



UNIVERSIDAD DE CONCEPCIÓN
DIRECCIÓN DE POSTGRADO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
PROGRAMA MAGISTER EN ESTADÍSTICA

Implementación de un método de estimación y predicción espacio-tiempo con datos faltantes basado en filtro de Kalman

Tesis para optar al grado de Magíster en Estadística

Leonardo José Padilla Sepúlveda

Profesor Guía: Bernardo Lagos Álvarez
Depto. de Estadística, Facultad de Ciencias Físicas y Matemáticas
Universidad de Concepción



UNIVERSIDAD DE CONCEPCIÓN
DIRECCIÓN DE POSTGRADO
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
PROGRAMA MAGISTER EN ESTADÍSTICA

Implementación de un método de estimación y predicción espacio-tiempo con datos faltantes basado en filtro de Kalman

Profesor Guía : Bernardo Lágos-Álvarez Firma

Profesor Colaborador : Guillermo Ferreira Cabezas Firma

Profesor Colaborador : Jorge Figueroa Zúñiga Soto Firma

Profesor Colaborador : Tarik Faouzi Nadim Firma

Nombre Memorante : Leonardo Padilla Sepúlveda Firma

Teléfono : +56 9 42349881

e-mail : leonardopadilla@udec.cl

Concepción, 2018

Dedicatoria

Al conocimiento.

Aportando con mi grano de arena, para un mundo mejor.



Agradecimientos

*A los que me rodean.
Por estar ahí, dándome apoyo incondicional.*



Tabla de contenido

Dedicatoria	iii
Agradecimientos	iv
Índice de figuras	vii
Índice de tablas	viii
Resumen	ix
1. Introducción	1
2. Marco Teórico	3
2.1. El modelo Espacio-Tiempo AR(p)	3
2.2. Estacionariedad	4
2.2.1. Estacionariedad fuerte	4
2.2.2. Estacionariedad débil o de segundo orden	5
2.2.3. Función de covarianza espacio-tiempo del modelo	5
2.3. Representación Espacio-Estado del modelo	8
2.4. Implementación e inferencia	9
2.4.1. Implementación secuencial	11
2.5. Inferencia para el proceso espacio-temporal	14
2.5.1. Modelos lineales/gaussianos	14
2.5.2. Filtro de Kalman	14
2.5.3. Suavizador de Kalman	17
2.5.4. Supuestos que sustentan el algoritmo	18

2.6. Valores iniciales	20
2.7. Estimación de parámetros del modelo	20
2.7.1. Estimación por Máxima Verosimilitud	21
2.7.2. Estimación mediante algoritmo Esperanza-Maximización Ge- neral	22
2.8. Interpolación espacial	26
3. Estudio de simulación	29
3.1. Desempeño del método de estimación	29
4. Aplicación a datos reales	34
4.1. Análisis exploratorio	39
4.2. Estimación de parámetros del modelo	42
4.3. Predicción	43
Conclusión	51
Bibliografía	52
Anexos	54

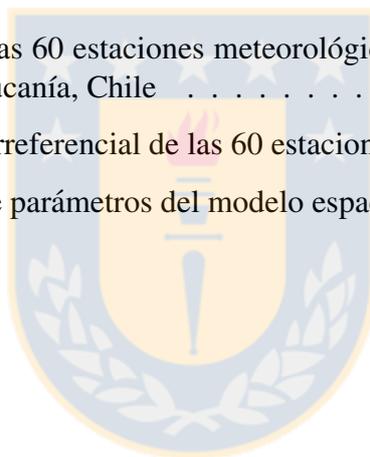


Índice de figuras

4.1. Región de estudio donde se ubican las 60 estaciones meteorológicas	35
4.2. Gráfico de dispersión para detectar componente sistemática	40
4.3. Imputación de información perdida	44
4.4. Mapa de predicción de temperatura promedio, año 2015	45
4.5. Mapa de predicción de temperatura promedio, año 2016	46
4.6. Mapa de predicción de temperatura promedio, año 2017	47
4.7. Mapa de varianza predicción de temperatura promedio, año 2015	48
4.8. Mapa de varianza predicción de temperatura promedio, año 2016	49
4.9. Mapa de varianza predicción de temperatura promedio, año 2017	50

Índice de tablas

3.1. Resultados de las estimaciones vía Máxima Verosimilitud para un modelo espacio-tiempo AR(1)	32
3.2. Resultados de las estimaciones vía algoritmo Esperanza-Maximización General para un modelo espacio-tiempo AR(1)	33
4.1. Resumen de las 60 estaciones meteorológicas en la región del Maule, Biobío y Araucanía, Chile	38
4.2. Resumen georreferencial de las 60 estaciones meteorológicas	41
4.3. Estimación de parámetros del modelo espacio-tiempo propuesto	43



Resumen

Mediante el uso del filtro de Kalman, se propuso un algoritmo para proporcionar un análisis estadístico formal de datos espacio-tiempo con una estructura autorregresiva en el tiempo. El filtro de Kalman, a través de la representación espacio-estado, permite capturar tanto la dependencia temporal como la estructura de correlación espacial, con el objetivo de realizar inferencia estadística en términos de estimación de parámetros y predicción en ubicaciones no observadas.

Este algoritmo se extendió para el caso en que la muestra presente datos faltantes, con la finalidad de lograr estimaciones lo más precisas posible de los parámetros involucrados en el modelo espacio-tiempo, así como predicciones confiables bajo estas condiciones.

Estos hallazgos se ilustraron a través de una aplicación sobre la temperatura promedio diaria del aire en algunas regiones del sur de Chile, donde el conjunto de datos presenta datos faltantes en diversos lugares de la muestra. Finalmente, logrando modelar y caracterizar el comportamiento, en términos de fuentes de variación espacio-tiempo, de este proceso ambiental dentro de esta zona de estudio.

Capítulo 1

Introducción

Considere un proceso estocástico espacio-temporal general

$$\mathbf{Z}_{\mathcal{D} \times \mathcal{T}} = \{Z(\mathbf{s}, t) : (\mathbf{s}, t) \in \mathcal{D} \times \mathcal{T} \subset \mathbb{R}^d \times \mathbb{R}\},$$

donde \mathcal{D} es el conjunto de índices de la ubicación espacial \mathbf{s} y \mathcal{T} es el conjunto de índices correspondiente al tiempo t .

El enfoque geoestadístico modela las observaciones como una realización parcial de un proceso estocástico espacio-temporal típicamente Gaussiano. Frecuentemente los datos son observados en instantes de tiempo equidistante, por ende podría ser suficiente modelar un proceso espacio-temporal en $\mathbb{R}^d \times \mathbb{Z}$. En efecto, en este proyecto se considerará \mathcal{D} como un subconjunto fijo del plano euclídeo, y \mathcal{T} será discreto. De esta manera, nos estaremos enfrentando ante un proceso estocástico espacio-temporal $\mathbf{Z}_{\mathcal{D} \times \mathcal{T}} = \{Z_t(\mathbf{s}) : (\mathbf{s}, t) \in \mathcal{D} \times \mathcal{T} \subset \mathbb{R}^2 \times \mathbb{Z}\}$.

Generalmente, los procesos espacio-temporales están asociados a fenómenos de índole ambiental, meteorológicos y geofísicos entre otros, como por ejemplo la temperatura del aire, precipitaciones, contaminantes, calidad del aire, recursos hídricos, vientos en la superficie, temperatura sobre el mar, etc. Lo fundamental radica en intentar

modelar y explicar estos fenómenos en términos de fuentes de variación identificables de la forma más exacta posible.

En la actualidad, esta información es recopilada en estaciones, donde lo ideal es esperar que toda la información se encuentre disponible. Sin embargo, hay situaciones la cual dificulta la recopilación de dicha información, ya sea por eventos propios de la naturaleza, por condiciones técnicas, recursos monetarias o humanos.

Bajo esta problemática, es necesaria la implementación de algoritmos adecuados que puedan lograr un cálculo confiable en la estimación de parámetros para el modelamiento espacio-temporal, además de un método confiable de predicción espacio-temporal, el cual sería el objetivo principal de cualquier estudio de esta índole.



Capítulo 2

Marco Teórico

2.1. El modelo Espacio-Tiempo AR(p)

Sea $\mathbf{Z}_{\mathcal{D} \times \mathcal{T}} := \{\mathbf{Z}_t(\mathbf{s}) : \mathbf{s} \in \mathcal{D}, t \in \mathcal{T}\}$ un proceso espacio-tiempo con $\mathcal{D} \subset \mathbb{R}^2$ y $\mathcal{T} \subset \mathbb{Z}^+$. Considere el siguiente modelo para el dato observado

$$\mathbf{Z}_t(\mathbf{s}) = \mu_t(\mathbf{s}) + \varepsilon_t(\mathbf{s}) + \omega_t(\mathbf{s}), \quad (2.1)$$

donde $\mu_t(\mathbf{s})$ es una componente sistemática que explica la mayor parte de la variación de $\mathbf{Z}_t(\mathbf{s})$, $\omega_{\mathcal{D} \times \mathcal{T}} = \{\omega_t(\mathbf{s}) : \mathbf{s} \in \mathcal{D}, t \in \mathcal{T}\}$ es un proceso ruido blanco Gaussiano de media cero y varianza σ_ω^2 , usualmente conocida como efecto pepita, ver Cressie and Wikle, 2015, el proceso de estado no observable $\varepsilon_{\mathcal{D} \times \mathcal{T}} = \{\varepsilon_t(\mathbf{s}) : \mathbf{s} \in \mathcal{D}, t \in \mathcal{T}\}$ se asume de media cero, espacial y temporalmente estacionario, y que evoluciona con una dinámica autorregresiva de orden p , como se especifica en Huang and Cressie, 1996, i.e.,

$$\varepsilon_t(\mathbf{s}) = \phi_1 \varepsilon_{t-1}(\mathbf{s}) + \phi_2 \varepsilon_{t-2}(\mathbf{s}) + \dots + \phi_p \varepsilon_{t-p}(\mathbf{s}) + \eta_t(\mathbf{s}), \quad (2.2)$$

donde los coeficientes $\phi_1, \phi_2, \dots, \phi_p$ son escogidos de tal forma que el valor absoluto de todas las raíces de $\lambda^p - \sum_{m=1}^p \phi_m \lambda^{p-m} = 0$ sean menores que 1 con la finalidad de lograr estacionariedad temporal. Finalmente $\boldsymbol{\eta}_{\mathcal{D} \times \mathcal{T}} = \{\eta_t(\mathbf{s}) : \mathbf{s} \in \mathcal{D}, t \in \mathcal{T}\}$ es un proceso Gaussiano de media cero, espacial y temporalmente estacionario, independiente de $\boldsymbol{\omega}_{\mathcal{D} \times \mathcal{T}}$, y es caracterizado por la función de covarianza espacio-temporal

$$\text{Cov}[\eta_t(\mathbf{s}), \eta_{t'}(\mathbf{r})] = \begin{cases} \mathcal{C}^\eta(\mathbf{s}, \mathbf{r}) & , \text{ si } t = t' \\ 0 & , \text{ si } t \neq t' \end{cases} \quad (2.3)$$

para $\mathbf{s}, \mathbf{r} \in \mathcal{D}$, donde la función de covarianza puramente espacial, $\mathcal{C}^\eta(\mathbf{s}, \mathbf{r})$, es especificada de una familia paramétrica de funciones de covarianza espaciales admisibles.

2.2. Estacionariedad

Sea el proceso estocástico espacio-temporal $\mathbf{Z}_{\mathcal{D} \times \mathcal{T}}$. Formalmente éste se define mediante el conjunto de distribuciones finito-dimensionales

$$F_{\mathbf{Z}(\mathbf{s}_1, t_1), \dots, \mathbf{Z}(\mathbf{s}_n, t_T)}(z_1, \dots, z_{n \times T}) \equiv \text{P}(\mathbf{Z}(\mathbf{s}_1, t_1) \leq z_1, \dots, \mathbf{Z}(\mathbf{s}_n, t_T) \leq z_{n \times T}),$$

para cualquier n puntos del espacio $\mathbf{s}_1, \dots, \mathbf{s}_n$, T instantes en el tiempo t_1, \dots, t_T y $z_i \in R_z \forall i = 1, \dots, n \times T$.

2.2.1. Estacionariedad fuerte

El proceso estocástico espacio-temporal es estrictamente estacionario en espacio y tiempo si

$$F_{\mathbf{Z}(\mathbf{s}_1, t_1), \dots, \mathbf{Z}(\mathbf{s}_n, t_T)}(z_1, \dots, z_{n \times T}) = F_{\mathbf{Z}(\mathbf{s}_1 + \mathbf{h}, t_1 + k), \dots, \mathbf{Z}(\mathbf{s}_n + \mathbf{h}, t_T + k)}(z_1, \dots, z_{n \times T}),$$

es decir, su función de distribución de probabilidad conjunta es invariante bajo operaciones de traslación respecto a un vector \mathbf{h} y a un resago k . Este supuesto, en la práctica es imposible de determinar y sólo se puede esperar inferir los primeros momentos de la distribución de $Z(\mathbf{s}, t)$, $s \in \mathcal{D}$ y $t \in \mathcal{T}$.

2.2.2. Estacionariedad débil o de segundo orden

El proceso estocástico espacio-temporal $\mathbf{Z}_{\mathcal{D} \times \mathcal{T}}$ es al menos, estacionario de segundo orden si su media es constante y su función de auto covarianza depende únicamente de las localidades \mathbf{s} y \mathbf{r} sólo a través de la distancia espacial euclídea $\mathbf{h} = \|\mathbf{s} - \mathbf{r}\| \in \mathbb{R}$, implicando, además, un proceso isotrópico (Cressie and Wikle, 2015), y en el tiempo a través del retraso temporal $k = |t - t'| \in \mathbb{N}$; esto es

$$E[Z(\mathbf{s}, t)] = \mu, \quad \forall t, \mathbf{s},$$

$$\text{Cov}[Z(\mathbf{s}, t), Z(\mathbf{r}, t')] = \mathcal{C}(\|\mathbf{s} - \mathbf{r}\|, |t - t'|), \quad \forall t, t', \mathbf{s}, \mathbf{r}.$$

Observe que cuando el proceso es Gaussiano, estacionariedad de segundo orden y estacionariedad fuerte coinciden, porque un proceso Gaussiano es caracterizado por su media y su función de covarianza.

