Cov}(X(t_1), Y(t_2))}{\sqrt{\text{Var}(X(t_1))\text{Var}(Y(t_2))}}$ .

### 3.2.2 Componentes principales

#### Componentes principales Multivariante

Las *componentes principales* son una herramienta básica en el entorno multivariante diseñada para explicar un conjunto de datos  $\mathbf{X} = (X_{ij})_{n \times p}$  mediante una combinación de variables ortonormales que cumplen la propiedad de maximizar la varianza. Es la solución del siguiente algoritmo:

- Encontrar el vector  $\xi_1$  de norma 1 tal que la combinación lineal  $\vec{f}_1 = \xi_1' \mathbf{X}$  maximice  $\sum_{i=1}^p f_{i1}^2$ .
- Repetir el proceso anterior, exigiendo además en el paso  $m$  que  $\xi_m \perp \xi_k, k < m$ .

#### Componentes principales para FDA (FPCA)

Algoritmo adaptado para datos funcionales:

- Encontrar la función  $\xi_1(s)$  de norma 1 ( $\int \xi_1(s)^2 ds = 1$ ) tal que  $f_1 = \int \xi_1(s) X_i(s)$  maximice  $\sum_{i=1}^p f_{i1}^2$ .
- Repetir el proceso anterior, exigiendo además en el paso  $m$  que  $\xi_m(s) \perp \xi_k(s), k < m$ .

La descomposición en esta base ortonormal permite escribir cada dato como:

$$\hat{X}_i(t) = \sum_{k=1}^K f_{ik} \xi_k(t)$$

donde  $f_{ik}$  es el valor de la componente principal  $\int X_i \xi_k$ .

Las características más notables de las componentes principales son las siguientes:

- Resumen rápidamente la información de la muestra.
- Permiten obtener una base ortonormal empírica adaptada a los datos.
- Pueden servir para detectar datos atípicos (aunque también pueden esconderlos).
- La rotación de las componentes pueden ayudar a encontrar mejores explicaciones de las componentes.
- Posibilidad de modificar el algoritmo para conseguir suavidad en las componentes.

### 3.3 Bases para datos funcionales

Una *base* es un conjunto de funciones conocidas e independientes  $\{\phi_k\}_{k \in \mathbb{N}}$  tales que cualquier función puede ser aproximada, tan bien como se quiera, mediante una combinación lineal de  $K$  de ellas. De esta forma, la observación funcional puede aproximarse como  $\chi(t) = \sum_{k \in \mathbb{N}} c_k \phi_k(t) \approx \sum_{k=1}^K c_k \phi_k(t)$ .

Si los elementos de la base son fácilmente diferenciables hasta orden  $q$  tenemos  $\chi^{(q)}(t) = \sum_{k \in \mathbb{N}} c_k \phi_k^{(q)}(t) \approx \sum_{k=1}^K c_k \phi_k^{(q)}(t)$ .

Las bases que se suelen usar para datos funcionales son las *bases de Fourier* si los datos son periódicos, *bases B-spline* para cálculos rápidos y flexibles, *bases de Wavelets* apropiada para modelizar discontinuidades, *exponencial, potencial, polonomial, ...*.

Bibliografía adecuada, con respecto a esto se encuentra en los libros de [30] y [26].

Una vez que se ha decidido usar la representación en bases de las funciones surgen las tres preguntas siguientes:

1. ¿Qué tipo de base de funcionaes son las más adecuadas?.
2. ¿Cuántas bases se deben seleccionar describir nuestros datos?.
3. ¿Cómo se determinan los coeficientes  $\mathbf{c}$  basados en las funciones parcialmente observada?.

En el libro [26] se encuentran detalladamente resueltas estas preguntas. A continuación se describen las bases más habituales y a lo largo del capítulo 4 se definen criterios útiles para la elección del número de bases de funciones. El método de mínimos cuadrados estándar se usa para estimar el vector de coeficientes  $c$ , aunque, una matriz de penalización también puede ser incluida en el problema de minimización.

#### 3.3.1 Bases de Fourier

Una *base de Fourier* es una base periódica de período  $\frac{2\pi}{\omega}$  que cuando se seleccionan datos  $\{t_j\}$  equiespaciados en  $T = [0, T]$  y  $\omega = \frac{2\pi}{T}$  está formada por las siguientes funciones ortonormales:

$$\begin{aligned}\phi_0(t) &= \frac{1}{\sqrt{T}}, \\ \phi_{2r-1}(t) &= \frac{\sin(r\omega t)}{\sqrt{\frac{T}{2}}}, y \\ \phi_{2r}(t) &= \frac{\cos(r\omega t)}{\sqrt{\frac{T}{2}}}.\end{aligned}$$

### 3.3.2 Bases B-Splines

Un *spline* es un conjunto de polinomios (de orden  $m$ ) definidos en subintervalos contruidos de tal modo que al final del polinomio en un subintervalo coincida con el inicio del polinomio del siguiente subintervalo (hasta derivada  $m - 2$ ). Los puntos de corte de subintervalos se llaman *nodos*.  $\tau = \{\tau_l\}_{l=0}^L$ . Los *B-splines* (Bases de splines) se calculan fácilmente con el algoritmo de Boor. Entre ellos los más utilizados son los cúbicos. El número de parámetros para definir una función spline es el número de nodos interiores  $(L - 1) +$  el orden del polinomio ( $m$ ).

$$S(t) = \sum_{k=1}^{m+L-1} c_k B_k(t, \tau)$$

### 3.3.3 Bases de Wavelets

La base de *Wavelets* se construye a partir de dos funciones. El wavelet *padre*  $\phi$  que verifica que  $\int \phi(t)dt = 1$  y el wavelet *madre*  $\psi$  que verifica que  $\int \psi(t)dt = 0$ . Los elementos de la base se obtienen a partir de estas dos funciones ortogonales por traslación y cambio de escala.

$$\begin{aligned}\phi_{j,k}(t) &= 2^{-\frac{j}{2}} \phi(2^{-j}t - k), \\ \psi_{j,k}(t) &= 2^{-\frac{j}{2}} \psi(2^{-j}t - k), \\ \int \phi_{j,k}(t) \phi_{j,k'}(t) dt &= \delta_{k,k'}, \\ \int \psi_{j,k}(t) \phi_{j,k'}(t) dt &= 0, \\ \int \psi_{j,k}(t) \psi_{j,k'}(t) dt &= \delta_{j,j'} \delta_{k,k'}.\end{aligned}$$

Elegida la base la aproximación ortogonal wavelet de una función  $f(t)$  viene dada por:

$$\begin{aligned}
f(t) &\approx \sum_k S_{J,k} \phi_{J,k}(t) + \sum_k d_{J,k} \psi_{J,k}(t) \\
&+ \sum_k d_{J-1,k} \psi_{J-1,k}(t) + \cdots + \sum_k d_{1,k} \psi_{1,k}(t)
\end{aligned}$$

Llamando:

$$\begin{aligned}
S_J(t) &= \sum_k S_{J,k} \phi_{J,k}(t) \\
D_J(t) &= \sum_k d_{J,k} \psi_{J,k}(t) \\
D_{J-1}(t) &= \sum_k d_{J-1,k} \psi_{J-1,k}(t) \\
D_1(t) &= \sum_k d_{1,k} \psi_{1,k}(t)
\end{aligned}$$

A la función  $S_j(t)$  se la conoce como *señal suave* y a las funciones  $D_j(t)$  como las *funciones detalle*. A esta descomposición se la llama *descomposición multiresolución*.

### 3.4 Suavización de datos funcionales

Supongamos que observamos  $y(t_j) = x(t_j) + \epsilon(t_j)$  donde el término  $\epsilon(t_j)$  representa el ruido originado al medir los datos. Para recuperar la señal original podemos usar un suavizador lineal, es decir,

$$\hat{x}(t_j) = \sum_{i=1}^N s_j(t_i) y(t_i) \Rightarrow \hat{\mathbf{X}} = \mathbf{S}\mathbf{Y}$$

- Suavización tipo Kernel ( $h$ ):  $s_j(t_i) = \frac{1}{h} K\left(\frac{t_i - t_j}{h}\right)$ .
- Representación truncada ( $K$ ) en una base:  $\mathbf{S} = \Phi (\Phi' \mathbf{W} \Phi)^{-1} \Phi' \mathbf{W}$ .
- Suavización penalizada ( $\lambda$ ):  $\mathbf{S} = \Phi (\Phi' \mathbf{W} \Phi + \lambda \mathbf{R})^{-1} \Phi' \mathbf{W}$ .

Grados de libertad del ajuste:  $df = \text{traza}(\mathbf{S}\mathbf{S}')$ .

### 3.5 Elección de la suavización

En general, se pretende minimizar el error cuadrático medio:

$$\begin{aligned} MSE[\hat{x}(t)] &= \mathbb{E} \left[ \{\hat{x}(t) - x(t)\}^2 \right] \\ &= \text{Sesgo}^2(\hat{x}(t)) + \text{Var}(\hat{x}(t)). \end{aligned}$$

Minimizar (por ejemplo), tenemos el criterio del GCV:

$$\mathbf{GCV}(\theta) = \frac{n}{n - df(\theta)} \frac{(y - \hat{x}_\theta) \mathbf{W} (y - \hat{x}_\theta)}{n - df(\theta)}.$$

La varianza de la predicción viene dada por:

$$\hat{y} = \mathbf{S}y \rightarrow \text{Var}(\hat{y}) = \mathbf{S}\Sigma_e\mathbf{S}'.$$

### 3.5.1 Regresión - representación en base

Observados los pares  $(X_1, y_1), \dots, (X_n, y_n)$ , podemos escribir:

$$\begin{aligned} X_i(t) &= \sum_k^{K_x} c_{ik} \psi_k(t) \rightarrow \mathbf{X} = \mathbf{C}\Psi(t), \\ \beta(t) &= \sum_k^{K_\beta} b_k \theta_k(t) \Rightarrow \beta = \theta' \mathbf{b} \end{aligned}$$

donde resulta

$$\mathbf{b} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'y, \hat{y} = \mathbf{C}\mathbf{J}_{\psi\theta}\mathbf{b}y = \mathbf{Z}\mathbf{b}y = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'y = \mathbf{S}y$$

con  $\mathbf{J}_{\psi\theta} = \int \psi(t)\theta'(t)dt$ .

Versión penalizada

$$\mathbf{b} = (\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{R}_0)^{-1} \mathbf{Z}'y, \hat{y} = \mathbf{C}\mathbf{J}_{\psi\theta}\mathbf{b}y = \mathbf{Z}\mathbf{b}y = \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{R}_0)^{-1} \mathbf{Z}'y = \mathbf{S}y$$

### 3.5.2 Regresión Lineal Funcional

Asúmase que  $(X, Y)$  es un par de variables aleatorias con  $(X, Y) \in L^2(T) \times L^2(T)$ . Suponiendo que ambas variables está centradas, es decir,  $\mathbb{E}[X(t)] = 0$  para  $t \in T$  y  $\mathbb{E}(Y(t)) = 0$  establecemos la siguiente relación entre  $X$  e  $Y$ .

$$Y_i(t) = \int_T X(s)\beta(s, t)ds + \epsilon_i(t),$$

donde ahora el parámetro  $\beta$  es de cuadrado integrable en  $S \times T$  y  $\epsilon_i(t)$  es una variable funcional con media cero.

Escribiendo  $\beta$  como una doble expansión en términos de dos bases  $\eta_k$  y  $\theta_l$  se tiene:

$$\beta(s, t) = \sum_{k=1}^k \sum_{l=1}^L b_{kl} \eta_k(s) \theta_l(t) = \eta(s)' \mathbf{B} \theta(t).$$

Observados los pares  $(X_1, Y_1), \dots, (X_n, Y_n)$  para estimar el modelo previo, se sugiere estimar  $\beta$  minimizando:

$$LMSSSE(\beta) = \int \sum_{i=1}^n \left( Y_i(t) - \int X_i(s) \eta(s)' \mathbf{B} \theta(t) ds \right)^2 dt,$$

Como antes debe restringirse el tamaño de las bases para que la estimación sea razonable.

$$Y(t) = \int X(s) \eta'(s) \mathbf{B} \theta(t) ds + \epsilon(t) = \mathbf{X} \mathbf{B} \theta(t) + \epsilon(t)$$

donde  $\mathbf{X}$  es una matriz  $n \times K$  de la forma  $\mathbf{X} = \int X(s) \eta'(s) ds$ .

Por tanto, de la ecuación anterior, usando la representación de  $\theta(t)$

$$\mathbf{X}' \mathbf{X} \mathbf{B} \int \theta(t) \theta'(t) dt = \mathbf{X}' \int Y(t) \theta'(t) dt,$$

que se puede reescribir:

$$[\mathbf{J}_{\theta\theta} \otimes (\mathbf{X}' \mathbf{X})] \text{vec}(\mathbf{B}) = \text{vec} \left( \mathbf{X}' \int Y(t) \theta'(t) dt \right),$$

donde  $\mathbf{J}_{\theta\theta} = \int \theta(t) \theta'(t) dt$ . Si además  $\mathbf{Y} = \mathbf{C} \phi$  entonces:

$$\text{vec}(\hat{\mathbf{B}}) = [\mathbf{J}_{\theta\theta} \otimes (\mathbf{X}' \mathbf{X})]^{-1} (\mathbf{J}_{\phi\theta} \otimes \mathbf{X}') \text{vec}(\mathbf{C})$$

donde  $\mathbf{J}_{\phi\theta} = \int \phi(t) \theta'(t) dt$ .

La penalización de estos estimadores puede hacerse tanto en  $s$  como en  $t$ . Sea

$$\begin{aligned} \mathbf{R} &= \int [L_s \eta_s] [L'_\eta(s)] ds, \\ \mathbf{S} &= \int [L_t \theta_t] [L'_\theta(t)] dt. \end{aligned}$$

Con términos de penalización pueden escribirse las anteriores ecuaciones como:

$$\mathbf{X}'\mathbf{X}\mathbf{B}\mathbf{J}_{\theta\theta} + \lambda_s\mathbf{R}\mathbf{B}\mathbf{J}_{\theta\theta} + \lambda_t\mathbf{J}_{\eta\eta}\mathbf{B}\mathbf{S} = \mathbf{X}' \int Y(t)\theta'(t)dt$$

donde  $\mathbf{J}_{\eta\eta} = \int \eta(t)\eta'(t)dt$ . Si como antes  $\mathbf{Y} = \mathbf{C}\phi$  entonces:

$$vec(\hat{\mathbf{B}}) = [\mathbf{J}_{\theta\theta} \otimes (\mathbf{X}'\mathbf{X}) + \lambda_s\mathbf{J}_{\theta\theta} \otimes \mathbf{R} + \lambda_t\mathbf{S} \otimes \mathbf{J}_{\eta\eta}]^{-1} (\mathbf{J}_{\phi\theta} \otimes \mathbf{X}') vec(\mathbf{C})$$

donde  $\mathbf{J}_{\phi\theta} = \int \phi(t)\theta'(t)dt$ .



## Capítulo 4

# Ejemplo real de Kriging con datos reales funcionales

En los tres capítulos anteriores se ha presentado el marco teórico necesario para poder llevar a cabo la aplicación a datos reales combinando ambas técnicas. Se usará para cada técnica el conjunto de datos de la temperatura descrito en la introducción.

Antes de describir las técnicas kriging usadas se presenta una estadística descriptiva básica y funcional de los datos junto con unos conceptos básicos de meteorología que ayudarán a interpretar los resultados.

En la parte superior de la Figura (4.1) se tiene la temperatura de Galicia a lo largo del año 2009 y en la inferior lo mismo salvo que destacando la temperatura en la Facultad de Matemáticas (Santiago de Compostela) punto no muestreado.

La Figura (4.2) presenta la temperatura en Galicia a lo largo del año 2009 distribuida a lo largo de las cuatro estaciones del año (primavera, verano, otoño e invierno) destacando siempre la temperatura en el punto no muestreado.

En el Capítulo 3 se ha tratado con detalle la metodología del Análisis de Datos Funcionales, pero como en cualquier análisis estadístico es básico conocer la media, la varianza, el mínimo, máximo, ... de los datos, que es lo que se presenta en la Figura (4.3). El cálculo de estos estadísticos queda desarrollado en dicho capítulo.

### 4.1 Conceptos básicos de meteorología

Se define el *tiempo atmosférico* como el estado de la atmósfera en un determinado momento. Para determinarlo se analizan algunas características como son la temperatura, humedad, presión atmosférica y el viento. El *clima* es el conjunto de valores medios de estas condiciones atmosféricas durante muchos años. Depende fundamentalmente de 3 factores que son:

- **Latitud.** Los rayos solares calientan más a latitudes próximas al Ecuador (los rayos caen más perpendicularmente) que aquellas que se encuentran en los polos.
- **Relieve o altitud.** La temperatura de la troposfera disminuye con la altitud (las capas altas de la atmósfera son menos densas y no retienen el calor). Cuando una corriente de aire

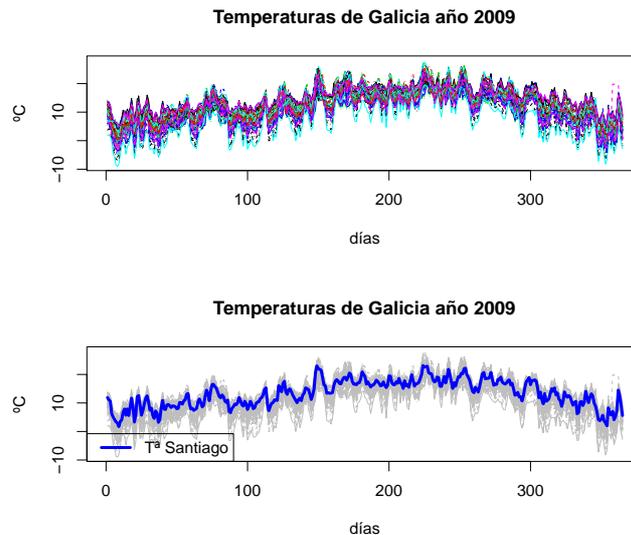


Figura 4.1: En la parte superior Temperatura de Galicia año 2009. En la inferior la de Santiago.

circula por la superficie y se encuentra una cadena montañosa, asciende por ella y la enfría. El vapor de agua que contiene el aire, se condensa y cae en forma de precipitaciones. En las montañas hay temperaturas bajas y llueve más.

- **Distancia al mar.** El agua tiene una elevada capacidad calorífica (tarda mucho en calentarse y luego tarda mucho también en enfriar). En el verano la Tierra se calienta con más intensidad (no transmite el calor, lo acumula en la superficie) con el mar. Por eso si el agua del mar está más fría, refresca el ambiente de las zonas costeras. En el invierno la Tierra enfría más que el mar, y el mar así modera el frío en las zonas costeras. (Por ejemplo: las brisas costeras: terrestres y marinas, nocturnas y diurnas)

Otros conceptos que cabe citar son:

- **Isotérmicas.** Son líneas que unen puntos de igual temperatura.
- **Isobaras.** Son líneas que unen puntos de igual presión atmosférica.
- **Vegetación.** La vegetación también influye en el clima, cuanta más vegetación, mayor cantidad de lluvia. (Hay que tener en cuenta que el movimiento del aire, viento, se produce de las zonas frías, alta presión atmosférica, a las calientes, poca presión atmosférica, a ras del suelo y al revés de las capas altas de la atmósfera.
- **Predicción meteorológica.** La predicción meteorológica se hace en base a los datos suministrados por: las *estaciones meteorológicas* (instalaciones que cuentan con instrumentos de medición, como barómetros, termómetros, pluviómetros y anemómetros), *radares meteorológicos* (instrumentos que permiten conocer en tiempo real donde llueve y con que

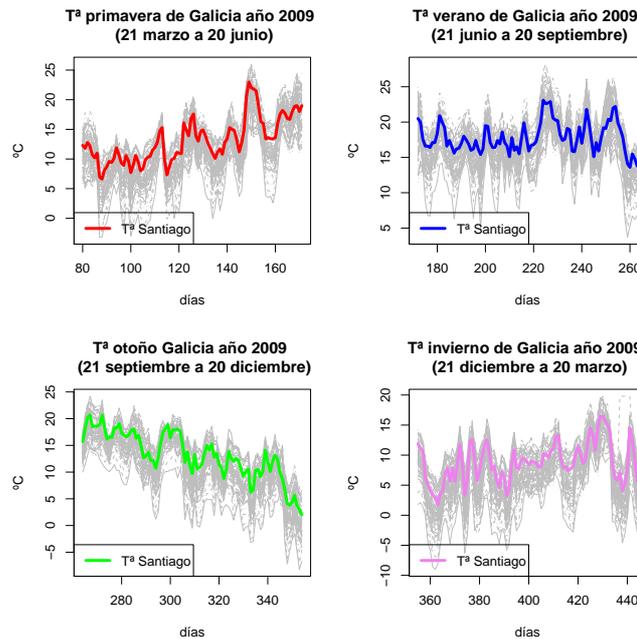


Figura 4.2: Temperatura de Galicia año 2009.

intensidad lo hace) y *satélites meteorológicos* (satélites artificiales que se sitúan en órbita alrededor de la Tierra y permiten el análisis de imágenes e información de la atmósfera desde gran altura).

Todos estos factores determinan que en Galicia predomine el clima **oceánico templado**. Este clima se caracteriza por unas temperaturas suaves en el verano y moderadamente frías en el invierno, así como por unas precipitaciones muy abundantes, sobre todo en el invierno. En la Tabla (4.1) se recogen la temperaturas mínimas, medias y máximas en las 66 estaciones de estudio.

	Primavera	Verano	Otoño	Invierno
Mínima	-3.2	3.7	-8.2	-9.0
Media	26.0	16.9	11.7	7.2
Máxima	11.7	28.2	24.2	19.8

Tabla 4.1: Temperatura mínima, media y máxima en las 66 estaciones. Año 2009.

Sin embargo, al estudiar con detalle las características climatológicas de nuestra Comunidad, podemos distinguir las siguientes variedades de clima: oceánico costero, oceánico continental, oceánico mediterráneo y oceánico de montaña.

