

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ  
ESCUELA DE POSGRADO



INFERENCIA BAYESIANA APROXIMADA DEL MODELO  
ESPACIO-TEMPORAL USANDO NNGP

TESIS PARA OPTAR EL GRADO ACADÉMICO DE MAGÍSTER EN  
ESTADÍSTICA QUE PRESENTA:

Omar Eduardo Benites Alfaro

ASESORA:

Dra. Zaida Jesús Quiroz Cornejo

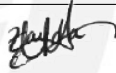
Enero, 2023

## Informe de Similitud

Yo Zaida Jesús Quiroz Cornejo docente de la Escuela de Posgrado de la Pontificia Universidad Católica del Perú, asesora de la tesis titulada *Inferencia bayesiana aproximada del modelo espacio-temporal usando NNGP*, del autor Omar Eduardo Benites Alfaro, dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de 13 %. Así lo consigna el reporte de similitud emitido por el software Turnitin el 31/01/2023.
- He revisado con detalle dicho reporte y la Tesis, y no se advierte indicios de plagio.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lima, 31 de enero de 2023

Apellidos y nombres de la asesora: Quiroz Cornejo Zaida Jesús	
DNI: 43704124	Firma: 
ORCID: <a href="https://orcid.org/0000-0003-3821-0815">https://orcid.org/0000-0003-3821-0815</a>	

## Resumen

Los modelos espacio-temporales nos permiten estudiar la distribución espacial de una variable en el tiempo. Por ejemplo, se puede estudiar la distribución espacial del material particulado en un país a través de los años, dado que las concentraciones de material particulado en estaciones cercanas pueden ser similares y la concentración en una estación en un año puede depender de la concentración en la misma estación el año anterior anterior. En esta tesis se propone usar un modelo espacio-temporal a través del proceso gaussiano de vecinos más cercanos. Para implementar este modelo y aplicarlo en grandes bases de datos se propone usar inferencia bayesiana a través del método de integración aproximada de Laplace (INLA). La bondad de ajuste del modelo y su eficiencia se estudia a través de simulaciones. Finalmente se aplica el modelo implementado a una base de datos reales.

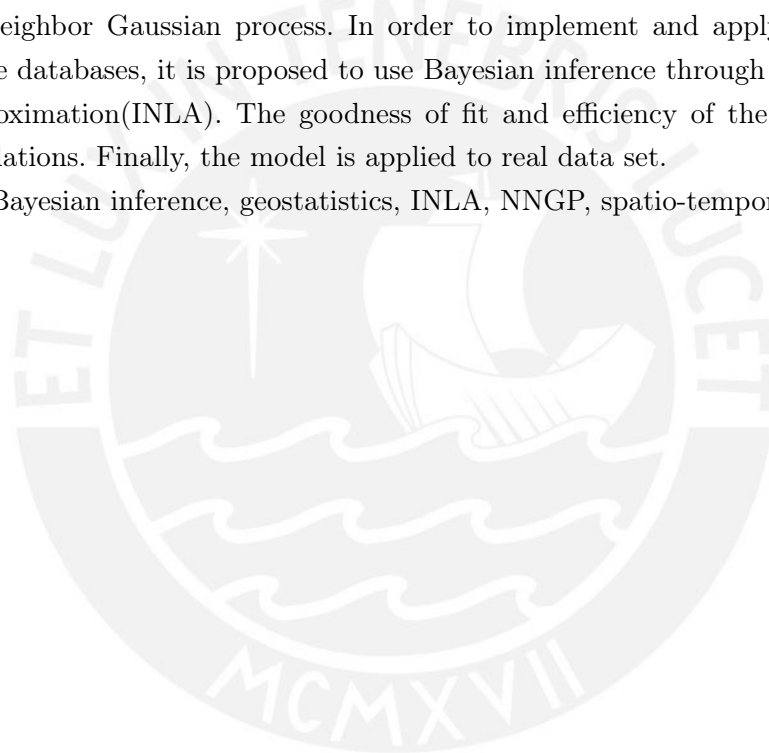
**Palabras-clave:** geostatística, inferencia bayesiana, INLA, modelo espacio-temporal, NNGP.



## Abstract

Spatio-temporal models allow us to study the spatial distribution of a variable over time. For example, we can study the spatial distribution of particulate matter in a country through the years, given that the concentrations of particulate matter in nearby stations can be similar and the concentration in a station in a year can depend on the concentration in the same station in the previous year. In this thesis, we proposed to use a spatio-temporal model through the nearest neighbor Gaussian process. In order to implement and apply the hierarchical model in large databases, it is proposed to use Bayesian inference through Integrated nested Laplace approximation(INLA). The goodness of fit and efficiency of the model is studied through simulations. Finally, the model is applied to real data set.

**Keywords:** Bayesian inference, geostatistics, INLA, NNGP, spatio-temporal model.



# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Objetivos	2
<b>2. Revisión de literatura</b>	<b>3</b>
2.1. Estadística espacial	3
2.1.1. Campo aleatorio	4
2.2. Campo aleatorio gaussiano	4
2.3. La función de covarianza y el variograma	6
2.3.1. Modelo lineal geostadístico gaussiano	7
2.4. Procesos temporales	8
2.4.1. Modelo temporal autoregresivo de orden 1 - AR(1)	8
2.5. Procesos espacio-temporales	9
2.6. Inferencia bayesiana	12
2.6.1. Estadística computacional para la inferencia Bayesiana	12
2.6.2. Aproximación de Laplace	13
<b>3. Modelo espacio-temporal usando NNGP</b>	<b>16</b>
3.1. Proceso espacio - temporal gaussiano de vecinos más cercanos	16
3.2. Inferencia bayesiana usando INLA	19
3.3. INLA para el modelo espacio-temporal NNGP	20
<b>4. Estudio de Simulación</b>	<b>22</b>
4.1. Simulación del proceso espacio-temporal	22
4.1.1. Estimación de parámetros e hiperparámetros	23
4.1.2. Estudio de simulación 1	24
4.1.3. Estudio de simulación 2	26
4.1.4. Estudio de simulación 3	29
<b>5. Aplicación</b>	<b>33</b>
5.1. Descripción del estudio	33
5.2. Modelamiento espacio-temporal NNGP para estimar la calidad del aire	35
5.2.1. Modelo 1 ( $m = 10$ vecinos)	35
5.2.2. Modelo 2 ( $m = 20$ vecinos)	36
5.2.3. Modelo 3 ( $m = 30$ vecinos)	37
5.2.4. Evaluación de los modelos	38

<i>ÍNDICE GENERAL</i>	VI
<b>6. Conclusiones</b>	<b>40</b>
<b>A. Resultados teóricos</b>	<b>41</b>
A.1. Matriz de precisión del efecto espacio temporal . . . . .	41
A.2. Reparametrización usada en el modelo espacio-temporal NNGP . . . . .	42
<b>7. Bibliografía</b>	<b>43</b>



# Capítulo 1

## Introducción

En diversas situaciones prácticas deseamos investigar y analizar variables que derivan de procesos espacio-temporales, es decir, variables cuyas observaciones dependen del espacio y el tiempo en el cual fueron medidas. Por ejemplo, la concentración de  $CO_2$  en la atmósfera ha sufrido variaciones durante muchos años y en diferentes partes del mundo. Por lo tanto, la medición de estos fenómenos no puede ser analizada de forma aislada, ya sea solo temporalmente o espacialmente, sino se debe brindar una respuesta en conjunto usando la interacción espacio-tiempo. Dicha interacción es estudiada por los modelos espacio-temporales (Wikle et al., 2019). En general, estos modelos sirven para modelar la incertidumbre asociada a procesos espacio-temporales, es decir, donde existe una autocorrelación espacial y temporal entre las observaciones de la variable en estudio.

El primer paso para la construcción de los modelos espacio temporales involucra un análisis exploratorio a través del estudio de evidencia de autocorrelación espacial y temporal. Otro aspecto importante es el modelamiento estadístico dinámico, el cual busca describir cómo un proceso espacial cambia a lo largo del tiempo (Wikle et al., 2019). Para la definición de un modelo espacio temporal se puede incluir un efecto aleatorio espacio-temporal, que en particular se asume que sea un efecto aleatorio espacial gaussiano que evoluciona en el tiempo. Usualmente, el efecto espacial es incorporado en el modelo a través de un proceso espacial Gaussiano para el cual es conocido que la matriz de covarianza o matriz de precisión limitan su eficiencia computacional debido a la dependencia espacial entre las  $n$ -observaciones. Por ejemplo, calcular la inversa y determinante de dichas matrices es del orden  $\mathcal{O}(n^3)$ .

Diversos enfoques permiten lidiar con este problema, por ejemplo ver Banerjee et al. (2014); Lindgren (2015); Bolin et al. (2019); Datta et al. (2016). Algunos enfoques buscan aproximar la matriz de covarianza a una matriz llena de ceros, o trabajar en una dimensión menor. Otros enfoques buscan aproximar la inversa de la matriz de covarianza, llamada matriz de precisión, a una matriz con más ceros. En particular, en esta tesis se propone usar el Proceso Gaussiano de Vecinos más Cercanos (Nearest Neighbor Gaussian Process - NNGP) propuesto por Datta et al. (2016), el cual sugiere usar un conjunto de vecinos más cercanos para construir un proceso espacial cuya matriz de precisión contiene más ceros.

El modelo espacio-temporal que usamos en esta tesis pertenece a los llamados modelos gaussianos latentes. Esta clase de modelos se caracterizan por tener dos propiedades: primero, un campo latente de gran dimensión que admite independencia condicional, y segundo, un número reducido de hiperparámetros (Rue et al., 2009). Además, estos modelos son muy

flexibles dado que pueden modelar variables de respuesta que pueden ser gaussianas o no-gaussianas. Dado que los modelos espacio-temporales requieren una cantidad significativa de tiempo en el procesamiento para realizar la inferencia, recurrimos al uso de métodos bayesianos modernos como la aproximación anidada integrada de Laplace (INLA), la cual sirve para aproximar las distribuciones marginales posteriores en un corto tiempo (horas o minutos), a diferencia de los métodos computacionales más comunes como los de Monte Carlo vía cadenas de Markov (MCMC) que pueden tardar días u horas (Rue et al., 2009). De esta forma, en esta tesis se propone extender el modelo espacial NNGP en INLA (Quiroz et al., 2021) al modelo espacio-temporal NNGP en INLA.

### 1.1. Objetivos

El objetivo general de la tesis es implementar el método de inferencia bayesiana INLA para estimar los parámetros de un modelo espacio-temporal definido a partir de un NNGP. De manera específica:

- Revisar la literatura acerca de los modelos y procesos espacio-temporales, estructuras de covarianza espaciales y espacio-temporales.
- Proponer, estudiar las propiedades, y definir el método de estimación los parámetros asociados al modelo espacio-temporal.
- Implementar el modelo espacio-temporal de inferencia bayesiana considerando el método INLA del modelo espacio-temporal.
- Realizar estudios de simulación acerca del modelo espacio-temporal.
- Ajustar el modelo espacio-temporal usando INLA al conjunto de datos de concentración de partículas de contaminación ( $PM_{10}$ ) de las estaciones de monitoreo ambiental de la Union Europea.



## Capítulo 2

### Revisión de literatura

Los modelos espacio-temporales son usados generalmente para modelar fenómenos complejos, los cuales tienen una serie de estados que van evolucionando en el espacio y el tiempo. Existen modelos determinísticos como las ecuaciones diferenciales, los cuales utilizan condiciones iniciales tanto para los estados y los coeficientes; sin embargo, esto no logra capturar la incertidumbre propia de estos fenómenos. Es por ello que se emplean los modelos probabilísticos en caso no haya información suficiente sobre el fenómeno observado (Hristopulos, 2020).

Uno de los conceptos principales para modelar estos fenómenos es el de *campo aleatorio*, el cual es una herramienta probabilística para modelar fenómenos complejos o macroscópicos que no pueden ser descritos por medio de funciones determinísticas. Los campos aleatorios nos servirán para caracterizar fenómenos espacio-temporales. A continuación se dará una serie de definiciones importantes para describir los modelos espacio-temporales.

#### 2.1. Estadística espacial

La estadística espacial es una disciplina que estudia y modela el comportamiento de un fenómeno tomando en cuenta la estructura espacial y el conjunto de locales en donde éste es observado. A diferencia de la estadística tradicional, la estadística espacial considera la existencia de la autocorrelación o asociación espacial de las observaciones, puesto que, espacialmente existirán observaciones más similares o disímiles a otras dada la distancia que exista entre ellas.

Dentro de la estadística espacial encontramos la rama de la geoestadística, desarrollada por Matheron (1963); la cual combina la estadística, métodos numéricos-computacionales y las ciencias ambientales. La geoestadística permite la generación de los datos geoestadísticos, los cuales representan observaciones o mediciones en una región espacial continua recolectadas en un conjunto de locales no aleatorios (Moraga, 2019). Por ejemplo, las mediciones de contaminación en el aire en diferentes estaciones de una ciudad.

Finalmente, la estadística espacial utiliza el concepto de campo aleatorio, el cual es una generalización de un proceso estocástico, que permite modelar fenómenos macroscópicos que incluyen efectos espaciales, temporales y espacio-temporales. En especial se tomará en cuenta los campos aleatorios gaussianos, de donde se obtendrán las realizaciones o estados para plantear un modelo estadístico; en especial, se planteará un modelo lineal geoestadístico (Banerjee et al., 2014).

### 2.1.1. Campo aleatorio

La definición de un campo aleatorio espacial se desprende de la definición de un campo aleatorio, el cual matemáticamente se define a continuación.

**Definición 1** *Es la colección de variables aleatorias indexadas, espacialmente asociadas y distribuidas tal que:*

$$Y(\mathbf{s}, \omega) = \{Y(\mathbf{s}, \omega) : \mathbf{s} \in D, \omega \in \Omega\},$$

donde  $\mathbf{s}$  es el vector de locales o índices,  $D \subset \mathbb{R}^d$  es un *espacio paramétrico continuo*; y  $\omega$  un resultado de  $\Omega$ , el *espacio muestral*. Si tomamos  $\mathbf{s}$  fijo, entonces  $Y(\mathbf{s}, \omega)$  es una variable aleatoria. En adelante, solamente se denotará por  $Y(s)$ . Si tomamos un  $\omega = \omega_0$  fijo, entonces  $Y(\mathbf{s}, \omega_0)$  es una función determinística llamada *realización* o *estado* del campo aleatorio, y será denotado por  $y(s)$ .

Para  $d \geq 2$ , nos referimos al aspecto geoespacial, al cual llamaremos *campo espacial aleatorio* (o *spatial random field*, en inglés), y si  $d = 1$  nos referimos al aspecto temporal, al cual llamaremos *proceso estocástico* (stochastic process, en inglés).

Teóricamente, un campo espacial aleatorio está definido para un conjunto infinito de variables aleatorias definidas en infinitos locales. Sin embargo, en la práctica, no contaremos con un conjunto infinito de locales, dado la dificultad para tomar las muestras. En efecto, a través de un conjunto finito de locales y variables aleatorias  $\mathbf{s} = (s_1, s_2, \dots, s_n)$  e  $Y(s_1), Y(s_2), \dots, Y(s_n)$  podemos también caracterizar a un campo aleatorio. Esto es garantizado por el teorema de existencia de Kolmogorov (Banerjee et al., 2014). Por lo tanto, sin pérdida de generalidad, la distribución conjunta finito-dimensional  $P(Y(s_1) \leq y(s_1), Y(s_2) \leq y(s_2), \dots, Y(s_n) \leq y(s_n))$  será la distribución conjunta que usaremos para la inferencia estadística espacial.

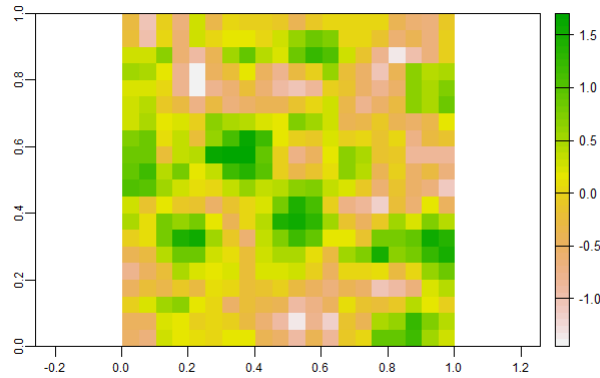
En particular, es importante notar la diferencia entre un campo aleatorio y un conjunto de observaciones aleatorias. En el primero, la dependencia genera un comportamiento o patrón dentro del campo espacial o espacio-temporal, mientras que el segundo no se establece alguna relación entre las observaciones como se muestra en la figura 2.1.

## 2.2. Campo aleatorio gaussiano

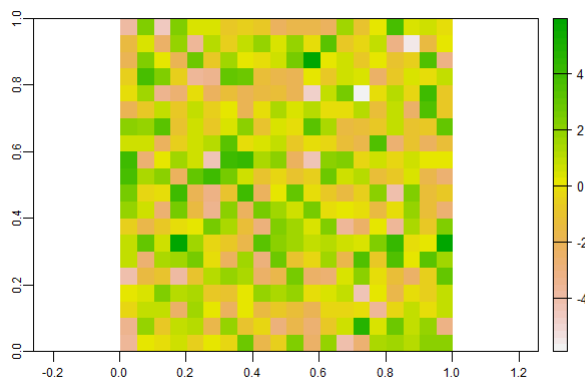
Podemos caracterizar a los procesos espaciales teniendo en cuenta diferentes distribuciones de probabilidad, en nuestro caso, usaremos la distribución gaussiana multivariada. En cuyo caso, el proceso se conoce como proceso espacial gaussiano o campo espacial gaussiano (Gaussian random field en inglés). Una definición formal de este proceso se detalla a continuación.