2.2.3. Función de covarianza espacio-tiempo del modelo

Debido a la estacionariedad temporal, vemos que al post-multiplicar (2.2) por $\varepsilon_{t-k}(\mathbf{r})$ y tomando esperanza, tenemos

$$C_k^\varepsilon(\mathbf{s}, \mathbf{r}) = \sum_{m=1}^p \phi_m C_{k-m}^\varepsilon(\mathbf{s}, \mathbf{r}) + \delta_{\{k=0\}} C^n(\mathbf{s}, \mathbf{r}), \quad k = 0, 1, 2, \dots \quad (2.4)$$

con $\delta_{\{t\}}$ la función indicadora, así obteniéndose las ecuaciones de Yule-Walker del proceso $\varepsilon_{\mathcal{D} \times \mathcal{T}}$. Observe que las primeras $(p + 1)$ ecuaciones pueden ser resueltas fácilmente

$$\begin{aligned} \mathcal{C}_0^\varepsilon(\mathbf{s}, \mathbf{r}) &= \sum_{m=1}^p \phi_m \mathcal{C}_{-m}^\varepsilon(\mathbf{s}, \mathbf{r}) + \mathcal{C}^\eta(\mathbf{s}, \mathbf{r}) \\ \mathcal{C}_1^\varepsilon(\mathbf{s}, \mathbf{r}) &= \sum_{m=1}^p \phi_m \mathcal{C}_{1-m}^\varepsilon(\mathbf{s}, \mathbf{r}) \\ \mathcal{C}_2^\varepsilon(\mathbf{s}, \mathbf{r}) &= \sum_{m=1}^p \phi_m \mathcal{C}_{2-m}^\varepsilon(\mathbf{s}, \mathbf{r}) \end{aligned} \tag{2.5}$$

⋮

$$\mathcal{C}_p^\varepsilon(\mathbf{s}, \mathbf{r}) = \sum_{m=1}^p \phi_m \mathcal{C}_{p-m}^\varepsilon(\mathbf{s}, \mathbf{r}),$$

las restante $p < t \leq T$ ecuaciones pueden obtenerse de forma recursiva. De esta forma se puede obtener la relación

$$\mathcal{C}_k^\varepsilon(\mathbf{s}, \mathbf{r}) = f(\phi_1, \dots, \phi_p, k) \mathcal{C}^\eta(\mathbf{s}, \mathbf{r}). \tag{2.6}$$

En particular, si consideramos un modelo espacio-tiempo AR(1), de (2.5) se tiene

$$\begin{aligned} \mathcal{C}_0^\varepsilon(\mathbf{s}, \mathbf{r}) &= \phi_1 \mathcal{C}_1^\varepsilon(\mathbf{s}, \mathbf{r}) + \mathcal{C}^\eta(\mathbf{s}, \mathbf{r}) \\ \mathcal{C}_1^\varepsilon(\mathbf{s}, \mathbf{r}) &= \phi_1 \mathcal{C}_0^\varepsilon(\mathbf{s}, \mathbf{r}). \end{aligned}$$

Así, la función de covarianza para el proceso espacio-temporal AR(1), temporalmente estacionario, viene dado por

$$\mathcal{C}_k^\varepsilon(\mathbf{s}, \mathbf{r}) = \frac{\phi^k}{(1 - \phi^2)} \mathcal{C}^\eta(\mathbf{s}, \mathbf{r}), \quad k = 0, 1, \dots \tag{2.7}$$

La función $\mathcal{C}^\eta(\mathbf{s}, \mathbf{r})$ debe ser definida positiva para garantizar una varianza de error de predicción no negativa. Esto es, para cualquier número finito n de ubicaciones es-

paciales y cualquier conjunto de números reales, a_1, a_2, \dots, a_m con $m \in \mathbb{Z}_+$, $\mathcal{C}^\eta(\mathbf{s}, \mathbf{r})$ debe satisfacer

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j \mathcal{C}^\eta(\mathbf{s}_i, \mathbf{s}_j) \geq 0, \quad \mathbf{s}_i, \mathbf{s}_j \in \mathcal{D},$$

Para asegurar este cumplimiento, con frecuencia se especifica una función de covarianza espacial que pertenezca a una familia paramétrica, en la cual todas las funciones cumplen dicha propiedad.

Así, desde (2.3), la función de covarianza $\mathcal{C}^\eta(\mathbf{s}, \mathbf{r}) \equiv \mathcal{C}^\eta(\mathbf{s}, \mathbf{r} | \sigma_\eta^2, \boldsymbol{\theta})$ puede ser escrita como el producto de un parámetro de varianza σ_η^2 y una función de correlación espacial definida positiva $\rho^\eta(\mathbf{s}, \mathbf{r} | \boldsymbol{\theta})$ para todo vector de parámetros $\boldsymbol{\theta}$.

Una opción usual es adoptar la función de correlación espacial Matérn (Matérn, 2013) dada por

$$\rho^\eta(\mathbf{s}, \mathbf{r}) = \frac{1}{2^{\nu-1} \Gamma(\nu)} \left(\frac{\|\mathbf{s} - \mathbf{r}\|}{\alpha} \right)^\nu \mathcal{K}_\nu \left(\frac{\|\mathbf{s} - \mathbf{r}\|}{\alpha} \right), \quad \alpha > 0, \nu > 0,$$

donde \mathcal{K} es la función de Bessel modificada de la segunda clase de orden ν . El parámetro α , generalmente llamado alcance, es un parámetro de escala, la cual controla el grado de decaimiento de la correlación espacial entre dos puntos, valores grandes de éste, relativo a la máxima distancia entre dos puntos dentro del polígono de estudio, indica que la alta correlación espacial entre dos puntos sigue siendo persistente a pesar de aumentar la distancia entre estos. Finalmente, el parámetro ν es un parámetro de suavizamiento que controla la diferenciabilidad en media cuadrática de $\eta_t(\cdot)$. Usualmente, para efecto de simplicidad de cómputo, ν se considera fijo y conocido, obteniéndose casos particulares de la familia Matérn. Uno de estos es el modelo exponencial

$$\rho^\eta(\mathbf{s}, \mathbf{r}) = \exp \left(-\frac{\|\mathbf{s} - \mathbf{r}\|}{\alpha} \right), \quad \alpha > 0,$$

la cual es recuperado cuando $\nu = 1/2$. Por otro lado, como caso límite, se tiene el

modelo Gaussiano

$$\rho^\eta(\mathbf{s}, \mathbf{r}) = \exp\left(-\left(\frac{\|\mathbf{s} - \mathbf{r}\|}{\alpha}\right)^2\right), \quad \alpha > 0,$$

la cual es recuperado cuando $\nu \rightarrow \infty$. Finalmente, cuando $\nu = 3/2$, se puede obtener el modelo

$$\rho^\eta(\mathbf{s}, \mathbf{r}) = (1 + \alpha\|\mathbf{s} - \mathbf{r}\|) \exp\left(-\frac{\|\mathbf{s} - \mathbf{r}\|}{\alpha}\right), \quad \alpha > 0.$$

Observe que la función de covarianza espacio-temporal (2.7) puede ser escrita como

$$C_k^\varepsilon(\mathbf{s}, \mathbf{r}) = \sigma_\varepsilon^2 \rho^T(k) \rho^S(\mathbf{s}, \mathbf{r}),$$

con $\sigma_\varepsilon^2 = \sigma_\eta^2 / (1 - \phi^2)$, $\rho^T(k) = \phi^k$ y $\rho^S(\mathbf{s}, \mathbf{r}) := \rho^\eta(\mathbf{s}, \mathbf{r})$, la cual corresponde a un modelo de covarianza espacio-temporal del tipo separable (Gneiting et al., 2006).

2.3. Representación Espacio-Estado del modelo

Sea $Z_t(\mathbf{s})$ la variable observada en la localidad \mathbf{s} y tiempo t . Sea $\boldsymbol{\xi}_t(\mathbf{s})$ un vector de $(p \times 1)$, este vector es conocido como el vector de estado, la cual es no observable.

La representación espacio-estado de la dinámica de $\mathbf{Z}_{\mathcal{D} \times \mathcal{T}}$ es dada por el siguiente sistema de ecuaciones:

$$\boldsymbol{\xi}_t(\mathbf{s}) = \mathbf{F}\boldsymbol{\xi}_{t-1}(\mathbf{s}) + \mathbf{V}_t(\mathbf{s}), \quad (2.8a)$$

$$Z_t(\mathbf{s}) = \boldsymbol{\beta}^\top \mathbf{X}_t(\mathbf{s}) + \mathbf{H}^\top \boldsymbol{\xi}_t(\mathbf{s}) + \omega_t(\mathbf{s}), \quad (2.8b)$$

donde la ecuación (2.8a) describe un proceso Markoviano de primer orden, la cual es comúnmente llamada *ecuación de estado*; y la ecuación (2.8b), *ecuación de observación*.

En la ecuación de observación, $Z_t(\mathbf{s})$ es la observación en el instante de tiempo t y localidad \mathbf{s} , $\boldsymbol{\beta} = (\beta_1, \dots, \beta_l)^\top$ es un vector de coeficientes l -dimensional asociado al vector l -dimensional de regresores no-estocásticos $\mathbf{X}_t(\mathbf{s}) = (X_t^{(1)}(\mathbf{s}), \dots, X_t^{(l)}(\mathbf{s}))^\top$, \mathbf{H} es un vector p -dimensional, usualmente llamado como operador de observación, la cual mapea el proceso a la observación, $\boldsymbol{\xi}_t(\mathbf{s})$ es un vector de estado p -dimensional, y $\omega_t(\mathbf{s})$ fue definido en la ecuación (2.1). Por otro lado, en la ecuación de estado, \mathbf{F} es una matriz $p \times p$ -dimensional, usualmente llamada operador de transición de estado, y $\mathbf{V}_t(\mathbf{s})$ es un vector de arreglo p -dimensional para la innovación espacial.

Hay muchas formas de modelar el sistema dinámico. La representación espacio-estado que reproducirá el modelo espacio-temporal a analizar tendrá la siguiente estructura

$$\boldsymbol{\xi}_t(\mathbf{s}) = \begin{pmatrix} \varepsilon_t(\mathbf{s}) \\ \varepsilon_{t-1}(\mathbf{s}) \\ \vdots \\ \varepsilon_{t-p}(\mathbf{s}) \\ \varepsilon_{t-(p-1)}(\mathbf{s}) \end{pmatrix}, \quad \mathbf{V}_t(\mathbf{s}) = \begin{pmatrix} \eta_t(\mathbf{s}) \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad s \in \mathcal{D}, t \in \mathcal{T} \quad (2.9)$$

$$\mathbf{H} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}. \quad (2.10)$$

2.4. Implementación e inferencia

Durante esta sección se asumirá que todos los parámetros son fijos y conocidos, y enfocaremos nuestra discusión en la implementación e inferencia del proceso de esta-

do. Así, al denotar distribuciones condicionales, se elegirá no mostrar los parámetros a menos que sea absolutamente necesario.

Denotando toda la información muestral en el instante t por el vector n_t -dimensional de observaciones espaciales \mathbf{Z}_t y la representación vectorial np -dimensional del proceso de estado para las n localidades por $\boldsymbol{\xi}_t = (\boldsymbol{\xi}_t(\mathbf{s}_1)^\top, \dots, \boldsymbol{\xi}_t(\mathbf{s}_n)^\top)^\top$. Una forma natural de extender nuestra representación espacio estado en (2.8a) y (2.8b) es dada por

$$\boldsymbol{\xi}_t = \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1} + \mathbf{V}_t, \quad (2.11a)$$

$$\mathbf{Z}_t = \mathbf{X}_t\boldsymbol{\beta} + \Lambda_t\boldsymbol{\xi}_t + \mathbf{W}_t, \quad (2.11b)$$

donde, desde la ecuación de estado (2.11a), podemos extender el operador de evolución como $\boldsymbol{\Phi} = \mathbf{I}_n \otimes \mathbf{F}$ y el vector de innovación espacial como $\mathbf{V}_t = (\mathbf{V}_t(\mathbf{s}_1)^\top, \dots, \mathbf{V}_t(\mathbf{s}_n)^\top)^\top$. Desde la ecuación de observación (2.11b), \mathbf{W}_t es un vector de error, de dimensión $(n_t \times 1)$, asociado a las observaciones en las n_t localidades observadas en el instante t , \mathbf{X}_t es la correspondiente $n_t \times l$ matriz de covariables observadas en el instante t , y, finalmente, para tratar los datos faltantes asociado a cada instante de tiempo, el operador de observación extendido Λ_t puede ser escrito como $\Lambda_t = \mathbf{L}_t^{\text{obs}} \otimes \mathbf{H}^\top$, donde

$$\mathbf{L}_t^{\text{obs}} := \left(\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \dots, \mathbf{e}_{i_{n_t}} \right)^\top,$$

de dimensión $(n_t \times n)$, corresponde a una parte de la matriz identidad de tamaño n , esto es, $\{\mathbf{e}_{i_j}, i_j \in I_t\}$ es un conjunto de vectores canónicos de dimensión $(1 \times n)$, con $I_t = \{i_j, j = 1, \dots, n_t\}$ el conjunto de índices espaciales asociado a las localidades $\mathbf{s}_{i_j} \forall j = 1, \dots, n_t$ donde hubo registro en el tiempo t .

Finalmente, implicando el modelo de datos $[\mathbf{Z}_1, \dots, \mathbf{Z}_T | \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T]$, donde, por simplicidad de notación, se se escribirá $\mathbf{Z}_{1:T} := \{\mathbf{Z}_1, \dots, \mathbf{Z}_T\}$ y $\boldsymbol{\xi}_{1:T} := \{\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T\}$.

2.4.1. Implementación secuencial

Como asumimos que el proceso de estado tiene evolución Markoviana de primer orden, especificamos la *distribución de evolución de estado*, $[\xi_t|\xi_{t-1}]$, por las densidades $f(\xi_1)$ y $f(\xi_t|\xi_{t-1})$ para $t = 2, \dots, T$, así

$$f(\xi_{1:T}) = f(\xi_1) \prod_{t=2}^T f(\xi_t|\xi_{t-1}). \quad (2.12)$$

Por otro lado, asumiremos que las observaciones en $\mathbf{Z}_{1:t}$ son condicionalmente independientes, dado $\xi_{1:t}$, y para un t fijo, la función de densidad condicional de \mathbf{Z}_t , dado $\xi_{1:t}$, sólo depende de ξ_t ; $t = 1, 2, \dots, T$. Así, la distribución del modelo de datos es especificada por las funciones de densidades

$$f(\mathbf{Z}_{1:T}|\xi_{1:T}) = \prod_{t=1}^T f(\mathbf{Z}_t|\xi_t), \quad (2.13)$$

donde $f(\mathbf{Z}_t|\xi_t)$ define lo que usualmente es llamada *distribución de medición*, $[\mathbf{Z}_t|\xi_t]$. Por simplicidad de notación, omitimos \mathbf{X}_t en $[\mathbf{Z}_t|\mathbf{X}_t, \xi_t]$, así $[\mathbf{Z}_t|\mathbf{X}_t, \xi_t] \equiv [\mathbf{Z}_t|\xi_t]$.

Bajo el supuesto (2.12) de la distribución de evolución y el supuesto (2.13) del modelo de datos, se puede usar la regla de Bayes para encontrar la *distribución de filtro y predicción*. Esto es, para predecir $\xi_{1:T}$ en (2.12), dada las observaciones $\mathbf{Z}_{1:T}$, podemos obtener la distribución a posteriori de $\xi_{1:T}$ como

$$f(\xi_{1:T}|\mathbf{Z}_{1:T}) \propto f(\mathbf{Z}_1|\xi_1)f(\xi_1) \prod_{t=2}^T f(\mathbf{Z}_t|\xi_t)f(\xi_t|\xi_{t-1}). \quad (2.14)$$

La clave del enfoque secuencial es que, a medida que se dispongan nuevos datos, uno busca actualizar la distribución de filtro previa. Esto es, la distribución a posteriori basada en los datos hasta el momento $t - 1$ se convierte en la priori para la distribución

de datos hasta el tiempo t . Específicamente, buscamos la *distribución de predicción*

$$[\boldsymbol{\xi}_t | \mathbf{Z}_{1:t-1}], \quad (2.15)$$

y la *distribución de filtro*

$$[\boldsymbol{\xi}_t | \mathbf{Z}_{1:t}], \quad (2.16)$$

para $t = 1, \dots, T$.

La *distribución de predicción* se puede obtener considerándola en términos de la distribución de filtro en el momento anterior:

$$f(\boldsymbol{\xi}_t | \mathbf{Z}_{1:t-1}) = \int f(\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1} | \mathbf{Z}_{1:t-1}) d\boldsymbol{\xi}_{t-1} \quad (2.17a)$$

$$= \int f(\boldsymbol{\xi}_t | \boldsymbol{\xi}_{t-1}) f(\boldsymbol{\xi}_{t-1} | \mathbf{Z}_{1:t-1}) d\boldsymbol{\xi}_{t-1}, \quad (2.17b)$$

donde la segunda igualdad deriva del supuesto de independencia condicional. La primera densidad en el integrando de (2.17b) está asociada a la *distribución de evolución* de proceso; y la segunda, a la *distribución de filtro* (posteriori) del estado en el tiempo previo. La distribución de predicción a k pasos en el tiempo puede ser obtenida similarmente.