- En el *oceánico costero* predomina en las zonas costeras de Pontevedra, A Coruña y Lugo, es decir, abarca toda la zona costera desde A Garda hasta Ribadeo. Se caracteriza por

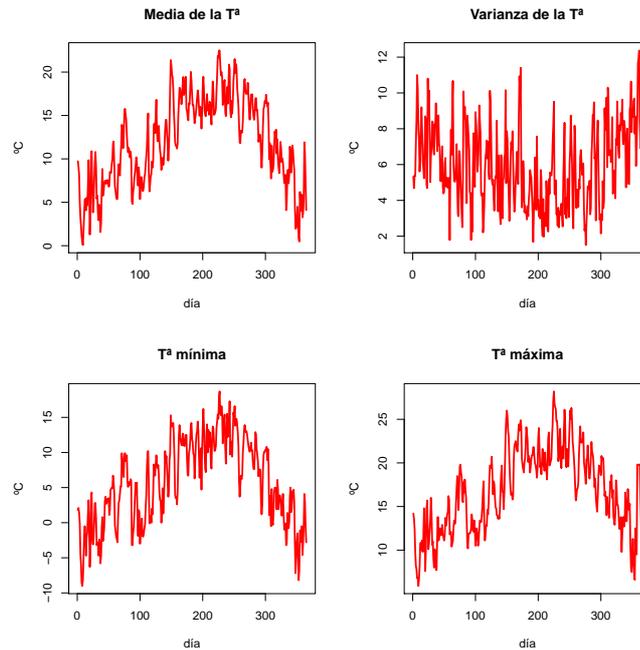


Figura 4.3: Estadísticos descriptivos para la temperatura de Galicia año 2009.

tener temperaturas suaves en el verano y en el invierno y precipitaciones abundantes, que pueden alcanzar los 1.000 o 1.500mm anuales.

- El *oceánico continental* se localiza en las zonas del interior. Se caracteriza por tener veranos cálidos e inviernos fríos con posibilidades de heladas. Las precipitaciones alrededor de los 1.000mm anuales, pudiendo ser de nieve en el invierno.
- El *oceánico mediterráneo* es propio de los valles del Miño y del Sil. Se caracteriza por tener temperaturas elevadas en el verano e suaves en el invierno. Pluviosidad más escasa, sobre todo en el verano, oscilando entre los 600 y 1.000mm anuales.
- El *oceánico de montaña* propio de las zonas montañosas como las sierras de los Ancares, El Caurel, Queixa, . . . . Se caracteriza por tener veranos frescos e inviernos menos fríos que en otras zonas montañosas más alejadas del mar. Precipitaciones abundantes, de hasta 1.500mm anuales, que en el invierno pueden ser de nieve.

## 4.2 Kriging ordinario para funciones-valores de datos espaciales

En esta sección se considera un primer enfoque con respecto al problema de predicción espacial de datos funcionales. Se propone un procedimiento de kriging funcional donde la curva a predecir es una combinación lineal de las curvas observadas y donde los coeficientes son números reales.

Este problema ha sido estudiado inicialmente por Goulard y Voltz (1993), (véase [11]). En su caso, las funciones eran conocidas solamente por un conjunto finito de puntos y el modelo paramétrico los ajustaba para reconstruir la curva entera. En este contexto, el modelo paramétrico se suponía conocido y tanto el número de puntos conocidos para cada función como el número de los parámetros en el modelo paramétrico, se asumía que eran pequeños.

En el caso que nos ocupa se propone aplicar un ajuste no paramétrico al pre-proceso de las funciones observadas por ello se usarán bases de Fourier para la suavización. Por ello se reconstruyen los datos a datos a continuo, es decir, representándolos en una base. Los valores medios de temperatura, por la naturaleza de los sistemas atmosféricos, son una función continua  $x(t)$  a lo largo período de tiempo  $t \in [1, 365]$ . En el seguimiento de la temperatura, el proceso continuo es observado y registrado como discreto en intervalos de  $t_j, j = 1, 2, \dots, J$ . Para cada día se tiene registrado 66 observaciones, una en cada estación. Es decir, para una observación de cualquiera en un día  $i$  se tienen 66 registros discretos,  $X_{M,I,1}, \dots, X_{m,I,66}$ . Estos puntos discretos se pueden utilizar para reconstruir una observación continua mediante técnicas de interpolación, de suavizado o el uso de bases, ([23]). Una **base** es un conjunto de funciones conocidas e independientes  $\phi_k, k \in \mathbb{N}$  tales que cualquier función puede ser aproximada, tan bien como se quiera mediante una combinación  $K$  de ellas. Mediante esta definición podemos expresar la temperatura en una base para un día  $i$  del modo que sigue:

$$x_{m,i}(t) = \sum_{k=1}^K c_{m,i,k} \phi_k(t)$$

Las bases más usadas son las B-splines y las de Fourier. Las primeras se usan para cálculos rápidos y flexibles y las segundas para datos periódicos tales como los ciclos diarios de la temperatura como es el caso que nos ocupa. Una vez se tiene un dato expresado en la forma de la base de Fourier se pueden expresar todos los datos como elementos de esa base. Si  $X_{M,I,1}, \dots, X_{m,I,66}$  son todas las observaciones de una de las variables, se puede obtener una representación de la base minimizando:

$$Q = \sum_{j=1}^{66} [X_{m,i,j} - x_{m,i}(t_j)]^2$$

con respecto a los  $c_{m,i,k}$ , (con  $K \leq 66$ , ya que cada día se tienen 66 observaciones). El número de bases  $K$  determina la suavidad resultante del ajuste, esto es, valores pequeños o grandes de  $K$  proporcionan más o menos suavidad en el ajuste. Hay que tener cuidado que no suceda se que se esté interpolando, ([23]).

Para cada estación y tanto para la temperatura se ha representado el 1 de enero del 2009 como elemento de una base de Fourier de 21, 33 y 65 elementos. Se ha determinado que en ambos casos la base de Fourier de 65 proporciona un nivel de ajuste aceptable para los datos diarios. Después de esta conversión el objeto de estudio es reconstruir no sólo el 1 de enero del 2009 sino los 365 días del año. En la Figura 4.4 puede observarse el proceso seguido.

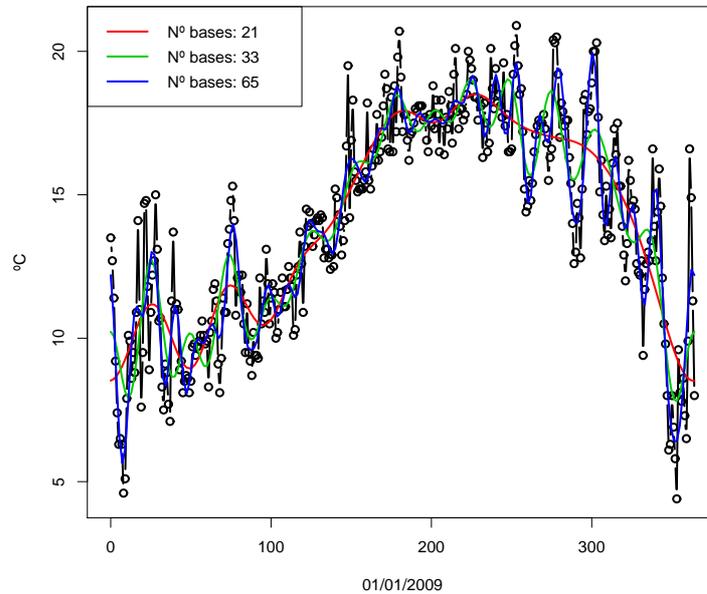


Figura 4.4: Representación en bases.

Este enfoque está totalmente de acuerdo con las tendencias actuales del FDA, y, en particular, con la metodología funcional de estimación no paramétrica. La propuesta de Giraldo et al. es hacer Kriging basado en predicción espacial de curvas aleatorias formalmente coincide con el Kriging funcional introducido por Goulard y Voltz, ([11]), pero la versión no paramétrica desarrolla notables diferencias (por ejemplo, la representación de los datos) y un problema adicional cómo es el de la elección de los parámetros de suavización (la clave de los métodos no paramétricos). El predictor está basado en la filosofía básica de los datos funcionales, esto es, las curvas son entidades singulares, más que una secuencia de observaciones individuales (Ramsay y Silverman, 2005). El desarrollo de las fórmulas en las tres técnicas pueden consultarse con detalle en los artículos de Giraldo et al.

#### 4.2.1 Predicción y estimación de los parámetros

Se considera un proceso aleatorio funcional  $\chi_s : s \in D \subseteq \mathbb{R}^d$ , generalmente  $d = 2$ , tal que  $\chi_s$  es una variable funcional para cualquier  $s \in D$ . Sean  $s_1, \dots, s_n$  puntos arbitrarios en  $D$ , y considerese que se puede observar una realización funcional del proceso aleatorio  $\chi_s$  en los  $n$  lugares,  $\chi_{s_1}, \dots, \chi_{s_n}$ . Se asume que se tiene un proceso aleatorio estacionario de segundo orden

e isotrópico, es decir, la media y la varianza son funciones constantes y la covarianza depende sólo de la distancia entre los puntos muestreados. Formalmente se asume que:

- $\mathbb{E}(\chi_s(t)) = m(t), \forall t \in T, s \in D.$
- $V(\chi_s(t)) = \sigma^2(t), \forall t \in T, s \in D.$
- $\text{Cov}(\chi_{s_i}(t), \chi_{s_j}(t)) = C(h, t) = C_{s_i s_j}(t), \forall s_i, s_j \in D, t \in T, \text{ donde } h = \|s_i - s_j\|.$
- $\frac{1}{2}V(\chi_{s_i}(t) - \chi_{s_j}(t)) = \gamma(h; t) = \gamma_{s_i s_j}(t), \forall s_i, s_j \in D, t \in T, \text{ donde } h = \|s_i - s_j\|.$

La función  $\gamma(h; t)$ , es una función de  $h$ , llamada **semivariograma** de  $\chi(t)$ . Además, se considera la familia de predictores lineales para  $\chi(t)_{s_0}$ :

$$\hat{\chi}_{s_0}(t) = \sum_{i=1}^n \lambda_i \chi_{s_i}(t), \lambda_1, \dots, \lambda_n \in \mathbb{R}$$

El predictor,  $\hat{\chi}_{s_0}(t)$ , tiene la misma expresión que un kriging ordinario, pero considerando curvas en vez de variables, es decir, la curva de predicción es una combinación lineal de las curvas observadas. Se asume por tanto que cada curva medida es un dato completo. Este enfoque trata la curva completa como una entidad singular. Los  $\lambda$ 's en la ecuación  $\hat{\chi}_{s_0}(t)$  muestran la influencia de las curvas que están alrededor de la localización no muestreada donde se llevará a cabo la predicción. Las curvas de las ubicaciones más cercanas al punto de predicción tendrán mayor influencia que las más alejadas. Este es un primer paso y natural en el modelado de datos espaciales funcionales.

Se considera la condición insesgidez para encontrar el mejor predictor lineal insesgado (*BLUP*). De la condición anterior, media constante, se exige que  $\sum_{i=1}^n \lambda_i = 1$ . En geostatística clásica univariante se asume que las observaciones son realizaciones de una muestra aleatoria  $\{Z(s) : s \in D, D \in \mathbb{R}^d\}$ . El Kriging se define como  $\sum_{i=1}^n \lambda_i Z(s_i)$  y el mejor predictor lineal insesgado (*BLUP*) se obtiene minimizando  $\sigma_{s_0}^2 = V(\hat{Z}(s_0) - Z(s_0))$  sujeto a  $\sum_{i=1}^n \lambda_i = 1$ . Por otro lado en geostatística multivariante, ([28]), los datos,  $\{Z(s_1), \dots, Z(s_n)\}$  se tienen de observaciones de un proceso espacial  $\{Z(s) : s \in D, Z(s) \in \mathbb{R}^m, D \in \mathbb{R}^d\}$ . En este contexto  $V(\hat{Z}(s_0) - Z(s_0))$  es una matriz y el *BLUP* de  $m$  variables en una localización no muestreada  $s_0$  se puede obtener minimizando  $\sigma_{s_0}^2 = \sum_{i=1}^m V(\hat{Z}(s_i) - Z(s_i))$  sujeto a las restricciones que garantizan la condición de insesgidez, esto es, minimizar la traza de la matriz del error cuadrático medio sujeto a alguna restricción dada por la condición de insesgidez ([16]). Extendiendo el criterio al contexto funcional, es decir, reemplazando el sumatorio por una integral, los  $n$  parámetros en el Kriging de  $\chi_{s_0}$  considerados se obtienen como solución del siguiente problema de optimización:

$$\begin{aligned} \min_{\lambda_1, \dots, \lambda_n} \int_T V(\hat{\chi}(s_0)(t) - \chi(s_0)(t)) dt, \\ \text{s.t. } \sum_{i=1}^n \lambda_i = 1 \end{aligned}$$

Donde  $\sum_{i=1}^n \lambda_i = 1$  son las restricciones de insesgadez. De las condiciones de insesgadez y del teorema de Fubini se obtiene:

$$\int_T V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)) = \mathbb{E} \left[ \int_T (\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))^2 dt \right]$$

Con lo que se necesita minimizar:

$$\int_T V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)) + 2\mu \left( \sum_{i=1}^n \lambda_i - 1 \right),$$

donde el predictor  $\hat{\chi}_{s_0}(t) = \sum_{i=1}^n \lambda_i \chi(t)_{s_i}(t)$  y  $\mu$  es el multiplicador de Lagrange usado para tener en cuenta la restricción de insesgadez. La integral en la ecuación anterior se puede escribir como:

$$\sigma_{s_0}^2 = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \int_T C_{ij}(t) dt + \int_T \sigma^2(t) - 2 \sum_{i=1}^n \lambda_i \int_T C_{i_0}(t) dt.$$

Así, la función objetivo se puede escribir:

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \int_T C_{ij}(t) dt + \int_T \sigma^2(t) - 2 \sum_{i=1}^n \lambda_i \int_T C_{i_0}(t) dt + 2\mu \left( \sum_{i=1}^n \lambda_i - 1 \right).$$

Minimizando la ecuación anterior con respecto a  $\lambda_1, \dots, \lambda_n$  y  $\mu$  se obtienen el siguiente conjunto de  $(n + 1)$  ecuaciones:

$$\begin{aligned} \sum_{j=1}^n \lambda_j \int_T C_{1j}(t) dt + \mu &= \int_T C_{10}(t) dt \\ \sum_{j=1}^n \lambda_j \int_T C_{2j}(t) dt + \mu &= \int_T C_{20}(t) dt \\ &\vdots \\ \sum_{j=1}^n \lambda_j \int_T C_{nj}(t) dt + \mu &= \int_T C_{n0}(t) dt \\ \sum_{i=1}^n \lambda_i &= 1 \end{aligned}$$

Las cuales podemos expresar en notación matricial como sigue:

$$\begin{bmatrix} \int_T C_{s_1 s_1}(t) dt & \dots & \int_T C_{s_1 s_n}(t) dt & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \int_T C_{s_n s_1}(t) dt & \dots & \int_T C_{s_n s_n}(t) dt & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \int_T C_{s_1 s_0}(t) dt \\ \vdots \\ \int_T C_{s_n s_0}(t) dt \\ 1 \end{bmatrix}$$

También se pueden obtener las estimaciones basadas en la traza del variograma. Asumiendo estacionariedad, se tiene que

$$\gamma_{s_i s_j}(t) = \sigma^2(t) - C_{ij}(t)$$

Entonces

$$\int_T C_{ij}(t) dt = \int_T \sigma^2(t) dt - \int_T \gamma_{s_i s_j}(t) dt$$

Reemplazando la ecuación anterior en el sistema de ecuaciones, se obtiene el sistema:

$$\begin{bmatrix} \int_T \gamma_{s_1 s_1}(t) dt & \dots & \int_T \gamma_{s_1 s_n}(t) dt & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \int_T \gamma_{s_n s_1}(t) dt & \dots & \int_T \gamma_{s_n s_n}(t) dt & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ -\mu \end{bmatrix} = \begin{bmatrix} \int_T \gamma_{s_1 s_0}(t) dt \\ \vdots \\ \int_T \gamma_{s_n s_0}(t) dt \\ 1 \end{bmatrix}$$

De las primeras  $n$  ecuaciones del primer sistema, se tiene la relación

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \int_T C_{ij}(t) dt = \sum_{i=1}^n \int_T C_{i_0}(t) dt - \sum_{i=1}^n \lambda_i \mu.$$

Reemplazando la ecuación anterior en la ecuación  $\sigma_{s_0}^2$ , se obtiene:

$$\sigma^2 = \int_T \sigma^2(t) dt - \sum_{i=1}^n \lambda_i \int_T C_{i_0}(t) dt - \mu.$$

Si además se considera la relación  $\int_T C_{ij}(t) dt = \int_T \sigma^2(t) dt - \int_T \gamma_{s_i s_j}(t) dt$ , la traza de la varianza, de la predicción del kriging ordinario funcional basado en la traza del variograma vendrá dado por

$$\sigma_{s_0}^2 = \sum_{i=1}^n \lambda_i \gamma(h) - \mu$$

El parámetro definido en la ecuación anterior debería ser considerado como una medida global de incerteza, en el sentido que es una versión integrada de la predicción clásica point-wise de la predicción de la varianza del Kriging Ordinario. Bajo alguna especificación en la traza del modelo del variograma, se podrán usar estimaciones de este parámetro para identificar esas zonas que presentan gran incerteza en las predicciones. Si se fija  $t \in T$  y se reemplaza la traza del variograma en la ecuación anterior por el variograma de una variable aleatoria  $\chi(t)$ ,  $t \in T$  se obtendría la varianza de predicción del Kriging Ordinario (modelo clásico), (véase en [14]). Este resultado se puede usar para calcular intervalos de confianza point-wise.

### 4.2.2 Estimación la traza del variograma

Para resolver el (4.2.1), es necesario un estimador de la traza del variograma. Ya que se asume que  $\chi_s(t)$  tiene una función de media constante  $m$  sobre  $D$ ,

$$V(\chi_{s_i}(t) - \chi_{s_j}(t)) = \mathbb{E} \left[ (\chi_{s_i}(t) - \chi_{s_j}(t))^2 \right].$$

Notar que, usando el teorema de Fubini,

$$\gamma(h) = \frac{1}{2} \mathbb{E} \left[ \int_T (\chi_{s_i}(t) - \chi_{s_j}(t))^2 dt \right], \text{ para } s_i, s_j \in D \text{ con } h = \|s_i - s_j\|$$

Entonces una adaptación del modelo clásico de los momentos (*MOM*) para esta cantidad da el siguiente estimador

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{i,j \in N(h)} \int_T (\chi_{s_i}(t) - \chi_{s_j}(t))^2 dt$$

donde  $N(h) = \{(s_i, s_j) : \|s_i - s_j\| = h\}$ , y  $|N(h)|$  es el número de elementos distintos en  $N(h)$ . Debido a la irregularidad de los datos espaciales, generalmente, no hay observaciones suficientes separadas exactamente a una distancia  $h$ . Entonces  $N(h)$  se modifica por  $N(h) = \{(s_i, s_j) : \|s_i - s_j\| \in (h - \epsilon, h + \epsilon)\}$ , con  $\epsilon > 0$  un valor pequeño. Una vez que se ha estimado la traza del variograma para una secuencia de  $K$  valores  $h_k$ , se propone un ajuste paramétrico del modelo  $\gamma_\alpha(h)$  (como el esférico, el Gaussiano, el exponencial o el Matérn podrían ser utilizados) en los puntos  $(h_k, \hat{\gamma}(h_k))$ ,  $k = 1, \dots, K$  como si se obtuvieran bajo los argumentos de la geostatística clásica. Generalmente, este tipo de ajuste, se hace por mínimos cuadrado ordinarios o mínimos cuadrados ponderados (ver, [6]). Notar que el ajuste paramétrico de la traza del variograma es siempre un variograma válido porque sus propiedades son las de un variograma paramétrico ajustado por un conjunto geostatístico univariante.

Un procedimiento diferente, como alternativa al ajuste paramétrico, consiste en aplicar técnicas de suavización (splines o regresión local lineal, ver Wasserman, 2006) al conjunto de datos  $(h_k, \hat{\gamma}(h_k))$ ,  $k = 1, \dots, K$  para ser capaz de evaluar aproximadamente  $\hat{\gamma}(h)$  mediante cualquier valor de  $h \in \mathbb{R}^+$ . Sin embargo, en este caso, si  $\tilde{\gamma}(h)$  denota la versión suavizada de  $\hat{\gamma}(h)$ , la condición de semifinida negativa de  $\tilde{\gamma}(h)$  merece más atención. Si  $\gamma_{\hat{\alpha}(h)}$  denota la traza del variograma estimado paraméricamente, esta forma funcional se usa para obtener los coeficientes  $\lambda_i$  del Kriging en el sistema (4.2.1), y para estimar la traza de la varianza de predicción a través de la ecuación  $\sigma_{s_0}^2$ .

### 4.2.3 Enfoque no paramétrico

En el artículo [11] tratan el mismo problema de geostatística de interpolación de curvas y específicamente consideran el predictor  $\hat{\chi}_{s_0}(t)$ . Consideran que las curvas son sólo conocidas por un conjunto finito de sus puntos:  $\chi_{s_i}(t_j) : j = 1, \dots, M, i = 1, \dots, n$  y muestran tres argumentos para la predicción de curvas en zonas no visitadas:

- Enfoque Multivariante 1: *Cokrige First, Fit later (CFP)*. El vector de valores observados  $(\chi_{s_i(t_1)}, \dots, \chi_{s_i(t_M)})$  se considera como la observación de una variable aleatoria  $M$ -dimensional en el sitio  $s_i$ . El Cokriging se aplica, entonces, para predecir los valores de este vector aleatorio en el sitio no muestreado  $s_0$ :  $(\hat{\chi}_{s_i(t_1)}, \dots, \hat{\chi}_{s_i(t_M)})$ . Por tanto, un modelo paramétrico  $\chi(\cdot; \theta)$ ,  $\theta \in \mathbb{R}^p$  se ajusta a los valores  $(\hat{\chi}_{s_i(t_1)}, \dots, \hat{\chi}_{s_i(t_M)})$  para reconstruir la función entera en  $s_0$ :  $\chi(\cdot; \hat{\theta}_{s_0})$ .
- Enfoque Multivariante 2: *Fit first, Cokrige later (FCP)*. Primero, el modelo paramétrico se ajusta a la curva observada:  $\chi(\cdot; \hat{\theta}_{s_i})$ ,  $i = 1, \dots, n$ . Los valores de los parámetros  $p$ -dimensionales se consideran como observaciones de la variable aleatoria. El Cokriging se aplica para predecir el valor del parámetro  $\theta$  en la posición  $s_0$ , decimos que  $\hat{\theta}_{s_0}^*$ , y  $\chi(\cdot; \hat{\theta}_{s_0}^*)$  es el resultado de predecir la función en  $s_0$ .
- Enfoque de una curva kriging (CKP). Dado que las funciones  $\chi_{s_i}$  son sólo conocidas para  $M$  valores, Goulard y Voltz (1993) proponen un modelo paramétrico  $\chi(\cdot; \theta)$ ,  $\theta \in \mathbb{R}^p$  para estos datos para obtener  $\chi(\cdot; \hat{\theta}_{s_i})$  como una aproximación de la función entera  $\chi_{s_i}$ . Por tanto, el predictor  $\hat{\chi}_{s_0}(t)$  se puede reescribir como

$$\hat{\chi}_{s_0} = \sum_{i=1}^n \lambda_i \chi(\cdot; \hat{\theta}_{s_i})$$

y las integrales en  $T$  para la estimación de los coeficientes  $\lambda_i$  se calculan usando  $\chi(\cdot; \hat{\theta}_{s_i})$  en vez de  $\chi_{s_i}$ .