**Definición 2** *Un proceso espacial  $Y(s)$  se dice que es un proceso aleatorio gaussiano, si para  $n > 1$  y para un conjunto de locales  $\{s_1, s_2, \dots, s_n\}$ , la distribución conjunta de  $\mathbf{Y} = (Y(s_1), Y(s_2), \dots, Y(s_n))^T$ , sigue una distribución normal multivariada.*

Existen dos tipos de procesos espaciales importantes, uno estrictamente estacionario y otro débilmente estacionario, ambos definidos como sigue.



(a) Campo aleatorio con autocorrelación



(b) Campo aleatorio sin autocorrelación

Figura 2.1: Simulación de un campo aleatorio con dependencia espacial y sin dependencia espacial entre las observaciones.

**Definición 3** *Un proceso espacial es estrictamente estacionario (fuertemente estacionario) si dado un conjunto de  $n > 1$  locales  $s_1, s_2, \dots, s_n$  y para cualquier  $h \in \mathbb{R}^d$  la distribución de  $\{Y(s_1), Y(s_2), \dots, Y(s_n)\}$  es la misma que  $\{Y(s_1 + h), Y(s_2 + h), \dots, Y(s_n + h)\}$ . Esto es:*

$$F_{s_1, s_2, \dots, s_n}(y_1, y_2, \dots, y_n) = F_{s_1+h, s_2+h, \dots, s_n+h}(y_1, y_2, \dots, y_n),$$

donde  $F(\cdot)$  representa la función de distribución acumulada.

El enunciado anterior indica que las características del proceso o campo aleatorio espacial se mantienen invariantes a la traslación  $h$  y su distribución sigue siendo la misma (Wackernagel, n.d.) dada la traslación. Esta condición implica que todos los momentos sean invariantes a la traslación.

Sin embargo, en la práctica es muy difícil conseguir tal estacionariedad estricta, por lo que se relaja las condiciones solamente a los dos primeros momentos de  $Y(s)$ , es decir, que la media y la covarianza sean constantes. A este proceso se le conoce como un proceso débilmente estacionario, el cual definimos seguidamente.

**Definición 4** Un proceso espacial es débilmente estacionario (o estacionario de segundo orden) si la media  $\mu(s) = \mu$  es constante y la covarianza  $Cov(Y(s), Y(s+h)) = C(h)$ , para todo  $h \in \mathbb{R}^d$  donde  $s$  y  $s+h$  pertenecen a  $D$ .

Es importante notar que procesos fuertemente estacionarios implican procesos débilmente estacionarios. Excepto, cuando el proceso espacial tiene una distribución normal multivariada, dado que un proceso estacionario débil implica uno fuerte y viceversa (Banerjee et al., 2014).

Un tercer tipo de estacionariedad es la *estacionariedad intrínseca*, bajo la cual asumimos las propiedades de la estacionariedad débil y queda definida tomando en cuenta solamente la diferencia entre dos locales  $s_i$  y  $s_j$  separados por un vector de separación  $h$  (*lag* en inglés) (Wackernagel, n.d.).

**Definición 5** Un proceso se dice que es intrínsecamente estacionario si además de tener una media constante  $\mu(s) = \mu$ , este también satisface:

$$E(Y(s+h) - Y(s))^2 = Var(Y(s+h) - Y(s)) = 2\gamma(h), \quad (2.1)$$

donde la función  $2\gamma(h)$  es conocida como *variograma* y  $\gamma(h)$  como *semivariograma*.

El semivariograma  $\gamma(h)$  es una función que mide la variabilidad de una variable aleatoria en dos locales distintos. Dicha función depende de una distancia  $d = \|h\|$ . De la ecuación 2.1 notamos que mientras  $h$  sea más pequeño,  $Y(s+h)$  y  $Y(s)$  son similares, es decir la variabilidad entre estas dos observaciones es pequeña. Mientras que si  $h$  es muy grande, existirá una mayor diferencia entre  $Y(s+h)$  y  $Y(s)$ , y por ende una mayor variabilidad.

### 2.3. La función de covarianza y el variograma

Por otro lado, se puede establecer una relación entre el variograma y la covarianza, mediante  $\gamma(h) = C(0) - C(h)$  (Banerjee et al., 2014). Existen diferentes formas paramétricas de modelar el variograma, por ejemplo de forma lineal, exponencial, Matérn, gaussiana, entre otras. Todas ellas usan a los siguientes parámetros principales: el *efecto pepita*  $\tau^2$  (o *nugget effect* en inglés) que representa el error de medición espacial o variabilidad intrínseca del proceso, la *meseta parcial* (partial sill)  $\sigma_s^2$  que representa la variabilidad espacial neta del proceso, la *meseta* (o sill) que se denota como la suma del efecto pepita y la variabilidad neta espacial  $\tau^2 + \sigma_s^2$ , además, representa el valor donde la semivarianza alcanza un valor casi constante (Chiles and Delfiner, 2012), y finalmente, el *rango*  $\phi$  (o *range* en inglés) que representa la distancia donde el semivariograma  $\gamma(d)$  alcanza la meseta (Banerjee et al., 2014).

Si el variograma o semivariograma depende solamente de la distancia  $d$ , entonces decimos que el variogram es *isotrópico* (Banerjee et al., 2014). Existen diferentes tipos de (semi) variogramas isotrópicos teóricos que ayudan a modelar de forma sencilla a la variabilidad espacial que depende de  $d$ , usando los parámetros espaciales anteriormente mencionados. Uno de los modelos de (semi)variograma mayormente usados es el llamado variograma *exponencial* (Banerjee et al., 2014), el cual está definido por:

$$\gamma(d) = \begin{cases} \tau^2 + \sigma_s^2(1 - \exp(-\phi d)) & \text{if } d > 0 \\ \tau^2 + \sigma_s^2 & d = 0. \end{cases} \quad (2.2)$$

El variograma exponencial está asociado a una función de covarianza exponencial,  $C(d)$ , de la forma:

$$C(d) = \begin{cases} \sigma_s^2 \exp(-\phi d) & \text{if } d > 0 \\ \tau^2 + \sigma_s^2 & d = 0 \end{cases} \quad (2.3)$$

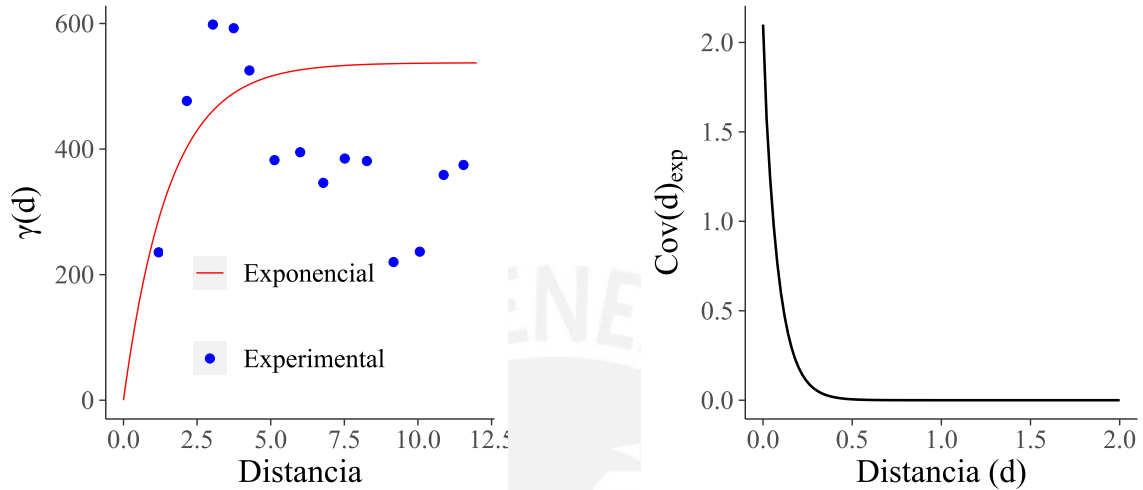


Figura 2.2: Variograma (izquierda) y función de covarianza (derecha) del modelo exponencial.

Por ejemplo, en la figura 2.2 se muestra a la izquierda (semi)variograma experimental (puntos de azul) y teórico (curva roja); y a la derecha la función de covarianza de un modelo exponencial bajo los parámetros mencionados en la ecuación (2.2). En capítulos posteriores la función de covarianza  $C(d)$  también será denotada por  $C(d|\boldsymbol{\theta})$ , con  $\boldsymbol{\theta}$  el vector parámetros de la función de covarianza. En el caso de la función de covarianza exponencial 2.3, el vector de parámetros esta compuesto de  $\boldsymbol{\theta} = (\tau^2, \phi, \sigma_s)$ .

### 2.3.1. Modelo lineal geostatístico gaussiano

La realización  $\mathbf{Y} = (Y(s_1), Y(s_2), \dots, Y(s_n))^T$  representa al conjunto de v.a.s que son medibles directamente a través del campo gaussiano aleatorio. Las mediciones observables de este campo aleatorio son el resultado de la suma de los efectos de: las covariables, del efecto de un proceso espacial gaussiano latente  $W(s)$ , cuyas realizaciones capturan la correlación espacial pero no son observables directamente; y finalmente, del efecto de un error de medición, que representa el error aleatorio o no-espacial.

Formalmente, para un conjunto de locales  $s_1, s_2, \dots, s_n$ , el modelo lineal geoestadístico es definido por:

$$Y(s_i) = \mu(s_i) + w(s_i) + \epsilon_i, \quad (2.4)$$

el cual también es equivalente a:

$$\begin{aligned} Y(s_i) &= X^\top(s_i)\beta + w(s_i) + \epsilon_i \\ &= \sum_{j=1}^k x_j(s_i)\beta_j + w(s_i) + \epsilon_i, \end{aligned}$$

donde  $w = (w(s_1), w(s_2), \dots, w(s_n))^\top \sim \mathcal{N}(0, C(\theta))$  es una realización del campo aleatorio espacial latente  $W(s)$ ,  $\mu(s_i) = X^\top(s_i)\beta$  es la media del modelo con  $k$ -covariables  $X(s_i)$ , y  $k$ -coeficientes de regresión  $\beta$ , y los errores  $\epsilon_i$  son v.a.s. independientes e idénticamente distribuidas y no correlacionadas con el efecto espacial, cuya distribución es normal,  $\epsilon_i \sim \mathcal{N}(0, \tau^2)$ .

## 2.4. Procesos temporales

**Definición 6** *Un proceso temporal o proceso estocástico es la colección de variables aleatorias ordenadas o indexadas en el tiempo  $t$ . Formalmente se define como:*

$$\{Y(\omega, t) : t \in D, \omega \in \Omega\},$$

donde  $t$  es el índice temporal y  $D$  es llamado el espacio paramétrico.

Asumiremos que el conjunto de  $D$  es un conjunto numerable discreto de la forma  $D = \{0, 1, 2, 3, \dots\}$ . Las variables aleatorias que forman parte del proceso temporal serán denotadas como  $Y_t$ , y las realizaciones (o observaciones) de las variables aleatorias serán denotadas como  $y_t$ .

### 2.4.1. Modelo temporal autoregresivo de orden 1 - AR(1)

Sea  $\mathbf{Y} = (Y_1, Y_2, Y_3, \dots, Y_n)^\top$  un conjunto de variables aleatorias indexadas en el tiempo. Un modelo autoregresivo asume que las v.a.s. de  $\mathbf{Y}$  están autocorrelacionadas en el tiempo. Un modelo AR de orden 1 se define por:

$$Y_t = aY_{t-1} + \epsilon_t, t = 2, 3, \dots, n \quad (2.5)$$

donde  $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$ ,  $|a| < 1$  es un parámetro de autocorrelación, y se asume que  $Y_1 \sim \mathcal{N}\left(0, \frac{\sigma_\epsilon^2}{1-a^2}\right)$ . En este modelo  $Y_t$  es explicada por  $Y_{t-1}$ . Además es posible definir una distribución condicional basada en los modelos AR(1), es decir, la distribución de  $Y_t$  dado  $Y_{t-1}$ :

$$Y_t|Y_{t-1} \sim \mathcal{N}(aY_{t-1}, \sigma_\epsilon^2).$$

Si tomamos la covarianza de la variable aleatoria temporal  $Y$  en dos tiempos distintos,  $t_j$  y  $t_{j'}$ , la covarianza entre  $Y_{t_j}$  y  $Y_{t_{j'}}$ , se define como:

$$Cov(Y_{t_j}, Y_{t_{j'}}) = a^{|t_j - t_{j'}|} \frac{\sigma_\epsilon^2}{1 - a^2}.$$

Así, podemos definir la ecuación(2.5) en forma multivariada por,

$$\mathbf{Y} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{T}}),$$

donde  $\boldsymbol{\Sigma}_{\mathbf{T}}$ , la matriz de covarianza, se define como  $\Sigma_{\mathbf{T}}(i, j) = a^{|t_i - t_j|} \frac{\sigma_{\epsilon}^2}{1 - a^2}$ .

Una vez definidos los conceptos de procesos temporales y espaciales por separado, podemos extender estos conceptos a los procesos espacio-temporales.

## 2.5. Procesos espacio-temporales

En la práctica el proceso espacio-temporal sirve para estudiar la distribución espacial de la variable en estudio en espacios continuos a través del tiempo. En esta sección se define formalmente un proceso espacio-temporal y las propiedades que requieren para ser válidos.

**Definición 7** *Un proceso estocástico espacio-temporal*

$$\{Y(s, t) : s \in D, t \in T\},$$

*es definido como un conjunto de variables aleatorias generalmente dependientes, para cada localización espacial  $s \in D$  y cada tiempo  $t \in T$ , donde el índice espacial varía en  $D \subset \mathbb{R}^d$  y el índice temporal en  $T \in \mathbb{R}$  o  $\mathbb{Z}$ . Sean los locales  $\mathbf{s}_i : i = 1, 2, 3, \dots, n$  y sean los tiempos  $t_j : j = 1, 2, 3, \dots, T$ , entonces el conjunto de variables  $Y(\mathbf{s}_i, t_j)$  representa una realización del proceso espacio temporal. Por lo tanto  $(\mathbf{s}_i, t_j)$  varía conjuntamente en  $D \times T$ .*

De la definición asumimos que la localización  $\mathbf{s}$  varía de forma continua sobre  $\mathbb{R}^d$ , mientras que el tiempo  $\mathbf{t}$  varía de forma discreta sobre  $\mathbb{Z}^+$ . Por ejemplo en la figura 2.3 se muestra una realización de un proceso espacio-temporal en  $\mathbb{R}^d \times T$ , donde se observa que los datos son observados en la misma ubicación  $\mathbf{s}$  con respecto al tiempo  $\mathbf{t}$ .

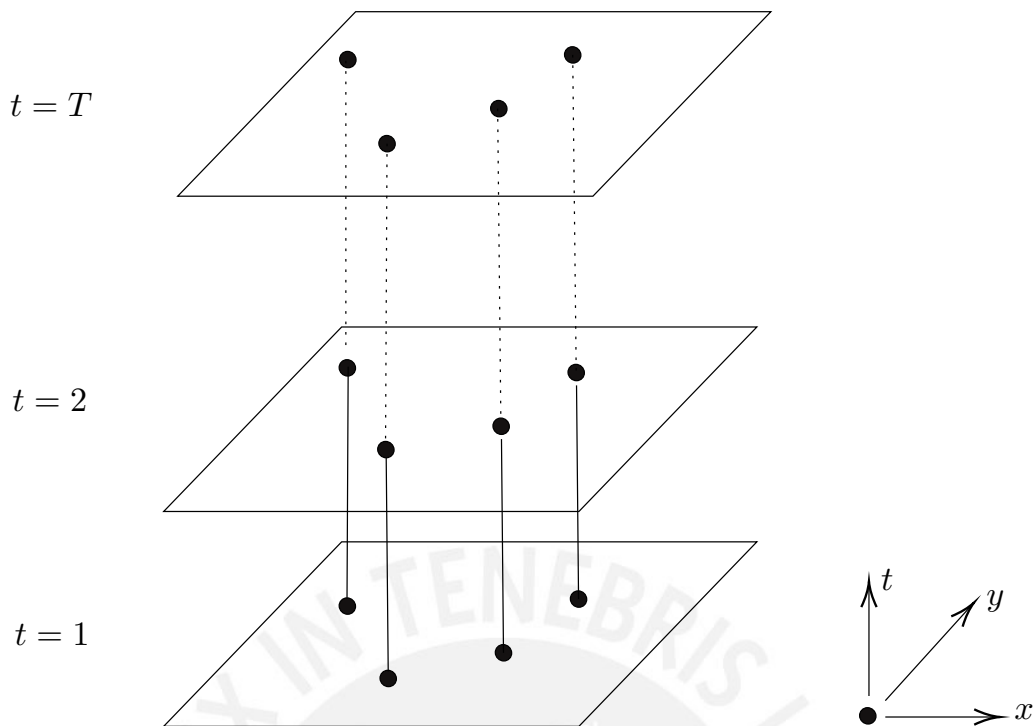


Figura 2.3: Realización de un proceso espacio-temporal con los locales  $s = (x, y)$  y el tiempo en forma discreto  $t = \{t_1, t_2, \dots, T\}$ .

Además, podemos caracterizar al proceso espacio-temporal través de una función de media y una función de covarianza:

$$\mu(s, t) = E(Y(s, t))$$

$$Cov(Y(s_i, t_j), Y(s_{i'}, t_{j'})) = C((s_i, t_j), (s_{i'}, t_{j'})).$$

Es importante resaltar que para que una función de covarianza sea válida, una condición necesaria y suficiente es que sea simétrica  $C(Y(s_i, t_j), Y(s_{i'}, t_{j'})) = C(Y(s_{i'}, t_{j'}), Y(s_i, t_j))$  y que sea definida positiva, esto es,

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j C(Y(s_i, t_i), Y(s_j, t_j)) > 0,$$

para cualquier  $n \in \mathbb{N}$ , y para cualesquiera  $(s_i, t_i) \in \mathbb{R}^2 \times \mathbb{Z}^+$  y  $a_i \in \mathbb{R}$ ,  $i = 1, \dots, n$ .