La *distribución de filtro* puede ser obtenida a través de la regla de Bayes:

$$f(\boldsymbol{\xi}_t | \mathbf{Z}_{1:t}) = f(\boldsymbol{\xi}_t | \mathbf{Z}_t, \mathbf{Z}_{1:t-1}) \quad (2.18a)$$

$$\propto f(\mathbf{Z}_t | \boldsymbol{\xi}_t, \mathbf{Z}_{1:t-1}) f(\boldsymbol{\xi}_t | \mathbf{Z}_{1:t-1}) \quad (2.18b)$$

$$= f(\mathbf{Z}_t | \boldsymbol{\xi}_t) f(\boldsymbol{\xi}_t | \mathbf{Z}_{1:t-1}), \quad (2.18c)$$

donde la primera densidad en el lado derecho de (2.18c) se obtiene del modelo de datos, y la segunda densidad está asociada a la distribución de predicción (2.17b). Si iteramos entre la distribución de filtro y la distribución de predicción a medida

que nueva información se encuentre disponible, se obtiene la siguiente secuencia de distribuciones:

$$[\xi_1], [\xi_1|\mathbf{Z}_1], [\xi_2|\mathbf{Z}_1], [\xi_2|\mathbf{Z}_{1:2}], \dots, [\xi_T|\mathbf{Z}_{1:T-1}], [\xi_T|\mathbf{Z}_{1:T}],$$

procedimiento secuencial conocido como filtraje.

El término *suavizamiento* hace referencia a obtener $[\xi_t|\mathbf{Z}_{1:T}]$, la distribución del estado en cualquier instante de tiempo $t \in \{1, 2, \dots, T\}$ dada toda la información desde el instante 1 hasta T , la cual incluye la colección de datos posteriores al instante t . Obviamente, para $t = T$, la distribución de suavizamiento corresponde a la distribución de filtro.

La *distribución de suavizamiento* puede ser obtenida escribiéndola como

$$f(\xi_t|\mathbf{Z}_{1:T}) = \int f(\xi_t|\xi_{t+1}, \mathbf{Z}_{1:T})f(\xi_{t+1}|\mathbf{Z}_{1:T})d\xi_{t+1}. \quad (2.19)$$

La primera distribución del lado derecho de (2.19) puede ser escrita como

$$f(\xi_t|\xi_{t+1}, \mathbf{Z}_{1:T}) = f(\xi_t|\xi_{t+1}, \mathbf{Z}_{1:t}), \quad (2.20)$$

ya que $\{\mathbf{Z}_{t+1}, \dots, \mathbf{Z}_T\}$ son independientes de ξ_t dado ξ_{t+1} . Ahora, note que con la regla de Bayes obtenemos,

$$f(\xi_t|\xi_{t+1}, \mathbf{Z}_{1:t}) \propto f(\xi_{t+1}|\xi_t, \mathbf{Z}_{1:t})f(\xi_t|\mathbf{Z}_{1:t}) \quad (2.21a)$$

$$= f(\xi_{t+1}|\xi_t)f(\xi_t|\mathbf{Z}_{1:t}), \quad (2.21b)$$

donde el primer término en el lado derecho de (2.21b) está asociado a la distribución de evolución, y el segundo término, a la de filtro para el instante de tiempo t .

2.5. Inferencia para el proceso espacio-temporal

2.5.1. Modelos lineales/gaussianos

En general, podría ser difícil obtener formulación analítica explícita para las distribuciones de predicción, filtro y suavizamiento (2.15), (2.16), y (2.19), respectivamente. Sin embargo, en el caso de especificar un operador de observación y de evolución que sean lineales, y distribuciones Gaussianas de los errores, podemos obtener las distribuciones de predicción, filtro y suavizamiento explícitamente. Estos resultados son los conocidos *Filtro de Kalman* y *Suavizador de Kalman*.

2.5.2. Filtro de Kalman

El filtro de Kalman es ideal para una actualización secuencial, donde los operadores son lineales y las distribuciones de los errores son Gaussianos (e.g., Kalman, 1960; Jazwinski, 1970; Anderson and Moore, 1979; West and Harrison, 1997; Wikle and Cressie, 1999). Se puede derivar de muchas perspectivas diferentes. Acá, se usará la fórmula Bayesiana general.

Denotando la esperanza condicional para las distribuciones de filtro y predicción como $\xi_{t|t} := E[\xi_t | \mathbf{Z}_{1:t}]$ y $\xi_{t|t-1} := E[\xi_t | \mathbf{Z}_{1:t-1}]$. Y las matrices de covarianza condicional de error para el filtro y predicción, respectivamente, como

$$\begin{aligned}\mathbf{P}_{t|t} &:= \text{Var}[\xi_t | \mathbf{Z}_{1:t}] = E[(\xi_t - \xi_{t|t})(\xi_t - \xi_{t|t})^\top | \mathbf{Z}_{1:t}], \\ \mathbf{P}_{t|t-1} &:= \text{Var}[\xi_t | \mathbf{Z}_{1:t-1}] = E[(\xi_t - \xi_{t|t-1})(\xi_t - \xi_{t|t-1})^\top | \mathbf{Z}_{1:t-1}].\end{aligned}$$

Considere la distribución de medición y de evolución dada, respectivamente, por los

modelos

$$\mathbf{Z}_t | \boldsymbol{\xi}_t \sim \text{Gau}(\mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\Lambda}_t \boldsymbol{\xi}_t, \sigma_\omega^2 \mathbf{I}_{n_t}), \quad (2.22a)$$

$$\boldsymbol{\xi}_t | \boldsymbol{\xi}_{t-1} \sim \text{SGau}(\boldsymbol{\Phi} \boldsymbol{\xi}_{t-1}, \mathbf{Q}), \quad (2.22b)$$

donde *SGau* hace referencia a la distribución gaussiana multivariada singular (Kwong and Iglewicz (1996)), cuya matriz de covarianzas, \mathbf{Q} , se puede escribir como $\mathbf{Q} = \sigma_\eta^2 \mathbf{R}$, con $\mathbf{R} = \boldsymbol{\rho}^\eta \otimes \mathbf{T}_p$, donde $\boldsymbol{\rho}^\eta$ y \mathbf{T}_p vienen dados por

$$\boldsymbol{\rho}^\eta = \begin{pmatrix} \rho^\eta(\mathbf{s}_1, \mathbf{s}_2) & \rho^\eta(\mathbf{s}_1, \mathbf{s}_2) & \dots & \rho^\eta(\mathbf{s}_1, \mathbf{s}_n) \\ \rho^\eta(\mathbf{s}_2, \mathbf{s}_1) & \rho^\eta(\mathbf{s}_2, \mathbf{s}_2) & \dots & \rho^\eta(\mathbf{s}_2, \mathbf{s}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \rho^\eta(\mathbf{s}_n, \mathbf{s}_1) & \rho^\eta(\mathbf{s}_n, \mathbf{s}_2) & \dots & \rho^\eta(\mathbf{s}_n, \mathbf{s}_n) \end{pmatrix}, \quad \mathbf{T}_p = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}.$$

Usando el operador de la esperanza y varianza condicional, se puede encontrar la *distribución de predicción* $[\boldsymbol{\xi}_t | \mathbf{Z}_{1:t-1}]$

$$\boldsymbol{\xi}_t | \mathbf{Z}_{1:t-1} \sim \text{Gau}(\boldsymbol{\xi}_{t|t-1}, \mathbf{P}_{t|t-1}), \quad (2.23)$$

donde el vector de medias y a matriz de covarianzas vienen, respectivamente, dados por

$$\boldsymbol{\xi}_{t|t-1} = \mathbf{E}[\mathbf{E}[\boldsymbol{\xi}_t | \boldsymbol{\xi}_{t-1}] | \mathbf{Z}_{1:t-1}] = \mathbf{E}[\boldsymbol{\Phi} \boldsymbol{\xi}_{t-1} | \mathbf{Z}_{1:t-1}] = \boldsymbol{\Phi} \boldsymbol{\xi}_{t-1|t-1}, \quad (2.24)$$

$$\begin{aligned} \mathbf{P}_{t|t-1} &= \text{Var}[\mathbf{E}[\boldsymbol{\xi}_t | \boldsymbol{\xi}_{t-1}] | \mathbf{Z}_{1:t-1}] + \mathbf{E}[\text{Var}[\boldsymbol{\xi}_t | \boldsymbol{\xi}_{t-1}] | \mathbf{Z}_{1:t-1}] \\ &= \text{Var}[\boldsymbol{\Phi} \boldsymbol{\xi}_{t-1} | \mathbf{Z}_{1:t-1}] + \mathbf{E}[\mathbf{Q} | \mathbf{Z}_{1:t-1}] \\ &= \boldsymbol{\Phi} \mathbf{P}_{t-1|t-1} \boldsymbol{\Phi}^\top + \mathbf{Q}. \end{aligned} \quad (2.25)$$

Por consiguiente, podemos obtener $[\mathbf{Z}_t | \mathbf{Z}_{1:t-1}]$, la cual está dada por

$$\mathbf{Z}_t | \mathbf{Z}_{1:t-1} \sim \text{Gau}(\mathbf{Z}_{t|t-1}, \Delta_{t|t-1}), \quad (2.26)$$

donde su vector de medias y matriz de covarianzas vienen, respectivamente, dadas por

$$\begin{aligned} \mathbf{Z}_{t|t-1} &= \mathbf{X}_t \boldsymbol{\beta} + \Lambda_t \boldsymbol{\xi}_{t|t-1}, \\ \Delta_{t|t-1} &= \Lambda_t \mathbf{P}_{t|t-1} \Lambda_t^\top + \sigma_\omega^2 \mathbf{I}_{n_t}. \end{aligned}$$

Por otro lado, dada nueva información en el sistema, digamos \mathbf{Z}_t , desde la distribución $[\boldsymbol{\xi}_t, \mathbf{Z}_t | \mathbf{Z}_{1:t-1}]$, la cual es Gaussiana con vector de medias y matriz de covarianzas dadas, respectivamente, por

$$\begin{pmatrix} \boldsymbol{\xi}_{t|t-1} \\ \mathbf{Z}_{t|t-1} \end{pmatrix}, \begin{pmatrix} \mathbf{P}_{t|t-1} & \mathbf{G}_{t|t-1} \\ \mathbf{G}_{t|t-1}^\top & \Delta_{t|t-1} \end{pmatrix},$$

con $\mathbf{G}_{t|t-1} := \text{Cov}[\boldsymbol{\xi}_t, \mathbf{Z}_t | \mathbf{Z}_{1:t-1}] = \mathbf{P}_{t|t-1} \Lambda_t^\top$, se logra obtener la *distribución de filtro*

$$\boldsymbol{\xi}_t | \mathbf{Z}_{1:t} \sim \text{Gau}(\boldsymbol{\xi}_{t|t}, \mathbf{P}_{t|t}), \quad (2.27)$$

donde el vector de medias y la matriz de covarianzas vienen dadas por

$$\boldsymbol{\xi}_{t|t} = \boldsymbol{\xi}_{t|t-1} + \mathbf{G}_{t|t-1} \Delta_{t|t-1}^{-1} (\mathbf{Z}_t - \mathbf{Z}_{t|t-1}), \quad (2.28a)$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{G}_{t|t-1} \Delta_{t|t-1}^{-1} \mathbf{G}_{t|t-1}^\top, \quad (2.28b)$$

(e.g, Searle, 1979, p.46), usualmente conocidas como ecuaciones de filtro-actualización.

Para el caso particular donde en el instante t no halla ninguna observación disponible, esto es, $n_t = 0$, el paso de filtro-actualización dado por (2.28a) y (2.28b) es simplemente constante, sin variabilidad, así las ecuaciones de actualización se simplifican a

$$\boldsymbol{\xi}_{t|t} = \boldsymbol{\xi}_{t|t-1}, \quad (2.29a)$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1}. \quad (2.29b)$$

2.5.3. Suavizador de Kalman

En algunas situaciones se tiene interés en la distribución de $\boldsymbol{\xi}_t$ en el momento $t \leq T$ dado todos los datos hasta T , esto es $[\boldsymbol{\xi}_t | \mathbf{Z}_{1:T}]$, la *distribución de suavizamiento* (posteriori) de $\boldsymbol{\xi}_t$. la distribución de suavizamiento es útil para un análisis retrospectivo.

Al igual que con la derivación del filtro de Kalman, sea $\boldsymbol{\xi}_{t|T} := E[\boldsymbol{\xi}_t | \mathbf{Z}_{1:T}]$ y $\mathbf{P}_{t|T} := \text{Var}[\boldsymbol{\xi}_t | \mathbf{Z}_{1:T}]$. Buscamos fórmulas secuenciales para $[\boldsymbol{\xi}_t | \mathbf{Z}_{1:T}]$ asumiendo que tenemos acceso a la distribución de filtro, $\boldsymbol{\xi}_t | \mathbf{Z}_{1:t} \sim \text{Gau}(\boldsymbol{\xi}_{t|t}, \mathbf{P}_{t|t})$, y la distribución de predicción, $\boldsymbol{\xi}_{t+1} | \mathbf{Z}_{1:t} \sim \text{Gau}(\boldsymbol{\xi}_{t+1|t}, \mathbf{P}_{t+1|t})$, para $t = 1, \dots, T$.

Bajo el supuesto de Gaussianidad en la *distribución de evolución y distribución de filtro*

$$\begin{aligned} \boldsymbol{\xi}_{t+1} | \boldsymbol{\xi}_t &\sim \text{SGau}(\Phi \boldsymbol{\xi}_t, \mathbf{Q}), \\ \boldsymbol{\xi}_t | \mathbf{Z}_{1:t} &\sim \text{Gau}(\boldsymbol{\xi}_{t|t}, \mathbf{P}_{t|t}), \end{aligned}$$

es posible encontrar la distribución asociada al primer término del lado derecho de (2.19), esto es

$$\boldsymbol{\xi}_t | \boldsymbol{\xi}_{t+1}, \mathbf{Z}_{1:t} \sim \text{Gau}(\boldsymbol{\xi}_{t|t} + \mathbf{J}_t(\boldsymbol{\xi}_{t+1} - \boldsymbol{\xi}_{t+1|t}), \mathbf{P}_{t|t} - \mathbf{J}_t \Phi \mathbf{P}_{t+1|t}),$$

con $\mathbf{J}_t := \mathbf{P}_{t|t} \Phi^\top \mathbf{P}_{t+1|t}^{-1}$.