Considérese ahora el caso común donde un gran número de valores  $M$  son observados para cada función muestreada y no hay un modelo paramétrico que los ajuste adecuadamente. En este contexto el segundo propósito, *CFP* (Cokrige first, Fit later) que puede llegar a ser extremadamente caro en términos de recursos computacionales. La razón es que *CFP* desarrolla un primer paso del Cokriging donde la dimensión de la observación multivariante es igual a  $M$ . El coste computacional de este paso es razonable si  $M$  está en decenas, pero es inaceptable si  $M$  es del orden de varios cientos o miles. Cuando un modelo paramétrico es adecuado para representar las funciones observadas, la alternativa *FCP* (Fit first, Cokrige later) es factible, porque en este caso el paso Cokriging tiene en cuenta  $p$  vectores dimensionales, donde  $p$  es el número de parámetros en el modelos paramétrico. Aquí el problema aparece cuando el modelo no paramétrico se considera para ajustar los valores observados. Entonces se podría hacer un ajuste no paramétrico, pero este proceso es esencialmente equivalente al ajuste paramétrico con número de parámetros  $p_M$  creciente con el número de valores observados  $M$ . Sin embargo, la flexibilidad extra proporcionada por un modelo no paramétrico es generalmente obtenida a expensas de permitir grandes valores  $p_M$  (incluso  $p_M \leq M$ ). Por tanto se concluye que el coste computacional del argumento *FCP* podría también ser prohibitivo. Así que nos limitamos a utilizar la curva de kriging predictor (*CKP*). Del sistema (4.2.1) y la sección anterior, es evidente que, el paso crucial para la estimación de la traza del variograma  $\gamma(h)$  y para calcular los coeficientes Kriging  $\lambda_i$  en el predictor  $\hat{\chi}_{s_0}(t)$ , está el cálculo de integrales a través de  $T$  con la forma genérica  $\int_T (\chi_{s_i}(t) - \chi_{s_j}(t))^2$ . Cuando se ajusta bien un modelo paramétrico

los valores de las funciones observadas están disponibles. Se trata de sustituir esta integral por  $\int_T \left( \chi(t; \hat{\theta}_{s_i}) - \chi(t; \hat{\theta}_{s_j}) \right)^2$ , que por lo general tiene una expresión analítica cerrada. Cabe plantearse qué es lo que se puede hacer si no existe un modelo paramétrico aceptable que ajuste las funciones observadas. La propuesta que se plantea es la de sustituir el paso paramétrico del ajuste por uno no paramétrico.

#### 4.2.4 Resultados de la aplicación

No es fácil proponer un modelo para estas curvas y que hay un cierto grado de ruido observacional. Suavizamos las funciones observadas usando bases de funciones de Fourier.

En las Figuras 4.5 y 4.6 podemos ver una comparación entre los datos sin suavizar y suavizados con 85 y 35 bases de Fourier.

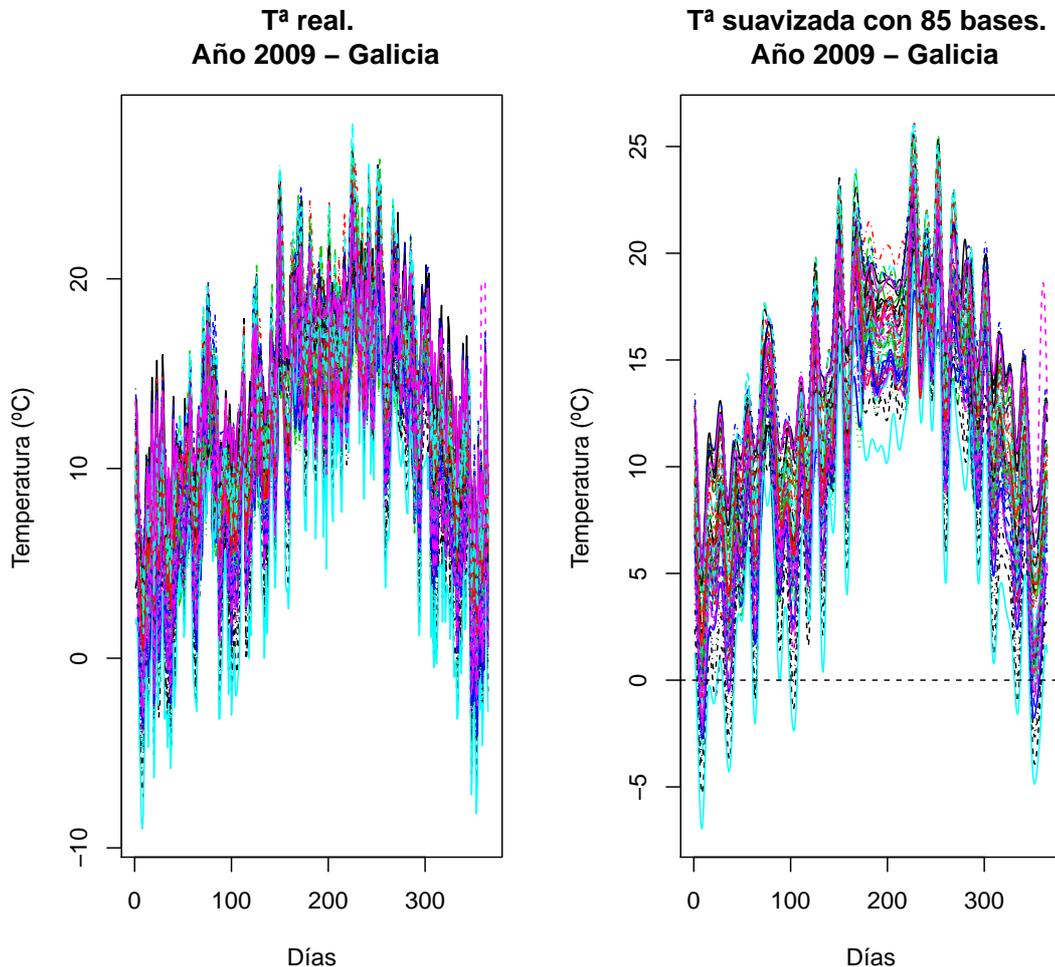


Figura 4.5: Estadísticos descriptivos para la temperatura de Galicia año 2009.

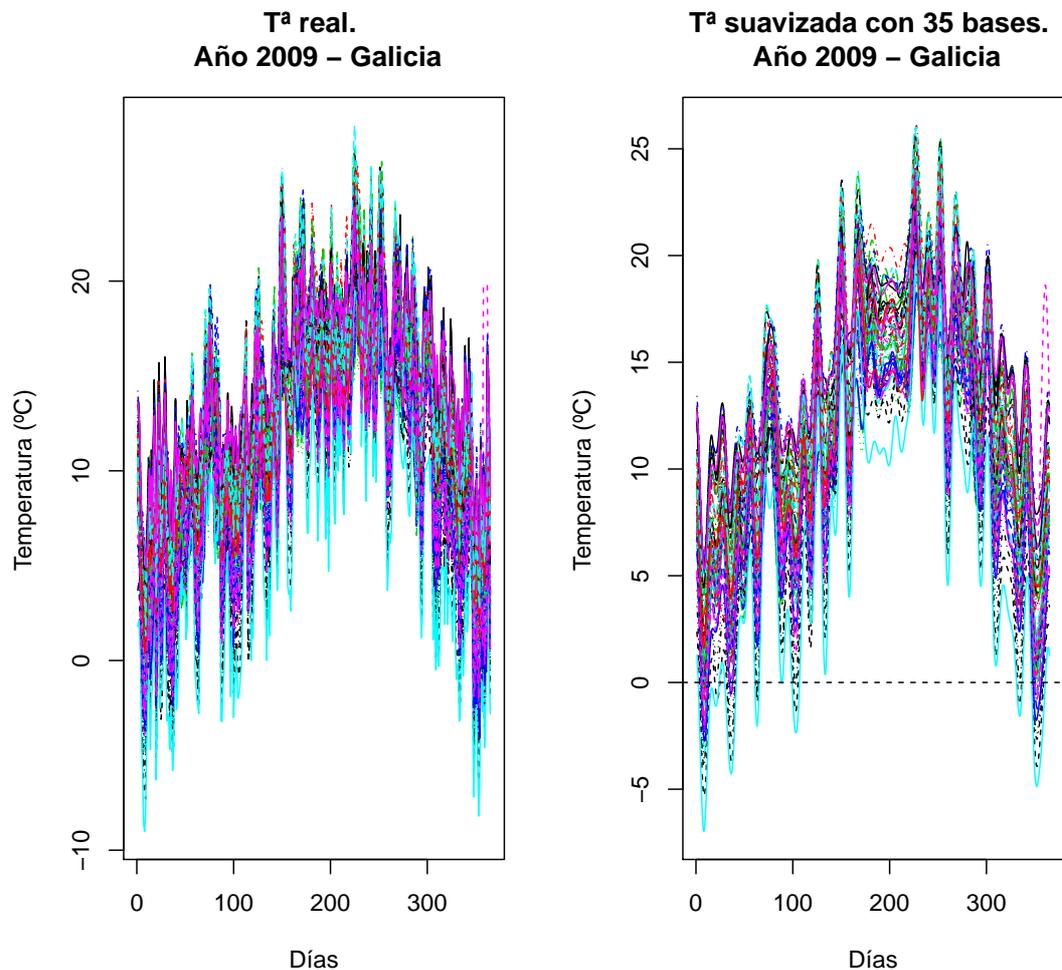


Figura 4.6: Estadísticos descriptivos para la temperatura de Galicia año 2009.

Visualmente los datos se representan bien con 35 bases de Fourier. Si tenemos muchas bases se produce un sobrestimamiento y además el coste computacional es mayor. Para decidir el número de bases más indicado se usó la librería *fda.usc* de Febrero-Bande, M. y Oviedo de la Fuente, M.. Se ha usado el método *min.np*. Este método realiza una suavización funcional de los datos usando una estimación no paramétrica del kernel con métodos de validación cruzada (CV) o mediante métodos de validación cruzada generalizada (GCV). En la sección siguiente se explicará más en detalle la elección del número de bases por tratarse de una técnica que expresa la solución basada en bases de funciones. El parámetro  $h$  de la ventana es 4.56. Cuando se han obtenido que 85 era un buen número para considerar de bases en el método de suavización se ha usado un método de validación cruzada y la matriz de estimación calculada mediante Regresión Local Lineal (LLR), y la traza de la matriz vale 87.49. En el caso de 35 bases el método de validación usado es de validación cruzada generalizada y la matriz se calcula mediante la

estimación del kernel de Nadaraya-Watson. En este caso la traza obtenida es 32.37.

### Criterio de evaluación del Kriging

En cualquier problema de predicción, el modelo ideal de evaluación consiste en dividir el conjunto de datos en dos partes: una muestra de entrenamiento (estimación) para el modelo de ajuste y otra muestra de test (validación). Este planteamiento, sin embargo, no es eficiente, salvo que el tamaño de la muestra sea grande. La idea que está detrás es el de validación cruzada, donde se conmutan los papeles de muestra test y entrenamiento. En el contexto de datos funcionales espacialmente correlacionados, donde la meta es predecir la curva entera de la función  $\chi_{s_0}(t)$  en el punto  $s_0$  no muestreado, la técnica de validación cruzada de dejar uno fuera trabaja del siguiente modo: cada dato de la localización se saca del conjunto de datos y mediante una función de suavización se predice en esa localización usando Kriging funcional como predictor. Llamamos a este procedimiento validación cruzada funcional (*FCV*). Calculamos el *SSE* de *FCV* por:

$$SSE_{FCV} = \sum_{i=1}^n SSE_{FCV}(i) = \sum_{i=1}^n \left\| \left( \chi_{s_i}(t_j) - \hat{\chi}_{s_i}^{(i)}(t_j) \right) \right\|,$$

donde  $\hat{\chi}_{s_i}^{(i)}(t_j)$  es la predicción en  $s_i$  evaluada en  $t_j, j = 1, \dots, M$ , sacando el sitio  $s_i$  temporalmente fuera de la muestra.

Para decidir que modelo de semivariograma usar se han probado mediante validación cruzada cuatro ejemplos de semivariogramas isotrópicos, el Matérn, Exponencial, Esférico y Circular descritos en el Capítulo 1. En la Figura 4.7 se muestran los resultados de predicción en la Facultad de Matemáticas obtenidos al calcular el Kriging Ordinario mediante validación cruzada para elegir el modelo más apropiado.

En la Tabla 4.2 se concluye que todos los modelos tienen un comportamiento similar. Por tanto se trabajará con el esférico de aquí en adelante. Suponemos que no se aprecian estas diferencias debido a que las distancias entre las estaciones no es muy grande, la máxima distancia de una estación a otra son 225.32Km y la máxima distancia de cualquier estación al punto no muestreado es 155.82Km.

Estadístico	Matérn	Exponencial	Esférico	Circular
Mínimo	687.7	687.7	687.6	687.3
1st Qu.	1544.0	1544.0	1545.0	1552.0
Mediana	2394.0	2394.0	2394.0	2394.0
Media	2872.0	2872.0	2872.0	2876.0
3rd Qu.	3132.0	3132.0	3132.0	3127.0
Máximo	16290.0	16290.0	16290.0	16290.0

Tabla 4.2: Principales tipos de kriging lineal y propiedades.

En la Figura 4.8 se muestra el variograma obtenido al calcular la predicción en el punto no muestreado con el método esférico, como se explicó con los otros modelos el resultado es

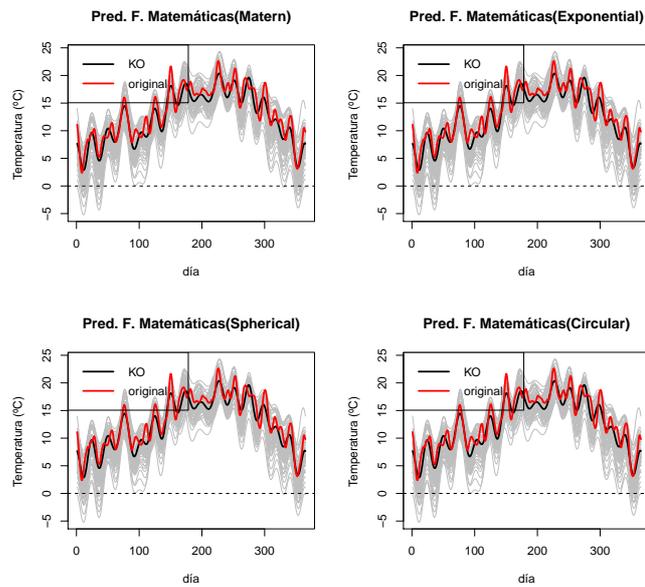


Figura 4.7: Predicción en función del tipo de modelo.

análogo.

Una vez que se tiene determinado que modelo se va a usar se procede a hacer la predicción. La Figura 4.9 muestra el variograma obtenido al calcular la predicción en la Facultad de Matemáticas.

En la Figura 4.10 se muestran los resultados de predicción comparados con el dato real, pues aunque consideramos como no muestreado la Facultad de Matemáticas, se tienen los resultados para el mismo año de la estación Santiago-EOAS situada a escasos metros de nuestro punto de interés. Se verifica que el  $\lambda$  mayor corresponde a la estación más cercana geográficamente al punto no muestreado. En la Figura 4.11 se presentan todos los  $\lambda$ 's.

Sergude, (está en el ayuntamiento de Boqueixón), es la estación con el mayor  $\lambda$  y la que tiene el menor  $\lambda$  está en Malpica. En la Tabla 4.3 se presenta un resumen de los resultados obtenidos.

	<b>Sergude</b>	<b>Río do Sol</b>	<b>Punta Candieira</b>	<b>Malpica</b>
$\lambda$	0.61	0.17	0.144	-0.02
Distancia en Km	9.78	26.89	101.08	56.13

Tabla 4.3:  $\lambda$ 's y distancias más representativos.

En la Figura 4.12 se ven las temperaturas de aquellas estaciones más representativas, por un lado se tiene la temperatura de Santiago, luego la de Sergude, Río do Sol y Punta Candieira, por ser las estaciones con mayores  $\lambda$ 's y por último Malpica la de menor  $\lambda$ . La temperatura media anual de estas estaciones se ve en la Tabla 4.4

Del análisis de validación cruzada se concluye que la estación óptima para la predicción en

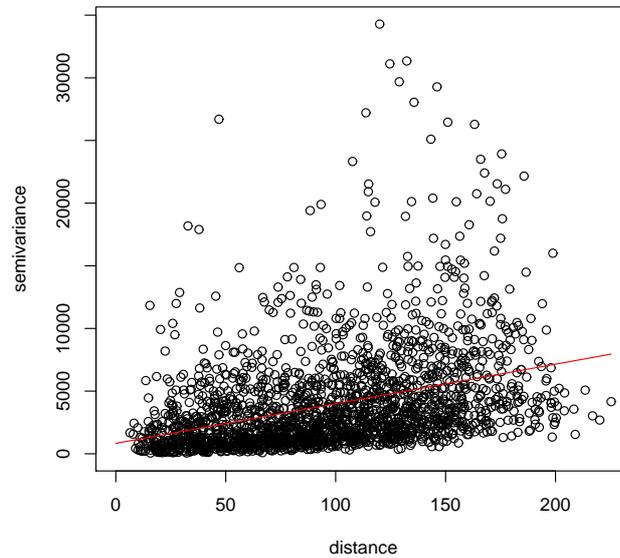


Figura 4.8: Predicción en función del tipo de modelo.

Santiago	Sergude	Río do Sol	Punta Candieira	Malpica
13.28	13.28	11.02	13.20	13.18

Tabla 4.4: Temperaturas medias en el año 2009.

la Facultad de Matemáticas ( $s_0$ ) es la de Sergude, estación más cercana a  $s_0$ . Los coeficientes  $\lambda_i$  del kriging se obtuvieron resolviendo el sistema de ecuaciones 4.2.1 con  $\gamma_h$  estimado mediante el variograma esférico  $\hat{\gamma}_h$ , como ha quedado justificado anteriormente. El método de validación cruzada funcional (FCV) usado para elegir el número de parámetros de suavización, también puede ser considerado una herramienta útil para comparar curvas observadas y predichas, esto es, define una medida de distancia entre estas dos curvas. De hecho el  $SSS(i)$  puede considerarse como una aproximación al error cuadrático medio integrado (MISE). Cuando se usa validación cruzada la idea es que los valores predichos deberían estar cerca de los observados (utilizando la técnica de dejar uno fuera en cada iteración). En la Figura 4.13 la parte de la izquierda muestra una comparación gráfica entre las curvas observadas (las reales) y la predicha (utilizando FCV).

La Tabla 4.5 muestra el resumen de los residuos de validación cruzada para el Kriging Ordinario y los datos suavizados mediante bases de Fourier.

Por último antes de cambiar a la técnica 2, debido a que no se aprecian cambios en el variograma ni en las estimaciones dependiendo del modelo considerado se ha intentado hacer lo mismo pero con Kriging Universal. Hemos creído que la altitud influye en la estimación de la temperatura e incluso la latitud. Por tanto se han creado 8 modelos diferentes con estas variables.

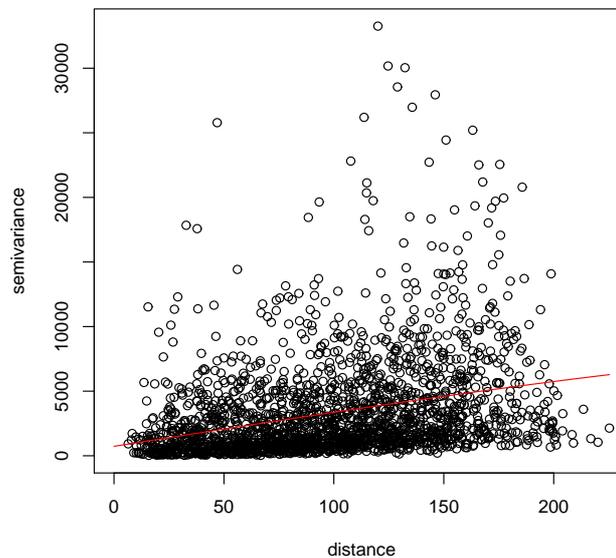


Figura 4.9: Variograma.

Los modelos serán combinaciones  $Temp_0(t) = \alpha_0(t) + \beta_1(t) \times Altitud_0 + \hat{\epsilon}_0(t)$  (*modelo1*) y  $Temp_0(t) = \alpha_0(t) + \beta_1(t) \times Altitud_0 + \beta_2(t) \times Latitud_0 + \hat{\epsilon}_0(t)$  (*modelo2*). Es decir el *modelo1* considerando la temperatura centrada o no y modelo con constante o sin ella y de modo análogo el *modelo2*. En la Tabla 4.6 se presenta un resumen del coeficiente de determinación para esos modelos.

A continuación se ha hecho la estimación del Kriging Universal primero con los modelos que usan como variable a la altitud. Los resultados obtenidos eran muy similares a los que se tenían en el Kriging Ordinario, por ello, se han omitido para no hacer muy extenso este trabajo. De la Tabla 4.6 los modelos que tienen mayor coeficiente de determinación (*R-Squared*) corresponden

Estadístico	Kriging Ordinario
Mínimo	1326.0
1st Qu.	2931.0
Mediana	3861.0
Media	4306.0
3rd Qu.	4847.0
Máximo	12880.0
Desviación típica	2279.5
Suma	284222.1

Tabla 4.5: Resumen de los residuos de validación cruzada para el Kriging Ordinario.

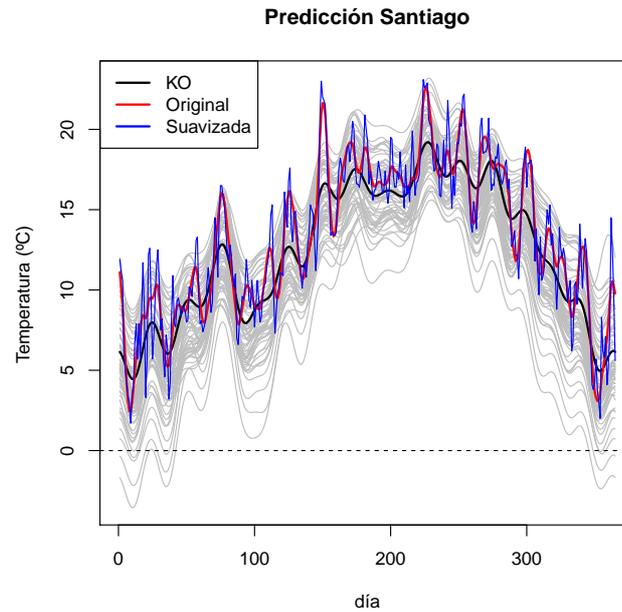


Figura 4.10: Predicción en función del tipo de modelo.

a aquellos que usan además la variable latitud. En la Tabla 4.7 se consideran los tres modelos anteriores mejores y se muestra el número de veces que las variables han resultado significativas.