Otra extensión importante, es que los procesos espacio-temporales también pueden definirse como procesos gaussianos.

**Definición 8** El proceso espacio-temporal  $Y(s, t)$  es un proceso gaussiano si para cualquier conjunto de localizaciones espacio-temporales  $(s_1, t_1), \dots, (s_n, t_n)$ , el vector aleatorio  $Y = (Y(s_1, t_1), \dots, Y(s_n, t_n))^T$  sigue una distribución normal multivariante con media  $\mu$  y covarianza  $C(\cdot)$ .

Por ejemplo la figura 2.4 muestra una simulación de un proceso espacio-temporal gaussiano usando un modelo autoregresivo de orden uno en cuatro años.



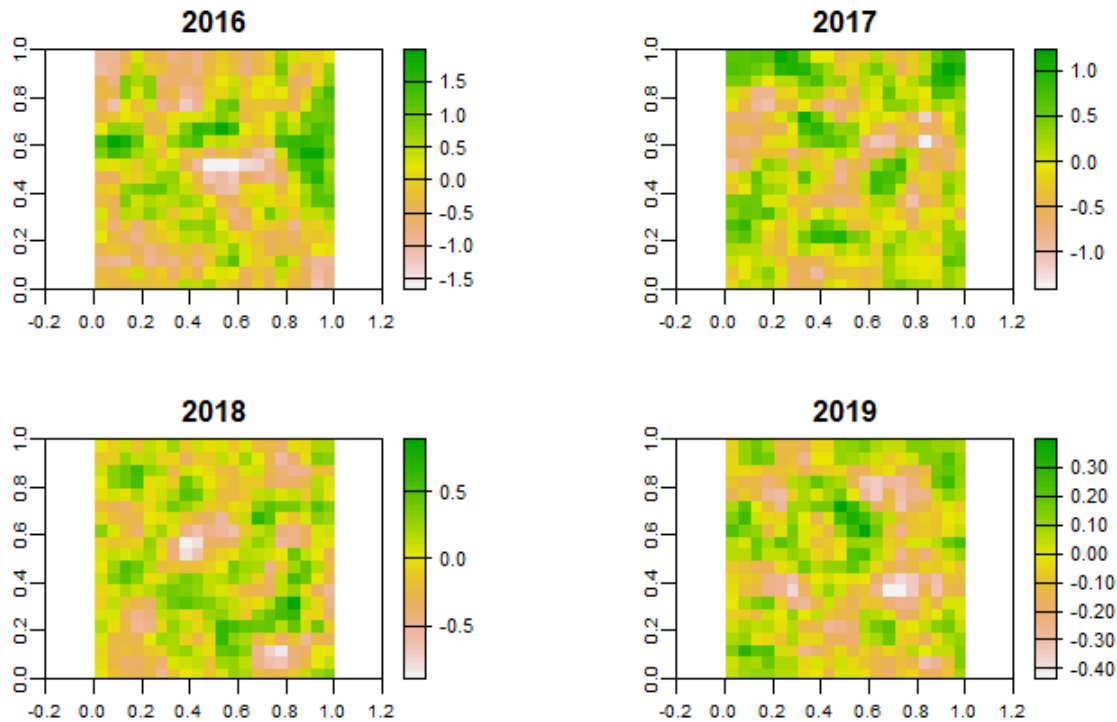


Figura 2.4: Simulación de un campo aleatorio espacio-temporal en cuatro tiempos (discretos) en una región (espacio continuo).

El concepto de proceso débilmente estacionario o de segundo orden en procesos espaciales también se extiende a los procesos espacio-temporales.

**Definición 9** Se dice que un proceso espacio-temporal es débilmente estacionario (o de segundo orden) si para cualquier localidad,  $\mathbf{s}$ , y tiempo,  $\mathbf{t}$ , se tiene que la media  $E(Y(\mathbf{s}, \mathbf{t})) = \mu$  es constante, y la covarianza  $Cov(Y(\mathbf{s} + \mathbf{h}, \mathbf{t} + u), Y(\mathbf{s}, \mathbf{t})) =: Cov(\mathbf{h}, u)$ , para todo vector de distancia geoespacial  $\mathbf{h} \in \mathbb{R}^d$  y tiempo  $t \in \mathbb{Z}^+$ .

De forma similar, la definición de procesos espaciales intrínsecamente estacionarios se generaliza para procesos espacio-temporales. En este caso se asume una estacionariedad de segundo orden y nos enfocamos en la variación tanto espacial como temporal en dos puntos  $(\mathbf{s}, \mathbf{t})$  y  $(\mathbf{s} + \mathbf{h}, \mathbf{t} + u)$  respectivamente.

**Definición 10** Se dice que un proceso espacio-temporal es intrínsecamente estacionario, si además de ser de segundo orden y tener media constante, satisface

$$E(Y(\mathbf{s} + \mathbf{h}, \mathbf{t} + u), Y(\mathbf{s}, \mathbf{t}))^2 = V(Y(\mathbf{s} + \mathbf{h}, \mathbf{t} + u) - Y(\mathbf{s}, \mathbf{t})) = 2\gamma(\mathbf{h}, u);$$

donde  $2\gamma(h, u)$  y  $\gamma(h, u)$  son el variograma y semivariograma espacio-temporal respectivamente. El variograma espacio-temporal, depende de  $\|\mathbf{h}\|$  y  $|t|$ , la distancia entre los locales y la diferencia temporal de forma respectiva. También mantiene la relación entre el variograma y la función de covarianza  $\gamma(\mathbf{h}, u) = C(0, 0) - C(\mathbf{h}, u)$ .

Finalmente, un proceso espacio-temporal  $Y(s, t)$  tiene función de covarianza separable o no separable.

**Definición 11** *La función de covarianza de un proceso espacio-temporal es llamada separable si existe una función de covarianza puramente espacial  $C_s$  y una función de covarianza puramente temporal  $C_t$  tales que  $C((s_i, t_j), (s_{i'}, t_{j'})) = C_s(s_i, s_{i'})C_t(t_j, t_{j'})$ , para cualquier par de localizaciones  $(s_i, t_j)$  y  $(s_{j'}, t_{j'}) \in \mathbb{R}^d \times \mathbb{R}$ .*

Esta función de covarianza es muy usada debido a que simplifica la inferencia y estimación de parámetros en el modelo espacio-temporal, desde un punto de vista computacional.

## 2.6. Inferencia bayesiana

La inferencia estadística busca inferir sobre los parámetros  $\theta$  de una población a través de una muestra representativa  $\mathbf{y}$  (Wasserman, 2004). En el caso de la inferencia bayesiana, se introduce un conocimiento previo de los parámetros  $\theta$  a través de una distribución a priori  $p(\theta)$ . Después de observar los datos, obtenemos la distribución muestral o verosimilitud  $p(\mathbf{y}|\theta)$ , la cual se actualiza a través de nuestro conocimiento a priori del parámetro obteniéndose una distribución a posteriori  $p(\theta|\mathbf{y})$  (Carlin et al., 2009). Esta actualización se calcula a través del teorema de bayes:

$$p(\theta|\mathbf{y}) = \frac{p(\theta)p(\mathbf{y}|\theta)}{p(\mathbf{y})}. \quad (2.6)$$

Recordemos que por probabilidad básica  $p(\mathbf{y}) = \int f(\mathbf{y}|\theta)p(\theta)d\theta$ , por lo tanto en la ecuación (2.6) reemplazamos el denominador, obteniéndose que:

$$p(\theta|\mathbf{y}) = \frac{p(\theta)p(\mathbf{y}|\theta)}{\int f(\mathbf{y}|\theta)p(\theta)d\theta}. \quad (2.7)$$

Como  $p(\theta|\mathbf{y})$  no depende del parámetro  $\theta$ , podemos escribir la ecuación (2.7) como:

$$p(\theta|\mathbf{y}) \propto p(\theta)p(\mathbf{y}|\theta). \quad (2.8)$$

Al denominador  $c = \frac{1}{p(\mathbf{y})}$  de la ecuación (2.8) también se le conoce como constante normalizadora (Wasserman, 2004).

Finalmente las ecuaciones (2.7) y (2.8) son la base para la inferencia bayesiana, la cual nos permitirá obtener la distribución de los parámetros y a partir de ella la media a posteriori, la varianza a posteriori, la moda a posteriori, los intervalos de credibilidad, cuantiles a posteriori, entre otras medidas resumen. La obtención de la distribución a posteriori es una de las principales ventajas que se obtiene a través de la inferencia bayesiana, dado que permite trabajar con modelos complejos (jerárquicos o anidados) y que poseen una gran cantidad de parámetros (Gelman et al., 2014).

### 2.6.1. Estadística computacional para la inferencia Bayesiana

La inferencia bayesiana conlleva el cálculo de integrales, como por ejemplo las de la ecuación 2.7. Del mismo modo, podemos observar el cálculo de integrales para estimar la media

o varianza. El cálculo de integrales dentro de la estadística computacional suele obtenerse de dos formas: la más usada involucra identificar una serie de estimaciones de la integral que convergen al valor real, por ejemplo Monte Carlo vía cadenas de Markov (MCMC) Peng (2012). La segunda forma consiste en identificar funciones alternativas (o similares) que mejor coincida a la función original de la integral que deseamos aproximar. Una de estas técnicas es la aproximación de Laplace Peng (2012). En este trabajo nos enfocaremos en el segundo caso, en especial, la aproximación de Laplace.

### 2.6.2. Aproximación de Laplace

El método de Laplace consiste en hallar una función de forma exponencial que aproxime a la función original que deseamos integrar, evaluada en su valor máximo. En general, el término aproximación de Laplace es usado para referirse al método de aproximar la distribución a posteriori a través de una distribución gaussiana centrada en la moda (máximo a posteriori). El costo computacional de la aproximación de Laplace tiende a ser relativamente bajo y bastante rápido (Tierney and Kadane, 1986; Rue et al., 2009).

En particular, supongamos que tenemos la función  $g(\theta)$  es la función de densidad de una variable aleatoria, tal que se desea calcular:

$$\int g(\theta)d\theta. \quad (2.9)$$

Sea  $h(\theta) = \log(g(\theta))$ , luego la ecuación (2.9) puede escribirse como

$$\int g(\theta)d\theta = \int \exp(h(\theta))d\theta. \quad (2.10)$$

Usando la serie de Taylor de segundo orden para  $h(\theta)$  alrededor de  $\theta_0$ , se define

$$h(\theta) = h(\theta_0) + h'(\theta_0)(\theta - \theta_0) + \frac{1}{2}h''(\theta_0)(\theta - \theta_0)^2. \quad (2.11)$$

Reemplazando la ecuación (2.11) en la integral de la ecuación (2.10), se obtiene que:

$$\int \exp(h(\theta))d\theta = \int \exp\left(h(\theta_0) + h'(\theta_0)(\theta - \theta_0) + \frac{1}{2}h''(\theta_0)(\theta - \theta_0)^2\right)d\theta.$$

Si  $\theta_0$  es igual a la moda  $\theta^* = \operatorname{argmax}_{\theta} h(\theta)$ , entonces éste alcanza su valor máximo, tal que la derivada  $h'(\theta_0) = h'(\theta^*) = 0$ , por lo tanto la aproximación será:

$$\int \exp(h(\theta))d\theta = \int \exp\left(h(\theta^*) + \frac{1}{2}h''(\theta^*)(\theta - \theta^*)^2\right)d\theta.$$

Dado que  $h(\theta^*)$  no depende de  $\theta$ , se puede simplificar y reducir la integral a:

$$\int \exp(h(\theta))d\theta = \exp(h(\theta^*)) \int \exp\left(-\frac{1}{2} \left[ \frac{(\theta - \theta^*)^2}{-h''(\theta^*)^{-1}} \right]\right)d\theta. \quad (2.12)$$

Se observa dentro de la integral en la ecuación (2.12) un núcleo similar al de una distribución normal, con media  $\theta^*$  y varianza  $-h''(\theta^*)^{-1}$ . Luego, reescribiendo en términos de

$g$ :

$$\int g(\theta)d\theta = g(\theta^*) \int \exp \left( -\frac{1}{2} \frac{(\theta - \theta^*)^2}{-\left(\frac{d^2 \log(g(\theta))}{d\theta^2} \Big|_{\theta=\theta^*}\right)^{-1}} \right) d\theta.$$

Se observa que la cantidad al interior de la integral es proporcional al núcleo de la función de densidad de la distribución normal, tal que

$$\tilde{g}(\theta) \sim \mathcal{N} \left( \theta^*, -\left(\frac{d^2 \log(g(\theta))}{d\theta^2} \Big|_{\theta=\theta^*}\right)^{-1} \right).$$

Para hallar la constante normalizadora de  $g(\theta)$ , se define  $\Phi(\cdot|\cdot)$  como la función de distribución acumulada condicional de  $\mathcal{N} \left( \theta^*, -\left(\frac{d^2 \log(g(\theta))}{d\theta^2} \Big|_{\theta=\theta^*}\right)^{-1} \right)$  evaluada en el intervalo  $[a, b]$ ,

$$\int_a^b g(\theta)d\theta = g(\theta^*) \sqrt{\frac{2\pi}{-\left(\frac{d^2 \log(g(\theta))}{d\theta^2} \Big|_{\theta=\theta^*}\right)^{-1}}} [\Phi(b|\cdot) - \Phi(a|\cdot)].$$

Si  $a = -\infty$  y  $b = \infty$ , la expresión entre los corchetes es igual a 1, por lo tanto la constante de normalización es:

$$\int_a^b g(\theta)d\theta = g(\theta^*) \sqrt{\frac{2\pi}{-\left(\frac{d^2 \log(g(\theta))}{d\theta^2} \Big|_{\theta=\theta^*}\right)^{-1}}}. \quad (2.13)$$

Por lo tanto la aproximación de Laplace de  $g(\theta)$  es aproximada con una función de distribución normal con media igual a la moda  $\theta^*$  y varianza  $-\left(\frac{d^2 \log(g(\theta))}{d\theta^2} \Big|_{\theta=\theta^*}\right)^{-1}$ , y cuya constante de normalización es definida en la ecuación (2.13).

Por otro lado, si usamos la aproximación de Laplace para hallar la distribución a posteriori  $\theta|y$ , tomamos  $g(\theta) = p(\theta)p(y|\theta)$ , la distribución condicional que depende solamente de  $\theta$  con  $y$  fijo; y además,  $p(y) = \int g(\theta) = \int p(\theta)p(y|\theta)d\theta$  la constante normalizadora, entonces

$$p(\theta|y) = \frac{1}{\sqrt{\frac{2\pi}{-\left(\frac{d^2 \log(g(\theta))}{d\theta^2} \Big|_{\theta=\theta^*}\right)^{-1}}}} \exp \left( -\frac{1}{2} \frac{(\theta - \theta^*)^2}{-\left(\frac{d^2 \log(g(\theta))}{d\theta^2} \Big|_{\theta=\theta^*}\right)^{-1}} \right)$$

$$p(\theta|y) = \frac{1}{\sqrt{\frac{1}{-\left(\frac{d^2 \log(g(\theta))}{d\theta^2}\right)\bigg|_{\theta=\theta^*}}} \sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(\theta - \theta^*)^2}{-\left(\frac{d^2 \log(g(\theta))}{d\theta^2}\right)\bigg|_{\theta=\theta^*}^{-1}}\right).$$

Habiendo hallado la función de densidad a posteriori, la distribución a posteriori  $\theta|y$  usando la aproximación de Laplace es normal, específicamente:

$$\theta|y \sim \mathcal{N}\left(\theta; \theta^*, -\left(\frac{d^2 \log(g(\theta))}{d\theta^2}\right)\bigg|_{\theta=\theta^*}^{-1}\right). \quad (2.14)$$

En el caso multivariado, la expresión anterior (ecuación 2.14) se puede reescribir como:

$$\boldsymbol{\theta}|y \sim \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\theta}^*, \mathbf{H}^{-1}),$$

donde  $\mathbf{H}^{-1} = -\nabla\nabla\log(g(\boldsymbol{\theta}))\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$  es la matriz Hessiana. Luego, la constante normalizadora multivariada es:

$$p(y) = \int g(\boldsymbol{\theta}) = g(\boldsymbol{\theta}^*) |2\pi\mathbf{H}^{-1}|^{1/2},$$

donde  $|\cdot|$  la determinante de la matriz.

## Capítulo 3

# Modelo espacio-temporal usando NNGP

Para definir el modelo espacio temporal, se define  $D_s$  como un dominio espacial continuo en  $\mathbb{R}^2$  y  $D_t$  como un dominio temporal discreto finito en  $\mathbb{N}$ . Sea  $Y(s_i, t)$  la variable respuesta e  $y(s_i, t)$  su valor observado en la localidad  $s_i \in D_s$  y tiempo  $t \in D_t$ , donde  $i = 1, \dots, N$  y  $t = 1, \dots, T$ . Luego el modelo espacio-temporal gaussiano se puede definir por:

$$Y(s_i, t) = Z_i^\top \boldsymbol{\beta} + f_{st}(s_i, t) + \epsilon(s_i, t), \quad (3.1)$$

donde  $Z_i$  es un vector de covariables,  $\boldsymbol{\beta}$  es un vector de coeficientes de regresión,  $\epsilon(s_i, t) \sim \mathcal{N}(0, \sigma_\epsilon^2)$  es un efecto aleatorio gaussiano no estructurado con  $\sigma_\epsilon^2$  igual a la varianza del error (o efecto pepita) y  $f_{st}(\cdot, \cdot)$  es un efecto aleatorio espacio-temporal. Este último se escribe como la suma de dos efectos, uno autoregresivo en el tiempo AR(1) y un efecto aleatorio espacial  $w$ , tal que:

$$f_{st}(s_i, t) = a f_{st}(s_i, t-1) + w(s_i), \quad (3.2)$$

donde el coeficiente  $|a| < 1$ ,  $w(s_i) \perp f_{st}(s_i, 1)$  ( $\perp$ : independiente de), y  $w(s_i)$  es independiente del tiempo.