Finalmente, podemos obtener la distribución de suavizamiento desde (2.19), específicamente

$$\boldsymbol{\xi}_t | \mathbf{Z}_{1:T} \sim \text{Gau}(\boldsymbol{\xi}_{t|T}, \mathbf{P}_{t|T}), \quad (2.30)$$

donde su vector de medias y matriz de covarianzas vienen, respectivamente, dadas por

$$\begin{aligned}
\xi_{t|T} &:= \mathbb{E}[\mathbb{E}[\xi_t | \xi_{t+1}, \mathbf{Z}_{1:t}] | \mathbf{Z}_{1:T}] \\
&= \mathbb{E}[\xi_{t|t} + \mathbf{J}_t(\xi_{t+1} - \xi_{t+1|t}) | \mathbf{Z}_{1:T}] \\
&= \xi_{t|t} + \mathbf{J}_t(\xi_{t+1|T} - \xi_{t+1|t}),
\end{aligned} \tag{2.31}$$

y

$$\begin{aligned}
\mathbf{P}_{t|T} &:= \mathbb{E}[\text{Var}[\xi_t | \xi_{t+1}, \mathbf{Z}_{1:t}] | \mathbf{Z}_{1:T}] + \text{Var}[\mathbb{E}[\xi_t | \xi_{t+1}, \mathbf{Z}_{1:t}] | \mathbf{Z}_{1:T}] \\
&= \mathbb{E}[\mathbf{P}_{t|t} - \mathbf{J}_t \Phi \mathbf{P}_{t|t} | \mathbf{Z}_{1:T}] + \text{Var}[\xi_{t|t} + \mathbf{J}_t(\xi_{t+1} - \xi_{t+1|t}) | \mathbf{Z}_{1:T}] \\
&= \mathbf{P}_{t|t} - \mathbf{J}_t \Phi \mathbf{P}_{t|t} + \mathbf{J}_t \text{Var}[\xi_{t+1} | \mathbf{Z}_{1:T}] \mathbf{J}_t^\top \\
&= \mathbf{P}_{t|t} - \mathbf{J}_t \mathbf{P}_{t+1|t} \mathbf{J}_t^\top + \mathbf{J}_t \mathbf{P}_{t+1|T} \mathbf{J}_t^\top \\
&= \mathbf{P}_{t|t} + \mathbf{J}_t (\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}) \mathbf{J}_t^\top.
\end{aligned} \tag{2.32}$$

2.5.4. Supuestos que sustentan el algoritmo

Para poder emplear tanto el filtro de Kalman como el suavizador de Kalman, se requiere asumir los siguientes supuestos estructurales.

El vector de innovación espacial $\mathbf{V}_t(\mathbf{s})$ no está autocorrelacionado en el tiempo, esto es

$$\text{Cov}[\mathbf{V}_t(\mathbf{s}), \mathbf{V}_{t'}(\mathbf{r})] = \begin{cases} \mathbf{Q}(\mathbf{s}, \mathbf{r}) & , \text{ si } t = t', \\ \mathbf{0} & , \text{ si } t \neq t'. \end{cases} \tag{2.33}$$

donde

$$\mathbf{Q}(\mathbf{s}, \mathbf{r}) = \begin{pmatrix} \mathcal{C}^\eta(\mathbf{s}, \mathbf{r}) & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}.$$

El error de medición, $\omega_t(\mathbf{s})$, no está autocorrelacionado en el tiempo, ni en el espa-

cio, esto es

$$\text{Cov}[\omega_t(\mathbf{s}), \omega_{t'}(\mathbf{r})] = \begin{cases} \sigma_\omega^2 & \text{si } t = t' \text{ y } \mathbf{s} = \mathbf{r}, \\ 0 & \text{en otro caso.} \end{cases} \quad (2.34)$$

Las perturbaciones $\omega_t(\mathbf{s})$ y $\mathbf{V}_t(\mathbf{r})$ son asumidas no correlacionadas para todo retraso temporal

$$\text{Cov}[\mathbf{V}_t(\mathbf{s}), \omega_{t'}(\mathbf{r})] = \mathbf{0} \quad \forall t, t', \mathbf{s}, \mathbf{r}. \quad (2.35)$$

Este sistema es generalmente usado para describir una serie finita de observaciones $\{Z_1(\mathbf{s}), \dots, Z_T(\mathbf{s})\}$ para la cual supuestos sobre valores iniciales del vector de estado $\xi_1(\cdot)$ son requeridos. Asumiremos que $\xi_1(\cdot)$ no está correlacionado con ninguna realización de $\mathbf{V}_t(\cdot)$ o $\omega_t(\cdot)$.

$$\text{Cov}[\mathbf{V}_t(\cdot), \xi_1(\cdot)] = \mathbf{0}, \quad \forall t = 1, 2, \dots, T, \quad (2.36a)$$

$$\text{Cov}[\omega_t(\cdot), \xi_1(\cdot)] = \mathbf{0}, \quad \forall t = 1, 2, \dots, T. \quad (2.36b)$$

Así (2.8a),(2.8b), (2.33), (2.34) y (2.35) implican que $\mathbf{V}_t(\cdot)$ no está correlacionado con valores rezagados de $\xi_t(\cdot)$

$$\text{Cov}[\mathbf{V}_t(\cdot), \xi_{t'}(\cdot)] = \mathbf{0}, \quad \forall t' = 1, \dots, t-1. \quad (2.37a)$$

$$\text{Cov}[\mathbf{V}_t(\cdot), Z_{t'}(\cdot)] = \mathbf{0}, \quad \forall t' = 1, \dots, t-1. \quad (2.37b)$$

Similarmente, $\omega_t(\cdot)$ no está correlacionado con ningún valor de $\xi(\cdot)$

$$\text{Cov}[\omega_t(\cdot), \xi_{t'}(\cdot)] = \mathbf{0}, \quad \forall t' = 1, \dots, T, \quad (2.38a)$$

$$\text{Cov}[\omega_t(\cdot), Z_{t'}(\cdot)] = 0, \quad \forall t' = 1, \dots, T. \quad (2.38b)$$

2.6. Valores iniciales

Dado que se necesitan valores iniciales para poder efectuar el filtraje, es necesario especificar el vector de estado inicial ξ_1 . Una propuesta es considerar

$$\xi_1 \sim \text{Gau}(\xi_{1|0}, \mathbf{P}_{1|0}),$$

donde su vector de medias y matriz de covarianzas no son más que la media y covarianza incondicional de ξ_1 , así $\xi_{1|0} := E[\xi_1] = \mathbf{0}$ y

$$\mathbf{P}_{1|0} := \text{Var}[\xi_1] = \begin{pmatrix} \mathcal{C}^\varepsilon(\mathbf{s}_1, \mathbf{s}_1) & \mathcal{C}^\varepsilon(\mathbf{s}_1, \mathbf{s}_2) & \cdots & \mathcal{C}^\varepsilon(\mathbf{s}_1, \mathbf{s}_n) \\ \mathcal{C}^\varepsilon(\mathbf{s}_2, \mathbf{s}_1) & \mathcal{C}^\varepsilon(\mathbf{s}_2, \mathbf{s}_2) & \cdots & \mathcal{C}^\varepsilon(\mathbf{s}_2, \mathbf{s}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{C}^\varepsilon(\mathbf{s}_n, \mathbf{s}_1) & \mathcal{C}^\varepsilon(\mathbf{s}_n, \mathbf{s}_2) & \cdots & \mathcal{C}^\varepsilon(\mathbf{s}_n, \mathbf{s}_n) \end{pmatrix},$$

con

$$\mathcal{C}^\varepsilon(\mathbf{s}, \mathbf{r}) = \begin{pmatrix} \mathcal{C}_0^\varepsilon(\mathbf{s}, \mathbf{r}) & \mathcal{C}_1^\varepsilon(\mathbf{s}, \mathbf{r}) & \cdots & \mathcal{C}_{p-1}^\varepsilon(\mathbf{s}, \mathbf{r}) \\ \mathcal{C}_1^\varepsilon(\mathbf{s}, \mathbf{r}) & \mathcal{C}_0^\varepsilon(\mathbf{s}, \mathbf{r}) & \cdots & \mathcal{C}_{p-2}^\varepsilon(\mathbf{s}, \mathbf{r}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{C}_{p-1}^\varepsilon(\mathbf{s}, \mathbf{r}) & \mathcal{C}_{p-2}^\varepsilon(\mathbf{s}, \mathbf{r}) & \cdots & \mathcal{C}_0^\varepsilon(\mathbf{s}, \mathbf{r}) \end{pmatrix},$$

donde $\mathcal{C}_k^\varepsilon(\mathbf{s}, \mathbf{r})$ se obtiene de 2.6.

2.7. Estimación de parámetros del modelo

El analista supone haber observado $\{\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_T\}$. Uno de los objetivos más fundamentales podría ser estimar los valores de cualquier parámetro desconocido del sistema, basándose en estas observaciones. Así, bajo el marco de la representación espacio-estado dada por (2.11a) y (2.11b), denotemos $\Theta = (\beta^\top, \phi^\top, \alpha, \sigma_\eta^2, \sigma_\omega^2)^\top$ como el vector de parámetros a estimar, con $\beta = (\beta_1, \dots, \beta_\ell)^\top$, $\phi = (\phi_1, \dots, \phi_p)^\top$, α el

alcance, σ_η^2 la varianza del proceso de innovación espacial, y σ_ω^2 el efecto pepita.

En lo que viene, se propondrán dos metodologías para estimar los parámetros involucrados en el modelo espacio-tiempo autorregresivo: Estimación por Máxima Verosimilitud (*ML*) y Estimación mediante algoritmo Esperanza-Maximización General (*GEM*).

2.7.1. Estimación por Máxima Verosimilitud

Dada la distribución $[\mathbf{Z}_t|\mathbf{Z}_{t-1}]$ calculada en (2.26), la cual es Gaussiana, podemos reescribirla, usando

$$\begin{aligned}\mathbf{Z}_{t|t-1} &= \mathbf{X}_t\boldsymbol{\beta} + \Lambda_t\Phi\xi_{t-1|t-1}, \\ \Delta_{t|t-1} &= \Lambda_t(\Phi\mathbf{P}_{t-1|t-1}\Phi^\top + \mathbf{Q})\Lambda_t^\top + \sigma_\omega^2\mathbf{I}_{n_t}.\end{aligned}\tag{2.39}$$

Su correspondiente función de densidad conjunta viene dada por

$$f(\mathbf{Z}_{1:T}|\Theta) = f(\mathbf{Z}_1|\Theta) \left(\prod_{t=2}^T f(\mathbf{Z}_t|\mathbf{Z}_{t-1}, \Theta) \right).$$

En efecto, el doble de la función log-verosimilitud, más una constante aditiva, puede ser obtenida

$$2\ell(\Theta|\mathbf{Z}_{1:T}) = - \sum_{t=1}^T \log |\Delta_{t|t-1}| - \sum_{t=1}^T (\mathbf{z}_t - \mathbf{z}_{t|t-1})^\top \Delta_{t|t-1}^{-1} (\mathbf{z}_t - \mathbf{z}_{t|t-1}).\tag{2.40}$$

De esta manera, la expresión (2.40) puede ser maximizada numéricamente con respecto a los parámetros desconocidos del modelo. La solución a las ecuaciones de verosimilitud, $\nabla_{\Theta}\ell = \mathbf{0}$, es aproximada particionando $\Theta = (\boldsymbol{\beta}^\top, \Theta^{*\top})^\top$, con $\Theta^* = (\boldsymbol{\phi}^\top, \alpha, \sigma_\eta^2, \sigma_\omega^2)^\top$. Así, se puede obtener una solución de forma cerrada para $\boldsymbol{\beta}$ dada por

$$\hat{\beta} = \left(\sum_{t=1}^T \mathbf{X}_t^\top \Delta_{t|t-1}^{-1} \mathbf{X}_t \right)^{-1} \sum_{t=1}^T \mathbf{X}_t^\top \Delta_{t|t-1}^{-1} (\mathbf{Z}_t - \Lambda_t \xi_{t|t-1}), \quad (2.41)$$

mientras se mantienen los parámetros restantes $\hat{\Theta}^*$ fijos en su valor más reciente. Como no hay solución de forma cerrada para Θ^* , una alternativa es usar el algoritmo de Newton-Raphson, donde la fórmula iterativa para $m \geq 1$ es definida por

$$\hat{\Theta}_m^* = \hat{\Theta}_{m-1}^* - \left(\nabla_{\Theta^*} \nabla_{\Theta^*}^\top \ell(\hat{\Theta}_{m-1} | \mathbf{Z}_{1:T}) \right)^{-1} \nabla_{\Theta^*} \ell(\hat{\Theta}_{m-1} | \mathbf{Z}_{1:T}), \quad (2.42)$$

donde $\nabla_{\Theta^*} \ell(\Theta | \mathbf{Z}_{1:t-1})$ es la función score y $\nabla_{\Theta^*} \nabla_{\Theta^*}^\top \ell(\Theta | \mathbf{Z}_{1:t-1})$ es la matriz hessiana de la función log-verosimilitud. Finalmente, dado un nivel de tolerancia, δ , escogido, el error relativo puede ser usado como criterio de convergencia

$$\frac{\|\hat{\Theta}_m - \hat{\Theta}_{m-1}\|}{\|\hat{\Theta}_m\|} < \delta.$$

2.7.2. Estimación mediante algoritmo Esperanza-Maximización General

Este algoritmo se utiliza para encontrar estimadores máximo verosímiles de parámetros en modelos probabilísticos que dependen de variables no observables. En este contexto, los datos observados son $\mathbf{Z}_{1:T}$ y los datos perdidos hacen referencia al proceso de estado $\xi_{1:T}$, la cual es no observable por definición. Así, la formulación de la densidad conjunta para este tipo de información es

$$f(\mathbf{Z}_{1:T}, \xi_{1:T} | \Theta) = \left(\prod_{t=1}^T f(\mathbf{Z}_t | \xi_t, \Theta) \right) \left(\prod_{t=2}^T f(\xi_t | \xi_{t-1}, \Theta) \right) f(\xi_1 | \Theta),$$

consiguiendo el doble de la log-verosimilitud completa, más una constante aditiva, por

$$2\ell(\Theta | \mathbf{Z}_{1:T}, \xi_{1:T}) = \ell^{(1)}(\beta, \sigma_\omega^2 | \mathbf{Z}_{1:T}, \xi_{1:T}) + \ell^{(2)}(\phi, \alpha, \sigma_\eta^2 | \xi_{1:T}) + \ell^{(3)}(\xi_1),$$

con

$$\ell^{(1)} := -\ln\{\sigma_\omega^2\} \sum_{t=1}^T n_t - \frac{1}{\sigma_\omega^2} \sum_{t=1}^T (\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_t)^\top (\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_t), \quad (2.43a)$$

$$\ell^{(2)} := -(T-1) (n\ln\sigma_\eta^2 + \ln|\mathbf{D}|) - \frac{1}{\sigma_\eta^2} \sum_{t=2}^T (\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1})^\top \mathbf{R}^- (\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}), \quad (2.43b)$$

$$\ell^{(3)} := -\ln|\mathbf{P}_{1|0}| - (\boldsymbol{\xi}_1 - \boldsymbol{\xi}_{1|0})^\top \mathbf{P}_{1|0}^{-1} (\boldsymbol{\xi}_1 - \boldsymbol{\xi}_{1|0}), \quad (2.43c)$$

donde $\ell^{(1)} \equiv \ell^{(1)}(\boldsymbol{\beta}, \sigma_\omega^2 | \mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T})$, $\ell^{(2)} \equiv \ell^{(2)}(\boldsymbol{\phi}, \alpha, \sigma_\eta^2 | \boldsymbol{\xi}_{1:T})$, $\ell^{(3)} \equiv \ell^{(3)}(\boldsymbol{\xi}_1)$, y donde la matriz \mathbf{D} , en (2.43b), se detalla en el Anexo.

Observe que tanto en (2.39) como en (2.43a), los datos faltantes no generan un mayor problema a la hora de calcular las estimaciones. Esto debido a que se ha definido un operador de observación extendido, $\boldsymbol{\Lambda}_t$, la cual mapea correctamente el proceso desde el vector de estado extendido $\boldsymbol{\xi}_t$, cuya dimensión es np para cualquier t , al vector de observación, \mathbf{Z}_t , cuya dimensión irá variando, digamos n_t , correspondiendo al tamaño del conjunto de datos disponibles en el instante t .