Si nos fijamos en la Tabla 4.7 los modelos de la dos primeras filas arrojan los mismos resultados. En el último modelo la variable latitud ha sido siempre significativa y debido a que el R-Squared es el mayor hemos optado por este modelo. Se va a trabajar en vez de con la temperatura con los residuos respecto a la media del modelo que tienen a su vez estructura de dato funcional. Luego a estos residuos se les aplicarán las técnicas del Kriging. La Figura 4.14 muestra los residuos del modelo con altitud y latitud junto con los residuos suavizados con una base de Fourier de 35 elementos.

	<b>R-Squared</b>
<i>modelo1 - no centrado sin constante</i>	0.48
<i>modelo1 - centrado sin constante</i>	0.17
<i>modelo1 - no centrado con constante</i>	0.56
<i>modelo1 - centrado con constante</i>	0.56
<i>modelo2 - no centrado con constante</i>	0.73
<i>modelo2 - centrado con constante</i>	0.73
<i>modelo2 - no centrado sin constante</i>	0.97
<i>modelo2 - centrado sin constante</i>	0.56

Tabla 4.6: R-Squared.

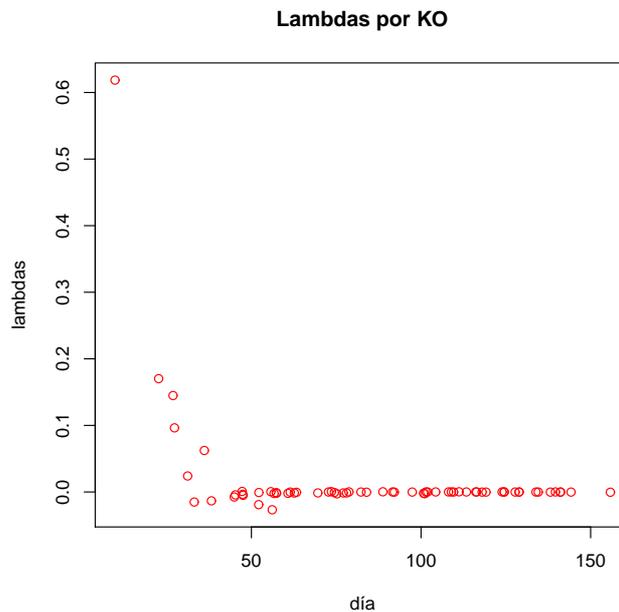


Figura 4.11: Lambdas obtenidos por Kriging Ordinario.

	<b>R-Squared</b>	<b>Constante</b>	<b>Altitud</b>	<b>Latitud</b>
<i>modelo2 - no centrado con constante</i>	0.73	305	342	270
<i>modelo2 - centrado con constante</i>	0.73	275	342	270
<i>modelo2 - no centrado sin constante</i>	0.97	-	318	365

Tabla 4.7: R-Squared y número de veces que las variables han sido significativas.

En la Figura 4.15 se tiene el variograma con los distintos modelos, y como en el caso del Kriging Ordinario se obtiene la misma estimación, con lo cuál se optará por el esférico nuevamente. Lo que si que se observa es que la varianza ha disminuido.

Se tiene la predicción para los residuos en la Figura 4.16. Hay que decir que las estaciones que tiene los  $\lambda$ 's representativos son las mismas que en el Kriging Ordinario, por tanto no se incidirá más en eso. El siguiente paso es deshacer el cambio en el modelo para ver la predicción auténtica en la Facultad de Matemáticas. Ya no se presentan estos resultados porque lo que se pretendía era ver si la el efecto de la altitud y/o la latitud ayudaban a definir que modelo usar.

También se ha aplicado Kriging Universal a los datos eliminándole a los residuos el efecto de la latitud y longitud. Los resultados obtenidos se pueden consultar en la tabla (4.14) y servirán para decidir al final con qué metodología nos quedamos.

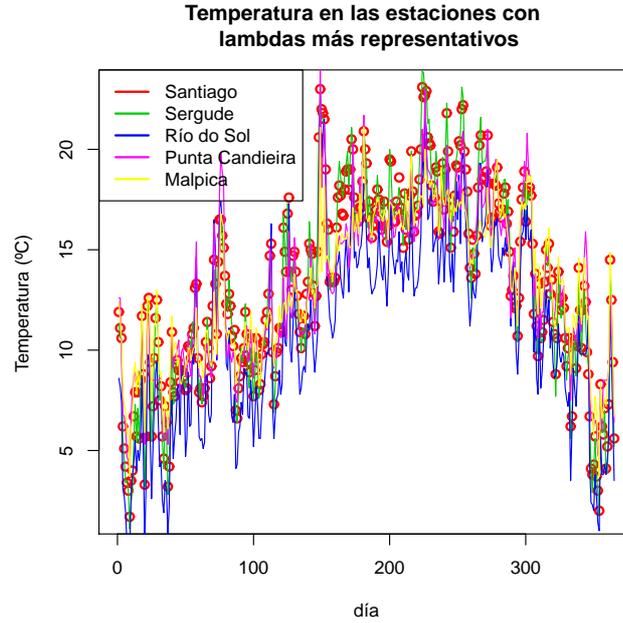


Figura 4.12: Estaciones con  $\lambda$ 's representativos.

### 4.3 Kriging variación de tiempo continua para la predicción espacial de datos funcionales

En esta técnica se considera el problema de predicción espacial de datos funcionales con ponderación de cada curva observada por un parámetro funcional. Esta técnica es una combinación del Kriging Ordinario y el modelo funcional lineal concurrente (punto-wise) tal y como se muestra en [26]. Se propone una solución basada en bases de funciones. Tanto las curvas como los parámetros funcionales se expanden en términos de un conjunto de bases de funciones. Entonces, el problema pasa a ser el de estimación de los coeficientes de estas bases de funciones para cada parámetro funcional. Para proporcionar una solución, se usa el modelo lineal de coregionalización para estimar la covarianza a través de los coeficientes de cada curva.

#### 4.3.1 Predictor y problema de minimización

Con esta técnica se propone usar la familia lineal para predictores point-wise  $\chi_{s_0}(t)$ ,  $t \in T$ , dada por

$$\hat{\chi}_{s_0}(t) = \sum_{i=1}^n \lambda_i(t) \chi_{s_i}(t), \lambda_1(t), \dots, \lambda_n(t) : T \rightarrow \mathbb{R}.$$

Este predictor fue mencionado en [11]. Se asumen las mismas afirmaciones dadas en la técnica anterior. Para cada  $t \in T$ , el predictor,  $\hat{\chi}_{s_0}(t)$ , tiene la misma expresión que el predictor del Kriging Ordinario. En el resto del proyecto a este predictor se le llamará Kriging Continuo

variable en el tiempo para datos funcionales (CTKFD). Este procedimiento del modelo es coherente con el modelo lineal funcional concurrente ([12], [26]) tal como se ve en la ecuación  $Y(t) = X(t)\beta(t) + \epsilon(t)$ , en la cuál la influencia de cada covariable en la respuesta es simultánea o point-wise. En este modelo, la respuesta  $Y_i(t)$  y cada covariable  $X_{ij}(t)$ ,  $j = 1, \dots, q$  son funciones con los mismos argumentos y  $X_j(t)$  sólo influye en  $Y(t)$  a través de su valor en el tiempo  $t$ . La estimación de parámetros funcionales  $\alpha(t)$ ,  $\beta_j(t)$ ,  $j = 1, \dots, q$ , se lleva a cabo resolviendo (Ramsay and Silverman, 2005)

$$\min_{\alpha(\cdot), \beta_1(\cdot), \dots, \beta_q(\cdot)} \mathbb{E} \left\| \hat{Y}(t) - Y(t) \right\|^2.$$

En este contexto, las covariables son las curvas observadas en  $n$  sitios de una región y la respuesta funcional es un función no observada en una zona no muestreada. Consecuentemente, el problema de optimización es

$$\min_{\lambda_1(\cdot), \dots, \lambda_n(\cdot)} \mathbb{E} \left\| \hat{\chi}_{s_0}(t) - \chi_{s_0}(t) \right\|^2,$$

o equivalentemente usando el teorema de Fubini

$$\min_{\lambda_1(\cdot), \dots, \lambda_n(\cdot)} \int_T \mathbb{E} (\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))^2 dt.$$

Si se consideran las afirmaciones estacionarias, el problema pasa a ser

$$\min_{\lambda_1(\cdot), \dots, \lambda_n(\cdot)} \int_T V (\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))^2 dt. \tag{4.1}$$

Como en la sección anterior, el problema de minimización (4.1) es una extensión del criterio de minimización dado por en [16] al contexto funcional, reemplazando el sumatorio por la integral y los vectores aleatorios  $[Z_1(s_0), \dots, Z_m(s_0)]$  y  $[\hat{Z}_1(s_0), \dots, \hat{Z}_m(s_0)]$  por las variables funcionales  $\chi(t)$  y  $\hat{\chi}(t)$  respectivamente, con  $t \in T$ . El predictor,  $\hat{\chi}_{s_0}(t)$  es insesgado si  $\mathbb{E}(\hat{\chi}_{s_0}(t)) = m(t)$ ,  $\forall t \in T$ , esto es, si  $\sum_{i=1}^n \lambda_i(t) = 1$ ,  $\forall t \in T$ . Consecuentemente, para encontrar el BLUP, los  $n$  parámetros funcionales en el predictor propuesto vendrán dados por la solución del siguiente problema de optimización

$$\min_{\lambda_1(\cdot), \dots, \lambda_n(\cdot)} \int_T V (\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))^2 dt, \text{ sujeto a : } \sum_{i=1}^n \lambda_i(t) = 1, \forall t \in T. \tag{4.2}$$

### 4.3.2 Una solución basada en bases de funciones

Se asume que cada función observada puede ser expresada en términos de  $K$  bases de funciones por

$$\chi_{s_i}(t) = \sum_{l=1}^K a_{il} B_l(t) = a_i^T B(t), i = 1, \dots, n.$$

Teniendo en cuenta que  $\chi_{s_i}(t)$ ,  $i = 1, \dots, n$  son funciones aleatorias con dependencia espacial, se asume que la matriz

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1K} \\ a_{21} & a_{22} & \dots & a_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nK} \end{bmatrix} = (\alpha_1, \dots, \alpha_K)_{(n \times K)}$$

constituye un vector aleatorio  $K$  multivariante con  $\mathbb{E}(\alpha_i) = v_{i(n \times 1)}$  y matriz de covarianzas

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \dots & \Sigma_{1K} \\ \Sigma_{21} & \Sigma_{22} & \dots & \Sigma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{K1} & \Sigma_{K2} & \dots & \Sigma_{KK} \end{bmatrix}_{(K \times n) \times (n \times K)}$$

donde  $\Sigma_{ij} = C(\alpha_i, \alpha_j)_{n \times n}$ . Los coeficientes  $a_{ij}$  se asumen como una realización del vector aleatorio espacial  $\alpha_j$ ,  $j = 1, \dots, K$ . Se propone usar geostatística multivariante y específicamente un modelo lineal de correogionalización (LMC) para la estimación de la matriz de covarianzas  $\Sigma$ . Para establecer la condición de insesgidez y para llevar a cabo la estimación de los parámetros en  $\hat{\chi}_{s_0}(t)$ , se expande cada parámetro funcional  $\lambda_i(t)$  por

$$\lambda_i(t) = \sum_{l=1}^K b_{il} B_l(t) = b_i^T B(t).$$

Por tanto, con las expresiones de  $\chi_{s_i}(t)$  y  $\lambda_i(t)$ , se obtienen para la ecuación del predictor  $\hat{\chi}_{s_0}(t)$ :

$$\hat{\chi}_{s_0}(t) = \sum_{i=1}^n b_i^T B(t) B(t)^T a_i.$$

Mediante  $\lambda_i(t)$  y la expansión de la función constante 1,  $\sum_{l=1}^K c_l B_l(t) = c^T B(t) = 1$ , la restricción de insesgidez se puede expresar como

$$\sum_{i=1}^n b_i^T B(t) = c^T B(t), \forall t, \Leftrightarrow \sum_{i=1}^n b_i = c,$$

o más específicamente por:

$$\sum_{i=1}^n b_{i1} = c_1, \dots, \sum_{i=1}^n b_{iK} = c_K.$$

Desarrollando la varianza en la función objetivo en el problema de minimización (4.2) se tiene:

$$\begin{aligned}
 V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)) &= V(\hat{\chi}_{s_0}(t)) + V(\chi_{s_0}(t)) - 2C(\hat{\chi}_{s_0}(t), \chi_{s_0}(t)) \\
 &= V\left(\sum_{i=1}^n b_i^T B(t) B^T(t) a_i\right) + B^T(t) V(a_0) B(t) \\
 &\quad - 2 \sum_{i=1}^n b_i^T B(t) B^T(t) C(a_i, a_0) B(t) \\
 &= \sum_{i=1}^n b_i^T B(t) B^T(t) V(a_i) B(t) B^T(t) b_i \\
 &\quad + 2 \sum_{i < j} b_i^T B(t) B^T(t) C(a_i, a_j) B(t) B^T(t) b_j \\
 &\quad + B^T(t) V(a_0) B(t) - 2 \sum_{i=1}^n b_i^T B(t) B^T(t) C(a_0, a_i) B(t).
 \end{aligned}$$

En la ecuación anterior,  $V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))$ , para  $i < j, i, j = 0, 1, \dots, n$  se tiene:

$$V(a_i) = \begin{bmatrix} \text{Var}(a_{i1}) & \text{Cov}(a_{i1}, a_{i2}) & \dots & \text{Cov}(a_{i1}, a_{iK}) \\ \text{Cov}(a_{i2}, a_{i1}) & \text{Var}(a_{i2}) & \dots & \text{Cov}(a_{i2}, a_{iK}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(a_{iK}, a_{i1}) & \text{Cov}(a_{iK}, a_{i2}) & \dots & \text{Var}(a_{iK}) \end{bmatrix}_{(K \times K)}$$

y

$$C(a_i, a_j) = \begin{bmatrix} \text{Cov}(a_{i1}, a_{j1}) & \text{Cov}(a_{i1}, a_{j2}) & \dots & \text{Cov}(a_{i1}, a_{jK}) \\ \text{Cov}(a_{i2}, a_{j1}) & \text{Cov}(a_{i2}, a_{j2}) & \dots & \text{Cov}(a_{i2}, a_{jK}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(a_{iK}, a_{j1}) & \text{Cov}(a_{iK}, a_{j2}) & \dots & \text{Cov}(a_{iK}, a_{jK}) \end{bmatrix}_{(K \times K)}$$

Si se definen:

$$\begin{aligned}
 Q_i &= \int_T (B(t) B^T(t) V(a_i) B(t) B^T(t)) dt \\
 Q_{ij} &= \int_T (B(t) B^T(t) C(a_i, a_j) B(t) B^T(t)) dt \\
 D &= \int_T B(t) V(a_0) B(t) dt \\
 J_i &= \int_T (B(t) B^T(t) C(a_0, a_i) B(t)) dt
 \end{aligned}$$

y considerando  $K$  multiplicadores de Lagrange  $m^T = (m_1, \dots, m_K)$  la función objetivo en el problema de minimización (4.2) puede expresarse como:

$$\min_{b_1, \dots, b_n, m} \sum_{i=1}^n b_i^T Q_i b_i + 2 \sum_{i < j} b_i^T Q_{ij} b_j + D - 2 \sum_{i=1}^n b_i^T J_i + 2m^T \left( \sum_{i=1}^n b_i - c \right)$$

Teniendo en cuenta  $\beta = (b_1^T, \dots, b_n^T, m^T)_{(K(n+1)x1)}^T$ , la expresión anterior aún se puede resumir a:

$$\min_{\beta} \beta^T Q \beta + D - 2\beta^T J \quad (4.3)$$

donde

$$Q = \begin{bmatrix} Q_{11} & Q_{12} & \dots & Q_{1n} & I \\ Q_{21} & Q_{22} & \dots & Q_{2n} & I \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Q_{n1} & Q_{n2} & \dots & Q_{nn} & I \\ I & I & \dots & I & 0 \end{bmatrix}, J = \begin{bmatrix} J_1 \\ J_2 \\ \vdots \\ J_n \\ c \end{bmatrix}$$

La matriz identidad en la ecuación anterior es de orden  $K$ . Minimizando la ecuación del problema de minimización (4.3) con respecto a  $\beta$  se obtiene:

$$\min_{\beta} \hat{\beta} = Q^{-1} J.$$

En la práctica, se empieza estimando ambos mediante un LMC para la variable aleatoria multivariante  $A = (\alpha_1, \dots, \alpha_K)$  y la matriz de covarianzas  $\Sigma$ . Consecuentemente, se pueden calcular las matrices  $Q$  y  $J$ . Reemplazando estas matrices en  $\hat{\beta} = Q^{-1} J$ , se pueden estimar  $b_i$ ,  $i = 1, \dots, n$  y los parámetros funcionales dados de  $\lambda_i(t)$ . Por otro lado, una estimación plug-in de la varianza de la predicción integrada  $\sigma_{s_0}^2 = \int_T V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)) dt$  vendrá dada por

$$\sigma_{s_0}^2 = \hat{\beta}^T Q \hat{\beta} + D - 2\hat{\beta}^T J,$$

donde la matriz  $D$  se calcula usando,  $\hat{V}(a_0)$  el cuál se obtiene por el ajuste LMC.

La predicción integrada de la varianza  $\sigma_{s_0}^2$  es una medida de incerteza en la predicción de toda la curva. Basados en los parámetros estimados y usando el desarrollo de la varianza,  $V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))$ , se puede estimar la varianza de predicción point-wise.

### 4.3.3 Resultados: predicción espacial de las curvas de temperatura en Galicia

En este apartado se muestra esta técnica aplicada al conjunto de los datos de temperatura de Galicia. Inicialmente se selecciona un número apropiado de bases funcionales. En un segundo paso se realiza la predicción en un punto no visitado usando el predictor propuesto y se describen los resultados desde un punto de vista práctico. Para evitar un gran trabajo computacional se han elegido el número de bases usando la metodología explicada en la técnica anterior y posteriormente se hace validación cruzada funcional con el  $K$  elegido, es decir  $K = 35$ .

Como ha quedado explicado con la otra técnica los datos de Galicia quedan bien representados en una base de 35. No se considera penalización robusta en este caso, es decir, se supone  $\eta = 0$  en:

$$c \in \mathbb{R}^k \sum_{j=1}^M (y_j - \chi(t_j))^2 + \eta \int_T (\chi'''(t) + \omega^2 \chi'(t))^2 dt.$$

Cuando los datos son periódicos, las bases de Fourier con un número impar de bases funcionales son la opción más apropiada ([26]). Por tanto una base de Fourier con 35 bases de funciones será una buena elección para los datos de Galicia. Aunque que se pueden expresar en función de una base de Fourier con número infinito de sinusoides, tomamos 365 como el límite porque este es el número de datos discretos de cada sitio en nuestro conjunto de datos. Frecuencias superiores a 365 en este caso distorsionarán la señal. Esto se conoce como el *problema de aliasing* (Lfeachor y Jervis, 1993).

Se trata de predecir la curva de la temperatura en un sitio no visitado con coordenadas 536.07 (Este) y 4747.02 (Norte), que corresponde a la Facultad de Matemáticas. La Facultad de Matemáticas está situada casi en el centro geográfico de Galicia. El clima de Santiago se aproxima al *oceánico costero*, que se caracteriza por tener temperaturas suaves en el verano y en el invierno y precipitaciones abundantes.

En esta sección se sigue la misma metodología de la sección anterior, es decir, se hace la predicción en el lugar no visitado y después un análisis de validación cruzada. Teniendo en cuenta que en los datos de Galicia para ajustar la media de la función no es constante, los datos suavizados están inicialmente sin tendencia usando un modelo de regresión funcional con respuesta funcional (suavizando las curvas de temperatura) y dos covariables escalares (las coordenadas: longitud y la latitud), es decir, se considera que el modelo de regresión funcional es

$$\chi_{s_i}(t) = \hat{\alpha}(t) + \hat{\beta}_1(t)Longitud + \hat{\beta}_2(t)Latitud + \epsilon_{s_i}(t). \quad (4.4)$$

Posteriormente, el residuo,  $\epsilon_{s_i}(t)$ , en un sitio no visitado se predice por *CTKFD* y por último, la predicción de la temperatura en  $s_0$  se obtiene mediante la adición a la predicción del kriging a la tendencia del modelo dado (4.4).

Los parámetros estimados para el modelo ajustado se muestran en la Figura 4.17.

Las variables *latitud* y *longitud* en el modelo (4.4) fueron estandarizados anteriormente. La Figura 4.18 muestra la influencia de la constante, de la longitud y de la latitud en el modelo. Los parámetros estimados muestran claramente que la temperatura está mucho más influenciada por las coordenadas en invierno que en verano para la longitud y al revés para la latitud.

Las Figuras 4.19, 4.20 y 4.21 muestran respectivamente los efectos de la variable longitud, latitud y residuos del modelo (4.4), al que se le aplicará el Kriging

En la Figura 4.22 se muestran las curvas de temperatura suavizadas. La Tabla (4.8) se presenta un resumen estadístico de las variables longitud y latitud después de centrarlas.

Estadístico	Longitud	Latitud
Mínimo	-88.74	-88.67
1st Qu.	-51.57	-41.98
Mediana	2.19	-12.93
Media	0.00	0.00
3rd Qu.	45.88	45.80
Máximo	91.10	111.20

Tabla 4.8: Resumen estadístico de las variables longitud y latitud.

En la Figura 4.24 se muestran los correspondientes residuos de este modelo, es decir los residuos obtenidos al aplicar el Kriging a los residuos del modelo de regresión funcional de la temperatura frente a las coordenadas, (longitud y latitud).

Deshaciendo el cambio, se obtiene la curva de predicción para la Facultad de Matemáticas (Figura 4.17). De las Figuras 4.17 y 4.24 se puede concluir que el modelo de regresión ajustado tiene un buen rendimiento.