Generalmente los procesos espacio-temporales involucran grandes conjuntos de datos, bajo los cuales el cálculo de la inversa y determinante de la matriz de covarianza  $\mathbf{C}(\cdot, \cdot | \boldsymbol{\theta})$  asociada a  $f_{st}(s, t)$  y con  $\boldsymbol{\theta}$  el vector de parámetros de la función de covarianza, resultan difíciles de calcular debido a su alto costo computacional. Es por ello que se propone aproximar el proceso espacial usando el proceso gaussiano de vecinos más cercanos como se detallará en la sección siguiente.

### 3.1. Proceso espacio - temporal gaussiano de vecinos más cercanos

En vez de calcular  $\mathbf{C}_w(\cdot, \cdot | \boldsymbol{\theta})^{-1}$ , matriz de precisión (la inversa de la matriz de covarianza) del proceso espacial, se propone usar una aproximación del proceso espacial mediante procesos gaussianos de vecinos más cercanos (NNGP - Nearest Neighbor Gaussian Process en inglés), proceso espacial propuesto por Datta et al. (2016). Este proceso tiene una matriz de precisión llena de ceros (dispersa), lo cual permite realizar los cálculos a un costo y complejidad computacional de orden más bajo.

Supongamos que se define un proceso espacio-temporal gaussiano (Gaussian Process-GP en inglés), tal que

$$\mathbf{w}(s) \sim GP(\mathbf{0}, \mathbf{C}_w(\cdot, \cdot | \boldsymbol{\theta})) \quad (3.3)$$

el cual tiene una media centrada en cero y una función de covarianza  $C_w(\cdot, \cdot | \boldsymbol{\theta})$ , donde  $\boldsymbol{\theta}$  denota el vector de parámetros de la función de covarianza del proceso gaussiano. Por ejemplo se puede usar la función de covarianza exponencial  $C_w(d | \boldsymbol{\theta}) = \sigma_s^2 \exp(-\phi d)$  (ver ecuación 2.3). Para derivar el NNGP se define  $\mathcal{S} = \{s_1, s_2, \dots, s_n\}$  llamado el *conjunto de referencia* y sea  $\mathbf{w}_{\mathcal{S}}(s) = (w(s_1), \dots, w(s_k))^{\top}$ . De la ecuación (3.3) podemos decir que  $\mathbf{w}_{\mathcal{S}}(s) \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{w}_{\mathcal{S}}}(\boldsymbol{\theta}))$  con matriz de covarianza  $\mathbf{C}_{\mathbf{w}_{\mathcal{S}}}(\boldsymbol{\theta})$ .

Se puede expresar la función de probabilidad o densidad (fdp) conjunta de  $\mathbf{w}_{\mathcal{S}}$  como el producto de fdp condicionales, tal que:

$$p(\mathbf{w}_{\mathcal{S}}) = p(w(s_1))p(w(s_2)|p(w(s_1))) \dots p(w(s_n)|p(w(s_{n-1}))), \dots, p(w(s_1))). \quad (3.4)$$

A partir de la ecuación (3.4), se tiene un gran cantidad de fdp condicionales las cuales se propuso reemplazar por un conjunto más pequeño de tamaño  $m \ll n$ . Así, para cada  $s_i \in \mathcal{S}$ , un conjunto condicionado  $N(s_i) \subset \mathcal{S} \setminus \{s_i\}$  (en un tiempo  $t$  fijo), es usado para definir una aproximación de  $p(\mathbf{w}_{\mathcal{S}})$  de la forma siguiente:

$$\tilde{p}(\mathbf{w}_{\mathcal{S}}) = \prod_{i=1}^n p(w(s_i) | w_{N(s_i)}), \quad (3.5)$$

donde  $\mathbf{w}_{N(s_i)}$  es un vector apilado de las realizaciones de  $\mathbf{w}_{\mathcal{S}}$  sobre  $N(s_i)$ , y  $N_{\mathcal{S}} = \{N(s_i); i = 1, 2, \dots, n\}$ , la colección de todos los (sub)conjuntos condicionados sobre  $\mathcal{S}$ . Podemos ver al par  $\{\mathcal{S}, N_{\mathcal{S}}\}$  como un grafo digirido  $\mathcal{G}$  donde  $\mathcal{S}$  es el conjunto de nodos y  $N_{\mathcal{S}}$  es el conjunto aristas directas. Para entender mejor esta definición note que si el grafo  $\mathcal{G}$  presenta ciclos se le conoce como *grafo cíclico digirido* (GCD), pero si no presenta ciclos, se le conoce como *grafo acíclico digirido* (GAD). Por ejemplo este tipo de grafos se muestra en la figura 3.1. Para definir una fdp válida se asume que  $\mathcal{G}$  es un GAD. Luego por cada dos nodos (localidades)  $s_i$  y  $s_j$  en un tiempo  $t$  fijo, decimos que  $s_j$  es un *vecino cercano* de  $s_i$ , si existe una arista que los conecte directamente. Por lo tanto  $N(s_i)$  denota el conjunto de vecinos directos (o solamente vecinos) de  $s_i$  en un tiempo  $t$  fijo.

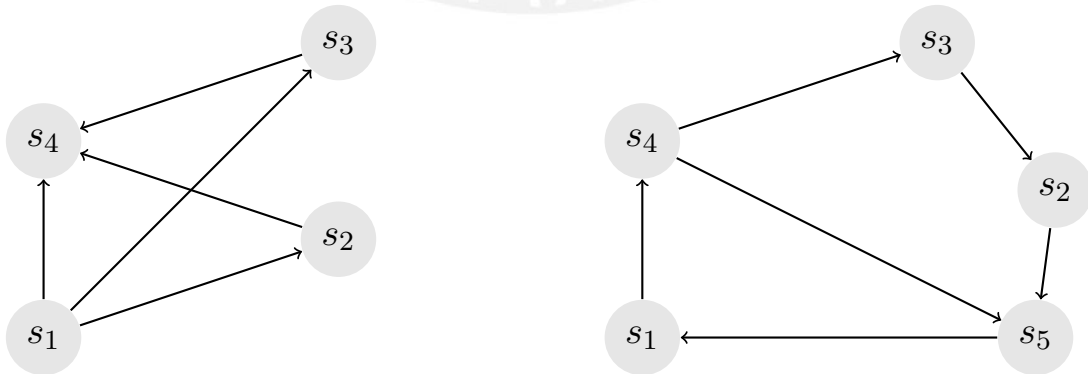


Figura 3.1: Izquierda: Grafo acíclico. Derecha: Grafo cíclico

Existe una relación de los GAD y las distribuciones de probabilidad, en este caso, como  $\mathcal{G}$  es un GAD y  $p(\mathbf{w}_{\mathcal{S}}(s_i))$  es una función de densidad, entonces  $\tilde{p}(\mathbf{w}_{\mathcal{S}})$  también es una función

de densidad que se deriva de  $\mathcal{G}$  (Lauritzen, 1996).

Denotamos por  $\mathbf{C}_{N(s_i)}$  a la covarianza de  $\mathbf{w}_{N(s_i)}$ , y  $\mathbf{C}_{s_i, N(s_i)}$  a la matriz de covarianza entre los vectores aleatorios  $\mathbf{w}_{N(s_i)}$  y  $\mathbf{w}_s(s_i)$  es decir, la covarianza entre los vecinos cercanos a  $s_i$ . Entonces, podemos reescribir la ecuación 3.4 como:

$$\tilde{p}(\mathbf{w}_S) = \prod_{i=1}^n \mathcal{N}(w_s(s_i) | B_{s_i} w_{N(s_i)}, F_{s_i}),$$

donde  $\mathbf{B}_{(s_i)} = \mathbf{C}_{s_i, N(s_i)} \mathbf{C}_N^{-1}(s_i)$  y  $\mathbf{F}_{s_i} = \mathbf{C}(s_i, s_i) - \mathbf{C}_{s_i, N(s_i)} \mathbf{C}_N^{-1}(s_i) \mathbf{C}_{N(s_i), s_i}$ .

Por lo tanto, se observa que  $\tilde{p}(\mathbf{w}_S)$  es una función de densidad multivariada gaussiana con matriz de covarianza  $\tilde{\mathbf{C}}_{\mathbf{w}_S} = \mathbf{B}_S^T \mathbf{F}_S^{-1} \mathbf{B}_S$ . Luego, para facilitar el computo de la matriz de covarianza, vamos a trabajar con la matriz dispersa donde  $\tilde{\mathbf{Q}}_{\mathbf{w}_S} = \tilde{\mathbf{C}}_{\mathbf{w}_S}^{-1}$ . Con todo lo expuesto anteriormente podemos construir un conjunto  $N(s_i)$  que representan los *vecinos más cercanos*  $s_i$ , y además  $\mathbf{w}_S$  y  $\tilde{p}(\mathbf{w}_S)$  será la *densidad de los vecinos más cercanos* de  $\mathbf{w}_S$ . Así, podemos aproximar la distribución de  $\mathbf{w}_S$  a través de un nuevo proceso espacial:

$$\tilde{\mathbf{w}}_S \sim NNGP(\mathbf{0}, \tilde{\mathbf{Q}}_{\mathbf{w}_S}^{-1}), \quad (3.6)$$

el cual tiene una matriz de precisión dispersa y por lo tanto es un campo aleatorio markoviano gaussiano.

Por otro lado, a partir de la aproximación del proceso espacial  $\tilde{\mathbf{w}}_S$  definido en la ecuación ecuación (3.6) se redefine las ecuaciones (3.1) y (3.2) de la siguiente forma:

$$Y(s_i, t) = Z_i^T \beta + \tilde{f}_{st}(s_i, t) + \epsilon(s_i, t), \quad (3.7)$$

$$\tilde{f}_{st}(s_i, t) = a \tilde{f}_{st}(s_i, t-1) + \tilde{w}(s_i), \quad (3.8)$$

donde se asume que  $\tilde{\mathbf{w}}_S(s) = (\tilde{w}(s_1), \dots, \tilde{w}(s_k))^T$ , y una distribución gaussiana para  $\tilde{\mathbf{f}}_{st}(\cdot, \mathbf{1}) = (f_{st}(s_1, 1), f_{st}(s_2, 1), \dots, f_{st}(s_n, 1))^T$  con media cero y matriz de covarianza  $\frac{\tilde{\mathbf{C}}_{\mathbf{w}_S}}{1-a}$ . Además para cualquier tiempo  $t$ ,  $\tilde{\mathbf{f}}_{st}(\mathbf{s}, \mathbf{t}) = (f_{st}(s_1, t), f_{st}(s_2, t), \dots, f_{st}(s_n, t))^T$  sigue una distribución gaussiana tal que:

$$\tilde{\mathbf{f}}_{st}(\mathbf{s}, \mathbf{t}) \sim N(\mathbf{0}, \mathbf{C}_{ST}),$$

donde  $\mathbf{C}_{ST}$  es una matriz con elementos definidos por:

$$\text{Cov}(\tilde{f}_{st}(s_i, t), \tilde{f}_{st}(s_j, t')) = \frac{a^{|t-t'|}}{1-a^2} \times \tilde{\mathbf{C}}_{\mathbf{w}_S(s_i, s_j)}. \quad (3.9)$$

Entonces, la matriz de covarianza asociada a  $\tilde{\mathbf{f}}_{st}(\mathbf{s}, \mathbf{t}) = (\tilde{\mathbf{f}}_{st}(\mathbf{s}, \mathbf{1}), \tilde{\mathbf{f}}_{st}(\mathbf{s}, \mathbf{2}), \dots, \tilde{\mathbf{f}}_{st}(\mathbf{s}, \mathbf{T}))^T$  se puede escribir como  $\mathbf{C}_{ST} = \mathbf{C}_T \otimes \tilde{\mathbf{C}}_{\mathbf{w}_S}$ , donde  $\otimes$  es un producto de Kronecker y  $\mathbf{C}_T$  es la matriz de correlación asociada al proceso temporal AR(1). Podemos reescribir en términos de la matriz de precisión  $\tilde{\mathbf{Q}}_{\mathbf{w}_S}$  donde obtendríamos que

$$\mathbf{Q}_{ST} = \mathbf{Q}_T \otimes \tilde{\mathbf{Q}}_{\mathbf{w}_S}. \quad (3.10)$$

La prueba de este resultado se muestra en el apéndice A.1.



### 3.2. Inferencia bayesiana usando INLA

La aproximación anidada de Laplace integrada (Integrated Nested Laplace Aproximation en inglés - INLA) fue introducida por Rue et al. (2009), como un método de aproximación alternativa al método de MCMC. INLA es un método para ajustar modelos bayesianos. Una de las motivaciones de INLA es resolver la dificultad computacional de MCMC en problemas con bastantes datos. En muchas aplicaciones, el proceso de muestreo de la distribución a posteriori usando MCMC puede tardar demasiado y en muchos casos no es viable con los recursos computacionales existentes.

La inferencia bayesiana a través del INLA es restringida para modelos gaussianos latentes (MGL) donde el campo aleatorio latente es un campo aleatorio markoviano gaussiano (CAMG), es decir tiene distribución gaussiana con una matriz de precisión esparsa. La aproximación mediante INLA permite realizar inferencia bayesiana a este tipo de modelos de forma eficiente y precisa.

Para poder aplicar INLA, tenemos que satisfacer tres condiciones: (i) ajustar un modelo gaussiano Latente, (ii) tener en el segundo nivel del MGL un CAMG y (iii) usar la aproximación de Laplace.

Antes de definir un MGL, definimos el CAMG.

**Definición 12** *El campo aleatorio markoviano gaussiano (CAMG) es un vector aleatorio finito dimensional, que sigue una distribución normal multivariada, la cual satisface la condición independencia condicional (Markov) (Rue and Held, 2005).*

Un MGL es un modelo jerárquico que está compuesto por los siguientes niveles:

- **Vector aleatorio ( $\mathbf{Y}$ ):** Sea  $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$  un vector de v.a, donde  $Y_i$ 's condicionadas a  $\mathbf{X}, \boldsymbol{\theta}$  son independientes, luego la f.d.p conjunta de  $Y$  es  $f_{\mathbf{Y}|\mathbf{X},\boldsymbol{\theta}}$  donde:

$$f_{\mathbf{Y}|\mathbf{X},\boldsymbol{\theta}}(y) = \prod_{i=1}^n f_{Y_i|\mathbf{X},\boldsymbol{\theta}}(y_i). \quad (3.11)$$

De forma general se tiene que:

$$g(E(Y)) = \eta = \beta_0 1 + A\boldsymbol{\beta} + f(\cdot) + \epsilon;$$

donde  $g(\cdot)$  es la función de enlace entre la media de  $Y$  y el predictor lineal  $\eta$ , además

- Efecto “fijo”: intercepto  $\beta_0$
  - A: covariables, Efectos “fixos”  $\boldsymbol{\beta}$ ,
  - Efectos estructurados:  $f(\cdot)$ ; por ejemplo estructuras espaciales, temporales, ...
  - Efecto no estructurado  $\epsilon$ .
- **Campo gaussiano latente ( $\mathbf{X}$ ):** El campo latente  $\mathbf{X} = (X_1, \dots, X_J)^\top$  condicionado a  $\boldsymbol{\theta}$  tiene una distribución  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ . En particular,  $\mathbf{X} = \{\beta_0, \boldsymbol{\beta}, \mathbf{f}\}$ .

**Definición 13** Sea  $G=(V,E)$ , un grafo donde  $V$  es el conjunto de vértices y  $E$  es un conjunto de aristas  $\{i,j\}$ . Un vector aleatorio  $X = (X_1, \dots, X_J)$  es un CAMG con respecto al grafo  $G=(V,E)$ , con media  $\mu$  y matriz de precisión  $Q > 0$  (definida positiva), si y solamente si la fdp conjunta de  $X$  es dada por

$$f_X(x) = (2\pi)^{(-n/2)} |Q|^{1/2} \exp\left(-\frac{1}{2}(x - \mu)^T Q (x - \mu)\right);$$

donde

$$Q_{ij} \neq 0 \iff i, j \in E, \forall i \neq j.$$

Es decir,  $Q_{ij}$  es diferente de cero si  $i$  y  $j$  no pertenecen a la misma arista (no son vecinos).

- **Hiperparámetros** Sea el vector de hiperparámetros  $\theta = (\theta_1, \dots, \theta_p) \in \Theta$ , compuesto por los parámetros desconocidos restantes. Por ejemplo:
  - Parámetros de escala de la distribución de  $Y$ .
  - Parámetros de los efectos estructurados o no estructurados, generalmente precisiones.

La fdp de  $\theta$  es:

$$\pi(\theta).$$

### 3.3. INLA para el modelo espacio-temporal NNGP

Definimos el modelo espacio-temporal NNGP como un modelo gaussiano latente de la siguiente forma:

- Vector observado: Del modelo planteado en las ecuaciones 3.7 y 3.8 tenemos que  $Y(s_i, t) | \mathbf{X}, \theta$  son variables independientes con distribución:

$$Y(s_i, t) | \mathbf{X}, \theta \stackrel{ind}{\sim} \mathcal{N}(Z_i \beta + \tilde{f}_{st}(s_i, t), \sigma_\epsilon^2),$$

donde la función de probabilidad conjunta de  $Y(s_i, t)$  es dada por:

$$f_{Y(s_i, t) | \mathbf{X}, \theta}(y) = \prod_{t=1}^T \prod_{i=1}^n f_{Y(s_i, t) | \mathbf{X}, \theta}$$

- Campo aleatorio markoviano gaussiano:  $\mathbf{X} = (\beta, f_{st})$

$$\beta \sim \mathcal{N}(0, \sigma_\beta^2)$$

$$f_{st} \sim \mathcal{N}(0, C_{ST})$$

donde  $C_{ST}$  es la matriz de covarianza del proceso espacio-temporal NNGP, con una función de covarianza de Matérn que depende de los parámetros  $\rho$  y  $\tau_\epsilon = \frac{1}{\sigma_\epsilon^2}$ .