Paso-E

Ahora, el paso-E del algoritmo EM requiere el cálculo de la esperanza,

$$\begin{aligned} q(\boldsymbol{\Theta} | \widehat{\boldsymbol{\Theta}}^{(i-1)}) &:= \mathbb{E} \left[2\ell(\boldsymbol{\Theta} | \mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}) \middle| \mathbf{Z}_{1:T}, \widehat{\boldsymbol{\Theta}}^{(i-1)} \right] \\ &= q^{(1)}(\boldsymbol{\beta}, \sigma_\omega^2 | \widehat{\boldsymbol{\Theta}}^{(i-1)}) + q^{(2)}(\boldsymbol{\phi}, \alpha, \sigma_\eta^2 | \widehat{\boldsymbol{\Theta}}^{(i-1)}) + q^{(3)}(\widehat{\boldsymbol{\Theta}}^{(i-1)}) \end{aligned} \quad (2.44)$$

donde

$$\begin{aligned} q^{(1)} &:= -\ln\{\sigma_\omega^2\} \sum_{t=1}^T n_t - \frac{1}{\sigma_\omega^2} \text{tr} \left\{ \sum_{t=1}^T \left[(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_{t|T}) (\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_{t|T})^\top + \boldsymbol{\Lambda}_t \mathbf{P}_{t|T} \boldsymbol{\Lambda}_t^\top \right] \right\}, \\ q^{(2)} &:= -(T-1) (n\ln\{\sigma_\eta^2\} + \ln|\mathbf{D}|) - \frac{1}{\sigma_\eta^2} \text{tr} \left\{ \mathbf{R}^- \left(\mathbf{S}_{22} - \mathbf{S}_{21} \boldsymbol{\Phi}^\top - \boldsymbol{\Phi} \mathbf{S}_{21}^\top + \boldsymbol{\Phi} \mathbf{S}_{11} \boldsymbol{\Phi}^\top \right) \right\}, \\ q^{(3)} &:= -\ln|\mathbf{P}_{1|0}| - \text{tr} \left\{ \mathbf{P}_{1|0}^{-1} \left((\boldsymbol{\xi}_{1|T} - \boldsymbol{\xi}_{1|0}) (\boldsymbol{\xi}_{1|T} - \boldsymbol{\xi}_{1|0})^\top + \mathbf{P}_{1|T} \right) \right\}, \end{aligned}$$

con $q^{(1)} \equiv q^{(1)}(\boldsymbol{\beta}, \sigma_\omega^2 | \widehat{\boldsymbol{\Theta}}^{(i-1)})$, $q^{(2)} \equiv q^{(2)}(\boldsymbol{\phi}, \alpha, \sigma_\eta^2 | \widehat{\boldsymbol{\Theta}}^{(i-1)})$, $q^{(3)} \equiv q^{(3)}(\widehat{\boldsymbol{\Theta}}^{(i-1)})$ y

$$\begin{aligned}\mathbf{S}_{22} &:= \sum_{t=2}^T [\mathbf{P}_{t|T} + \boldsymbol{\xi}_{t|T} \boldsymbol{\xi}_{t|T}^\top], \\ \mathbf{S}_{11} &:= \sum_{t=2}^T [\mathbf{P}_{t-1|T} + \boldsymbol{\xi}_{t-1|T} \boldsymbol{\xi}_{t-1|T}^\top], \\ \mathbf{S}_{21} &:= \sum_{t=2}^T [\mathbf{P}_{t,t-1|T} + \boldsymbol{\xi}_{t|T} \boldsymbol{\xi}_{t-1|T}^\top],\end{aligned}$$

donde $\mathbf{P}_{t,t-1|T} := \mathbb{E} [(\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t|T})(\boldsymbol{\xi}_{t-1} - \boldsymbol{\xi}_{t-1|T})^\top | \mathbf{Z}_{1:T}]$ es la comúnmente llamada *lag-one covarince smoother* (Shumway and Stoffer, 2006, Property P6.3), la cual es obtenida calculando

$$\mathbf{P}_{T,T-1|T} = (\mathbf{I}_{np} - \mathbf{P}_{T|T-1} \boldsymbol{\Lambda}_T^\top (\boldsymbol{\Lambda}_T \mathbf{P}_{T|T-1} \boldsymbol{\Lambda}_T^\top + \sigma_\omega^2 \mathbf{I}_{n_T})^{-1} \boldsymbol{\Lambda}_T) \boldsymbol{\Phi} \mathbf{P}_{T-1|T-1},$$

luego, para $t = T, T-1, \dots, 3$

$$\mathbf{P}_{t-1,t-2|T} = \mathbf{P}_{t-1|t-1} \mathbf{J}_{t-2}^\top + \mathbf{J}_{t-1} (\mathbf{P}_{t,t-1|T} - \boldsymbol{\Phi} \mathbf{P}_{t-1|t-1}) \mathbf{J}_{t-2}^\top.$$

El lado derecho de (2.44) depende de $\widehat{\boldsymbol{\Theta}}^{(i-1)}$, sin embargo no se hace explícito por economía de la notación. Por ejemplo, $\boldsymbol{\xi}_{t|T} := \mathbb{E}[\boldsymbol{\xi}_t | \mathbf{Z}_{1:T}] \equiv \mathbb{E}[\boldsymbol{\xi}_t | \mathbf{Z}_{1:T}, \widehat{\boldsymbol{\Theta}}^{(i-1)}]$.

Paso-M

El paso-M procede maximizando (2.44) con respecto a los parámetros $\boldsymbol{\Theta}$, esto es

$$\widehat{\boldsymbol{\Theta}}^{(i)} = \underset{\boldsymbol{\Theta}}{\text{Argmax}} \left\{ q(\boldsymbol{\Theta} | \widehat{\boldsymbol{\Theta}}^{(i-1)}) \right\}.$$

Así, la solución de $\nabla q = \mathbf{0}$ es aproximada particionando $\boldsymbol{\Theta} = (\boldsymbol{\Theta}^{*\top}, \alpha)^\top$, donde $\boldsymbol{\Theta}^* = (\boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \sigma_\eta^2, \sigma_\omega^2)^\top$. Para $\boldsymbol{\Theta}^*$ se puede obtener una solución de forma cerrada, mientras se mantiene $\widehat{\alpha}$ fijo en su valor actual. Así, las formas cerradas vienen dadas

por

$$\widehat{\boldsymbol{\beta}}^{(i)} = \left(\sum_{t=1}^T \mathbf{X}_t^\top \mathbf{X}_t \right)^{-1} \sum_{t=1}^T \mathbf{X}_t^\top \left(\mathbf{Z}_t - \boldsymbol{\Lambda}_t \boldsymbol{\xi}_{t|T} \right),$$

$$\widehat{\boldsymbol{\phi}}^{(i)} = \mathcal{H}_1 \left\{ (\boldsymbol{\rho}^{\eta(i)} \otimes \mathbf{I}_n)^{-1} \mathbf{S}_{11} \right\}^{-1} \left[\mathcal{H}_3 \left\{ \left(\mathbf{R}^{(i)} \right)^{-} \mathbf{S}_{21} \right\} - \mathcal{H}_2 \left\{ (\boldsymbol{\rho}^{\eta(i)} \otimes \mathbf{I}_n)^{-1} \mathbf{S}_{11} \right\} \widehat{\boldsymbol{\phi}}^{(i-1)} \right],$$

$$\widehat{\sigma}_\eta^{2(i)} = \frac{1}{(T-1)n} \text{tr} \left\{ \left(\mathbf{R}^{(i)} \right)^{-} \left(\mathbf{S}_{22} - \mathbf{S}_{21} \boldsymbol{\Phi}^{(i)} - \boldsymbol{\Phi}^{(i)} \mathbf{S}_{21}^\top + \boldsymbol{\Phi}^{(i)} \mathbf{S}_{11} \boldsymbol{\Phi}^{\top(i)} \right) \right\},$$

$$\widehat{\sigma}_\omega^{2(i)} = \frac{1}{\sum_{t=1}^T n_t} \text{tr} \left\{ \sum_{t=1}^T \left(\mathbf{Z}_t - \mathbf{X}_t \widehat{\boldsymbol{\beta}}^{(i)} - \boldsymbol{\Lambda}_t \boldsymbol{\xi}_{t|T} \right) \left(\mathbf{Z}_t - \mathbf{X}_t \widehat{\boldsymbol{\beta}}^{(i)} - \boldsymbol{\Lambda}_t \boldsymbol{\xi}_{t|T} \right)^\top + \boldsymbol{\Lambda}_t \mathbf{P}_{t|T} \boldsymbol{\Lambda}_t^\top \right\},$$

con $\boldsymbol{\Phi}^{(i)} \equiv \boldsymbol{\Phi} \left(\widehat{\boldsymbol{\phi}}^{(i)} \right)$, $\mathbf{R}^{(i)} \equiv \mathbf{R} \left(\widehat{\boldsymbol{\alpha}}^{(i)} \right)$, $\boldsymbol{\rho}^{\eta(i)} \equiv \boldsymbol{\rho}^\eta \left(\widehat{\boldsymbol{\alpha}}^{(i)} \right)$ y donde las funciones \mathcal{H}_1 , \mathcal{H}_2 y \mathcal{H}_3 se detallan en el Anexo. Si se particulariza la estructura temporal, y se considera una dinámica autorregresiva de primer orden, entonces el estimador del coeficiente ϕ viene dado por

$$\widehat{\phi}^{(i)} = \text{tr} \left\{ \boldsymbol{\rho}^\eta \left(\widehat{\boldsymbol{\alpha}}^{(i)} \right)^{-1} \mathbf{S}_{21} \right\} \text{tr} \left\{ \boldsymbol{\rho}^\eta \left(\widehat{\boldsymbol{\alpha}}^{(i)} \right)^{-1} \mathbf{S}_{11} \right\}^{-1}. \quad (2.46)$$

Como no hay solución de forma cerrada para el alcance, α , se puede usar el algoritmo de Newton-Raphson, donde la fórmula iterativa para $m \geq 1$ es definida por

$$\widehat{\alpha}_m^{(i)} = \widehat{\alpha}_{m-1}^{(i)} - \left(\frac{\partial^2 q(\boldsymbol{\Theta} | \widehat{\boldsymbol{\Theta}}^{*(i-1)}, \widehat{\alpha}_{m-1}^{(i)})}{\partial \alpha \partial \alpha} \right)^{-1} \frac{\partial q(\boldsymbol{\Theta} | \widehat{\boldsymbol{\Theta}}^{*(i-1)}, \widehat{\alpha}_{m-1}^{(i)})}{\partial \alpha}. \quad (2.47)$$

Finalmente, el algoritmo EM general converge de acuerdo un nivel de tolerancia, δ , escogido. Así, el error relativo puede ser determinado como criterio de convergencia

$$\frac{|\widehat{\alpha}_m^{(i)} - \widehat{\alpha}_{m-1}^{(i)}|}{|\widehat{\alpha}_{m-1}^{(i)}|} < \delta, \quad \frac{\|\widehat{\boldsymbol{\Theta}}^{(i)} - \widehat{\boldsymbol{\Theta}}^{(i-1)}\|}{\|\widehat{\boldsymbol{\Theta}}^{(i)}\|} < \delta.$$

2.8. Interpolación espacial

Suponga que se quiere hacer inferencia en una localidad $\mathbf{s}_0 \neq \mathbf{s}_i; i = 1, \dots, n$, con la finalidad de lograr predicciones para $Z_t(\mathbf{s}_0)$ dado $\mathbf{Z}_{1:t}; t = 1, \dots, T$. Desde la distribución $[\xi_t(\mathbf{s}_0), \xi_t | \mathbf{Z}_{1:t-1}]$, la cual es Gaussiana con vector de medias

$$\begin{pmatrix} \xi_{t|t-1}(\mathbf{s}_0) \\ \xi_{t|t-1} \end{pmatrix} = \begin{pmatrix} \mathbf{F} & \mathbf{0} \\ \mathbf{0} & \Phi \end{pmatrix} \begin{pmatrix} \xi_{t-1|t-1}(\mathbf{s}_0) \\ \xi_{t-1|t-1} \end{pmatrix},$$

y matriz de covarianzas

$$\begin{pmatrix} \mathbf{P}_{t|t-1}(\mathbf{s}_0, \mathbf{s}_0) & \mathbf{P}_{t|t-1}(\mathbf{s}_0) \\ \mathbf{P}_{t|t-1}(\mathbf{s}_0)^\top & \mathbf{P}_{t|t-1} \end{pmatrix} = \begin{pmatrix} \mathbf{F} & \mathbf{0} \\ \mathbf{0} & \Phi \end{pmatrix} \begin{pmatrix} \mathbf{P}_{t-1|t-1}(\mathbf{s}_0, \mathbf{s}_0) & \mathbf{P}_{t-1|t-1}(\mathbf{s}_0) \\ \mathbf{P}_{t-1|t-1}(\mathbf{s}_0)^\top & \mathbf{P}_{t-1|t-1} \end{pmatrix} \\ \times \begin{pmatrix} \mathbf{F} & \mathbf{0} \\ \mathbf{0} & \Phi \end{pmatrix}^\top + \begin{pmatrix} \mathbf{Q}(\mathbf{s}_0, \mathbf{s}_0) & \mathbf{Q}(\mathbf{s}_0) \\ \mathbf{Q}(\mathbf{s}_0)^\top & \mathbf{Q} \end{pmatrix}.$$

con

$$\mathbf{P}_{t|t-1}(\mathbf{s}_0, \mathbf{s}_0) := \text{Cov}[\xi_t(\mathbf{s}_0), \xi_t(\mathbf{s}_0) | \mathbf{Z}_{1:t-1}],$$

$$\mathbf{P}_{t|t-1}(\mathbf{s}_0) := \text{Cov}[\xi_t(\mathbf{s}_0), \xi_t | \mathbf{Z}_{1:t-1}],$$

$$\mathbf{Q}_t(\mathbf{s}_0, \mathbf{s}_0) := \text{Var}[\mathbf{V}_t(\mathbf{s}_0), \mathbf{V}_t(\mathbf{s}_0)],$$

$$\mathbf{Q}_t(\mathbf{s}_0) := \text{Var}[\mathbf{V}_t(\mathbf{s}_0), \mathbf{V}_t].$$

Así, podemos obtener las ecuaciones de predicción y matriz de covarianza del error de predicción

$$\xi_{t|t-1}(\mathbf{s}_0) = \mathbf{F}\xi_{t-1|t-1}(\mathbf{s}_0),$$

$$\mathbf{P}_{t|t-1}(\mathbf{s}_0, \mathbf{s}_0) = \mathbf{F}\mathbf{P}_{t-1|t-1}(\mathbf{s}_0, \mathbf{s}_0)\mathbf{F}^\top + \mathbf{Q}(\mathbf{s}_0, \mathbf{s}_0),$$

$$\mathbf{P}_{t|t-1}(\mathbf{s}_0) = \mathbf{F}\mathbf{P}_{t-1|t-1}(\mathbf{s}_0)\Phi^\top + \mathbf{Q}(\mathbf{s}_0).$$