Altos residuos en la Figura 4.24 se deben a malas estimaciones en las estaciones se trata de Cabeza de Manzaneda (ayuntamiento de Manzaneda (Ourense)) y Pedro Murias (ayuntamiento de Ribadeo (Lugo), 7124.53 y 6509.01, respectivamente. Cabeza de Manzaneda está al sur este de Galicia, la diferencia entre la temperatura máxima y la mínima es de  $28.5^{\circ}C$ . El clima se encuadra dentro de lo que se ha definido por *oceánico de montaña* que se caracteriza por tener veranos frescos e inviernos menos fríos. Por el contrario en la estación de Pedro Murias en Ribadeo, noreste de Galicia, posee clima *oceánico costero*, que se caracteriza por tener temperaturas suaves en el verano y en el invierno y precipitaciones abundantes. La diferencia de temperaturas máxima y mínima es esta estación es  $25.5^{\circ}C$ .

Por el contrario las estaciones de las Vigo-Campus (ayuntamiento de Vigo (Pontevedra)) y Melide (A Coruña), son las estaciones con menores residuos por validación cruzada, 254.13 y 397.53, respectivamente. La diferencias entre el máximo y el mínimo son de  $21.5^{\circ}C$  y  $23.4$ , para cada una. En Vigo-Campus el clima es *oceánico costero* y en Melide *oceánico continental*, se caracteriza por tener veranos cálidos e inviernos fríos con posibilidades de heladas.

La Figura 4.23 muestra un gráfico de los parámetros funcionales estimados. El mayor parámetro funcional estimado corresponde a la estación de Sergude (ayuntamiento de Boqueixón (A Coruña)) que es la estación más cercana al punto a estimar.

Otras estaciones cercanas a la Facultad de Matemáticas, y por tanto con influencia en la predicción, son Muralla (ayuntamiento de Lousame (A Coruña)) (valores en torno a 0.16), Río do Sol (ayuntamiento de Coristante (A Coruña)) (alrededor de 0.13 pesos), Fontecada (ayuntamiento de Santa Comba (A Coruña)) valores en torno a 0.009, en la Figura 4.23. La suma de los parámetros funcionales estimados es igual a 1 para todo  $t$  (Figura 4.23), lo que indica que la restricción de insesgadez se cumple. La estación que proporciona menos información es la de Malpica, igual que sucede con la técnica 1. La predicción en la Facultad de Matemáticas 4.17, es consistente con los valores reales registrados.

Una comparación entre la predicción en validación cruzada y las curvas suavizadas (Figuras 4.17 y 4.24) muestran que las predicciones tienen el mismo comportamiento en el tiempo que las curvas suavizadas.

Se puede observar que hay algunas estaciones con grandes curvas residuales positivas o negativa. Se aprecia que la desviación estándar residual es menor en la primavera y en el verano donde las curvas suavizadas y pronosticadas tienen menor variación y se comprueba con los datos de los promedios de la desviación típica en la Tabla 4.9.

	Primavera	Verano	Otoño	Invierno
Promedio de la desviación típica	2.77	2.42	2.91	2.91

Tabla 4.9: Promedios de la desviación típica.

La media residual varía alrededor de cero, lo que indica que las predicciones son insesgadas. En la Tabla 4.10 se obtienen los promedios por estación del año.

	<b>Primavera</b>	<b>Verano</b>	<b>Otoño</b>	<b>Invierno</b>
Promedios	5.6	9.06	1.25	1.62

Tabla 4.10: Promedios de la media.

En la Tabla (4.11) se muestran los estadísticos de la suma de los errores al cuadrado por validación cruzada para las 66 estaciones.

<b>Estadístico</b>	<b>CTKFD</b>
Mínimo	865.3
1st Qu.	2373.8
Mediana	2768.0
Media	3212.3
3rd Qu.	3793.2
Máximo	10061.0
Desviación típica	1474.8
Suma	212009.0

Tabla 4.11: Resumen estadístico de la suma de los cuadrados de los errores de validación cruzada.

## 4.4 De multivariante a geostatística funcional

En esta técnica, se propone un predictor Cokriging para hacer la predicción univariante (como en el sentido del Cokriging multivariante), pero considerando como información auxiliar curvas en vez de observaciones de vectores aleatorios. Del mismo modo, se amplía el Kriging multivariante de vectores aleatorios al contexto funcional definiendo un Kriging predictor funcional, que permite hacer predicciones de la curva entera en la estación no visitada.

Esto mismo ha sido también estudiado por Nerini y Monestiez (2008), [17]. Ellos proponen una solución basada en bases de funciones ortonormales, condición no requerida en este trabajo.

En estadística espacial y específicamente en geoestadística, tanto en el análisis Cokriging ([? ], [4]) como en el Kriging multivariante ([13]) se utilizan para modelar observaciones del vector aleatorias. En este trabajo se adaptan estas metodologías al contexto funcional. Se propone un predictor Cokriging haciendo una predicción univariante (como en el sentido del Cokriging multivariante), pero considerando como información auxiliar funciones en lugar de observaciones de vectores aleatorios. Del mismo modo, se extiende el Kriging multivariante de vectores aleatorios al contexto funcional mediante la definición de un predictor Kriging funcional que permite hacer una predicción de la curva completa de la zona no muestreada mediante el uso de la información de las curvas muestreadas a sitios cercanos al sitio de predicción. En ambos casos (Cokriging basado en funciones y Kriging funcional) da una solución no-paramétrica basada en bases de funciones, y se prueba que ambos propósitos coinciden cuando

se utiliza este enfoque. Las metodologías propuestas con esta técnica son del mismo modo a las soluciones dadas en la técnica anterior, es decir, cada curva observada es ponderada por un parámetro funcional. Sin embargo aquí la flexibilidad aumenta porque se estima el doble índice de los parámetros funcionales. Ahora, cada curva se pondera por un parámetro funcional para llevar a cabo la realización de la predicción en cada momento. Esta metodología sigue la filosofía básica del modelo lineal funcional de respuesta funcional (modelo total) de la ecuación  $Y_i(t) = \beta_0(t) + \int_T X_i(v)\beta(v, t) dv + \epsilon_i(t)$ , donde se debe estimar el coeficiente de regresión bivalente (Malfait y Ramsay, 2003).

#### 4.4.1 Cokriging basado en datos funcionales

Sean  $\{\chi_s(t), t \in T, s \in D \subset \mathbb{R}^d\}$  una función aleatoria definida en algún conjunto compacto  $T$  de  $\mathbb{R}$ . Supongamos una muestra de curvas  $\chi_{s_1}(t), \dots, \chi_{s_n}(t)$  definidas  $\forall t \in T, s_i \in D, i = 1, \dots, n$ . Se asume que estas curvas pertenecen a un espacio separable Hilbert  $H$  de funciones integrables definidas en  $T$ . Considérese que para cada  $t \in T$  se tiene un segundo proceso aleatorio isótropo y estacionario, es decir, la media y la varianza son funciones constantes y la covarianza sólo depende de la distancia entre los puntos de muestreo. Se quiere predecir una única variable en un único lugar de una muestra de datos funcionales espacialmente correlados  $\chi_{s_1}(t), \dots, \chi_{s_n}(t)$ . Sea  $\chi_{s_0}(v)$  la variable aleatoria a predecir en una localización no observada  $s_0$  en  $v \in T$ . Para llevar a cabo esta tarea se extiende el Cokriging predictor  $\hat{Z}_k(s_0) = \sum_{i=1}^n \sum_{j=1}^m \lambda_{ij} Z_j(s_i)$ , reemplazando los  $n \times m$  parámetros  $\lambda_{ij}^k$  por  $n$  parámetros funcionales  $\lambda_i^v(t)$  y las  $n \times m$  variables aleatorias  $Z_j(s_i)$  por  $n$  variables funcionales  $\chi_{s_i}(t), i = 1, \dots, n, j = 1, \dots, m$ . En el esquema siguiente se ven claramente estas transformaciones:

- Parámetros

<b>Cokriging multivariante</b>	$\Rightarrow$	<b>Cokriging BC</b>
$\lambda_{11} \cdots \lambda_{1m}$	$\Rightarrow$	$\lambda_1^v(t), t \in T$
$\lambda_{21} \cdots \lambda_{2m}$	$\Rightarrow$	$\lambda_2^v(t), t \in T$
$\ddots$	$\vdots$	$\vdots$
$\lambda_{n1} \cdots \lambda_{nm}$	$\Rightarrow$	$\lambda_n^v(t), t \in T$

- Variables

<b>Cokriging multivariante</b>	$\Rightarrow$	<b>Cokriging BC</b>
$Z_1(s_1) \cdots Z_m(s_1)$	$\Rightarrow$	$\chi_{s_1}(t), t \in T$
$Z_1(s_2) \cdots Z_m(s_2)$	$\Rightarrow$	$\chi_{s_2}(t), t \in T$
$\ddots$	$\vdots$	$\vdots$
$Z_1(s_1) \cdots Z_m(s_n)$	$\Rightarrow$	$\chi_{s_n}(t), t \in T$

Por lo tanto, el predictor Cokriging de  $\chi_{s_0}(v)$  basado en datos funcionales (CBFD) viene dado por

$$\hat{\chi}_{s_0}(v) = \sum_{i=1}^n \int_T \lambda_i^v(t) \chi_{s_i}(t) dt$$

Para cada  $v \in T$  específico, los parámetros funcionales  $\lambda_i^v(t)$ ,  $i = 1, \dots, n$  se estiman teniendo en cuenta las restricciones clásicas de geoestadística, es decir, insesgadez y mínima varianza de predicción. Este problema se resuelve utilizando un enfoque basado bases de funciones. Ampliamos las variables funcionales mediante la expresión

$$\chi_{s_i}(t) = a_i^T B(t), i = 1, \dots, n$$

y los parámetros funcionales por

$$\lambda_i^v(t) = b_{iv}^T B(t)$$

Por lo tanto el predictor se expresa como

$$\hat{\chi}_{s_i}(v) = \sum_{i=1}^n b_{iv}^T W a_i$$

donde

$$W = \int_T B(t) B^T(t) dt$$

Para cualquier base ortonormal como la base de Fourier, la matriz  $W$  de Gram es la matriz identidad. Para otras bases de funciones tales como las bases  $B$ -Splines,  $W$  debe ser calculada usando integración numérica. Asumiendo la hipótesis de estacionariedad de las funciones aleatorias la matriz:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1K} \\ a_{21} & a_{22} & \dots & a_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nK} \end{bmatrix} = (\alpha_1, \dots, \alpha_K)_{(n \times K)}$$

constituyen un vector aleatorio multivariante  $K$  con

$$\mathbb{E}(\alpha_j) = \begin{bmatrix} \mathbb{E}(a_{1j}) \\ \mathbb{E}(a_{2j}) \\ \vdots \\ \mathbb{E}(a_{nj}) \end{bmatrix} = \begin{bmatrix} \vartheta_j \\ \vartheta_j \\ \vdots \\ \vartheta_j \end{bmatrix}, j = 1, \dots, k,$$

y matriz de covarianzas

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \dots & \Sigma_{1K} \\ \Sigma_{21} & \Sigma_{22} & \dots & \Sigma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{K1} & \Sigma_{k2} & \dots & \Sigma_{KK} \end{bmatrix}$$

donde  $\Sigma_{ij} = C(\alpha_i, \alpha_j)_{n \times n}$ , obteniéndose así:

$$\mathbb{E}(a_i) = \begin{bmatrix} \mathbb{E}(a_{i1}) \\ \mathbb{E}(a_{i2}) \\ \vdots \\ \mathbb{E}(a_{ik}) \end{bmatrix} = \begin{bmatrix} \vartheta_1 \\ \vartheta_2 \\ \vdots \\ \vartheta_K \end{bmatrix} = \vartheta$$

Entonces, la media del predictor está dada por:

$$\mathbb{E}(\hat{\chi}_{s_0}(v)) = \sum_{i=1}^n b_{iv}^T W \vartheta$$

Por otro lado, la media de la función en el sitio  $s_0$  no observado en el instante  $v$  es

$$\mathbb{E}(\chi_{s_0}(v)) = B^T(v)\mathbb{E}(a_0) = B^T(v)\vartheta$$

Consecuentemente el predictor propuesto es insesgado si

$$\sum_{i=1}^n b_{iv}^T = B^T(v)W^{-1}$$

La matriz Gram es en general semidefinida positiva, la matriz  $W$  Gram,  $W = \int_T B(t)B^T(t)dt$ , es definida positiva porque las funciones  $B_l(t), l = 1, \dots, K$ , son linealmente independientes. Consecuentemente el predictor propuesto está bien definido. Para encontrar el mejor predictor lineal insesgado (BLUP), los  $n$  parámetros funcionales propuestos en el predictor se dan como solución al siguiente problema de optimización:

$$\min_{\lambda_1^v(\cdot), \dots, \lambda_n^v(\cdot)} V(\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)) \text{ sujeto a : } \mathbb{E}(\hat{\chi}_{s_0}(v)) = \mathbb{E}(\chi_{s_0}(v)) \quad (4.5)$$

Desarrollando la varianza en la función objetivo se tiene la siguiente expresión:

$$\begin{aligned} V(\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)) &= \sum_{i=1}^n b_{iv}^T W V(a_i) W^T b_{iv} + 2 \sum_{i < j} b_{iv}^T W C(a_i, a_j) W^T b_{jv} \\ &+ B^T(v) V(a_0) B(v) - 2 \sum_{i=1}^n b_{iv}^T W C(a_0, a_i) B(v) \end{aligned}$$

$V(a_i)$ ,  $C(a_i, a_j)$  y  $C(a_0, a_i)$  se pueden calcular si  $\Sigma$ , la matriz de covarianzas, ha sido previamente estimada. Podemos usar geostatística multivariante ([28]) y específicamente un modelo lineal de correogionalización (LMC) para estimar estas matrices. Si se definen:

$$\begin{aligned} M_i &= (WV(a_i)W^T)_{(K \times K)} \\ M_{ij} &= (WC(a_i, a_j)W^T)_{(K \times K)} \\ N_i(v) &= (WC(a_0, a_i)B(v))_{(K \times 1)} \\ D(v) &= B^T(v)V(a_0)B(v) \end{aligned}$$

la función objetivo está dada por

$$V(\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)) = \sum_{i=1}^n b_{iv}^T M_i b_{iv} + 2 \sum_{i<j} b_{iv}^T M_{ij} b_{jv} + D(v) - 2 \sum_{i=1}^n b_{iv}^T N_i(v) \quad (4.6)$$

De (4.6) y considerando los  $K$  multiplicadores de Lagrange  $m_v^T = (m_{1v}, \dots, m_{Kv})$  el problema de optimización (4.5) puede ser expresado como:

$$\min_{b_{1v}, \dots, b_{nv}, m_v} \sum_{i=1}^n b_{iv}^T M_i b_{iv} + 2 \sum_{i<j} b_{iv}^T M_{ij} b_{jv} + D(v) - 2 \sum_{i=1}^n b_{iv}^T N_i(v) + 2m_v^T \left( \sum_{i=1}^n b_{iv} - W^{-1}B(v) \right)$$

Tomando  $\beta_v = (b_{1v}^T, \dots, b_{nv}^T, m_v^T)^T_{(K(n+1) \times 1)}$ , se puede reescribir la ecuación anterior de la forma:

$$\min_{\beta_v} \beta_v^T M \beta_v + D(V) - 2\beta_v^T N(V) \quad (4.7)$$

donde

$$M = \begin{bmatrix} M_{11} & M_{12} & \dots & M_{1n} & I \\ M_{21} & M_{22} & \dots & M_{2n} & I \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ M_{n1} & M_{n2} & \dots & M_{nn} & I \\ I & I & \dots & I & 0 \end{bmatrix}_{[K(n+1)] \times [K(n+1)]}$$

y

$$N(v) = \begin{bmatrix} N_1(v) \\ N_2(v) \\ \vdots \\ N_n(v) \\ W^{-1}B(v) \end{bmatrix}_{[K(n+1)] \times 1}$$

Finalmente, si se minimiza la ecuación reducida respecto a  $\beta_v$ , se obtiene:

$$\hat{\beta}_v = M^{-1}N(v)$$

Una estimación plug-in de la varianza de predicción  $\sigma_{s_0}^2(v) = V(\hat{\chi}_{s_0}(v) - \chi_{s_0}(v))$  está dada por

$$\hat{\sigma}_{s_0}^2(v) = \hat{\beta}_v^T M \hat{\beta}_v + D(v) - 2\hat{\beta}_v^T N(v)$$

donde la matriz  $D(v)$  se calcula usando una estimación de  $\hat{V}(a_0)$  obtenida mediante el ajuste de medias del modelo linal de correogeneralización (LMC).

#### 4.4.2 Kriging funcional: modelo total

Para definir el Kriging funcional (modelo total) predictor (FKTM), se asumen los mismos supuestos de estacionariedad e isotropía que en el Cokriging basado en datos funcionales. El Cokriging predictor definido allí,  $\hat{\chi}_{s_0}(v) = \sum_{i=1}^n \int_T \lambda_i^v(t) \chi_{si}(t) dt$ , está definido para un  $v \in T$  específico. Si se quiere predecir la curva completa en  $s_0$ , el parámetro funcional  $\lambda_i^v(t)$  se reemplaza por un parámetro que depende de dos índices  $\lambda_i(t, v)$ . Así, el predictor para toda la curva está dada por

$$\hat{\chi}_{s_0}(v) = \sum_{i=1}^n \int_T \lambda_i(t, v) \chi_{si}(t) dt, v \in T,$$

tal que  $\lambda_1(t, v), \dots, \lambda_n(t, v) : TxT \rightarrow \mathbb{R}$ . Notar que de acuerdo a los reemplazamientos dados en los esquemas tanto para parámetros como para variables en el caso del Cokriging basado en datos funcionales, el predictor  $\sum_{i=1}^n b_{iv}^T = B^T(v)W^{-1}$  se extiende al contexto funcional mediante la sustitución del vector  $(\hat{Z}_1(s_0), \dots, \hat{Z}_m(s_0))^T$  por  $\hat{\chi}_{s_0}(v), v \in T$ , y el vector de parámetros  $(\lambda_{11}^1 \dots \lambda_{1m}^1 \dots \lambda_{11}^m \dots \lambda_{1m}^m), \dots, (\lambda_{n1}^1 \dots \lambda_{nm}^1 \dots \lambda_{n1}^m \dots \lambda_{nm}^m)$  por  $((\lambda_1^1(t) \dots \lambda_1^m(t)), \dots, (\lambda_n^1(t) \dots \lambda_n^m(t))), t \in T$ , respectivamente.

Además, teniendo en cuenta que  $v$  varía de forma continua en  $T$ , el conjunto discreto de parámetros funcionales  $(\lambda_i^1(t) \dots \lambda_i^m(t)), i = 1, \dots, n$  se sustituye por el doble índice del parámetro funcional  $\lambda_i(t, v), t, v \in T$ . Es evidente que para un fijo  $v \in T$  la expresión del predictor FKTM es igual a la del Cokriging basado en datos funcionales (CBFD). El parámetro funcional  $\lambda_i(t, v)$  determina el impacto de la  $i$ -ésima función observada en el tiempo  $t$  en la función no observada al tiempo  $v$ . Este modelo es coherente con el modelo funcional lineal con respuesta funcional (modelo total):

$$Y_i(t) = \beta_0(t) + \int_T X_i(v) \beta(v, t) dv + \epsilon_i(t).$$

La estimación funcional de los parámetros se lleva a cabo mediante la resolución (Ramsay y Silverman, 2005) de:

$$\min_{\alpha(\cdot), \beta_1(\cdot, \cdot), \dots, \beta_q(\cdot, \cdot)} \mathbb{E} \left\| \hat{Y}(v) - Y(v) \right\|^2.$$

En nuestro contexto las covariables son las curvas observadas en  $n$  lugares de una región y la respuesta funcional es una función no observada en un lugar no visitado. Por lo tanto la función objetivo es  $\mathbb{E} \|\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)\|^2$ , dependiendo de  $\lambda_1(\cdot, \cdot), \dots, \lambda_n(\cdot, \cdot)$  o usando el Teorema de Fubini  $\int_T \mathbb{E} (\hat{\chi}_{s_0}(v) - \chi_{s_0}(v))^2 dv$ . Teniendo en cuenta la estacionalidad de la función objetivo se tiene,  $\int_T V (\hat{\chi}_{s_0}(v) - \chi_{s_0}(v))^2 dv$ .

Una vez más, los parámetros funcionales  $\lambda_i(t, v)$  de  $\hat{\chi}_{s_0}(v)$  se calculan teniendo en cuenta las restricciones de insesgadez y la predicción con mínima varianza. Por lo tanto el problema de optimización es

$$\min_{\lambda_1(\cdot, \cdot), \dots, \lambda_n(\cdot, \cdot)} \int_T V (\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)) \text{ sujeto a : } \mathbb{E} (\hat{\chi}_{s_0}(v)) = \mathbb{E} (\chi_{s_0}(v)), \forall v \in T. \quad (4.8)$$

Resolvemos el problema utilizando un enfoque basado en bases de funciones. Se extienden las variables funcionales de la ecuación:

$$\chi_{si}(t) = a_i^T B(t), i = 1, \dots, n$$

y los parámetros funcionales bivariantes:

$$\lambda_i(t, v) = B^T(t)C_i B(v)$$

donde

$$C_i = \begin{bmatrix} c_{11}^i & c_{12}^i & \cdots & c_{1K}^i \\ c_{21}^i & c_{22}^i & \cdots & c_{2K}^i \\ \vdots & \vdots & \ddots & \vdots \\ c_{K1}^i & c_{K2}^i & \cdots & c_{KK}^i \end{bmatrix}_{(K \times K)}$$

Con estas consideraciones el predictor  $\hat{\chi}_{s_0}(v)$  se expresa:

$$\hat{\chi}_{s_0}(v) = B^T(v)\hat{a}_0,$$

donde la matriz de producto interior  $W$ , definida,  $W = \int_T B(t)B^T(t)dt$ . El predictor  $\hat{\chi}_{s_0}(v)$  también ha sido considerado por Nerini y Monestiez (2008). En su trabajo de Nerini y Monestiez (2008) se supone que  $W$  es una matriz de identidad ya que consideran una solución basada en expansiones de bases ortonormales, en este caso no es una condición necesaria. Se consideran las propiedades de insesgaredad y mínima varianza del predictor propuesto, se supone que los coeficientes  $a_i$  en la ecuación anterior son un vector aleatorio multivariante estacionario. En consecuencia, el valor esperado de la la curva en el sitio no visitado  $s_0$  viene dado por

$$\mathbb{E}(\chi_{s_0}(v)) = B^T(v)\vartheta \forall v \in T.$$

Por otro lado, teniendo en cuenta el valor esperado para el estimado se tiene:

$$\mathbb{E}(\hat{\chi}_{s_0}(v)) = B^T(v) \sum_{i=1}^n C_i^T W \vartheta \forall v \in T.$$

Consecuentemente de las ecuaciones  $\mathbb{E}(\chi_{s_0}(v))$  y  $\mathbb{E}(\hat{\chi}_{s_0}(v))$ , se observa que el predictor  $\hat{\chi}_{s_0}(v)$  es insesgado si y sólo si

$$B^T(v) \sum_{i=1}^n C_i^T W \vartheta = B^T(v)\vartheta \forall v \in T,$$

esto es, si y sólo si,

$$\sum_{i=1}^n C_i^T W \vartheta = \vartheta.$$

Esta condición es equivalente a  $\sum_{i=1}^n C_i = W^{-1}$ .