- Hiperparámetros:  $\theta = (\tau_\epsilon, \rho, \sigma_s^2, a)$  donde  $\tau_\epsilon = \sigma_\epsilon^2$ . Y se asume como distribuciones a priori:

$$\tau_\epsilon \sim \text{gamma}(,)$$

$$\sigma_s^2 \sim \text{gamma}(,)$$

$$\rho \sim \text{Uniforme}(a, b).$$

$$a \sim \text{Uniforme}(-1, 1).$$

Internamente en la implementación se usaron las reparametrizaciones detalladas en el apéndice A.2 y los logaritmos de las fdp de estas reparametrizaciones respectivamente.

De esta forma, la función de densidad conjunta a posteriori es dada por:

$$\pi(x, \theta | y) \propto \pi(\theta) \pi(x | \theta) \prod_{i=1}^n \pi(y_i | x, \theta),$$

$$\pi(x, \theta | y) \propto \pi(\theta) \pi(x | \theta) \exp \left( \sum_{t=1}^T \sum_{i=1}^N \log [\pi(y(s_i, t) | x, \theta)] \right).$$

Como el modelo propuesto cumple los tres supuestos que se requiere para usar INLA, podemos hallar las aproximaciones de las distribuciones a posteriori usando aproximación de Laplace.

En particular, INLA permite obtener aproximaciones de las densidades marginales a posteriori:

$$\pi(x_j | y) = \int \pi(x_j | \theta, y) \pi(\theta | y) d\theta$$

$$\pi(\theta_p | y) = \int \pi(\theta | y) d\theta_{-p}$$

para  $j = 1, \dots, J$  e  $p = 1, \dots, P$ . Como las integrales no pueden ser calculadas analíticamente, pueden ser aproximadas numéricamente. Para ello, INLA propone:

- Aproximar  $\pi(\theta | y) : \tilde{\pi}(\theta | y)$ .
- Explorar  $\tilde{\pi}(\theta | y)$  y usar integración numérica para aproximar  $\pi(\theta_p | y)$ :

$$\pi(\theta_p | y) \approx \int_{\Theta_{-p}} \tilde{\pi}(\theta | y) = \sum_k \tilde{\pi}(\theta^{(pk)} | y) w_{pk}.$$

- Aproximar  $\pi(x_j | \theta, y) : \tilde{\pi}(x_j | \theta, y)$ .
- Usar integración Numérica para aproximar  $\pi(x_j | y)$ :

$$\pi(x_j | y) \approx \int_{\Theta} \tilde{\pi}(x_j | \theta, y) \tilde{\pi}(\theta | y) = \sum_k \tilde{\pi}(x_j | \theta^{(k)}, y) \tilde{\pi}(\theta^{(k)} | y) w_k$$

## Capítulo 4

### Estudio de Simulación

En el presente capítulo se desarrollan los estudios de simulación de un proceso espacio-temporal, con el objetivo de estudiar cómo una distribución espacial va evolucionando en diferentes tiempos. Se desarrollan tres casos estudios de simulación de un modelo espacio-temporal usando el proceso NNGP planteado en la sección 3.1. Cada estudio se divide en tres escenarios, variando el número de vecinos ( $m = 10, 20$  y  $30$ ) y el número de tiempos ( $T = 4, 6$ , y  $8$ ) respectivamente. Bajo los estudios de simulación buscamos simular las variables de respuesta del proceso espacio-temporal y verificar la correcta estimación de los parámetros e hiperparámetros usando inferencia bayesiana aproximada a través del INLA.

#### 4.1. Simulación del proceso espacio-temporal

Se examinó mediante estudios de simulación el ajuste del modelo, para ello, se consideró tres estudios con tres escenarios cada uno, dando un total de nueve simulaciones. La tabla 4.1 muestra todas las simulaciones que fueron ejecutadas. Estas simulaciones consideran un dominio espacial con  $n = 500$  locales que evolucionan en el tiempo considerando los siguientes escenarios: el primero con cuatro tiempos  $T = 4$ , el segundo con seis tiempos  $T = 6$ ; y finalmente, el tercero con ocho tiempos  $T = 8$ . En todos estos escenarios se varía la cantidad de vecinos  $m = \{10, 20, 30\}$  en cada estudio. Todas las simulaciones fueron implementadas en el programa libre R.

Cuadro 4.1: Tabla de estudios y escenarios de simulación.

Estudio	Escenario	n	T	$\sigma_s^2$	$\sigma_e^2$	$\nu$	$\phi$	$a$	$\beta_0$	$\beta_1$	m
1	1	500	4	2	0.1	0.5	12	0.8	1	5	10
1	2	500	4	2	0.1	0.5	12	0.8	1	5	20
1	3	500	4	2	0.1	0.5	12	0.8	1	5	30
2	1	500	6	2	0.1	0.5	12	0.8	1	5	10
2	2	500	6	2	0.1	0.5	12	0.8	1	5	20
2	3	500	6	2	0.1	0.5	12	0.8	1	5	30
3	1	500	8	2	0.1	0.5	12	0.8	1	5	10
3	2	500	8	2	0.1	0.5	12	0.8	1	5	20
3	3	500	8	2	0.1	0.5	12	0.8	1	5	30

Ya definido el número de locales y los tiempos, se procedió a simular el proceso espacio-temporal. Para el proceso espacial, usamos la grilla de las localidades y usamos la función de

covarianza espacial exponencial definida en la ecuación 2.2 con parámetros  $\sigma_s^2 = 2$  y  $\phi = 12$ . Luego se simularon los efectos aleatorios  $w(s_i)$  para  $i = 1, \dots, n$  a partir de un proceso gaussiano con media cero y función de covarianza exponencial de acuerdo a estos parámetros. Luego, para la simulación del del proceso temporal autoregresivo AR(1) consideramos un parámetro de autocorrelación  $a = 0.8$ . Luego, procedemos a simular los valores de los efectos espacio-temporales estructurados  $f_{ST}$  definidos en la ecuación 3.9. Para obtener las covariables,  $\mathbf{Z} = [Z_1, Z_2]_{2 \times 500}$ , se definió  $Z_1 = \mathbf{1}^t$  (el intercepto) y se generó la segunda covariable a través de  $Z_2 \sim N(0, 1)$ . Además se definieron los coeficientes de regresión  $\beta_0 = 5$  y  $\beta_1 = 1$  respectivamente. Finalmente, se generaron los errores aleatorios a partir de una distribución normal con media cero y  $\sigma_e^2 = 0.1$ . La suma de estos efectos fijos y aleatorios nos permite simular la variable observada  $Y(s, t)$ . Cada valor de la variable observada  $Y(s, t)$  estará indexada por su localidad (latitud y longitud) y el tiempo. En la figura 4.1 se aprecia los datos simulados del proceso espacio-temporal considerando  $T=6$  tiempos.

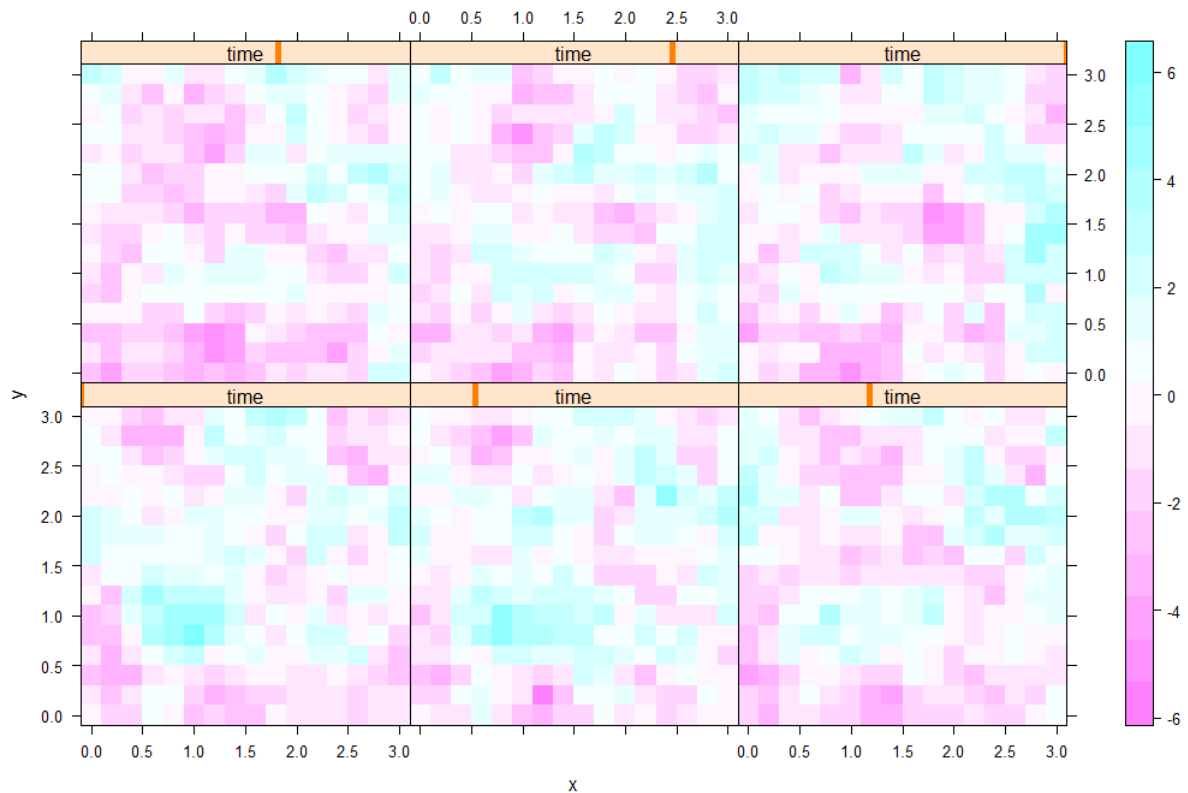


Figura 4.1: Simulación de un proceso espacio-temporal para  $T=6$  tiempos. En la primera fila se muestran las distribuciones para los tres primeros tiempos y en la fila de abajo los siguientes tres tiempos.

#### 4.1.1. Estimación de parámetros e hiperparámetros

Para los estudios de simulación se implementó el proceso espacio-temporal NNGP en el paquete *R-INLA*. El algoritmo nos permite estimar los valores de los parámetros e hiperparámetros a posteriori de los efectos espacio-temporales y los coeficientes de regresión del modelo latente (planteado en el capítulo 3). Internamente se usó una reparametrización, puesto que

numéricamente el algoritmo es más eficiente y estable. Es importante mencionar que los estudios de simulación fueron realizados usando *R version 4.1.3* bajo el sistema operativo *Linux Ubuntu 20.04.4 LTS*. Además, cada escenario fue simulado 50 veces.

#### 4.1.2. Estudio de simulación 1

En el primer estudio de simulación se presentan tres escenarios cuando el número de tiempos es  $T = 4$ . Luego se ajustó el modelo espacio-temporal usando NNGP con diferente número de vecinos  $m = \{10, 20, 30\}$ . La tabla 4.2 presenta los resultados obtenidos durante el proceso de simulación. Se muestra la media de las medias a posteriori de las 50 réplicas, la media de las desviaciones estándar a posteriori (D.est) de las 50 réplicas, la media del límite inferior de los intervalos de credibilidad de las 50 réplicas, la media del límite superior de los intervalos de credibilidad de las 50 réplicas, estos dos últimos se resumen en la tabla como IC 95 %.

Cuadro 4.2: Estudio de simulación número 1.

Escenario	m	T	Parámetro	Real	Media	D.est	IC 95 %
1	10	4	$\beta_0$	1.0	1.15	0.33	0.5-1.81
1	10	4	$\beta_1$	5.0	5.01	0.02	4.98-5.04
1	10	4	$\phi$	12.0	13.27	1.13	11.12-15.56
1	10	4	$a$	0.8	0.79	0.01	0.76-0.81
1	10	4	$\sigma_s^2$	2.0	2.01	0.15	1.74-2.32
1	10	4	$\sigma_\epsilon^2$	0.1	0.08	0.02	0.05-0.12
2	20	4	$\beta_0$	1.0	1.15	0.33	0.5-1.8
2	20	4	$\beta_1$	5.0	5.01	0.02	4.98-5.04
2	20	4	$\phi$	12.0	13.40	1.10	11.31-15.6
2	20	4	$a$	0.8	0.79	0.01	0.76-0.81
2	20	4	$\sigma_s^2$	2.0	2.02	0.14	1.77-2.31
2	20	4	$\sigma_\epsilon^2$	0.1	0.08	0.02	0.05-0.12
3	30	4	$\beta_0$	1.0	1.15	0.33	0.5-1.81
3	30	4	$\beta_1$	5.0	5.01	0.02	4.98-5.04
3	30	4	$\phi$	12.0	13.29	1.15	11.1-15.59
3	30	4	$a$	0.8	0.79	0.01	0.76-0.81
3	30	4	$\sigma_s^2$	2.0	2.01	0.15	1.74-2.32
3	30	4	$\sigma_\epsilon^2$	0.1	0.09	0.02	0.05-0.12

En la tabla 4.2 observamos que los parámetros e hiperparámetros reales están contenidos dentro del IC 95 %. Además la media de las medias a posteriori para las 50 réplicas se aproxima al valor real simulado de los parámetros. En general se comprueba que usando cuatro tiempos el ajuste del modelo es adecuado. Sin embargo, el parámetro que es más difícil de recuperar es el  $\phi$ , el cual está asociado al rango efectivo. Probablemente este resultado se debe a cómo se definen los vecinos del NNGP..

Gráficamente esto también se aprecia en las figuras 4.9 e 4.6, donde la línea roja punteada simboliza el parámetro real que cae dentro de las bandas del IC 95%; además los puntos azul, amarillo y naranja simbolizan las estimaciones de la media a posteriori usando los tres escenarios distintos, los cuales resultan muy cercanos al valor real.

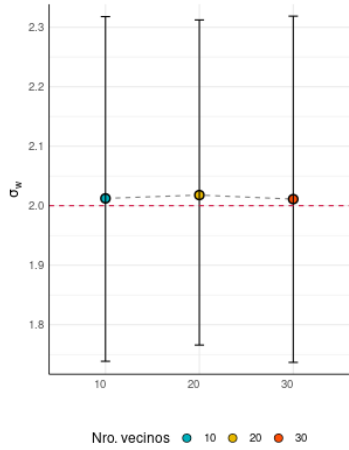


Figura 4.2: Varianza espacial posterior  $\sigma_s^2$

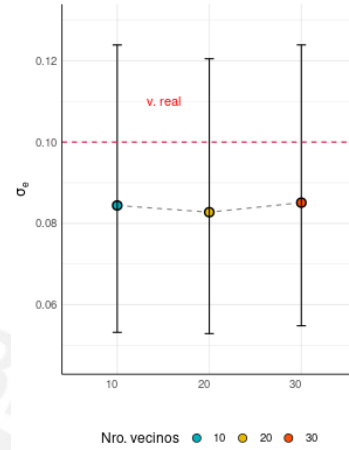


Figura 4.3: Efecto pepita posterior  $\sigma_e^2$

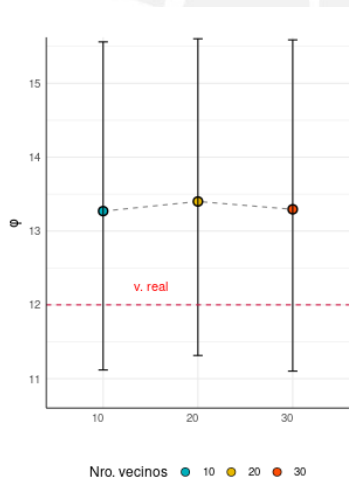


Figura 4.4: Rango a posteriori  $\phi$ .

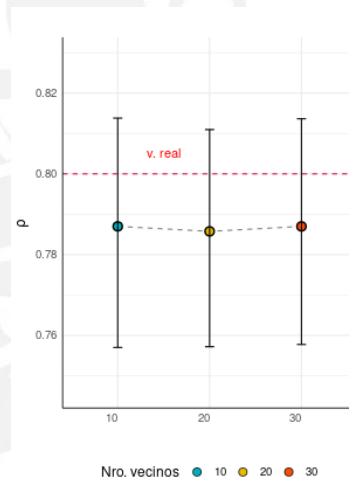


Figura 4.5: Autocorrelación temporal  $\alpha$ .

Figura 4.6: Estudio de simulación 1. Gráficos de las estimaciones de la media posteriori de los hiperparámetros con sus IC 95%.