Por otro lado, dado que la distribución $[\xi_t(\mathbf{s}_0), \xi_t, \mathbf{Z}_t | \mathbf{Z}_{1:t-1}]$ es Gaussiana con vector de medias y matriz de covarianzas dadas, respectivamente, por

$$\begin{pmatrix} \xi_{t|t-1}(\mathbf{s}_0) \\ \xi_{t|t-1} \\ \mathbf{Z}_{t|t-1} \end{pmatrix}, \begin{pmatrix} \mathbf{P}_{t|t-1}(\mathbf{s}_0, \mathbf{s}_0) & \mathbf{P}_{t|t-1}(\mathbf{s}_0) & \mathbf{G}_{t|t-1}(\mathbf{s}_0) \\ \mathbf{P}_{t|t-1}(\mathbf{s}_0)^\top & \mathbf{P}_{t|t-1} & \mathbf{G}_{t|t-1} \\ \mathbf{G}_{t|t-1}(\mathbf{s}_0)^\top & \mathbf{G}_{t|t-1}^\top & \Delta_{t|t-1} \end{pmatrix}.$$

con $\mathbf{G}_{t|t-1}(\mathbf{s}_0) := \text{Cov}[\xi_t(\mathbf{s}_0), \mathbf{Z}_t | \mathbf{Z}_{1:t-1}] = \mathbf{P}_{t|t-1}(\mathbf{s}_0) \mathbf{\Lambda}^\top$. Consecuentemente, la distribución $[\xi_t(\mathbf{s}_0), \xi_t | \mathbf{Z}_{1:t}]$ es Gaussiana con vector de medias

$$\begin{pmatrix} \xi_{t|t}(\mathbf{s}_0) \\ \xi_{t|t} \end{pmatrix} = \begin{pmatrix} \xi_{t|t-1}(\mathbf{s}_0) \\ \xi_{t|t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{G}_{t|t-1}(\mathbf{s}_0) \\ \mathbf{G}_{t|t-1} \end{pmatrix} \Delta_{t|t-1}^{-1} (\mathbf{Z}_t - \mathbf{Z}_{t|t-1})$$

y matriz de covarianzas

$$\begin{pmatrix} \mathbf{P}_{t|t}(\mathbf{s}_0, \mathbf{s}_0) & \mathbf{P}_{t|t}(\mathbf{s}_0) \\ \mathbf{P}_{t|t}(\mathbf{s}_0)^\top & \mathbf{P}_{t|t} \end{pmatrix} = \begin{pmatrix} \mathbf{P}_{t|t-1}(\mathbf{s}_0, \mathbf{s}_0) & \mathbf{P}_{t|t-1}(\mathbf{s}_0) \\ \mathbf{P}_{t|t-1}(\mathbf{s}_0)^\top & \mathbf{P}_{t|t-1} \end{pmatrix} - \begin{pmatrix} \mathbf{G}_{t|t-1}(\mathbf{s}_0) \\ \mathbf{G}_{t|t-1} \end{pmatrix} \Delta_{t|t-1}^{-1} \begin{pmatrix} \mathbf{G}_{t|t-1}(\mathbf{s}_0)^\top & \mathbf{G}_{t|t-1}^\top \end{pmatrix},$$

(e.g., Searle, 1979, p.46). Por lo tanto, tenemos las ecuaciones de actualización para $\xi_t(\mathbf{s}_0)$ dadas por

$$\xi_{t|t}(\mathbf{s}_0) = \xi_{t|t-1}(\mathbf{s}_0) + \mathbf{G}_{t|t-1}(\mathbf{s}_0) \Delta_{t|t-1}^{-1} (\mathbf{Z}_t - \mathbf{Z}_{t|t-1}), \quad (2.48a)$$

$$\mathbf{P}_{t|t}(\mathbf{s}_0, \mathbf{s}_0) = \mathbf{P}_{t|t-1}(\mathbf{s}_0, \mathbf{s}_0) - \mathbf{G}_{t|t-1}(\mathbf{s}_0) \Delta_{t|t-1}^{-1} \mathbf{G}_{t|t-1}(\mathbf{s}_0)^\top, \quad (2.48b)$$

$$\mathbf{P}_{t|t}(\mathbf{s}_0) = \mathbf{P}_{t|t-1}(\mathbf{s}_0) - \mathbf{G}_{t|t-1}(\mathbf{s}_0) \Delta_{t|t-1}^{-1} \mathbf{G}_{t|t-1}^\top. \quad (2.48c)$$

Para el caso particular donde en el instante t no halla ninguna observación disponible, esto es, $n_t = 0$, el paso de filtro-actualización dado por (2.48a), (2.48b) y (2.48c) es

simplemente saltado, ya que las ecuaciones se simplifican a

$$\boldsymbol{\xi}_{t|t}(\mathbf{s}_0) = \boldsymbol{\xi}_{t|t-1}(\mathbf{s}_0), \quad (2.49a)$$

$$\mathbf{P}_{t|t}(\mathbf{s}_0, \mathbf{s}_0) = \mathbf{P}_{t|t-1}(\mathbf{s}_0, \mathbf{s}_0), \quad (2.49b)$$

$$\mathbf{P}_{t|t}(\mathbf{s}_0) = \mathbf{P}_{t|t-1}(\mathbf{s}_0), \quad (2.49c)$$

Finalmente, para lograr una predicción espacio-temporal en $Z_t(\mathbf{s}_0)$ dado $\mathbf{Z}_{1:t}$, con su respectiva varianza de predicción, calculamos la distribución $[Z_t(\mathbf{s}_0)|\mathbf{X}_t(\mathbf{s}_0), \mathbf{Z}_{1:t}]$, dada por

$$Z_t(\mathbf{s}_0)|\mathbf{X}_t(\mathbf{s}_0), \mathbf{Z}_{1:t} \sim \text{Gau}(Z_{t|t}(\mathbf{s}_0), \Delta_{t|t}(\mathbf{s}_0)),$$

con

$$Z_{t|t}(\mathbf{s}_0) = \boldsymbol{\beta}^\top \mathbf{X}_t(\mathbf{s}_0) + \mathbf{H}^\top \boldsymbol{\xi}_{t|t}(\mathbf{s}_0),$$

$$\Delta_{t|t}(\mathbf{s}_0) = \mathbf{H}^\top \mathbf{P}_{t|t}(\mathbf{s}_0, \mathbf{s}_0) \mathbf{H} + \sigma_\omega^2.$$

Capítulo 3

Estudio de simulación

3.1. Desempeño del método de estimación

Se realizó un estudio de simulación Monte-Carlo para revisar el desempeño de la metodología de estimación propuesta en la sección 2.7. Para ello se consideró una muestra finita y el modelo

$$\begin{aligned} \mathbf{Z}_t(\mathbf{s}) &= \beta_0 + \varepsilon_t(\mathbf{s}) + \omega_t(\mathbf{s}); \quad \omega_{\mathcal{D} \times \mathcal{T}} \sim \mathcal{GWN}(0, \sigma_\omega^2), \\ \varepsilon_t(\mathbf{s}) &= \phi \varepsilon_{t-1}(\mathbf{s}) + \eta_t(\mathbf{s}), \end{aligned} \tag{3.1}$$

asociado a un modelo exponencial para la estructura de la covarianza del proceso de innovación espacial $\eta_{\mathcal{D} \times \mathcal{T}}$, es decir

$$C^\eta(\mathbf{s}, \mathbf{r}) = \sigma_\eta^2 \exp\left(-\frac{\|\mathbf{s} - \mathbf{r}\|}{\alpha}\right).$$

El esquema de simulación es como sigue: se consideraron 6 escenarios donde en cada uno se simularon 1000 muestras aleatorias desde un campo aleatorio Gaussiano espacio-temporal de media cero y varianza unitaria. Cada muestra consta de $N = 25$ localidades espaciales distribuidas de forma regular en el cuadrado unitario y $T = 400$

instantes de tiempo. En cada escenario, la varianza del error de medición se consideró de tal forma que explique el 10 % de la varianza total del proceso, esto es $\sigma_\omega^2 = 0.1$. En los escenarios del 1 al 3 se consideró alta continuidad espacial ($\alpha = 0.8$); y del 4 al 6, baja continuidad espacial ($\alpha = 0.4$). El valor del coeficiente autorregresivo se alternó de tal forma que en el escenario 1 y 4, $\phi = 0.7$; en el 2 y 5, $\phi = 0.5$; y en el 3 y 6, $\phi = 0.3$. Finalmente, la varianza de la innovación espacial, σ_η^2 , se calculó de tal forma que satisfaga la relación $\sigma_\eta^2 = 0.9(1 - \phi^2)$. Estos escenarios fueron elegidos con el propósito de estudiar el desempeño de la metodología de estimación en casos donde los niveles de continuidad espacial y la correlación temporal se combinen.

Así, el conjunto de parámetros involucrados en el modelo (3.1) viene dado por $\Theta = (\beta, \phi, \alpha, \sigma_\eta^2, \sigma_\omega^2)^\top$, donde las estimaciones de éstos se evaluaron a través de la media (Mean), desviación estándar (Sd), sesgo (B) y error cuadrático medio (MSE), definidos como:

$$\text{Mean}(\hat{\theta}) = \frac{1}{1000} \sum_{i=1}^{1000} \hat{\theta}_i, \quad \text{Sd}(\hat{\theta}) = \left(\frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_i - \text{Mean}(\hat{\theta}))^2 \right)^{\frac{1}{2}},$$

$$\text{B}(\hat{\theta}) = \text{Mean}(\hat{\theta}) - \theta, \quad \text{MSE}(\hat{\theta}) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\theta}_i - \theta)^2,$$

donde $\hat{\theta}_{ij}$ es la i -ésima estimación de $\theta_j \in \Theta$ de la i -ésima realización. Las Tablas 3.1 y 3.2 reportan los resultados de las simulaciones, bajo las dos metodologías de estimación analizadas en la sección 2.7.

Para el caso de estimación por *ML* (3.1) se puede observar que la media de los parámetros estimados es bastante cercana a su verdadero valor, con una leve subestimación en la mayoría de los casos, característica muy común en este tipo de estimador, la cual se ve evidenciado en el sesgo donde reporta valores negativos, pero muy cercanos a 0. El error cuadrático medio es muy pequeño, la cual nos garantiza una buena calidad del estimador en general.

Cabe destacar que el algoritmo de Newton-Raphson, como método para maximizar

numéricamente la verosimilitud, es muy sensible a los valores iniciales, así una leve modificación de estos hacen que simplemente el algoritmo diverja, lo cual se traduce en una desventaja a la hora de querer estimar, ya que hay que estar probando muchos valores iniciales como candidatos, haciendo que el método pierda poder de eficiencia.

Por otro lado, para el caso de estimación mediante *GEM* (3.2), se puede observar que la media de los parámetros estimados es muy cercano a su verdadero valor, en algunos casos con una despreciable subestimación o sobreestimación, tanto la desviación como el sesgo y el error cuadrático medio son muy pequeños, la cual nos garantiza una buena calidad del estimador en general.

Una ventaja notable del algoritmo *GEM* es que siempre converge a la hora de querer estimar, independiente de los valores iniciales, haciendo que el método sea muy robusto. Además, a pesar de necesitar emplear algún método numérico, como el algoritmo de Newton-Raphson, para estimar los parámetros asociados a la función de covarianza puramente espacial, como por ejemplo el alcance (α), se pueden obtener formas cerradas para los estimadores del resto de los parámetros del modelo $\{\beta, \phi, \sigma_{\eta}^2, \sigma_{\omega}^2\}$, lo cual nos da la facilidad de poder generalizarlo con mayor libertad.

Finalmente, desde las Tablas 3.1 y 3.2, se puede evidenciar que ambos métodos producen estimadores con un sesgo muy despreciable, sin embargo los del algoritmo *GEM* son más eficientes, así la calidad de estimación es superior, además de reportar mejores resultados en todos los escenarios bajo los 4 criterios usados. Como conclusión, para efectos prácticos, el método de estimación de parámetros vía algoritmo *GEM* se podría sugerir como una alternativa preferible entre ambos.

<i>Case 1</i>	$\beta_0 = 0$	$\phi = 0.7$	$\alpha = 0.8$	$\sigma_\eta^2 = 0.459$	$\sigma_\omega^2 = 0.1$
Mean	-0.00474	0.69763	0.79423	0.45896	0.09961
Sd	0.04685	0.01435	0.05698	0.01990	0.00860
B	-0.00474	-0.00237	-0.00577	-0.00262	-0.00039
MSE	0.00222	0.00021	0.00328	0.00040	0.00007
<i>Case 2</i>	$\beta_0 = 0$	$\phi = 0.5$	$\alpha = 0.8$	$\sigma_\eta^2 = 0.675$	$\sigma_\omega^2 = 0.1$
Mean	-0.00118	0.49772	0.78996	0.67334	0.09767
Sd	0.04839	0.01752	0.05887	0.02873	0.01542
B	-0.00118	-0.00228	-0.01004	-0.00166	-0.00233
MSE	0.00234	0.00031	0.00356	0.00083	0.00024
<i>Case 3</i>	$\beta_0 = 0$	$\phi = 0.3$	$\alpha = 0.8$	$\sigma_\eta^2 = 0.819$	$\sigma_\omega^2 = 0.1$
Mean	0.00031	0.29848	0.79295	0.81424	0.09942
Sd	0.04473	0.01590	0.05348	0.03371	0.01161
B	0.00031	-0.00152	-0.00705	-0.00476	-0.00058
MSE	0.00200	0.00025	0.00291	0.00116	0.00013
<i>Case 4</i>	$\beta_0 = 0$	$\phi = 0.7$	$\alpha = 0.4$	$\sigma_\eta^2 = 0.459$	$\sigma_\omega^2 = 0.1$
Mean	-0.00340	0.69907	0.39763	0.45755	0.09949
Sd	0.03333	0.01396	0.03095	0.02317	0.01919
B	-0.00340	-0.00093	-0.00237	-0.00145	-0.00051
MSE	0.00112	0.00020	0.00096	0.00054	0.00037
<i>Case 5</i>	$\beta_0 = 0$	$\phi = 0.5$	$\alpha = 0.4$	$\sigma_\eta^2 = 0.675$	$\sigma_\omega^2 = 0.1$
Mean	-0.00162	0.49856	0.39680	0.67435	0.09794
Sd	0.03605	0.01869	0.02972	0.03251	0.02668
B	-0.00162	-0.00144	-0.00320	-0.00065	-0.00206
MSE	0.00130	0.00035	0.00089	0.00106	0.00072
<i>Case 6</i>	$\beta_0 = 0$	$\phi = 0.3$	$\alpha = 0.4$	$\sigma_\eta^2 = 0.819$	$\sigma_\omega^2 = 0.1$
Mean	-0.00238	0.29916	0.39674	0.81831	0.09696
Sd	0.03524	0.01759	0.02910	0.03604	0.03027
B	-0.00238	-0.00084	-0.00326	-0.00069	-0.00304
MSE	0.00125	0.00031	0.00086	0.00130	0.00092

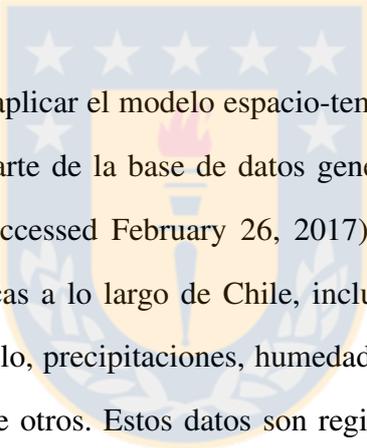
Tabla 3.1: Resultados de las estimaciones vía Máxima Verosimilitud para un modelo espacio-tiempo AR(1). Las localidades observadas están en el cuadrado $[0, 1]^2$. Fuente: Elaboración propia.