Los  $n$  parámetros funcionales en el predictor,  $\hat{\chi}_{s_0}(v)$ , vienen dados por la solución del siguiente problema de optimización

$$\min_{C_1, \dots, C_n} \int_T V(B^T(v)\hat{a}_0 - B^T(v)a_0) dv \text{ sujeto a : } \sum_{i=1}^n C_i = W^{-1}. \quad (4.9)$$

Desarrollando la integral en la función objetivo se obtiene que

$$\int_T B^T(v)V(\hat{a}_0 - a_0)B(v)dv = TrV(\hat{a}_0 - a_0)W.$$

La varianza en la expresión anterior es

$$V(\hat{a}_0 - a_0) = \sum_{i=1}^n C_i^T WV(a_i)WC_i + 2 \sum_{i < j}^n C_i^T WV(a_i, a_j)WC_j + V(a_0) - 2 \sum_{i=1}^n C_i^T WC(a_i, a_0).$$

Definiendo las siguientes matrices  $Q_{ii} = (WV(a_i)W)$ ,  $Q_{ij} = (WC(a_i, a_j)W)$ ,  $N_i = WC(a_i, a_0)$ . El problema de optimización (4.3) puede ser expresado como

$$\min_{C_1, \dots, C_n, m} \sum_{i=1}^n Tr(C_i^T Q_{ii} C_i W) + 2 \sum_{i < j}^n Tr(C_i^T Q_{ij} C_j W) \quad (4.10)$$

$$+ Tr(V(a_0)W) - 2 \sum_{i=1}^n Tr(C_i^T N_i W) \quad (4.11)$$

$$+ 2m \left( \sum_{i=1}^n C_i - W^{-1} \right) \quad (4.12)$$

Derivando con respecto a  $C_i, i = 1, \dots, n$  y  $m$ :

$$\frac{\partial}{\partial C_i} = 2 \sum_{j=1}^n Q_{ij} C_j W - 2N_i W + 2m$$

$$\frac{\partial}{\partial m} = \sum_{i=1}^n C_i - W^{-1}.$$

La solución del problema se obtiene igualando a cero estas derivadas. Esta solución en notación matricial queda:

$$\begin{bmatrix} Q_{11} & Q_{12} & \dots & Q_{1n} & I \\ Q_{21} & Q_{22} & \dots & Q_{2n} & I \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Q_{n1} & Q_{n2} & \dots & Q_{nn} & I \\ I & I & \dots & I & 0 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \\ m^* \end{bmatrix} = \begin{bmatrix} N_1 \\ N_2 \\ \vdots \\ N_n \\ W^{-1} \end{bmatrix}$$

donde  $m^* = mW^{-1}$ . De  $V(\hat{a}_0 - a_0)$  y de las matrices  $Q_{ii}$ ,  $Q_{ij}$  y  $N_i$ , una estimación de la predicción de la varianza integrada

$$\sigma_{int}^2 = \int_T V(\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)) dv$$

viene dada por

$$\begin{aligned} \hat{\sigma}_{int}^2 &= \sum_{i=1}^n Tr(\hat{C}_i^T Q_{ii} \hat{C}_i W) + 2 \sum_{i < j}^n Tr(\hat{C}_i^T Q_{ij} \hat{C}_j W) \\ &+ Tr(V(a_0)W) - 2 \sum_{i=1}^n Tr(\hat{C}_i^T N_i W) \end{aligned}$$

donde las matrices  $\hat{C}_1, \dots, \hat{C}_n$  se obtienen resolviendo el sistema matricial anterior.

En el resto de esta sección, se discute sobre la conexión entre los dos métodos introducidos en el proyecto. La relación entre Cbfd y FKTM es análoga al análisis del Cokriging y la predicción espacial multivariante, en el sentido de que la predicción obtenida por Cbfd es a la vez idéntica a la predicción obtenida por FKTM al mismo tiempo. Las expresiones de estos predictores, las restricciones de insesgadez y la respectiva función objetivo, son equivalentes para cada  $v$  fijo. De hecho, la siguiente proposición establece la equivalencia entre ambas alternativas cuando, se usa la expansión bases.

**Proposición 1:**

Supongamos que la representación  $\chi_{s_i}(t) = a_i^T B(t)$  es correcta para  $\chi_{s_i}(t)$ . Permitamos  $\chi_{s_0}^C(t), v \in T$  sea el conjunto de predictores para  $\chi_{s_0}(v)$  derivados de las  $\chi_{s_i}(t) = a_i^T B(t)$  y  $\lambda_i^v(t) = b_{iv}^T B(t)$ , donde los coeficientes  $b_{iv}, i = 0, 1, \dots, n$  son una solución al problema (4.7) para cada  $v \in T$ . Sea  $\hat{\chi}_{s_0}^F(v), v \in T$  el predictor dado por las ecuaciones  $\hat{\chi}_{s_0}(v) = \sum_{i=1}^n \int_T \lambda_i(t, v) \chi_{s_i}(t) dt$  y  $\lambda_i(t, v) = B^T(t) C_i B(v)$ , donde los coeficientes de las matrices  $C_i, i = 0, 1, \dots, n$  son la solución al problema (4.12). Entonces

$$\hat{\chi}_{s_0}^C(v) = \hat{\chi}_{s_0}^F(v)$$

para todo  $v \in T$  y para todo  $s_0 \in D$ .

La diferencia entre resolver un número infinito predicciones punto a punto (point-wise) mediante Cbfd y hacerlo sólo mediante una predicción de una curva completa por FKTM se debe a la utilización de un doble expansión en términos de bases de funciones para los parámetros funcionales de FKTM. Una característica distintiva entre estas metodologías se da en términos de la varianza de predicción. La predicción de la varianza estimada por FKTM en la expresión de  $\hat{\sigma}_{int}^2$ , se puede utilizar como una medida global de la incertidumbre en la predicción de una curva completa, mientras que la varianza de predicción estimada de Cbfd  $\hat{\sigma}_{s_0}^2(v)$  se puede utilizar en un sentido clásico, es decir, podemos, por ejemplo, calcular intervalos de confianza para la predicción.

### 4.4.3 Resultados: Predicción espacial de la temperatura en las curvas Galicia

En la sección anterior se concluyó que una base con 35 funciones era adecuada para suavizar este conjuntos de datos, por tanto ahora usamos ese mismo número de bases de funciones. De

la Proposición 1 sabemos que las predicciones obtenidas por CBFDF coinciden con aquellas dadas por FKTM. Así que sólo muestran los resultados obtenidos por FKTM por no hacer muy extensivo este trabajo. Para cada caso, se estimó una LMC y se utiliza para el cálculo de las matrices  $Q_{ij}, N_i, i, j = 1, \dots, 66$  en la ecuación matricial

$$\begin{bmatrix} Q & I \\ I & 0 \end{bmatrix} \begin{bmatrix} C \\ m^* \end{bmatrix} = \begin{bmatrix} N \\ W^{-1} \end{bmatrix}$$

y consecuentemente para la estimación de los parámetros indicados en dicho sistema. Para ajustar el LMC todos los variogramas simples (directos) y variogramas cruzados fueron modelados como una combinación lineal de modelos con efecto pepita (nugget) y modelo exponencial. Se ha usado para ello la librería *gstat*, (Pebesma, 2004), del lenguaje R.

La Figura 4.25 muestra la predicción en Facultad de Matemáticas (punto no muestreado), junto con una comparación gráfica entre curvas suavizadas y predichas, lo que muestra un buen resultado de las predicciones.

La Figura 4.26 presenta los residuos obtenidos por validación cruzada, la media y la desviación típica. Se puede ver en esta Figura que hay buenas predicciones en una alta proporción de los sitios (teniendo residuos cercanos a cero). Residuos grandes, negativos o positivos se obtienen en tan sólo un número pequeño de estaciones. No se observa, dependiendo de la estación del año en la que estemos que los residuos sean más grandes o pequeños que en otras estaciones. A final del año se aprecian residuos superiores a  $6^\circ C$ , se podría decir que se tiene una mayor incertidumbre en la predicción durante este período, pero como no se nota al principio y es un dato periódico más bien se trate de un atípico. Si se recuerda en la Figura 1 parece que se entrevé una curva que sobresal al final del año, lo que se sugiere que pueda tratarse de un atípico.

Se aprecia que la desviación estándar residual es menor en el otoño y en el invierno donde las curvas suavizadas y pronosticadas tienen menor variación y se comprueba con los datos de los promedios de la desviación típica en la Tabla 4.12.

	Primavera	Verano	Otoño	Invierno
Promedio de la desviación típica	2.07	2.03	1.91	1.92

Tabla 4.12: Promedios de la desviación típica.

La media residual varía alrededor de cero, lo que indica que las predicciones son insesgadas. En la Tabla 4.13 se obtienen los promedios por estación del año.

	Primavera	Verano	Otoño	Invierno
Promedios	0.00	-0.11	0.05	0.12

Tabla 4.13: Promedios de la media.

Las estaciones Lardeira (ayuntamiento de Carballeda de Valdeorras (Ourense)) y Alto do Rodicio (ayuntamiento de Maceda (Ourense)) están al sur este de Galicia son las estaciones con mayores residuos por validación cruzada, 10474.70 y 10227.14, respectivamente. Aunque la temperatura media diaria en estas estaciones tiene un comportamiento muy similar todo el tiempo, en Lardeira la diferencia entre el máximo y el mínimo es de  $28.1^\circ C$  y en Alto do

Rodicio es de  $27.7^{\circ}C$ . Esto genera residuos altos en ambos casos. En estas dos estaciones el clima que tienen se encuadra dentro de lo que se ha definido por *oceánico de montaña* que se caracteriza por tener veranos frescos e inviernos menos fríos que en otras zonas montañosas más alejadas del mar.

Por el contrario las estaciones de las Islas Cíes (ayuntamiento de Vigo (Pontevedra)) e Isla de Ons (ayuntamiento de Bueu (Pontevedra)) están al sur oeste de Galicia son las estaciones con menores residuos por validación cruzada, 1040.85 y 1053.69, respectivamente. La diferencias entre el máximo y el mínimo son de  $9.3^{\circ}C$  y  $18.7$ , para cada una.

Con respecto a la predicción en la Facultad de Matemáticas, observamos en la Figura 4.26 que la curva predicha muestra un comportamiento estacional similar a las curvas suavizadas y es consistente con los valores reales registrados en la estación de Santiago.

Tres de los parámetros estimados para hacer esta predicción se muestran en las Figuras 4.27 y 4.28, corresponden a las estaciones de A Pontenova (Lugo), Monte Aloia (ayuntamiento de Tui (Pontevedra)), Malpica (A Coruña) y Sergude (ayuntamiento de Boqueixón (A Coruña)).

De la Figura 4.28 se puede destacar que los parámetros de Malpica son casi nulos, son los más pequeños y por el contrario en Sergude son los más grandes. Otra vez se pone de manifiesto la cercanía al punto no muestreado. Por tanto esto quiere decir que Sergude tiene una mayor influencia sobre la predicción que Malpica (cuyos parámetros funcionales son casi nulos). Además la estimación funcional de los parámetros revela que hay un efecto a corto temporal en la predicción. Ambos resultados eran de esperar. Por un lado, el primer resultado es coherente con la filosofía de geoestadística, es decir, los sitios más cerca de la ubicación de predicción tienen una mayor influencia que otros más alejados. Al igual que en el caso de Malpica, otras estaciones muy separadas de la Facultad de Matemáticas tienen una baja influencia en la predicción.

Finalmente se ha aplicado esta técnica a los datos de la temperatura de Galicia (2009), eliminándole a los residuos el efecto de la latitud y longitud. Los resultados obtenidos se pueden consultar en la tabla (4.14) y servirán para concluir en la comparación con qué metodología nos quedamos.

## 4.5 Comparación de las tres técnicas

Para llevar a cabo la comparación se usa el resumen estadístico de  $SSE_{FCV}(i)$ :

$$SSE_{FCV} = \sum_{i=1}^n SSE_{FCV}(i) = \sum_{i=1}^n \sum_{j=1}^M \left( \chi_{si}(t_j) - \hat{\chi}_{si}^{(i)}(t_j) \right)^2$$

para comparar las metodologías propuestas en el proyecto. Evaluamos en  $j = 1, \dots, 365$  las predicciones obtenidas por validación cruzada para OKFD, CTKFD y FKTM. Para incluir el predictor Cbfd en este análisis calculamos  $\hat{\chi}_{s0}(v) = \sum_{i=1}^n \int_T \lambda_i^v(t) \chi_{si}(t) dt$ . Los valores del estadístico resumen  $SSE_{FCV}$  obtenidos por Cbfd y FKTM coinciden porque las predicciones por Cbfd se calculan con los mismos argumentos donde se evalúan las curvas predichas por FKTM. Para el conjunto de datos, que inicialmente se elimina la tendencia de los datos mediante un modelo de regresión funcional. Aplicamos los métodos de predicción espacial propuestos

para los residuos de la regresión. Las predicciones de temperatura para estos conjunto de datos se obtienen agregando las predicciones residual a la tendencia.

Estadístico	OKFD	OKFD*	CTKFD	FKTM	FKTM*
Mínimo	1326.0	812.6	865.3	1041.0	805.3
1st Qu.	2931.0	2017.0	2373.8	1971.0	1951.0
Mediana	3861.0	2626.0	2768.0	2554.0	2499.0
Media	4306.0	2990.0	3212.3	3030.0	3009.0
3rd Qu.	4847.0	3403.0	3793.2	3261.0	3468.0
Máximo	12880.0	11090.0	10061.0	10470.0	10180.0
Desviación típica	2279.5	1551.3	1474.8	1852.1	1780.6
Suma	284222.1	197345.3	212009.0	199959.1	198569.1

Tabla 4.14: Resumen estadístico de la suma de los cuadrados de los errores de validación cruzada. (Nota: \*, método de la columna anterior sin los efectos de la longitud y la latitud.)

Se observa en la Tabla (4.14) que donde existen las mayores diferencias entre los métodos son en términos de los valores mínimos o máximos y también influenciadas por el efecto de la longitud y la latitud.

Los predictores tienen un comportamiento similar cuando las curvas son relativamente homogéneas y las diferencias entre los mismos con los datos de temperatura para Galicia son fundamentalmente debidas a su desempeño en las estaciones con temperaturas extremas como se explicó anteriormente.

Los resultados mostrados en la Tabla (4.14) indican que la inclusión de un doble índice funcional en los parámetros (efecto temporal en el conjuntos de datos) no reflejan cambios sustanciales en las predicciones del análisis. Es decir, el trabajo de estimar tantos parámetros no revierte mejores resultados, aunque, si a los datos se les elimina el efecto de la latitud y la longitud se obtienen mejores

En principio si las curvas son homogéneas podríamos usar cualquiera de los tres enfoques. Sin embargo, si lo observado en los datos tiene una variabilidad alta CTKFD podría ser la mejor opción.

En resumen, podemos concluir que OKFD sin los efectos de la longitud y la latitud es la mejor opción para llevar a cabo la predicción espacial de datos funcionales en Galicia. Por lo tanto todos ellos son buenas alternativas para hacer la predicción espacial de los datos funcionales. Teniendo en cuenta que es más simple OKFD sin los efectos de la longitud y la latitud que otros desde un punto de vista práctico y computacional en aplicaciones con grandes conjuntos de datos este método puede ser preferible.

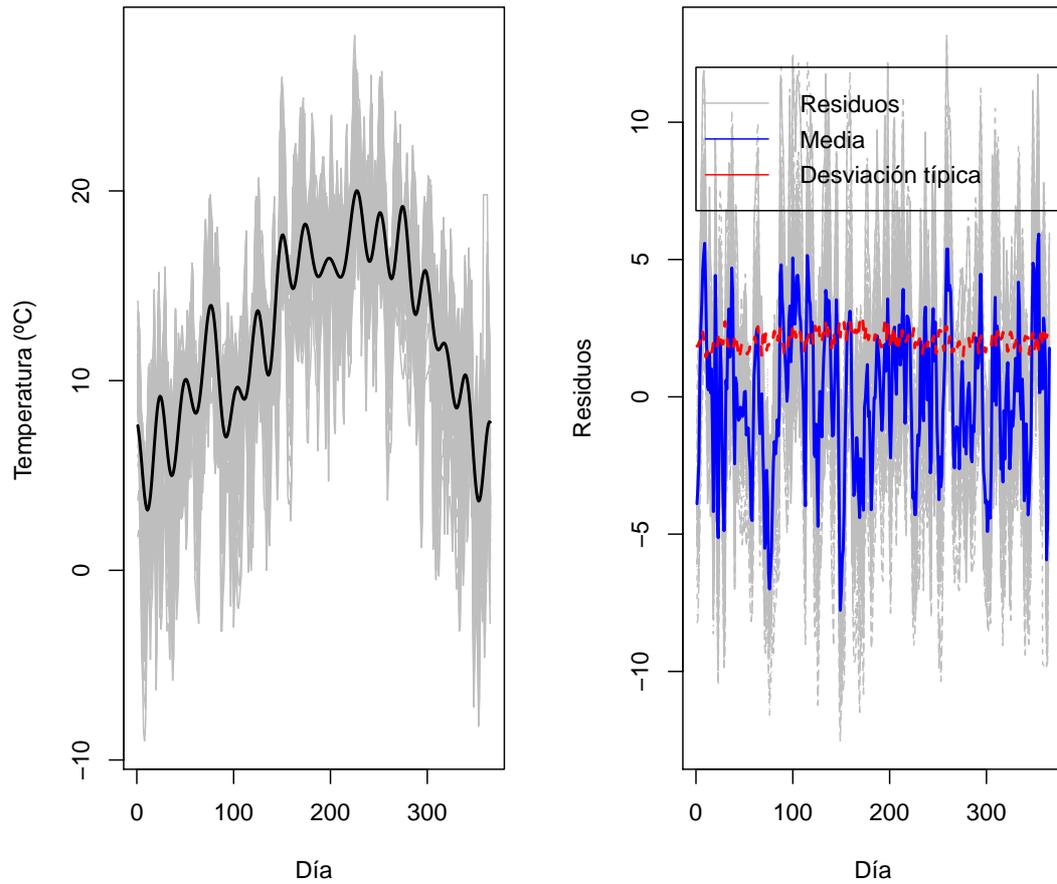


Figura 4.13: Izquierda: muestra una comparación gráfica entre las curvas observadas (las reales) y la predicha (utilizando FCV). Derecha: la de la derecha lo residuos de validación cruzada funcional, media residual y desviación típica residual.

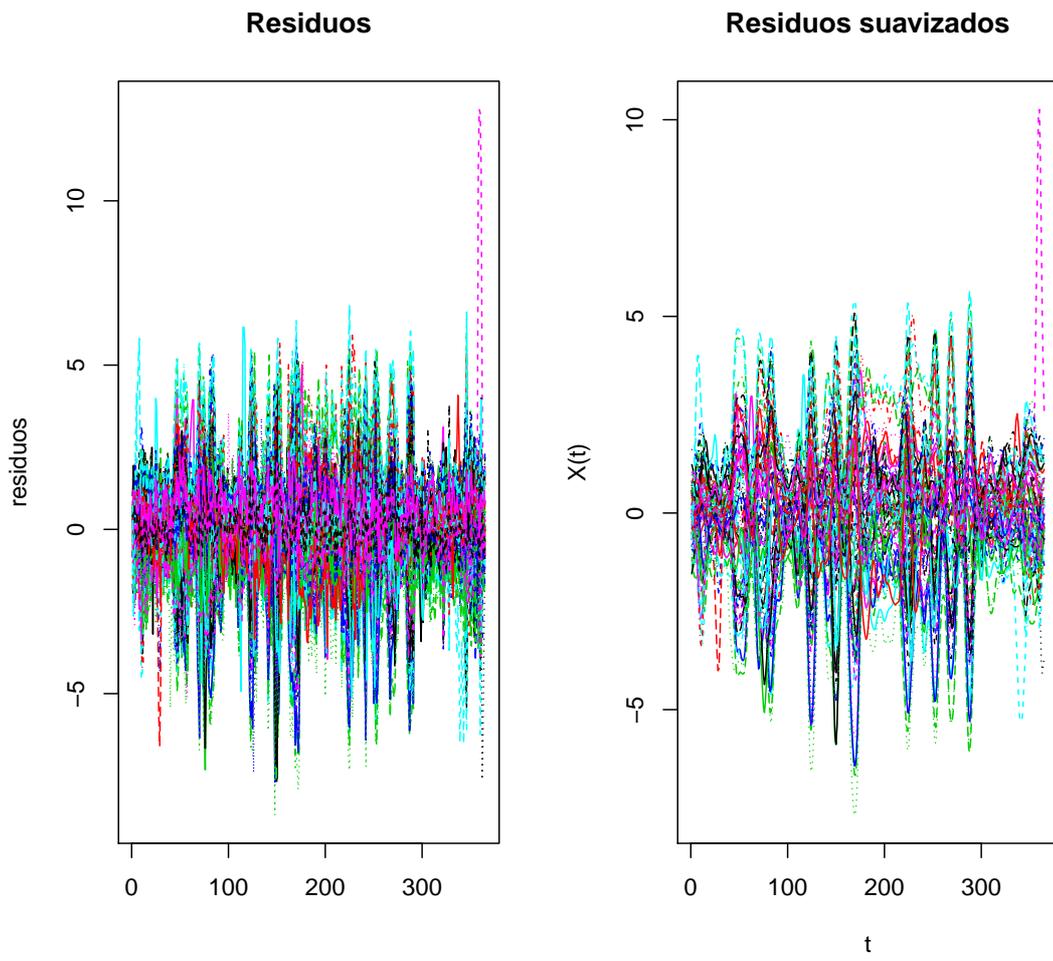


Figura 4.14: Residuos del modelo con altitud y latitud.

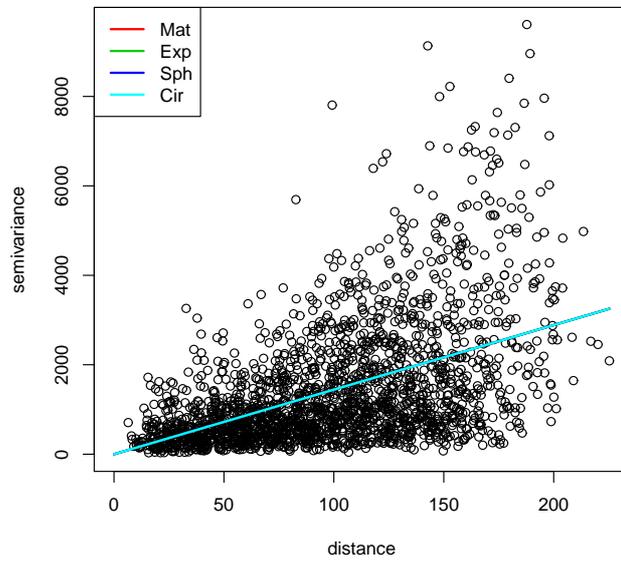


Figura 4.15: Variograma para los residuos.