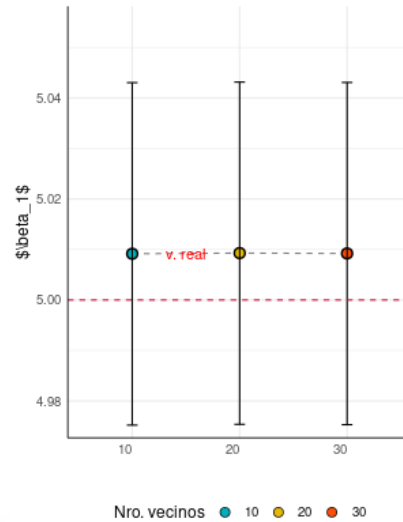
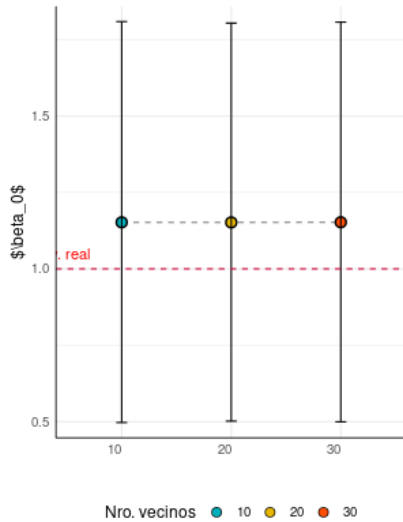


Figura 4.7: Intercepto coeficiente posterior  $\beta_0$

Figura 4.8: Coeficiente  $\beta_1$  posteriori

Figura 4.9: Estudio de simulación 1. Gráficos de las estimaciones de la media a posteriori de los parámetros con sus IC 95 %.

Para seleccionar qué escenario tuvo un mejor ajuste y explica mejor el modelo según el número de vecinos del NNGP, usamos el Criterio de Información de Watanabe (WAIC), el cual se muestra en la tabla 4.3 conjuntamente con los tiempo en segundos. El escenario 2 obtuvo el menor  $WAIC = 2045.1$  usando  $n = 20$  vecinos y con un tiempo de ejecución de  $CPU = 19.77$  minutos (o 1168.220 segundos). Según la tabla 4.2, el escenario 2 también presenta la menor amplitud en los intervalos de credibilidad donde recae los parámetros e hiperparámetros originales. Y como se esperaba conforme aumenta el número de vecinos aumenta también el tiempo de estimación.

Cuadro 4.3: Estudio de simulación 1. Tabla de resultados.

m	WAIC	CPU
10	2063.226	379.6175
20	2045.124	1168.2220
30	2058.270	1851.0905

#### 4.1.3. Estudio de simulación 2

En el segundo estudio de simulación se presentan 3 escenarios cuando el número de tiempos es  $T = 6$ . Luego se ajustó el modelo espacio-temporal usando NNGP con diferente número de vecinos  $m = \{10, 20, 30\}$ . La tabla 4.4 presenta los resultados obtenidos durante el proceso de simulación.



Cuadro 4.4: Estudio de simulación número 2.

Escenario	m	T	Parámetro	Real	Media	D.est	IC 95 %
1	10	6	$\beta_0$	1.0	1.16	0.32	0.54-1.79
1	10	6	$\beta_1$	5.0	5.01	0.01	4.98-5.03
1	10	6	$\phi$	12.0	12.88	0.87	11.23-14.63
1	10	6	$a$	0.8	0.79	0.01	0.77-0.82
1	10	6	$\sigma_s^2$	2.0	1.98	0.11	1.77-2.22
1	10	6	$\sigma_\epsilon^2$	0.1	0.10	0.01	0.07-0.13
2	20	6	$\beta_0$	1.0	1.17	0.32	0.54-1.79
2	20	6	$\beta_1$	5.0	5.01	0.01	4.98-5.03
2	20	6	$\phi$	12.0	12.90	0.89	11.19-14.67
2	20	6	$a$	0.8	0.79	0.01	0.77-0.82
2	20	6	$\sigma_s^2$	2.0	1.98	0.12	1.76-2.22
2	20	6	$\sigma_\epsilon^2$	0.1	0.10	0.01	0.07-0.13
3	30	6	$\beta_0$	1.0	1.17	0.32	0.54-1.79
3	30	6	$\beta_1$	5.0	5.01	0.01	4.98-5.03
3	30	6	$\phi$	12.0	12.89	0.88	11.2-14.66
3	30	6	$a$	0.8	0.79	0.01	0.77-0.82
3	30	6	$\sigma_s^2$	2.0	1.98	0.12	1.76-2.22
3	30	6	$\sigma_\epsilon^2$	0.1	0.10	0.01	0.07-0.13

En la tabla 4.4, tanto los parámetros e hiperparámetros reales están contenidos dentro del IC 95%. Gráficamente esto también se aprecia en las figuras 4.17 e 4.14, donde la línea roja punteada simboliza el parámetro real que cae dentro de las bandas del IC; y además, los puntos azul, amarillo y naranja simbolizan las estimaciones de la media de las medias a posteriori usando las 50 réplicas, los cuales resultan muy cercanos al valor real. Los resultados obtenidos son muy similares al estudio de simulación 1. La principal diferencia con el estudio de simulación 1 se nota justamente en el parámetro de autocorrelación en el tiempo  $a$ , pues la media de las medias a posteriori de las 50 réplicas para este parámetro es igual al valor del parámetro simulado igual a 0.8, y el IC 95% tiene una amplitud menor. Esto se debe a que conforme se aumenta el número de tiempos, el modelo consigue estimar mejor este parámetro. Además, el escenario 3 también presenta ligeramente la menor amplitud en los intervalos de credibilidad donde recae los parámetros e hiperparámetros originales.

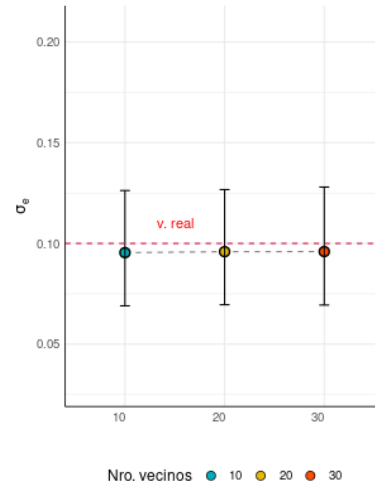
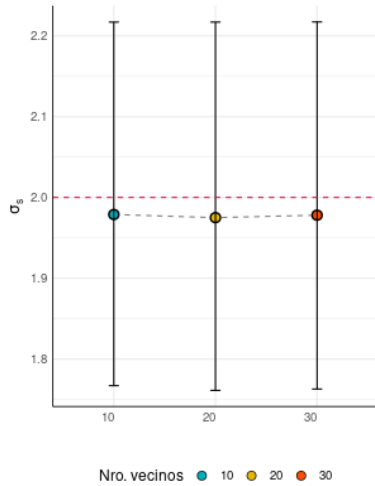


Figura 4.10: Varianza espacial posterior  $\sigma_s^2$

Figura 4.11: Efecto pepita posterior  $\sigma_e^2$

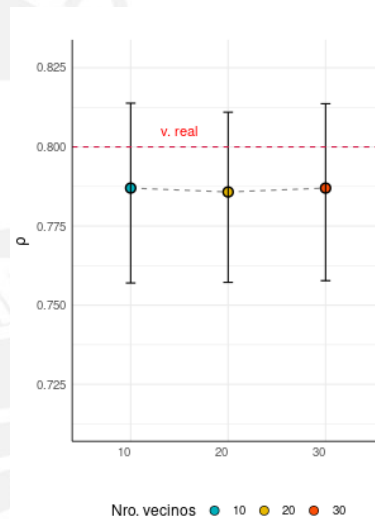
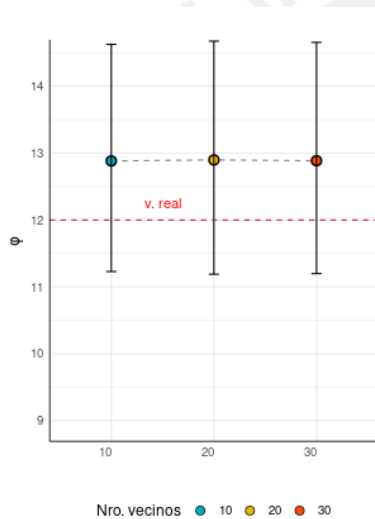


Figura 4.12: Rango posteriori  $\phi$

Figura 4.13: Autocorrelación temporal posterior  $\rho$

Figura 4.14: Estudio de simulación 2. Gráficos de las estimaciones de la media posteriori de los hiperparámetros con sus IC.

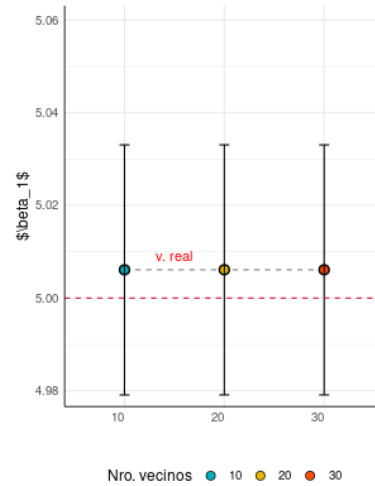
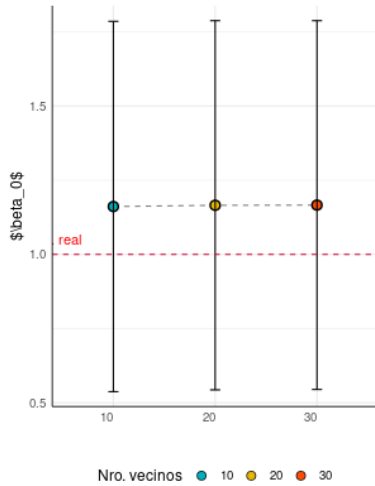


Figura 4.15: Intercepto coeficiente posterior  $\beta_0$

Figura 4.16: Coeficiente  $\beta_1$  aposteriori

Figura 4.17: Estudio de simulación 2. Gráficos de las estimaciones de la media a posteriori de los parámetros con sus IC.

Para seleccionar el mejor escenario nuevamente usamos el WAIC para seleccionar el modelo que se ajuste mejor a los datos, el cual se muestra en la tabla 4.5 conjuntamente con los tiempo en segundos. El escenario 3 obtuvo el menor  $WAIC = 3462.432$  usando  $n = 30$  vecinos, con un tiempo de ejecución de  $CPU = 37.3$  minutos (o 2235.870 segundos).

Cuadro 4.5: Tabla de resultados del estudio de simulación 2.

m	WAIC	CPU
10	3467.418	940.8706
20	3470.467	2235.8701
30	3462.432	3331.1880

#### 4.1.4. Estudio de simulación 3

En el tercer estudio de simulación se presentan 3 escenarios cuando el número de tiempos es  $T = 8$ . Luego se ajustó el modelo espacio-temporal usando NNGP con diferente número de vecinos  $m = \{10, 20, 30\}$ . La tabla 4.6 presenta los resultados obtenidos durante el proceso de simulación.

Cuadro 4.6: Estudio de simulacion del estudio número 3.

Escenario	m	T	Parámetro	Real	Media	D.est	IC 95 %
1	10	8	$\beta_0$	1.0	1.14	0.30	0.54-1.74
1	10	8	$\beta_1$	5.0	5.00	0.01	4.98-5.02
1	10	8	$\phi$	12.0	12.52	0.79	10.97-14.08
1	10	8	$a$	0.8	0.80	0.01	0.78-0.82
1	10	8	$\sigma_s^2$	2.0	1.98	0.10	1.79-2.19
1	10	8	$\sigma_\epsilon^2$	0.1	0.10	0.01	0.08-0.13
2	20	8	$\beta_0$	1.0	1.15	0.30	0.55-1.74
2	20	8	$\beta_1$	5.0	5.00	0.01	4.98-5.02
2	20	8	$\phi$	12.0	12.55	0.78	11.08-14.13
2	20	8	$a$	0.8	0.80	0.01	0.78-0.82
2	20	8	$\sigma_s^2$	2.0	1.98	0.10	1.79-2.19
2	20	8	$\sigma_\epsilon^2$	0.1	0.10	0.01	0.08-0.12
3	30	8	$\beta_0$	1.0	1.15	0.30	0.55-1.74
3	30	8	$\beta_1$	5.0	5.00	0.01	4.98-5.02
3	30	8	$\phi$	12.0	12.56	0.80	11.05-14.18
3	30	8	$a$	0.8	0.80	0.01	0.78-0.82
3	30	8	$\sigma_s^2$	2.0	1.98	0.10	1.79-2.19
3	30	8	$\sigma_\epsilon^2$	0.1	0.10	0.01	0.08-0.13

En la tabla 4.6, observamos que los parámetros e hiperparámetros reales están contenidos dentro del IC 95%. Gráficamente esto también se aprecia en las figuras 4.25 e 4.22, donde la línea roja punteada simboliza el parámetro real que cae dentro de las bandas del IC; además los puntos azul, amarillo y naranja simbolizan las estimaciones de la media a posteriori usando tres escenarios distintos, los cuales resultan muy cercanos al valor real. Los resultados obtenidos son muy similares al estudio de simulación 2. La principal diferencia con el estudio de simulación 2 se nota justamente en el parámetro de autocorrelación en el tiempo  $a$ , pues el IC 95% tiene una amplitud menor. Esto se debe a que conforme se aumenta el número de tiempos, el modelo consigue estimar mejor este parámetro.

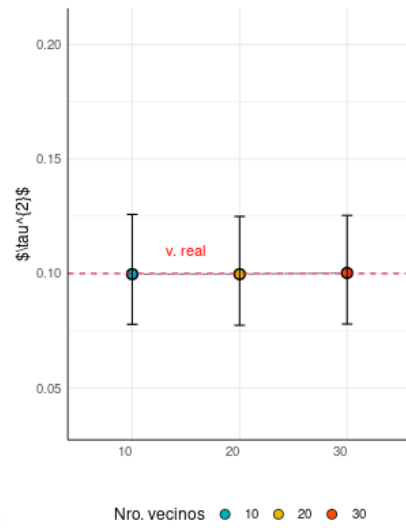
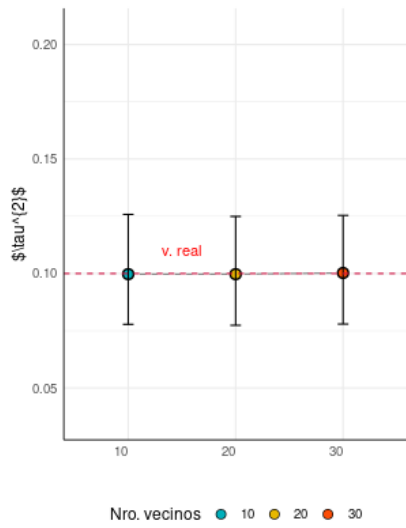


Figura 4.18: Varianza espacial posterior  $\sigma_s^2$

Figura 4.19: Efecto pepita posterior  $\sigma_e^2$

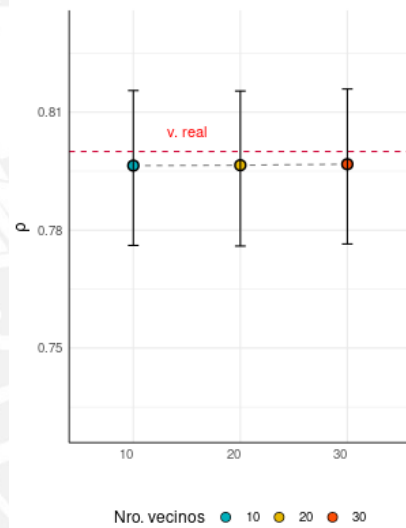
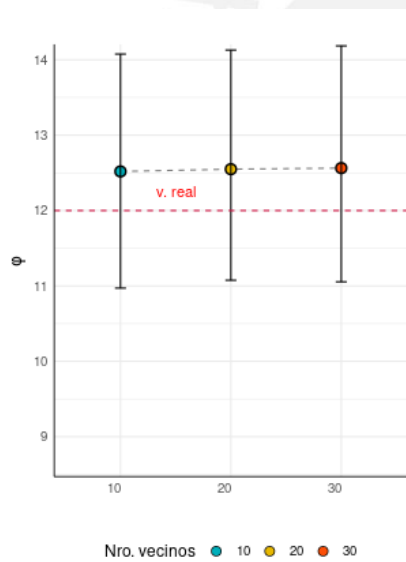


Figura 4.20: Rango posteriori  $\phi$

Figura 4.21: Autocorrelación temporal posterior  $\rho$

Figura 4.22: Estudio de simulación 3. Gráficos de las estimaciones de la media posteriori de los hiperparámetros con sus IC.

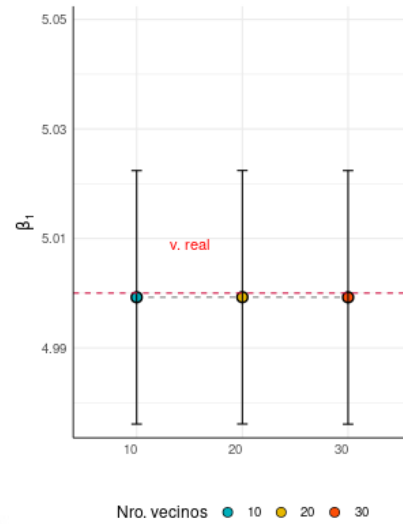
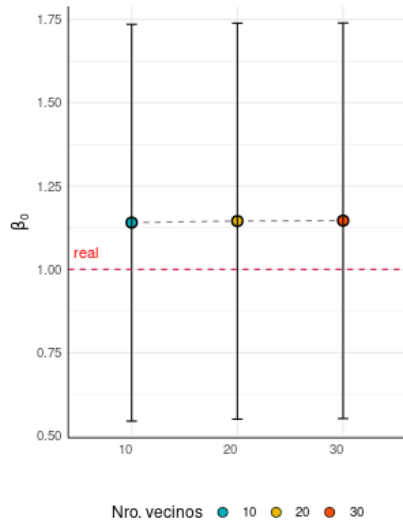


Figura 4.23: Intercepto coeficiente posterior  $\beta_0$

Figura 4.24: Coeficiente  $\beta_1$  posteriori

Figura 4.25: Estudio de simulación 3. Gráficos de las estimaciones de la media a posteriori de los parámetros con sus IC.