<i>Case 1</i>	$\beta_0 = 0$	$\phi = 0.7$	$\alpha = 0.8$	$\sigma_\eta^2 = 0.459$	$\sigma_\omega^2 = 0.1$
Mean	0.00001	0.70006	0.80032	0.45896	0.10002
Sd	0.00049	0.00805	0.01596	0.00701	0.00139
B	0.00001	0.00006	0.00032	-0.00004	0.00002
MSE	0.00000	0.00006	0.00025	0.00005	0.00000
<i>Case 2</i>	$\beta_0 = 0$	$\phi = 0.5$	$\alpha = 0.8$	$\sigma_\eta^2 = 0.675$	$\sigma_\omega^2 = 0.1$
Mean	0.00001	0.50007	0.80050	0.67461	0.09996
Sd	0.00068	0.01015	0.01539	0.01108	0.00111
B	0.00001	0.00007	0.00050	-0.00039	-0.00004
MSE	0.00000	0.00010	0.00024	0.00012	0.00000
<i>Case 3</i>	$\beta_0 = 0$	$\phi = 0.3$	$\alpha = 0.8$	$\sigma_\eta^2 = 0.819$	$\sigma_\omega^2 = 0.1$
Mean	0.00004	0.29926	0.80020	0.81854	0.10003
Sd	0.00092	0.01160	0.01568	0.01412	0.00093
B	0.00004	-0.00074	0.00020	-0.00046	0.00003
MSE	0.00000	0.00014	0.00025	0.00020	0.00000
<i>Case 4</i>	$\beta_0 = 0$	$\phi = 0.7$	$\alpha = 0.4$	$\sigma_\eta^2 = 0.459$	$\sigma_\omega^2 = 0.1$
Mean	0.00000	0.69986	0.39970	0.45870	0.09997
Sd	0.00051	0.00772	0.00863	0.00735	0.00097
B	0.00000	-0.00014	-0.00030	-0.00030	-0.00003
MSE	0.00000	0.00006	0.00007	0.00005	0.00000
<i>Case 5</i>	$\beta_0 = 0$	$\phi = 0.5$	$\alpha = 0.4$	$\sigma_\eta^2 = 0.675$	$\sigma_\omega^2 = 0.1$
Mean	0.00006	0.49973	0.39999	0.67467	0.09999
Sd	0.00082	0.00981	0.00881	0.01113	0.00069
B	0.00006	-0.00027	-0.00001	-0.00033	-0.00001
MSE	0.00000	0.00010	0.00008	0.00012	0.00000
<i>Case 6</i>	$\beta_0 = 0$	$\phi = 0.3$	$\alpha = 0.4$	$\sigma_\eta^2 = 0.819$	$\sigma_\omega^2 = 0.1$
Mean	0.00000	0.29990	0.40076	0.81917	0.09998
Sd	0.00119	0.01076	0.00944	0.01455	0.00049
B	0.00000	-0.00010	0.00076	0.00017	-0.00002
MSE	0.00000	0.00012	0.00009	0.00021	0.00000

Tabla 3.2: Resultados de las estimaciones vía algoritmo Esperanza-Maximización General para un modelo espacio-tiempo AR(1). Las localidades observadas están en el cuadrado $[0, 1]^2$. Fuente: Elaboración propia.

Capítulo 4

Aplicación a datos reales



Con la finalidad de aplicar el modelo espacio-temporal a través de la metodología propuesta, se analizó parte de la base de datos generada por la red integrada Agromet, Agromet (2017 (accessed February 26, 2017)). Esta red contiene más de 100 estaciones meteorológicas a lo largo de Chile, incluyendo datos de temperatura del aire, temperatura de suelo, precipitaciones, humedad, radiación solar, velocidad y dirección del viento, entre otros. Estos datos son registrados y actualizados por el Instituto de Investigaciones Agropecuarias (INIA), la cual se encuentran disponibles en <http://agromet.inia.cl/>.

El área geográfica de estudio a analizar contempla tres regiones de Chile: Región del Maule, Biobío y Araucanía, desplegadas en la Figura (4.1), de norte a sur, respectivamente, representando una porción del centro-sur de Chile. Hacia el este se encuentra la cordillera de los Andes; y al oeste, la cordillera de la costa, seguido por el océano pacífico, características que hacen de esta zona un lugar naturalmente aislado. Acorde al censo nacional de agricultura del 2007, esta área tiene una superficie de 99.206km^2 . La región del Maule concentra el 17.2 % del área nacional de cultivo, la cual se usa principalmente en la plantación de bosques, seguido por cereales, árboles de frutas, plantas de forraje, viñedo y parronal. La región del Biobío representa el 28 % del área

nacional de cultivo, donde el uso principal es destinado para plantaciones forestales, cultivo de cereales y, con una pequeña participación, plantas de forraje. Finalmente, la región de la Araucanía cubre el 20.6 % del área nacional de cultivo, y su uso principal corresponde a las plantaciones forestales, seguido por granos, y plantas de forraje.

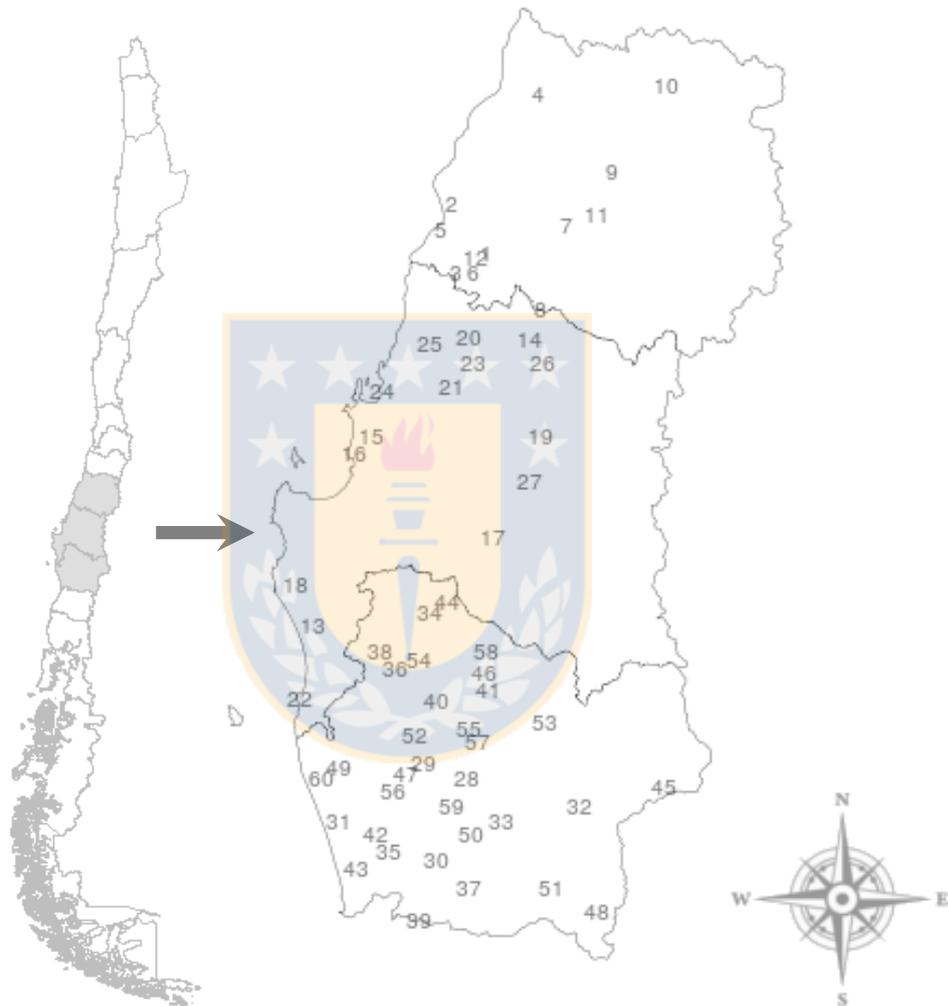


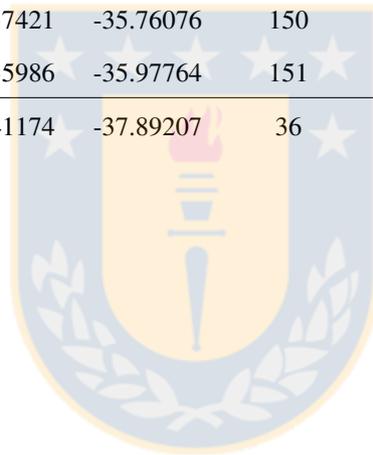
Figura 4.1: *Región de estudio donde se ubican las 60 estaciones meteorológicas. El polígono contempla las regiones del Maule, Biobío y Araucanía, Chile. Cada estación meteorológica se encuentra representada dentro del polígono con su respectiva ID. Fuente: Elaboración propia.*

Notemos que el área de estudio se usa principalmente para la agricultura y silvicult-

tura, por lo tanto la temperatura del aire es un factor predominante en el crecimiento de los cultivos, ya que las temperaturas extremas afectan tanto la producción como la calidad de éstos. Las heladas causan un deterioro de la producción, mermando la calidad de estos, implicando una reducción de la actividad de la industria agrícola en las zonas centro-sur de nuestro país, generando pérdidas de miles de dólares, además de la paralización y baja actividad en la exportación hacia los mercados externos. De este modo, es de gran interés estudiar la variabilidad espacio-temporal de este proceso meteorológico, y así generar una propuesta que ayude a explicar esta variabilidad en términos de fuentes identificables y predecir de manera coherente.

Particularmente, se estudió el comportamiento de la variabilidad espacio-tiempo de la temperatura promedio diaria del aire. La información fue obtenida de 60 estaciones meteorológicas ubicadas entre las regiones del Maule, Biobío y Araucanía, medida entre el 1 de enero de 2015 y 31 de mayo de 2018. La tabla 4.1, resume las estaciones meteorológicas con sus respectivas coordenadas geográficas, medidas en latitud y longitud; la elevación, medida en metros sobre el nivel del mar, el promedio y la desviación estándar muestral de la temperatura junto al porcentaje de datos perdidos.

ID	long	lat	elevation	avg	sd	% NA
1	-72.28994	-35.95597	162	14.5	5.5	2.0
2	-72.51119	-35.70658	60	12.8	2.8	0.0
3	-72.47769	-36.05909	172	14.0	5.2	0.0
4	-71.95595	-35.12782	48	14.3	3.7	76.0
5	-72.58749	-35.83545	172	12.2	2.9	76.0
6	-72.37144	-36.06211	167	14.3	5.3	0.0
7	-71.77502	-35.81805	124	17.0	4.6	83.4
8	-71.93429	-36.24480	548	13.9	5.2	26.8
9	-71.47951	-35.53065	221	14.0	5.4	0.0
10	-71.12058	-35.09180	310	13.9	4.6	76.0
11	-71.57421	-35.76076	150	13.7	5.2	2.8
12	-72.35986	-35.97764	151	13.8	5.3	0.0
13	-73.41174	-37.89207	36	14.3	5.0	0.2



14	-72.00673	-36.40930	162	13.4	5.0	3.0
15	-73.03546	-36.91349	23	13.8	4.0	0.0
16	-73.14041	-37.00482	57	12.9	2.9	0.0
17	-72.24398	-37.43356	195	13.0	4.8	0.0
18	-73.52217	-37.68254	194	11.4	2.9	0.9
19	-71.93560	-36.90730	314	12.9	4.9	0.0
20	-72.39520	-36.39811	91	14.9	5.3	2.3
21	-72.51349	-36.64890	82	13.2	3.9	0.0
22	-73.49849	-38.27529	143	11.8	2.5	29.3
23	-72.36827	-36.53204	109	14.4	5.3	0.8
24	-72.96468	-36.67239	190	13.2	3.4	2.0
25	-72.64778	-36.43104	36	13.2	3.9	79.8
26	-71.91643	-36.53520	194	13.4	5.1	0.0
27	-72.01176	-37.14138	265	13.4	5.1	2.8
28	-72.41722	-38.69306	200	11.2	4.1	0.6
29	-72.70106	-38.60670	60	11.9	2.9	94.9
30	-72.60528	-39.11556	115	11.9	4.2	0.0
31	-73.24028	-38.91861	93	11.9	2.9	0.0
32	-71.68250	-38.83417	527	11.0	4.8	3.8
33	-72.18351	-38.91318	293	11.4	4.1	48.8
34	-72.65966	-37.82781	81	12.4	3.3	92.6
35	-72.92198	-39.07115	56	11.5	3.4	75.6
36	-72.88350	-38.11692	76	12.5	4.2	75.6
37	-72.39889	-39.26333	266	11.1	4.2	3.6
38	-72.98200	-38.03000	76	12.0	4.5	29.8
39	-72.71852	-39.42813	86	11.9	3.8	75.6
40	-72.61139	-38.28944	254	11.7	4.3	0.2
41	-72.27333	-38.22778	388	10.9	4.3	5.5
42	-72.99778	-38.97694	24	11.5	3.7	0.0
43	-73.13294	-39.16541	8	11.4	2.9	77.4
44	-72.54611	-37.77250	88	13.2	4.8	12.3
45	-71.14083	-38.73167	1084	7.6	5.5	4.7
46	-72.29833	-38.14555	379	10.4	3.1	93.0

47	-72.80889	-38.66389	55	11.5	4.0	7.2
48	-71.57250	-39.38194	393	11.1	4.8	5.2
49	-73.24417	-38.63972	322	11.1	3.2	0.0
50	-72.38162	-38.98547	168	11.2	3.5	91.8
51	-71.86810	-39.26171	279	11.1	4.5	74.7
52	-72.75147	-38.46692	197	11.5	4.1	5.1
53	-71.90250	-38.40167	549	10.4	4.5	0.0
54	-72.72418	-38.07067	173	12.8	4.2	75.6
55	-72.39902	-38.43506	299	10.3	3.0	92.6
56	-72.89139	-38.75611	50	12.0	4.0	0.0
57	-72.34361	-38.50472	263	11.7	4.1	4.7
58	-72.28528	-38.03250	379	11.4	4.5	4.9
59	-72.50638	-38.83295	130	11.1	3.2	92.6
60	-73.35333	-38.69083	64	11.9	3.2	0.0

Tabla 4.1: Resumen de las 60 estaciones meteorológicas en la región del Maule, Biobío y Araucanía, Chile. Además elevación en metros sobre el nivel del mar, promedio y desviación estándar muestral de la temperatura del aire medida desde el 1 enero de 2015 hasta el 31 de mayo de 2018, y % de datos faltantes. Estaciones: 1 Cauquenes; 2 Chanco; 3 Coronel de Maule; 4 Deuca; 5 Lomas; 6 Los Despachos; 7 Miraflores; 8 Monte Flor-Tucapel; 9 San Clemente; 10 San Jorge Los Niches; 11 Santa Amada; 12 Santa Sofía; 13 Cañete; 14 CE Arroz; 15 Chiguayante; 16 Coronel; 17 Human; 18 Lebu; 19 Navidad; 20 Ninhue; 21 Nueva Aldea; 22 Ponotro; 23 Portezuelo; 24 Punta Parra; 25 Puralihue; 26 Santa Rosa; 27 Yungay; 28 Carillanca; 29 Collimallin; 30 Cuarta Faja; 31 Dominguez; 32 El Membrillo; 33 El Quincho; 34 El Vergel; 35 Faja Maisan; 36 Gaby-Ranquilco; 37 Huiscaipi; 38 La Isla; 39 La Paz; 40 La Providencia; 41 Las Palmas; 42 Llollinco; 43 Los Arrayanes; 44 Manzanares; 45 Marimenuco; 46 Pailahuenque; 47 Perales; 48 Puala; 49 Quiripio; 50 Radal; 51 San Enrique; 52 San Fabian; 53 San Luis; 54 San Rafael; 55 San Sebastian; 56 Santa Adela; 57 Santa Inés; 58 Surco y Semilla; 59 Taplon; 60 Tranapuenta. Fuente: Elaboración propia.