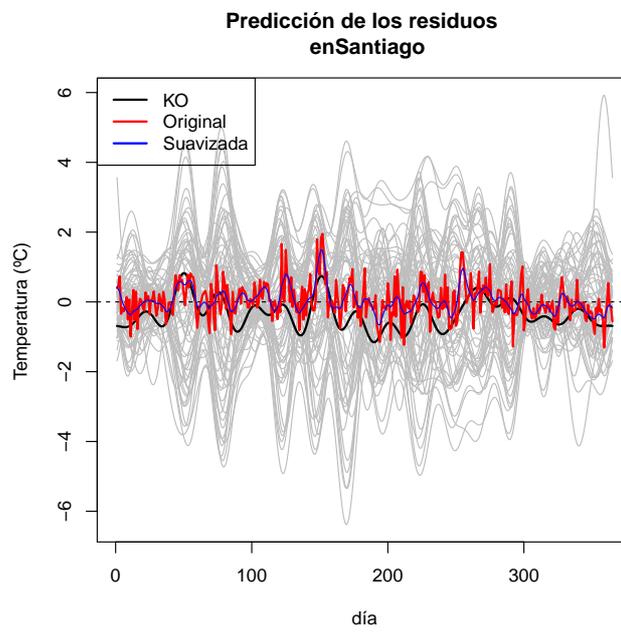


Figura 4.16: Predicción para los residuos.

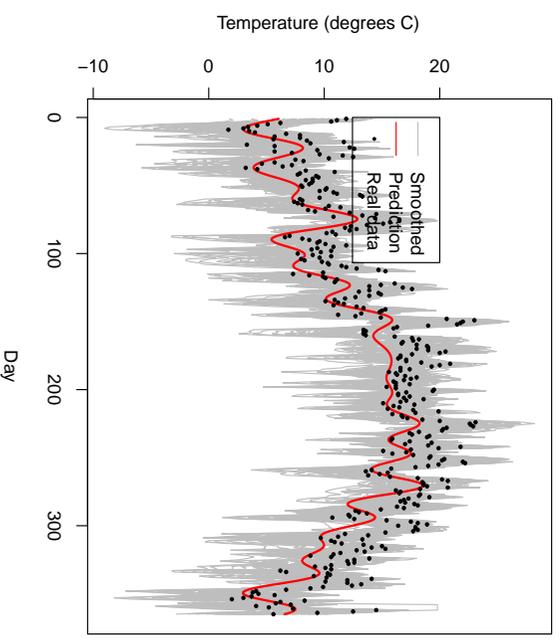


Figura 4.17: Datos suavizados, predicción en Santiago y datos reales en Santiago.

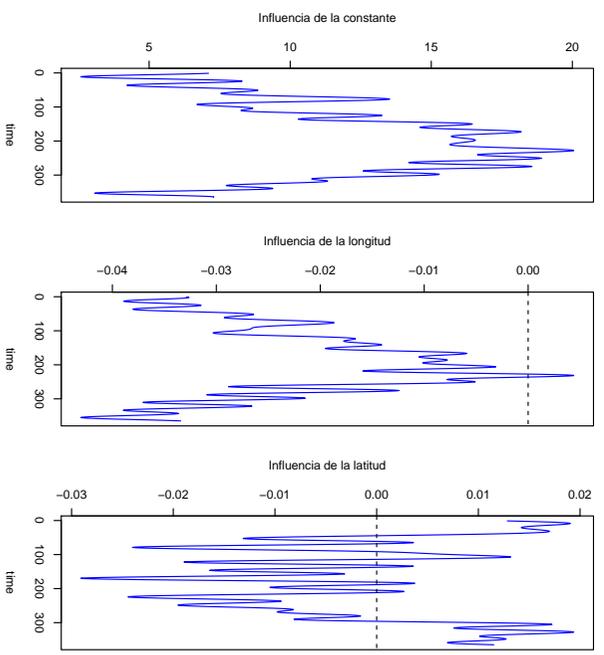


Figura 4.18: Influencia de la constante, de la longitud y latitud en el modelo.

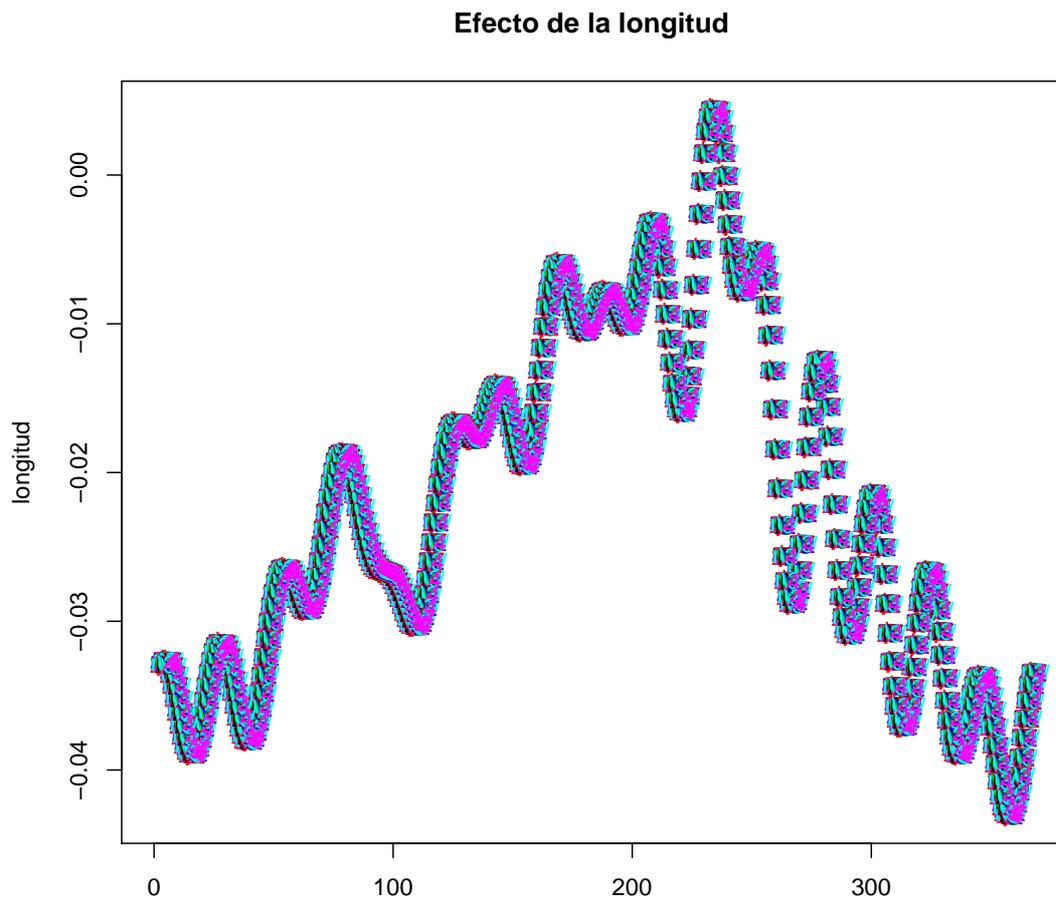


Figura 4.19: Efecto de la variable longitud en el modelo (4.4), al que se le aplicará el Kriging.

### Efecto de la latitud

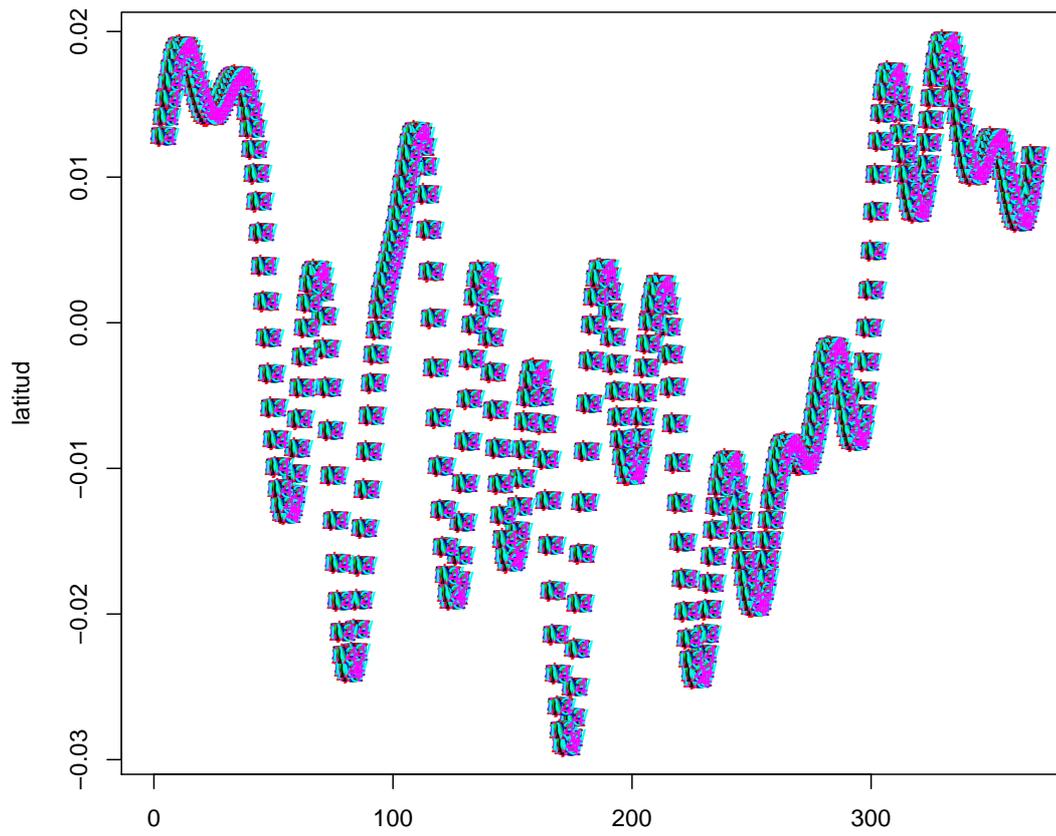


Figura 4.20: Efecto de la variable latitud en el modelo (4.4), al que se le aplicará el Kriging.

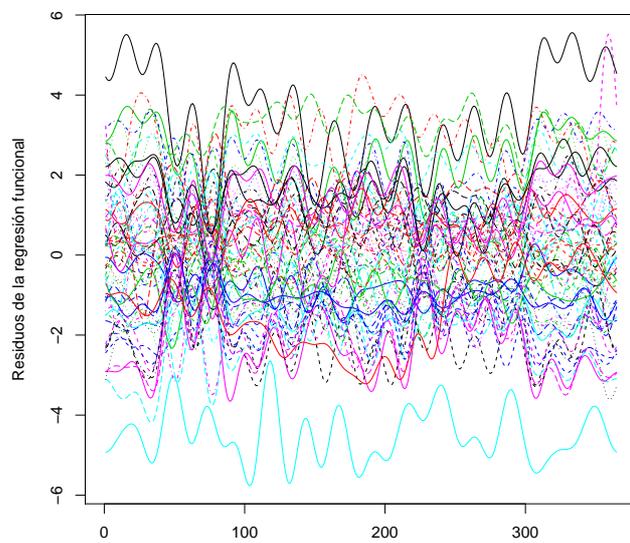


Figura 4.21: Residuos del modelo (4.4), al que se le aplicará el Kriging.

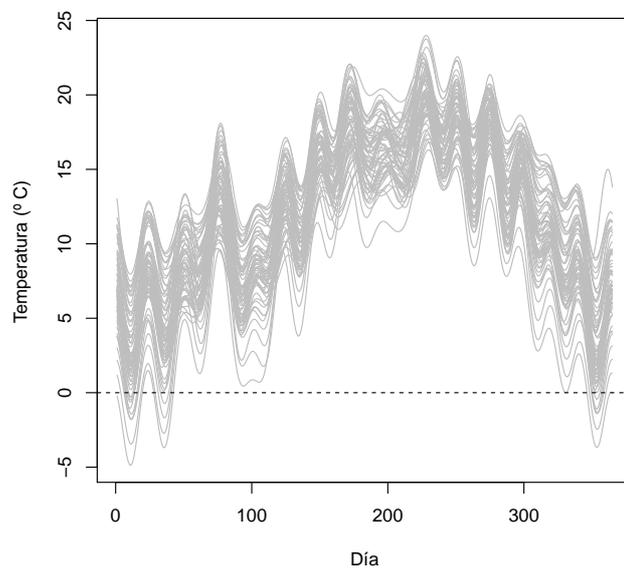


Figura 4.22: Datos suavizados, predicción en Santiago y datos reales en Santiago.

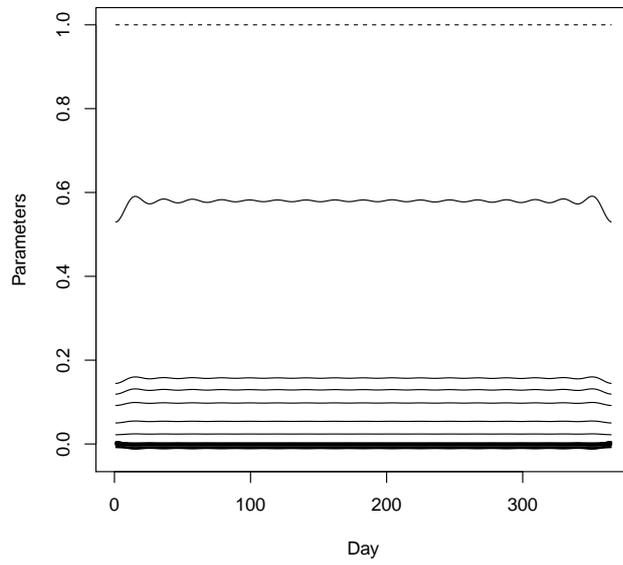


Figura 4.23: Parámetros funcionales estimados.

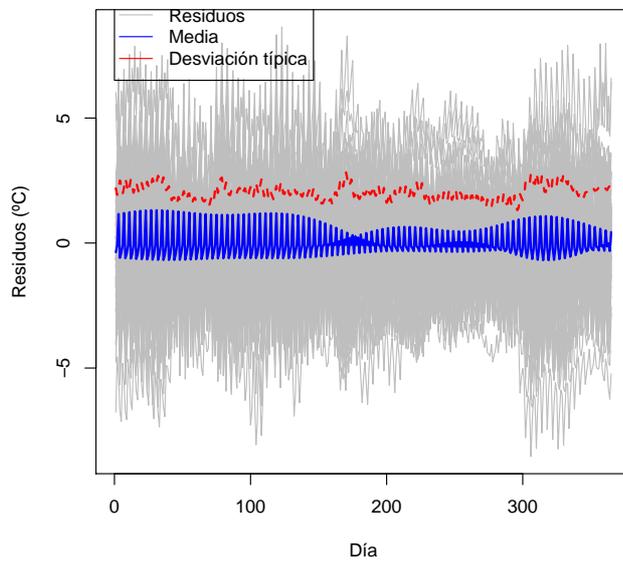


Figura 4.24: Residuos, media y desviación típica con la técnica 2 (solución basada en bases de funciones).

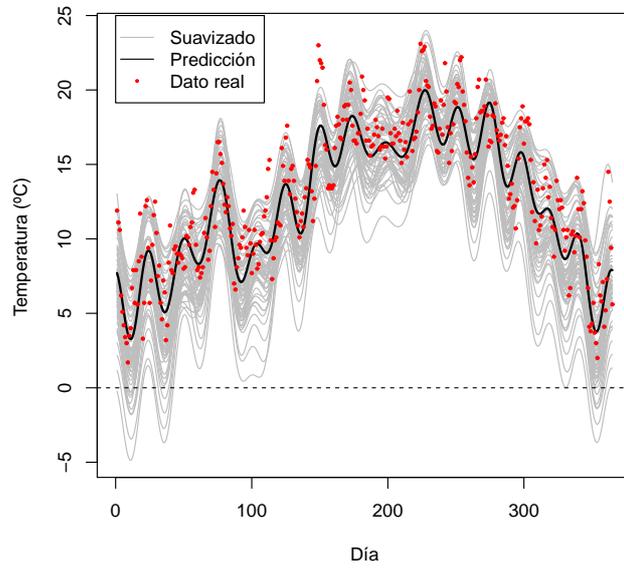


Figura 4.25: Datos suavizados, predicción en Santiago y datos reales en Santiago.

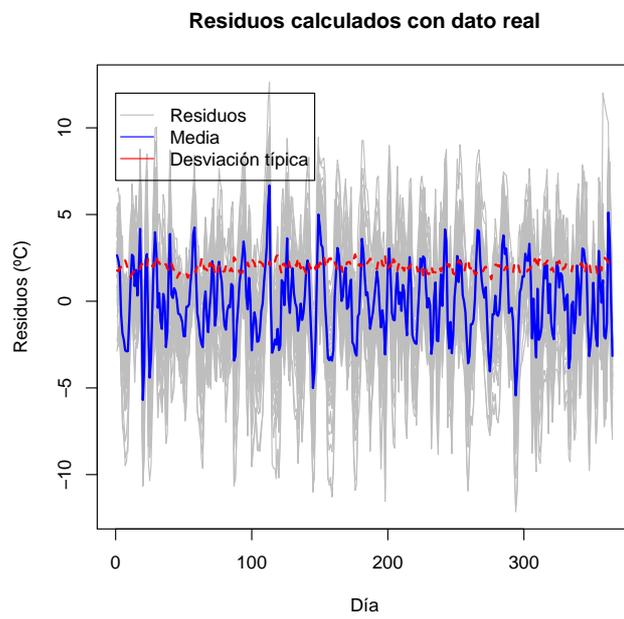
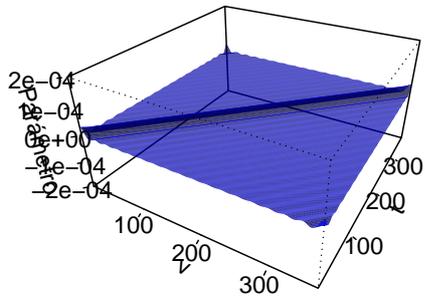


Figura 4.26: Residuos, media y desviación típica obtenidos al aplicar FKTM.

**Pontenova (A)**



**Monte Aloia**

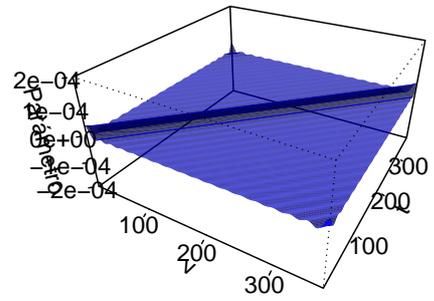
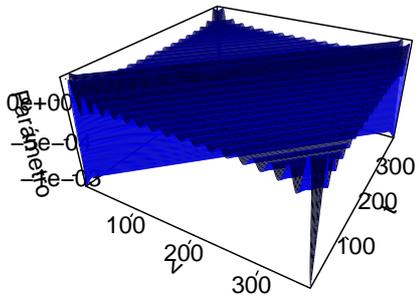


Figura 4.27: Parámetros funcionales estimados en A Pontenova y Monte Aloia.

**Mapica**



**Sergude**

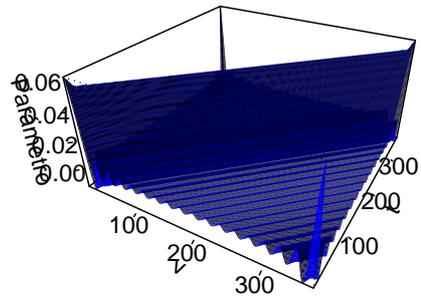


Figura 4.28: Parámetros funcionales estimados en Malpica y Sergude.



## **Apéndice A**

### **Histogramas y boxplots de la Temperatura de Galicia año 2009**

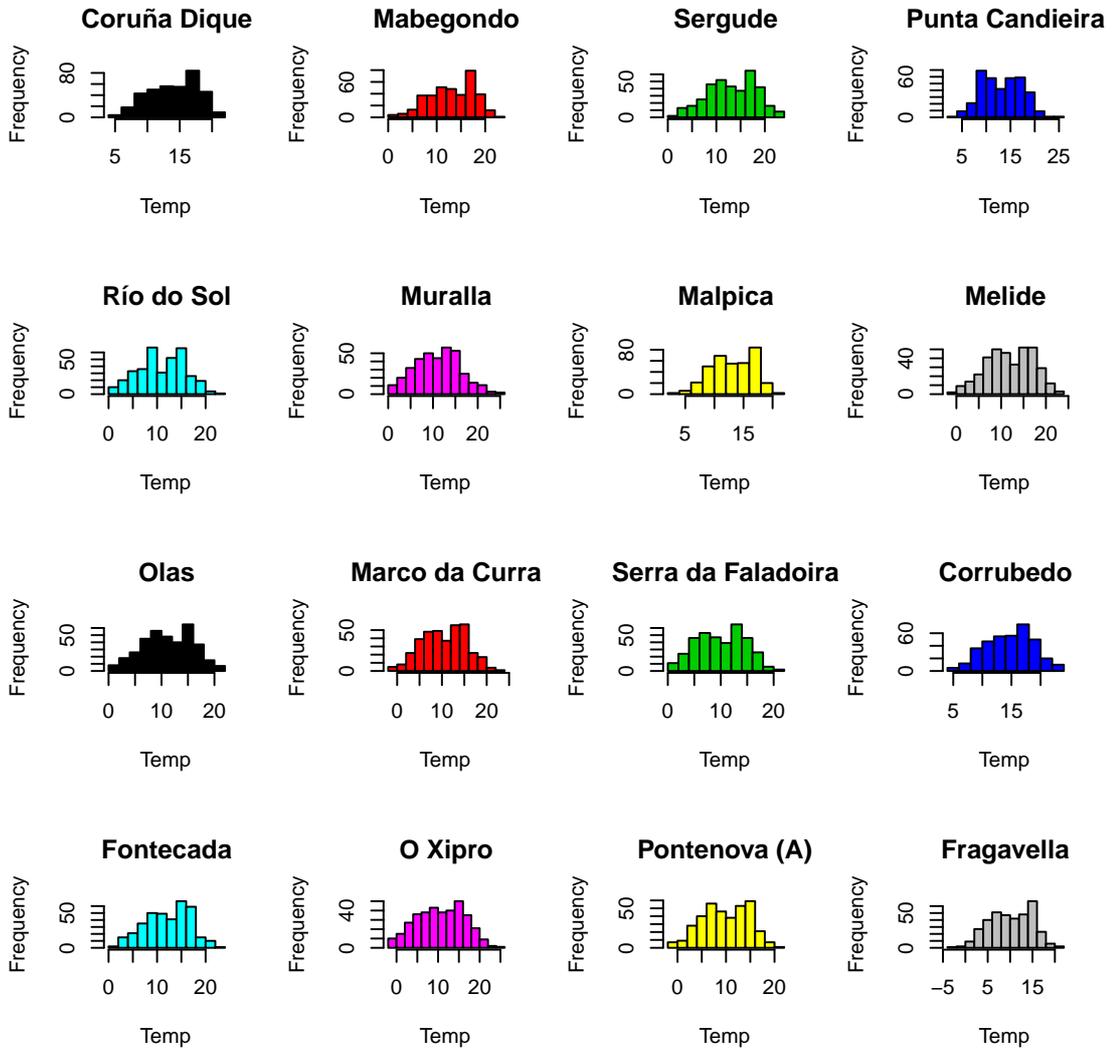


Figura A.1: Histogramas de la Temperatura de Galicia año 2009.