Usamos nuevamente el WAIC para seleccionar el mejor modelo, los resultados se muestran en la tabla 4.7 conjuntamente con los tiempo en segundos. El escenario 1 obtuvo el menor  $WAIC = 4766.025$  usando  $n = 10$  vecinos y con un tiempo de ejecución de  $CPU = 29.4$  minutos (o 1769.477 segundos). Aunque se podría esperar que más vecinos implica mejor ajuste, no siempre este es el caso dependiendo por ejemplo de cómo se seleccionan los vecinos debido al GAD usado. Finalmente si observamos todos los escenarios, vemos que conforme se tienen más tiempos también aumenta significativamente el tiempo de estimación, debido a ello, este tipo de modelos espacio-temporales usando NNGP permiten la inferencia cuando se cuenta con muestras grandes.

Cuadro 4.7: Estudio de simulación 3. Tabla de resultados.

m	WAIC	CPU
10	4766.025	1769.477
20	4778.707	4333.470
30	4786.633	5880.146

## Capítulo 5

### Aplicación

En este capítulo se presentan los resultados del modelo espacio-temporal usando NNGP aplicado a datos de concentración de partículas de contaminación en el aire ( $PM_{10}$  en  $ug/m^3$ ) Cameletti et al. (2013). El caso comprende un total de 4368 observaciones, las cuales se encuentran distribuidas espacialmente en 24 estaciones (locales) en una región espacial continua y que están temporalmente distribuidas en 182 días (tiempo discreto). En la aplicación se muestran la información utilizada y los análisis realizados en el desarrollo de la tesis; llevando a cabo las siguientes actividades: descripción de la problemática de los fenómenos medioambientales relacionados a la concentración de contaminación ambiental, presentación del modelo bayesiano espacio-temporal NNGP usando INLA y su relevancia respecto a los datos medioambientales, y finalmente el modelamiento, la estimación de los parámetros y la evaluación del modelo.

#### 5.1. Descripción del estudio

Para algunos países del sur de Europa la contaminación del aire es una emergencia medioambiental debido a los efectos adversos que los altos niveles de concentración de contaminantes pueden tener sobre la salud humana y los ecosistemas. En cuanto al  $PM_{10}$ , la situación es especialmente crítica en la cuenca del río Po Cameletti et al. (2013), situada en el norte de Italia, entre los Alpes y los Apeninos. En esta zona el límite anual y diario supera los valores máximo establecidos por la Unión Europea para la protección de la salud humana. Como consecuencia, la población está expuesta a niveles de contaminación que pueden provocar multitud de consecuencias nocivas, que van desde afectaciones menores en el sistema cardiorrespiratorio hasta la muerte prematura.

Bajo dicho contexto, se tiene el conjunto de datos que proviene de estudios realizados por la Unión Europea entre el año 2005 y 2006, los cuales provienen de 24 estaciones en la región de Piemonte (Italia) Cameletti et al. (2013), en las cuales monitorean la concentración de partículas de contaminación durante un periodo de 182 días. Estos datos son accesibles a través del portal de la Unión Europea para la protección de la salud. Los datos están distribuidos espacialmente en una región continua y con el tiempo discretizado. Por lo tanto, el estudio en mención cumple los supuestos básicos para utilizar el modelo espacio-temporal usando NNGP e INLA.

Las variables del estudio de partículas de contaminación Blangiardo (2015) son:

- $PM_{10}$  la concentración de partículas de contaminación  $PM_{10}$  medida en los locales

$i = 1, \dots, n$  y tiempo  $t = 1, \dots, T$

- $\log(PM_{10})$  denota el logaritmo de  $PM_{10}$ .
- UTMX y UTM Y son las coordenadas de latitud y longitud de las estaciones en formato UTM.5.1.
- PREC precipitación diaria total (en mm).
- TEMP temperatura diaria promedio (en Kelvin).
- EMI tasas de emision diaria de aerosoles primarios.
- HMIX máxima altura de mezcla diaria.
- WS velocidad promedio del viento diario ( $m/s$ ).

De forma gráfica, elaboramos un mapa donde las 24 estaciones de monitoreo (figura 5.1), las cuales estan distribuidas en una región continua. El mapa contiene los puntos georeferenciados en coordenadas UTM, los cuales capturan los datos repetidos temporalmente por 182 días entre el año 2005 y 2006.

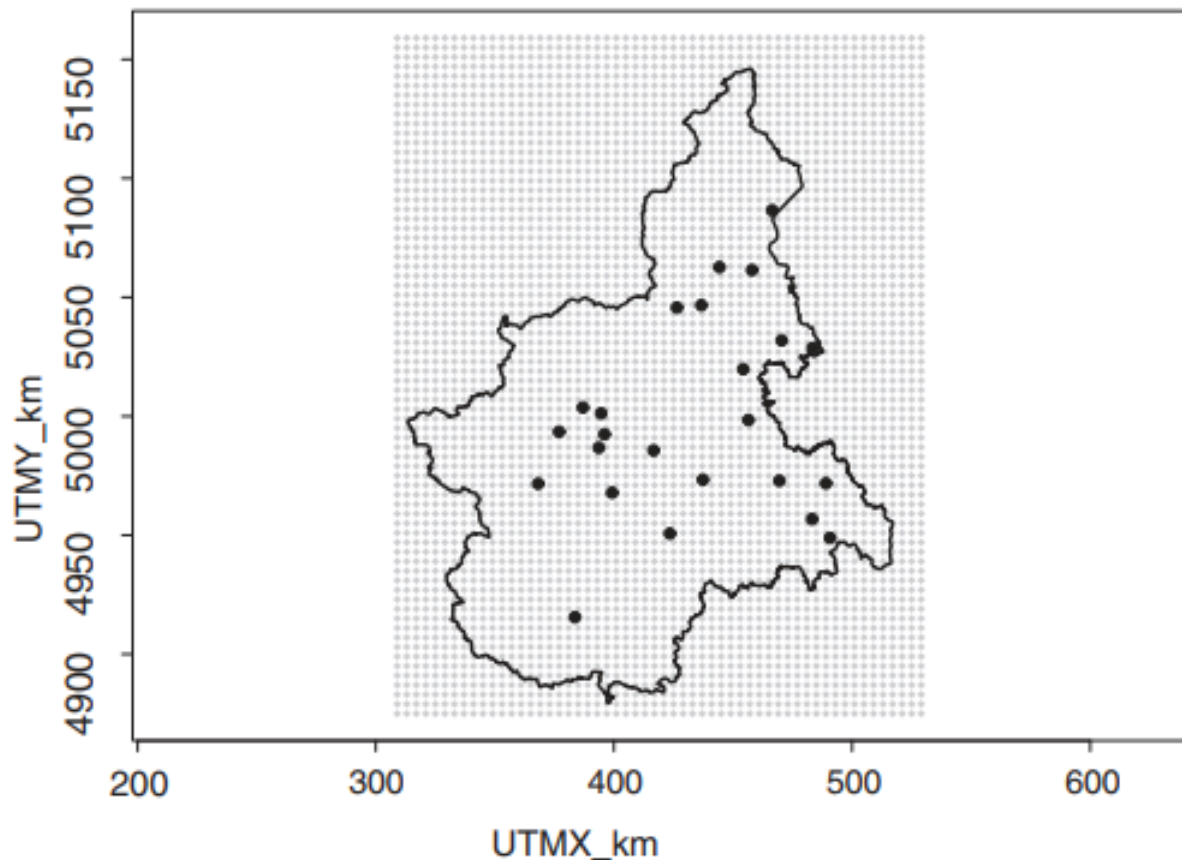


Figura 5.1: Gráfico de las estaciones de monitoreo en coordenadas UTM.



## 5.2. Modelamiento espacio-temporal NNGP para estimar la calidad del aire

Sea  $Y(s_i, t)$  variable aleatoria que representa el logaritmo de la concentración de  $PM_{10}$  medida en la estación  $i = 1, \dots, n = 24$  y día  $t = 1, \dots, T = 182$ . Ajustamos el modelo definido en la sección 3.1 con  $M$  covariables que en la aplicación son las covariables del estudio WS, TEMP, HMIX, PREC, EMI, UTMX, UTMY.

El modelo es ajustado con un número de vecinos  $m = 10, 20, 30$ . Posteriormente se compara las propuestas de modelos para identificar qué modelo presentó un mejor ajuste según el criterio de la raíz del error cuadrático medio (RSME) y el criterio de información de Watanabe (WAIC).

### 5.2.1. Modelo 1 ( $m = 10$ vecinos)

La tabla 5.1 muestra los resultados cuando el número de vecinos es  $m = 10$ . Específicamente, muestra la media a posteriori de los parámetros e hiperparámetros  $\beta_i$  con  $i = 0, \dots, 9$ ,  $\sigma_\epsilon^2$ ,  $\sigma_s^2$ ,  $\phi$  y  $a$ ; conjuntamente con sus intervalos de credibilidad (límite inferior y superior) al 95 %.

Cuadro 5.1: Resumen de resultados para el modelo usando  $m=10$  vecinos

Parámetro	Media	D.est	0.025	0.975
$\beta_0$ (Intercepto)	3.962	0.021	3.921	4.004
$\beta_1$ (A)	-0.145	0.028	-0.200	-0.089
$\beta_2$ (UTMX)	-0.104	0.030	-0.162	-0.045
$\beta_3$ (UTMY)	-0.070	0.022	-0.113	-0.027
$\beta_4$ (WS)	-0.074	0.007	-0.087	-0.060
$\beta_5$ (TEMP)	-0.056	0.017	-0.090	-0.022
$\beta_6$ (HMIX)	-0.086	0.008	-0.103	-0.070
$\beta_7$ (PREC)	-0.108	0.006	-0.120	-0.095
$\beta_8$ (EMI)	0.099	0.019	0.062	0.136
$\sigma_\epsilon^2$	0.008	0.003	0.003	0.015
$\sigma_s^2$	0.130	0.007	0.119	0.146
$\phi$	197.669	40.508	128.558	281.199
$a$	0.738	0.013	0.710	0.759

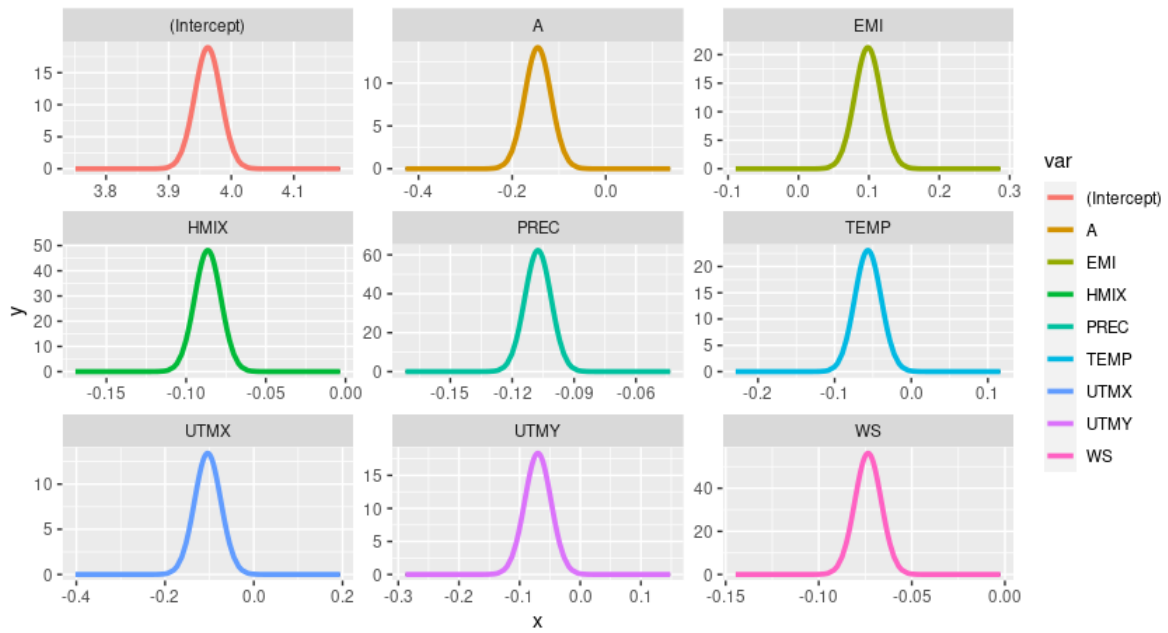


Figura 5.2: Distribución de los efectos fijos de las covariables usando 10 vecinos

**5.2.2. Modelo 2 ( $m = 20$  vecinos)**

La tabla 5.2 muestra los resultados cuando el número de vecinos es  $m = 20$ . Específicamente, muestra la media a posteriori de los parámetros e hiperparámetros  $\beta_0, \beta_1, \sigma_\epsilon^2, \sigma_s^2, \phi$  y  $a$ ; conjuntamente con sus intervalos de credibilidad (limite inferior y superior) al 95%.

Cuadro 5.2: Resumen de resultados para el modelo usando  $m=20$  vecinos

Parámetro	Media	D.est	0.025	0.975
$\beta_0$ (Intercepto)	3.963	0.021	3.922	4.004
$\beta_1$ (A)	-0.146	0.028	-0.200	-0.090
$\beta_2$ (UTMX)	-0.104	0.030	-0.162	-0.046
$\beta_3$ (UTMY)	-0.070	0.022	-0.112	-0.028
$\beta_4$ (WS)	-0.072	0.007	-0.086	-0.059
$\beta_5$ (TEMP)	-0.058	0.017	-0.092	-0.025
$\beta_6$ (HMIX)	-0.085	0.008	-0.101	-0.069
$\beta_7$ (PREC)	-0.107	0.006	-0.119	-0.094
$\beta_8$ (EMI)	0.098	0.019	0.062	0.135
$\sigma_\epsilon^2$	0.007	0.004	0.002	0.018
$\sigma_s^2$	0.132	0.007	0.120	0.146
$\phi$	218.651	61.367	99.335	323.291
$a$	0.737	0.014	0.708	0.763

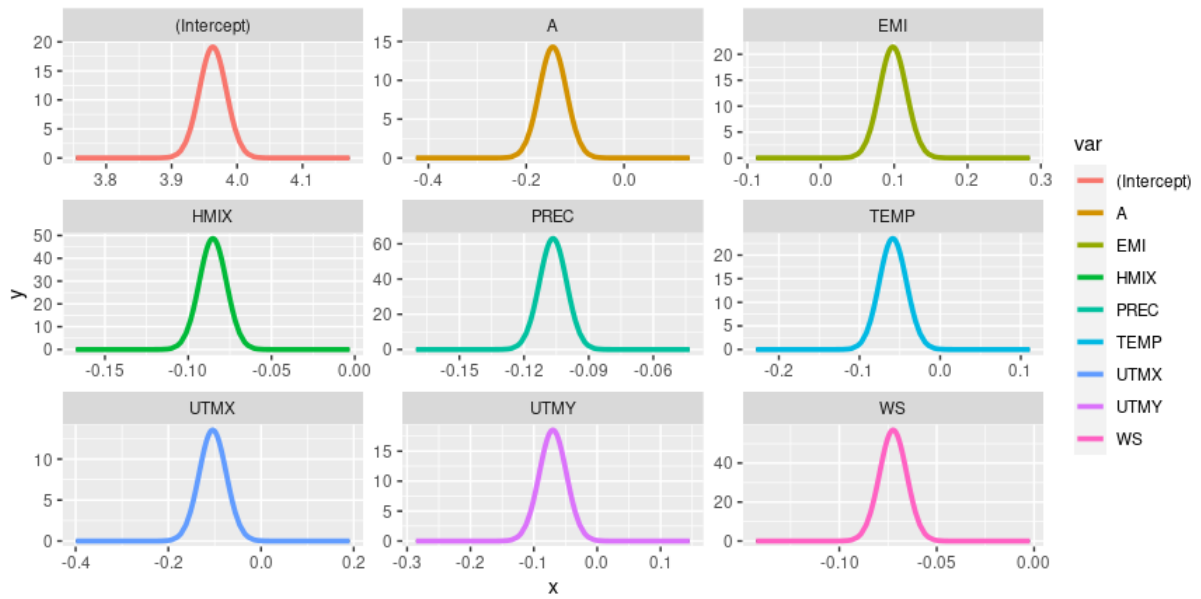


Figura 5.3: Distribución de los efectos fijos de las covariables usando 20 vecinos

**5.2.3. Modelo 3 ( $m = 30$  vecinos)**

La tabla 5.3 muestra los resultados cuando el número de vecinos es  $m = 30$ . Específicamente, muestra la media a posteriori de los parámetros e hiperparámetros  $\beta_0, \beta_1, \sigma_\epsilon^2, \sigma_s^2, \phi$  y  $a$ ; conjuntamente con sus intervalos de credibilidad (limite inferior y superior) al 95%.

Cuadro 5.3: Resumen de resultados para el modelo usando  $n=30$  vecinos

Parámetro	Media	D.est	0.025	0.975
$\beta_0$ (Intercepto)	3.963	0.021	3.921	4.004
$\beta_1$ (A)	-0.145	0.028	-0.200	-0.090
$\beta_2$ (UTMX)	-0.104	0.030	-0.162	-0.046
$\beta_3$ (UTMY)	-0.070	0.022	-0.113	-0.027
$\beta_4$ (WS)	-0.073	0.007	-0.087	-0.059
$\beta_5$ (TEMP)	-0.057	0.017	-0.091	-0.024
$\beta_6$ (HMIX)	-0.086	0.008	-0.102	-0.070
$\beta_7$ (PREC)	-0.107	0.006	-0.119	-0.094
$\beta_8$ (EMI)	0.099	0.019	0.062	0.135
$\sigma_\epsilon^2$	0.005	0.002	0.001	0.010
$\sigma_s^2$	0.135	0.006	0.125	0.149
$\phi$	184.121	92.407	28.414	334.738
$a$	0.732	0.014	0.702	0.756

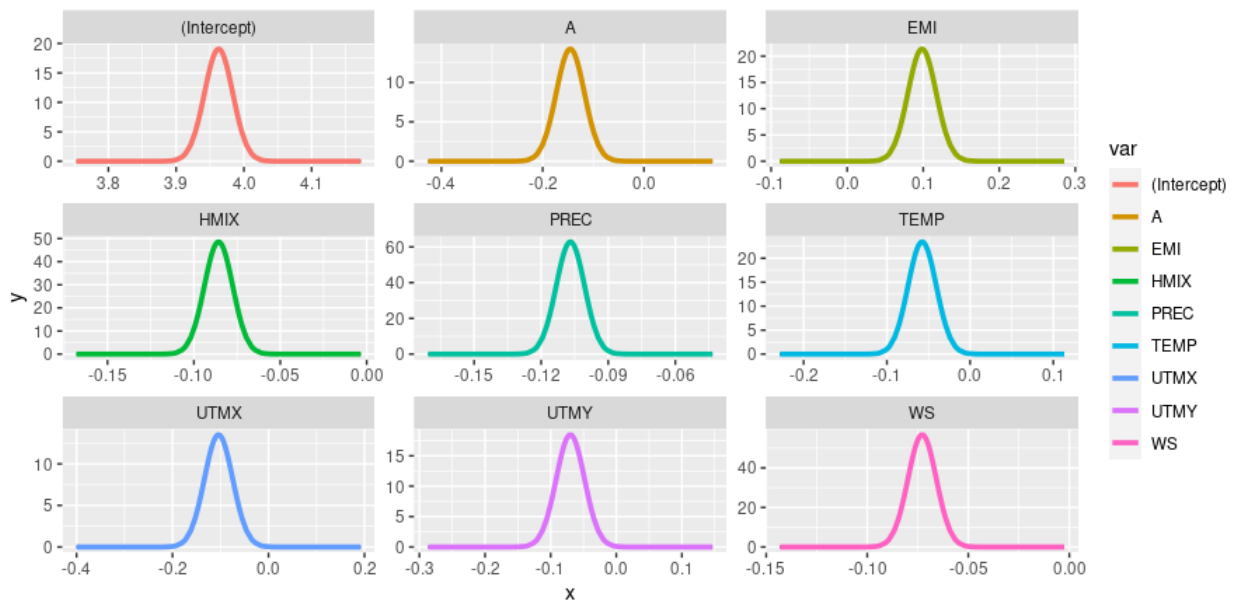


Figura 5.4: Distribución de los efectos fijos de las covariables usando 30 vecinos

### 5.2.4. Evaluación de los modelos

Para evaluar el ajuste de los modelos se utiliza el WAIC y el RMSE que se define como la diferencia cuadrática promedio entre los valores observados y las estimaciones. La tabla 5.4 muestra el RMSE y el WAIC de los modelos 1, 2 y 3 con  $m = 10, 20$  y  $30$  vecinos, respectivamente. En general, el RMSE es pequeño para todos los modelos mostrando un buen ajuste. Los modelos con 10 y 30 vecinos muestran el menor RMSE,  $0.002886$  y  $0.002896$ , respectivamente. Por otro lado, los modelos con 20 y 30 vecinos presentan el menor WAIC,  $-7.467.0030$  y  $-6.686.5220$ , respectivamente. Es importante remarcar, que el tiempo para estimar los parámetros se incrementó conforme se aumentó el número de vecinos. Tomando en cuenta que la muestra total es de  $n \times T = 24 \times 182 = 4368$  observaciones, el tiempo de ejecución menor a un minuto en todos los casos es razonablemente bajo para ajustar modelos espacio-temporales. En el caso de  $m = 30$  vecinos, el tiempo para estimar los parámetros fue aproximadamente menos de un minuto. Además a pesar de que aumentamos relativamente la cantidad de vecinos, el tiempo ejecución total no varía mucho, con apenas segundos de diferencia. Por lo tanto, de acuerdo a los criterios de selección planteados podemos seleccionar como mejor modelo el que ajustamos con  $m = 30$  vecinos.

Cuadro 5.4: Criterios de selección de modelos.

Modelo	Nro.vecinos (m)	RMSE	WAIC	CPU Time
1	10	0.002886	-5.549.3480	36.5871
2	20	0.002903	-7.467.0030	43.4534
3	30	0.002896	-6.686.5220	57.1805

De las tablas resumen, de los modelos 1, 2 y 3, podemos decir que los resultados para los tres modelos fueron similares. Las mayores diferencias como se esperaba se observan en las estimaciones de los parámetros espaciales, en particular del rango efectivo. Este resultado tiene sentido, porque si se incluyen más vecinos en teoría se incluye más información de la distancia hasta la cual dependen las observaciones.

A continuación, se interpretan los resultados obtenidos para el modelo 3, con  $m = 30$  vecinos, que fue el modelo que se seleccionó. La media del intercepto es 3.96, lo cual significa que el nivel promedio de polución está alrededor de  $53\mu\text{g}/\text{m}^3$ , considerando que no hay efecto de las covariables. Además de los tres escenarios, podemos concluir que existe una relación significativa y positiva por parte de la variable  $EMI$  de emisiones de  $CO_2$  con la concentración de  $PM_{10}$ . La variable  $A$ , la altitud de donde es observada la concentración de partículas, tiene un efecto significativo en la reducción de la concentración de partículas de  $PM_{10}$ . De igual modo, las variables climáticas  $WS$ ,  $TEMP$ ,  $HMIX$  y  $PREC$  influyen de forma significativa y están negativamente relacionadas a la concentración de  $PM_{10}$ , siendo la precipitación  $PREC$  la que más reduce la concentración de partículas  $PM_{10}$ .

Referente a los hiperparámetros del modelo, observamos que la autocorrelación temporal (0.732) es cercana a uno, mostrando que existe una autocorrelación moderada positiva entre el  $PM_{10}$  de un año y el  $PM_{10}$  del año anterior, es decir, existe una asociación lineal directa del  $PM_{10}$  en un año  $t$  y  $t - 1$ , por lo tanto si el  $PM_{10}$  en un local en el año  $t - 1$  es elevado, el  $PM_{10}$  en el mismo local en el año  $t$  también es elevado. Además este resultado indica que es adecuado incluir parámetros de tiempo en el modelo.

Respecto a los parámetros espaciales, el rango efectivo implica que el  $PM_{10}$  en un local depende en mayor medida del  $PM_{10}$  en otro local que se encuentra a una distancia de 184 km. Una gran parte de la variabilidad en la concentración de partículas  $PM_{10}$  es explicada por la varianza espacial  $\sigma_s^2 = 0.135$ , y muy poca debido al efecto pepita o varianza no estructurada  $\sigma_\epsilon^2 = 0.005$ .

## Capítulo 6

### Conclusiones

En la presente tesis se ha estudiado un modelo espacio-temporal usando NNGP para estudiar fenómenos que evolucionan tanto en el espacio como en el tiempo. Es importante mencionar que no hay literatura en cómo implementar los espacios-temporales usando NNGP e INLA. Así este trabajo busca estudiar la estimación de los parámetros del modelamiento de estos fenómenos espacio-temporales en campos aleatorios gaussianos usando NNGP a través de simulaciones usando diferente número de vecinos y diferentes escenarios. Esta es otra de las contribuciones de este trabajo, debido a que el modelo espacio-temporal generalmente toman un tiempo considerable en poder ser ejecutado; en ese sentido se utilizó INLA como herramienta proporcionando resultados de forma eficiente en términos de tiempo de ejecución y estimación. Según los resultados obtenidos en las simulaciones la mayor diferencia fue el tiempo de ejecución, conforme los vecinos aumentaban el tiempo de ejecución aumentaba muy poco, pero siempre manteniendo la precisión en la estimaciones. Asimismo, conforme se aumentó el número de tiempos, el parámetro de autocorrelación temporal se estimó mejor, pues evidentemente se tiene mayor información temporal en los datos.

Como aplicación se propone modelar la concentración de partículas de contaminación en un región determinada en un conjunto de estaciones monitoreadas por un periodo de tiempo. El mejor modelo según los criterios usados resultó ser el modelo con  $m = 30$  vecinos. El tiempo de ejecución de este modelo fue aproximadamente menos de un minuto. Dada la complejidad del modelo, este tiempo de ejecución es muy bueno. Cabe resaltar que los resultados obtenidos en esta aplicación usando NNGP fueron muy similares a los que mostraron Cameletti et al. (2013) usando ecuaciones diferenciales parciales estocásticas. De esta forma se puede comprobar que el modelo propuesto se ajustó adecuadamente.

Como trabajos futuros, se puede implementar el modelo usando c++ dentro de la implementación en el R-INLA, esto mejoraría aún más el tiempo de ejecución del modelo; y por otro lado también se puede aplicar el modelo implementado a datos no gaussianos.

## Apéndice A

### Resultados teóricos

#### A.1. Matriz de precisión del efecto espacio temporal

La matriz de covarianza de un proceso temporal AR(1) es definida por:

$$C_T = \frac{1}{1-a^2} \begin{pmatrix} 1 & a & a^2 & \dots & \dots & a^{|T-2|} & a^{|T-1|} \\ a & 1 & a & \dots & \dots & a^{|T-3|} & a^{|T-2|} \\ a^2 & a & 1 & \dots & \dots & a^{|T-4|} & a^{|T-3|} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ a^{|T-2|} & a^{|T-3|} & \dots & \dots & a & 1 & a \\ a^{|T-1|} & a^{|T-2|} & a^{|T-3|} & \dots & a^2 & a & 1 \end{pmatrix}.$$

A partir de la cual se puede calcular su inversa obteniendo la matriz de precisión de un proceso temporal AR(1) definida por:

$$Q_T = \begin{pmatrix} 1 & -a & 0 & \dots & \dots & 0 & 0 \\ -a & (1+a^2) & -a & \dots & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \dots & -a & (1+a^2) & -a \\ 0 & 0 & \dots & \dots & 0 & -a & 1 \end{pmatrix}$$

Dado  $\tilde{f}_{st}(s_i, t) = a\tilde{f}_{st}(s_i, t-1) + \tilde{w}(s_i)$ .

Luego, sea  $\tilde{Q}_{w_s} = \tilde{Q}_s$ , la matriz de precisión de un NNGP, entonces:

$$C_{ST} = \begin{pmatrix} \frac{1}{1-a^2} \tilde{Q}_s^{-1} & \frac{a^2}{1-a^2} \tilde{Q}_s^{-1} & \frac{a^3}{1-a^2} \tilde{Q}_s^{-1} & \dots & \dots & \frac{a^{|T-2|}}{1-a^2} \tilde{Q}_s^{-1} & \frac{a^{|T-1|}}{1-a^2} \tilde{Q}_s^{-1} \\ \frac{a^2}{1-a^2} \tilde{Q}_s^{-1} & \frac{1}{1-a^2} \tilde{Q}_s^{-1} & \frac{a}{1-a^2} \tilde{Q}_s^{-1} & \dots & \dots & \frac{a^{|T-3|}}{1-a^2} \tilde{Q}_s^{-1} & \frac{a^{|T-2|}}{1-a^2} \tilde{Q}_s^{-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \frac{a^{|T-2|}}{1-a^2} \tilde{Q}_s^{-1} & \frac{a^{|T-3|}}{1-a^2} \tilde{Q}_s^{-1} & \dots & \dots & \frac{a^3}{1-a^2} \tilde{Q}_s^{-1} & \frac{a^2}{1-a^2} \tilde{Q}_s^{-1} & \frac{a}{1-a^2} \tilde{Q}_s^{-1} \\ \frac{a^{|T-1|}}{1-(a^k)^2} \tilde{Q}_s^{-1} & \frac{a^{|T-2|}}{1-a^2} \tilde{Q}_s^{-1} & \frac{a^{|T-3|}}{1-a^2} \tilde{Q}_s^{-1} & \dots & \frac{a^2}{1-a^2} \tilde{Q}_s^{-1} & \frac{a}{1-a^2} \tilde{Q}_s^{-1} & \frac{1}{1-a^2} \tilde{Q}_s^{-1} \end{pmatrix},$$

$$C_{ST} = \frac{1}{1-a^2} \begin{pmatrix} 1 & -a & 0 & \dots & \dots & 0 & 0 \\ -a & (1+a^2) & -a & \dots & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \dots & -a & (1+a^2) & -a \\ 0 & 0 & \dots & \dots & 0 & -a & 1 \end{pmatrix} \otimes [\tilde{Q}_s]^{-1},$$

$$C_{ST} = \begin{pmatrix} 1 & -a & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ -a & (1+a^2) & -a & \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & -a & (1+a^2) & -a \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & -a & 1 \end{pmatrix}^{-1} \otimes [\tilde{Q}_s]^{-1}.$$

Luego,  $C_{ST} = [Q_T]^{-1} \otimes [\tilde{Q}_s]^{-1}$ , y por propiedad de matrices invertibles,

$$\tilde{Q} = [\tilde{\Sigma}]^{-1} = ([Q_T]^{-1} \otimes [\tilde{Q}_s^{-1}])^{-1} = ([Q_T]^{-1})^{-1} \otimes ([\tilde{Q}_s]^{-1})^{-1} = Q_T \otimes \tilde{Q}_s.$$

## A.2. Reparametrización usada en el modelo espacio-temporal NNGP

Internamente en la implementación se usaron estas reparametrizaciones en función de los parámetros  $\theta_1$ ,  $\theta_2$  y  $\theta_3$  asociados a  $\sigma_s^2$ ,  $\phi$  y  $a$ , respectivamente. Luego, se puede recuperar la estimación a posteriori de los parámetros. Específicamente:

- Para la varianza marginal espacial  $\sigma_s^2$ :

$$\theta_1 = -\log(\sigma_s^2),$$

luego

$$\sigma_s^2 = \exp(-\theta_1).$$

- Para el rango  $\phi$ :

$$\theta_2 = \log\left(\frac{\phi - a}{b - \phi}\right),$$

luego

$$\phi = b - \frac{(b - a)}{1 + \exp(\theta_2)}.$$

- Para la autocorrelación temporal  $a$ :

$$\theta_3 = \log\left(\frac{1 + a}{1 - a}\right),$$

luego

$$a = \frac{2 \exp(\theta_3)}{1 + \exp(\theta_3)} - 1.$$



## Bibliografía

- Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*, second edition edn, Chapman and Hall/CRC.
- Blangiardo (2015). *Spatial and Spatio-temporal Bayesian Models with R - INLA*.
- Bolin, D., Wallin, J. and Lindgren, F. (2019). Latent gaussian random field mixture models, **130**: 80–93.  
**URL:** <https://linkinghub.elsevier.com/retrieve/pii/S0167947318301907>
- Cameletti, M., Lindgren, F., Simpson, D. and Rue, H. (2013). Spatio-temporal modeling of particulate matter concentration through the SPDE approach, **97**(2): 109–131.  
**URL:** <http://link.springer.com/10.1007/s10182-012-0196-3>
- Carlin, B. P., Louis, T. A. and Carlin, B. P. (2009). *Bayesian methods for data analysis*, Chapman & Hall/CRC texts in statistical science series, 3rd ed edn, CRC Press. OCLC: ocn227205756.
- Chiles, J.-P. and Delfiner, P. (2012). *Geostatistics: modeling spatial uncertainty*, Wiley series in probability and statistics, 2nd ed edn, Wiley.
- Datta, A., Banerjee, S., Finley, A. O. and Gelfand, A. E. (2016). Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets, **111**(514): 800–812.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2014). *Bayesian Data Analysis*, third edition edn, CRC Press, Taylor & Francis Group. OCLC: 909477393.  
**URL:** <http://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=1438153>
- Hristopulos, D. (2020). *Random Fields for Spatial Data Modeling: A Primer for Scientists and Engineers*, Advances in Geographic Information Science, Springer Netherlands.  
**URL:** <https://www.springer.com/gp/book/9789402419160>
- Lauritzen, S. L. (1996). *Graphical models*, number 17 in *Oxford statistical science series*, Clarendon Press ; Oxford University Press.
- Lindgren, F. (2015). Computation fundamentals of discrete GMRF representations of continuous domain spatial models, p. 10.
- Moraga, P. (2019). *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*, 1 edn, Chapman and Hall/CRC.  
**URL:** <https://www.taylorfrancis.com/books/9781000732030>
- Peng, R. D. (2012). *Advanced Statistical Computing*, first edition edn, LeanPub.  
**URL:** <https://leanpub.com/advstatcomp>
- Quiroz, Z. C., Prates, M. O., Dey, D. K. and Rue, H. (2021). Fast bayesian inference of block nearest neighbor gaussian process for large data.  
**URL:** <http://arxiv.org/abs/1908.06437>

- Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*, number 104 in *Monographs on statistics and applied probability*, Chapman & Hall/CRC.
- Rue, H., Martino, S. and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations, **71**(2): 319–392.  
**URL:** <http://doi.wiley.com/10.1111/j.1467-9868.2008.00700.x>
- Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities, **81**(393): 82–86.  
**URL:** <http://www.tandfonline.com/doi/abs/10.1080/01621459.1986.10478240>
- Wackernagel, H. (n.d.). *Multivariate Geostatistics*, Springer Berlin Heidelberg.  
**URL:** <http://link.springer.com/10.1007/978-3-662-05294-5>
- Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*, Springer Texts in Statistics, Springer New York.  
**URL:** <http://link.springer.com/10.1007/978-0-387-21736-9>
- Wikle, C. K., Zammit-Mangion, A. and Cressie, N. A. C. (2019). *Spatio-temporal statistics with R*, Chapman & Hall/CRC the R series, CRC Press, Taylor and Francis Group.