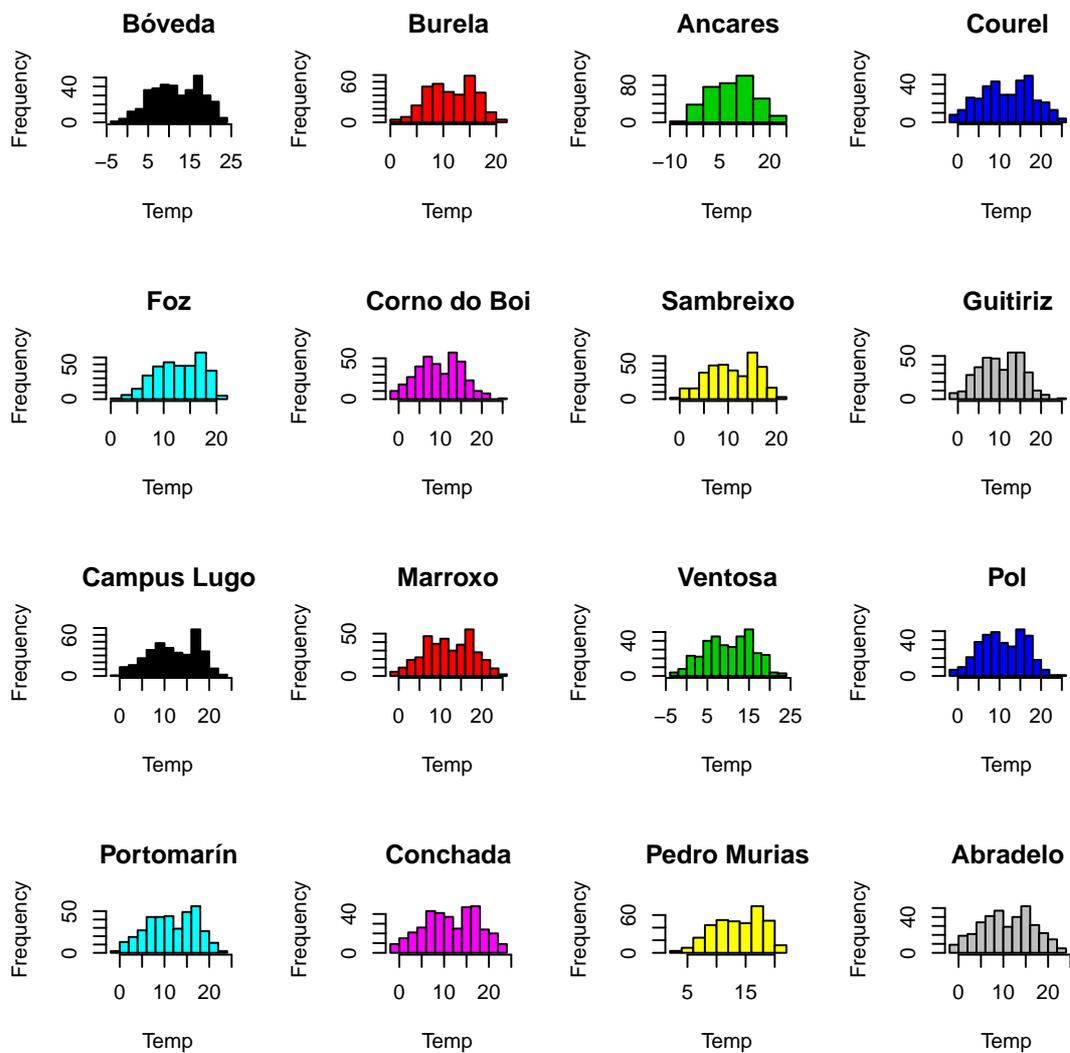


Figura A.2: Histogramas de la Temperatura de Galicia año 2009.

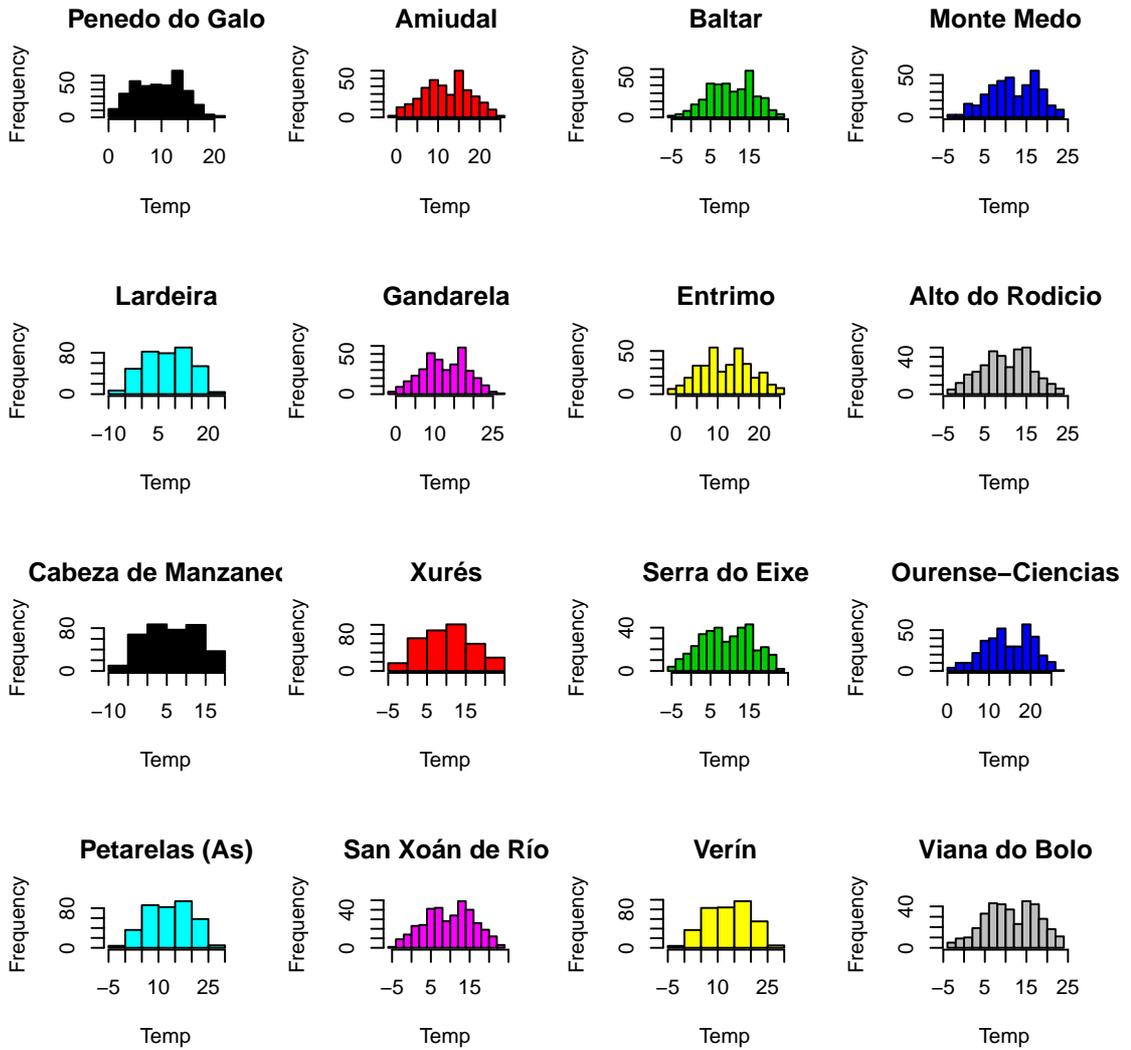


Figura A.3: Histogramas de la Temperatura de Galicia año 2009.

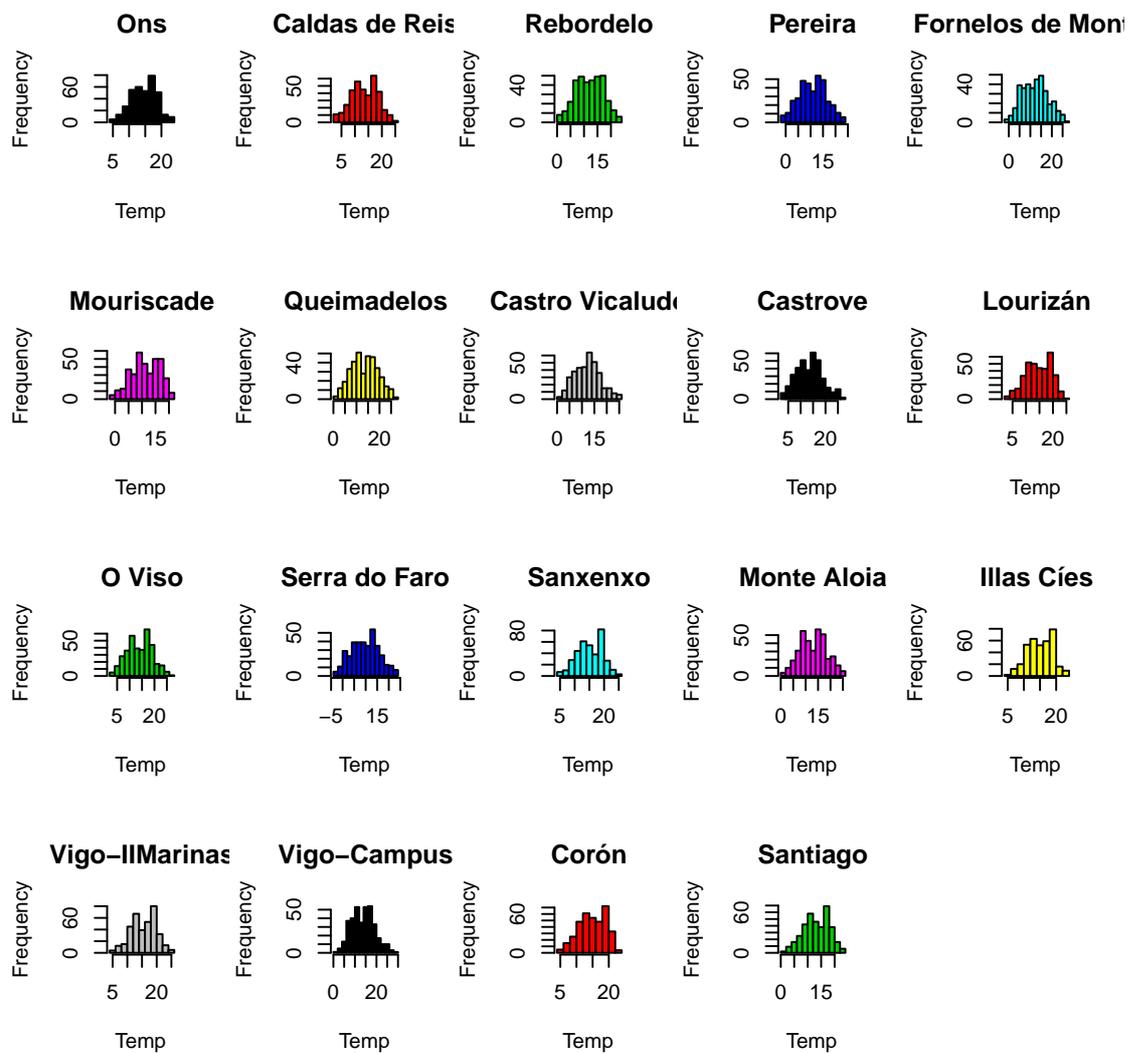


Figura A.4: Histogramas de la Temperatura de Galicia año 2009.

**Box-Plot de la temperatura Galicia año 2009**

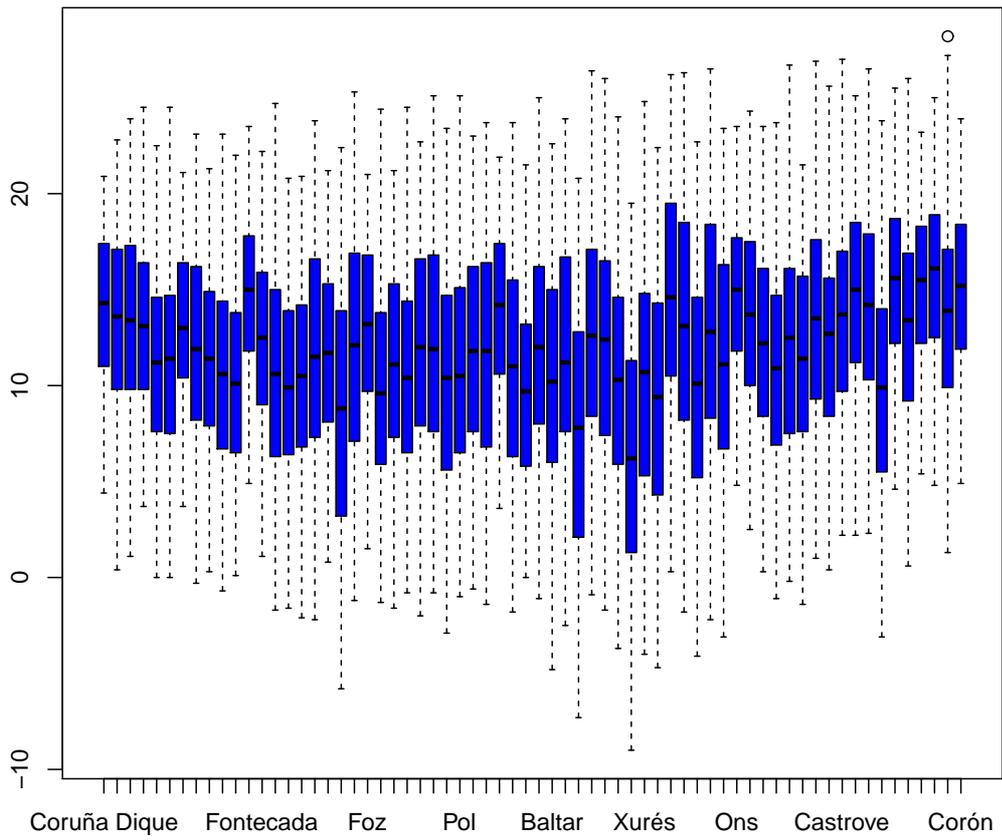


Figura A.5: Boxplot de la Temperatura de Galicia año 2009.

# Bibliografía

- [1] Alias Abdul-Rahman and Morakot Pilouk. *Spatial Data Modelling for 3D GIS*. Springer-Verlag, 2008.
- [2] Maria João Pereira Amílcar Soares and Roussos Dimitrakopoulos, editors. *geoENV VI GEOSTATISTICS FOR ENVIRONMENTAL APPLICATIONS*. Springer-Verlag, 2008. Proceedings of the Sixth European Conference on Geostatistics for Environmental Applications.
- [3] Pebesma Edzer J. Gómez-Rubio Virgilio Bivand, Roger S. *Applied Spatial Data Analysis with R*. Use R. Springer-Verlag, 2008.
- [4] Bogaert. Comparison of kriging techniques in a space-time context. *Mathematical Geology*, 28(1):73–86, 1996.
- [5] George Christakos. *Modern Spatiotemporal Geostatistics*. Studies in Mathematical Geology. Oxford University Press, Inc, 2000.
- [6] Noel A. C. Cressie. *Statistics for Spatial Data*. Probability and Statistics. Wiley, 1993.
- [7] Claude Mantéa David Nerini, Pascal Monestiez. Cokriging for spatial functional data. *Journal of Multivariate Analysis, Elsevier*, 101:409–418, 2010.
- [8] Frédéric Ferraty and Philippe Vieu. *Nonparametric Functional Data Analysis Theory and Practice*. Statistics. Springer-Verlag, 2006.
- [9] Matheron G. *The theory of regionalized variables and its applications*. 1971.
- [10] Ramón Giraldo. Análisis exploratorio de variables regionalizadas con métodos funcionales.
- [11] M. Goulard and M. Voltz. Geostatistical interpolation of curves: A case study un soil science. *Academic Press*, 2:805–816, 1993.
- [12] T. Hastie and R. Tibshiran. Varying-coefficient models. *Journal of the Royal Statistical Society Series B*, 55:755–796, 1993.
- [13] J. Ver Hoef and N. Cressie. Multivariable spatial prediction. *mathematical geology*. 25 (2):219–240, 1993.

- [14] E. Isaaks and M. Srivastava. *Applied geostatistics*. Oxford University Press, 1987.
- [15] R. Flanagan J. O. Ramsay, X. Wang. A functional data analysis of the pinch force of human fingers. *The Royal Statistical Society*, 44(1):17–39, 1995.
- [16] Myers. Matrix formulation of co-kriging. *Mathematical Geology*, 14 (3):249–257, 1982.
- [17] D. Nerini and P. Monestiez. A cokriging method for spatial functional data with applications in oceanology. 2008.
- [18] C. Comas P. Delicado, R. Giraldo and J. Mateu. Statistics for spatial functional data. *Environmetrics*, 2009.
- [19] C. Comas P. Delicado, R. Giraldo and J. Mateu. Statistics for spatial functional data: some recent contributions. *Environmetrics*, DOI: 10.1002/env.1003, 2009.
- [20] Paulo J. Rebeiro Jr. Peter J. Diggle. *Model-based Geostatistics*. Statistics. Springer-Verlag, 2007.
- [21] P. Delicado R. Giraldo and J. Mateu. Geostatistics for functional data: An ordinary kriging approach. *Environmental and Ecological Statistics*, 2007.
- [22] Pedro Delicado Ramón Giraldo and Jorge Mateu. Continuous time-varying kriging for spatial prediction of functional data: An environmental application. *Journal of Agricultural, Biological and Environmental Statistics*, 2008.
- [23] Ramsay and Silverman. *Functional data analysis*. Springer, 1997.
- [24] J. O. Ramsay and C.J. Dalzell. Some tools for functional data analysis. *The Royal Statistical Society*, 53(3):539–572, 1991.
- [25] James O. Ramsay and Bernard W. Silverman. *Applied Functional Data Analysis: Methods and Case Studies*. Springer-Verlag, 2002.
- [26] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Statistics. Springer-Verlag, 2nd edition, 2005.
- [27] J. M. Yarus T. C. Coburn and R. L. Chambers, editors. *Stochastic Modeling and Geostatistics: Principles, Methods, and Case Studies*, volume II. 2007.
- [28] H. Wackernagel. *Multivariable geostatistics: An introduction with applications*. Springer Verlag, 1995.
- [29] Hans Wackrenagel. *Multivariate Geostatistics*. Springer-Verlag, 2nd edition, 1998.
- [30] Larry Wasserman. *All of Nonparametric Statistics*. Springer, 2006.

# Glosario

$Z(u)$	Proceso aleatorio, 6
$u$	Localización, 7
$\mathbb{R}^d$	Espacio Euclídeo d-dimensional, 7
$2\gamma(\cdot)$	Variograma, 9
$\gamma(\cdot)$	Semivariograma, 9
$\ h\ $	Longitud, 9
$\Gamma(\cdot)$	Función Gamma, 12
$\gamma_0$	Semivariograma isotrópico, 13
$\lambda_\alpha$	Ponderaciones de los valores originales, 22
$C_{i,j}(h)$	Covariograma cruzado, 29
$\langle X(t), Y(t) \rangle$	Producto interior, 39
$\ X(t)\ $	Norma, 40
$L^2(T)$	Es el espacio de las funciones de cuadrado integrable, 40
$\xi_m \perp \xi_k$	$\xi_m$ y $\xi_k$ son ortogonales, 41
<i>FDA</i>	Análisis de datos funcionales, 41
<i>MoM</i>	Modelo clásico de los momentos, 58
<i>CV</i>	Métodos de validación cruzada, 61
<i>GCV</i>	Métodos de validación cruzada generalizada, 61
<i>LLR</i>	Regresión Local Lineal, 61
<i>Km</i>	Kilómetros, 62
<i>FCV</i>	Validación cruzada funcional, 64
<i>MISE</i>	Error cuadrático medio integrado, 64
<i>R – Squared</i>	Coefficiente determinación, 65
<i>W</i>	Matriz de Gram, 77
<i>BLUP</i>	Mejor predictor lineal insesgado, 78
<i>LMC</i>	Modelo linal de correogionalización, 79

# Índice

- análisis de datos funcionales, 1, 40
- B-spline, 38
- base, 38
- Cauchy-Schwarz, 8, 30
- Cokriging ordinario, 34
- Cokriging Simple, 31
- componentes principales, 41
- Condición de inestabilidad, 36
- condición de inestabilidad, 20
- covariograma, 8
- error cuadrático medio, 24, 55
- estacionariamente débil, 8
- estacionario de segundo orden, 8
- Estrictamente Estacionario, 8
- FDA, 41
- Fourier, 38
- georreferenciados, 5, 6
- Geostatística, 5
- Hole, 11
- intervalos de confianza, 83
- kriging funcional, 53
- Kriging Ordinario, 21
- kriging ordinario, 34, 55
- Kriging Simple, 20
- Kriging universal, 25
- lattice data, 5, 6
- Matérn, 12
- Modelo esférico, 11
- Modelo Exponencial, 11
- Modelo Gaussiano, 11
- Modelo lineal, 11
- Modelo potencial, 11
- nugget, 10
- Nugget effect, 10
- penalización, 42
- Pepita, 10
- point processes data, 5, 6
- proceso estocástico, 7
- proceso estrictamente estacionario, 8
- procesos puntuales, 5, 6
- pseudosemivariograma cruzado, 30
- rejilla, 5
- semivariograma, 9
- semivariogramas cruzados, 30
- umbral, 12
- Validación Cruzada, 24
- Variable regionalizada, 7
- variograma, 9
- Wavelet, 38