4.1. Análisis exploratorio

Previo al ajuste del modelo identificado, debemos capturar la componente sistemática para que el modelo base propuesto se encargue de capturar la correlación espacio-temporal, así el estudio de una posible incidencia de alguna covariable, como por ejemplo la latitud, longitud y elevación del índice observado. Una sospecha intuitiva es que la latitud afecte directamente sobre la temperatura, esto es, a medida que nos desplazamos hacia el norte, la temperatura asciende, entre esto está evidenciado en la Figura 4.2. El registro promedio de la temperatura es más elevado en latitudes mayores. Por otro lado, la elevación es otra covariable que también afecta directamente en la temperatura. El gradiente térmico es variable en función de la altitud sobre el nivel medio del mar, la temperatura disminuye más o menos uniforme con la altitud a una razón promedio de $6.5^{\circ}C$ por kilómetro (Nadal and Muñuzuri, 2006), esto es, $1^{\circ}C$ por cada $154m$ de altura, aproximadamente. En la Figura 4.2 se puede evidenciar tal hecho, en lo particular un outlier correspondiente a la estación de Marimenuco, la cual se encuentra ubicada a $1084m$ sobre el nivel medio del mar y que ha reportado una temperatura promedio de $7.6^{\circ}C$, bastante más baja en comparación a las demás estaciones.

Del total de estaciones meteorológicas, las más cercanas entre si son Santa Sofía (12) y Cauquenes (1), la cual están separadas por $6.75km$ aproximadamente, mientras que las más distantes son La Paz (39) y San Jorge Los Niches (10), separadas por $501.89km$ aproximadamente. La estación más elevada es Marimenuco (45), ubicada en la cordillera de los Andes en región de la Araucanía, a una altitud de 1084 metros sobre el nivel medio del mar, mientras que la menos elevada es la estación de Los Arrayanes (43), ubicada en la zona costera de la región de la Araucanía, a una altitud de 8 metros sobre el nivel medio del mar. La Tabla 4.2 resume parte de la información georeferencial de las 60 estaciones meteorológicas utilizadas en el estudio.

Busqueda de tendencia interna

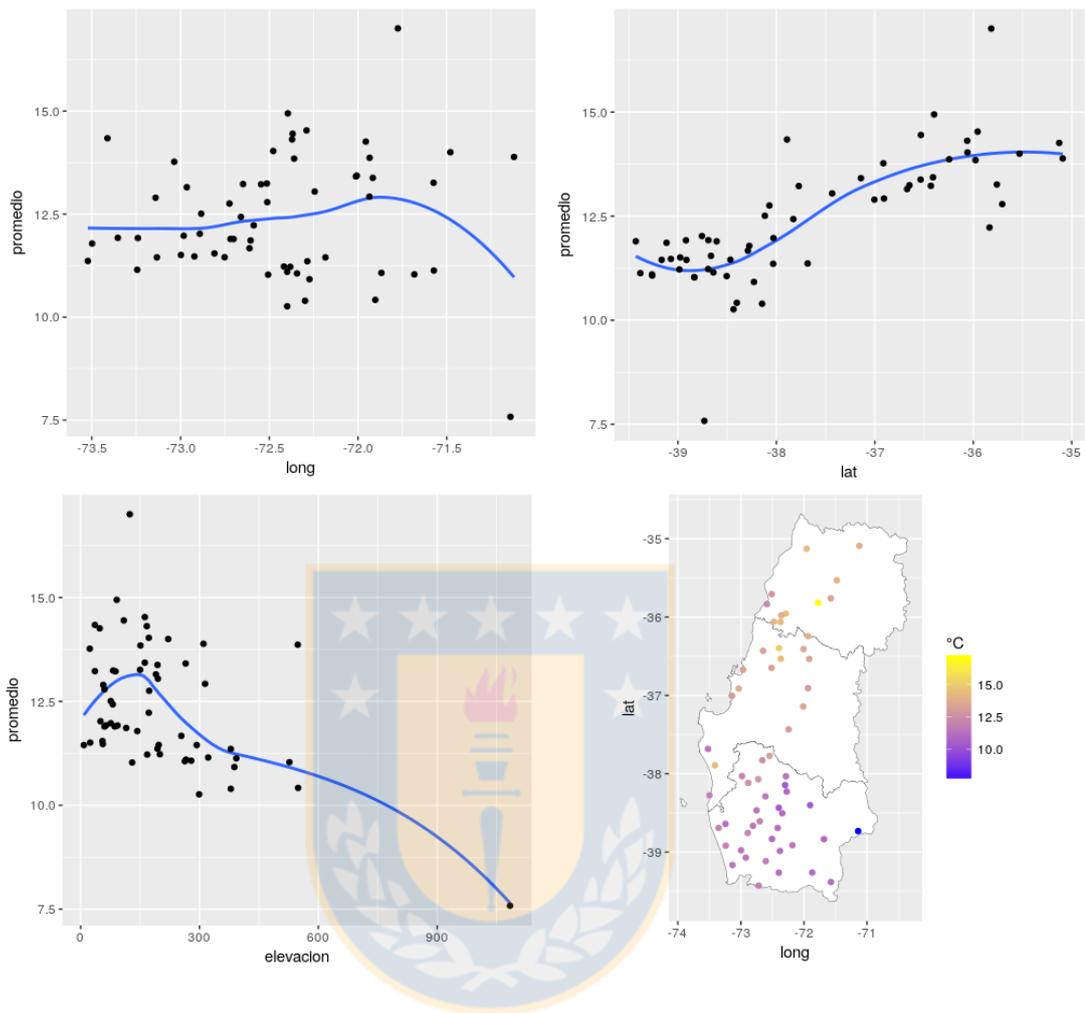


Figura 4.2: Gráfico de dispersión para detectar componente sistemática. Arriba a la izquierda, longitud versus valor promedio del registro de temperaturas por cada estación; arriba a la derecha, latitud versus valor promedio del registro de temperaturas por cada estación; abajo a la izquierda, elevación versus valor promedio del registro de temperaturas por cada estación; abajo a la derecha, distribución espacial del valor promedio del registro de temperaturas por cada estación. Fuente: Elaboración propia.

	Min	Max
<i>longitud</i>	-73.52217	-39.42813
<i>latitud</i>	-71.12058	-35.09180
<i>altitud</i>	8m	1084m
<i>distancia</i>	0.07320688 (6754.983m)	4.62138722 (501887.7m)

Tabla 4.2: Resumen georreferencial de las 60 estaciones meteorológicas. Fuente: Elaboración propia.

De esta forma, para modelar la componente sistemática, se consideraron como co-variables el índice del tiempo, para capturar el efecto estacional causado por las estaciones del año, la latitud y la elevación. Las altitudes fueron medidas con un datum vertical EGM96, medida sobre el nivel medio del mar. Desde el modelo digital de elevación ASTER GDEM del año 2011, remuestreado a un tamaño del pixel de 30 metros usando un método de interpolación bilineal. Esta última con el objetivo de poder entender el comportamiento del gradiente térmico a un nivel más local, así

$$\mu_t(\mathbf{s}) = \beta_0 + \beta_1 \cos\left(\frac{2\pi t}{365.25}\right) + \beta_2 \sin\left(\frac{2\pi t}{365.25}\right) + \beta_3 \text{lat} + \beta_4 h(\mathbf{s}),$$

donde $\mathbf{s} = (\text{long}, \text{lat})^\top$, y $h(\mathbf{s})$ hace referencia a la elevación en la localidad \mathbf{s} .

4.2. Estimación de parámetros del modelo

Con la finalidad de explicar la variabilidad espacio-temporal del conjunto de datos, se propuso el siguiente modelo

$$\begin{aligned} \mathbf{Z}_t(\mathbf{s}) &= \mu_t(\mathbf{s}) + \varepsilon_t(\mathbf{s}) + \omega_t(\mathbf{s}), \quad \omega_{\mathcal{D} \times \mathcal{T}} \sim \mathcal{GWN}(0, \sigma_\omega^2), \\ \mu_t(\mathbf{s}) &= \beta_0 + \beta_1 \cos\left(\frac{2\pi t}{365.25}\right) + \beta_2 \sin\left(\frac{2\pi t}{365.25}\right) + \beta_3 \text{lat} + \beta_4 h(\mathbf{s}), \end{aligned} \quad (4.1)$$

$$\varepsilon_t(\mathbf{s}) = \phi \varepsilon_{t-1}(\mathbf{s}) + \eta_t(\mathbf{s}),$$

asociado a un modelo exponencial para la estructura de la covarianza del proceso de innovación espacial $\eta_{\mathcal{D} \times \mathcal{T}}$

$$C^\eta(\mathbf{s}, \mathbf{r}) = \sigma_\eta^2 \exp\left(-\frac{\|\mathbf{s} - \mathbf{r}\|}{\alpha}\right),$$

donde la componente sistemática, $\mu_t(\mathbf{s})$, explica la variabilidad térmica causada por las estaciones del año, la latitud y la elevación. El proceso de estado, $\varepsilon_{\mathcal{D} \times \mathcal{T}}$, especificado por un modelo autorregresivo de primer orden en el tiempo.

La estimación de los parámetros del modelo, vía algoritmo EM General, se resume en la tabla 4.3. Entre las estimaciones involucradas en la componente sistemática, se puede observar que por cada grado de latitud hacia el norte, con respecto al ecuador, (desplazamiento aproximado de 111.12 km hacia el norte), la temperatura promedio media aumenta 0.856°C ($\hat{\beta}_3 = 0.856$), y por cada 250 m de altura, la temperatura promedio media disminuye aproximadamente 1°C ($\hat{\beta}_4 = -0.004$). Por otro lado, se puede observar que el coeficiente autorregresivo estimado es alto, $\hat{\phi} = 0.869$, la cual sugiere que la temperatura promedio está altamente correlacionada a la temperatura

promedio del día anterior. Adicionalmente, se puede observar que el alcance estimado es moderadamente alto, $\hat{\alpha} = 2.749$, con respecto al esquema de muestreo, ya que la distancia máxima entre las estaciones es de $h_{max} = 4.62$, reportada en la Tabla 4.2, indicando que el proceso tiene una alta continuidad espacial, la cual se evidencia en las Figuras 4.4, 4.5 y 4.6, donde las temperaturas predichas para cada mes presentan un suavizamiento en su variabilidad. Finalmente, el efecto pepita es pequeño, $\hat{\sigma}_{\omega}^2 = 0.148$, explicando el 1.3 % de la varianza total del proceso, valor que nos garantiza una alta precisión a la hora de realizar predicciones espacio-temporales.

$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\phi}$	$\hat{\alpha}$	$\hat{\sigma}_{\eta}^2$	$\hat{\sigma}_{\omega}^2$
44.713	4.741	2.259	0.856	-0.004	0.869	2.749	2.844	0.148

Tabla 4.3: *Estimación de parámetros del modelo espacio-tiempo propuesto. Estimaciones realizadas a través del algoritmo Esperanza-maximización General. Fuente: Elaboración propia.*

4.3. Predicción

Dado que la mayoría de las estaciones meteorológicas presenta datos faltantes, uno de los principales intereses del investigador podría ser obtener la predicción más confiable posible en esos períodos de tiempo donde no hubo registro, para cada estación. Así, podemos utilizar el modelo (4.1) y la metodología propuesta e imputar la información perdida. La Figura 4.3 muestra la imputación lograda de 3 estaciones meteorológicas: Monte Flor-Tucapel (8), San Jorge de Niches (10) y El Quincho (33). Se puede observar que las predicciones en San Jorge de Niches presenta intervalos de confianza más grandes con respecto a las otras dos, esto debido a la lejanía de esta estación al resto de las otras.

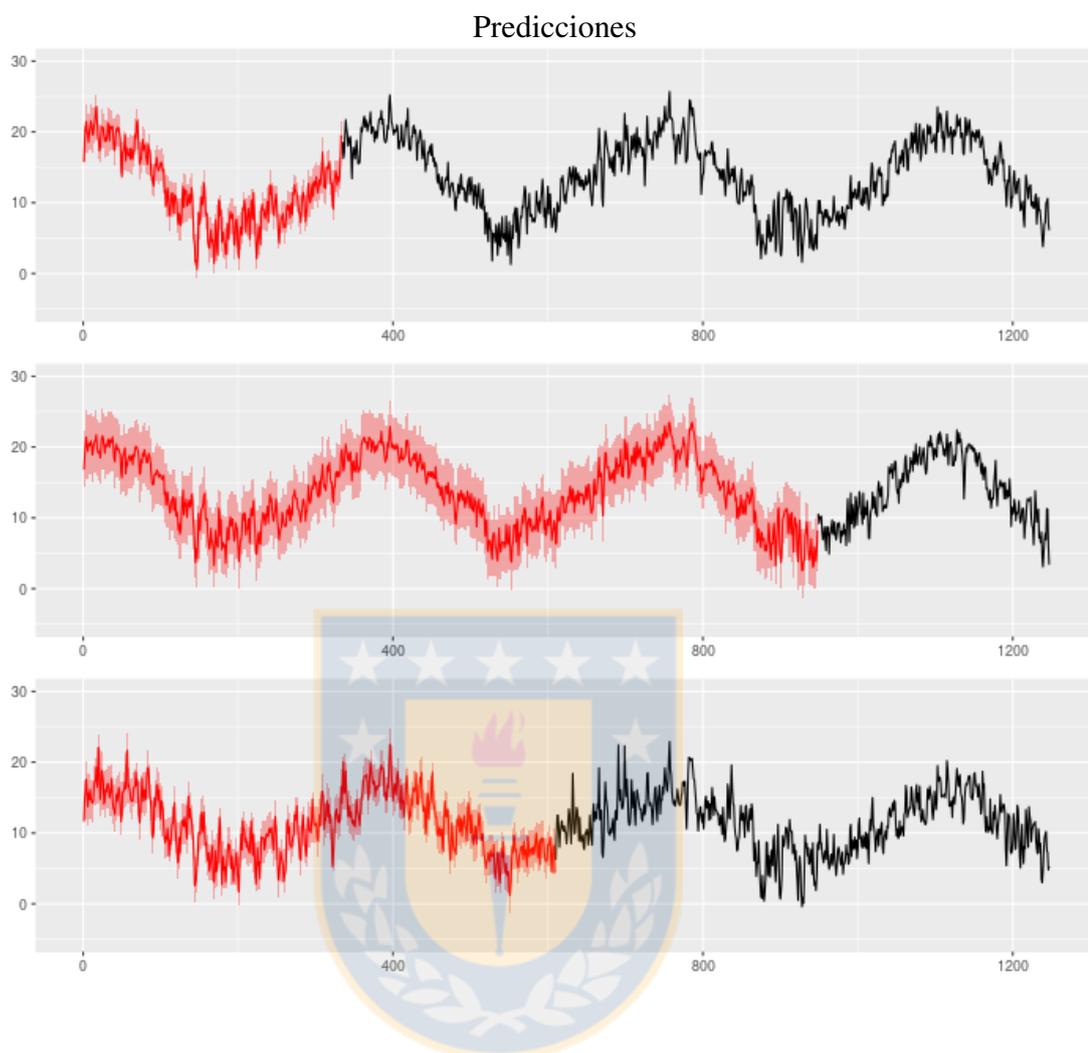


Figura 4.3: *Imputación de información perdida. De arriba hacia abajo: la primera, Monte Flor-Tucapel (8); la segunda, San Jorge los Niches (10); la tercera, El Quincho (33). Fuente: Elaboración propia.*

Con el objetivo de realizar mapas de predicciones para un día en específico de la temperatura promedio dentro de la VII, VII y IX región de Chile, se consideró una grilla de 5955 nodos dentro del polígono de estudio y con esto se obtuvieron predicciones espacio-temporales desde el 1 de Enero de 2015 hasta el 31 de Diciembre de 2017, conformando un total de 1247 mapas de predicción. Con un fin ilustrativo, se seleccionaron los mapas de predicción asociados a las quincenas de cada mes, mostrados en las Figuras 4.4, 4.5 y 4.6 junto a sus respectivas varianzas de error predicción.

Predicción año 2015

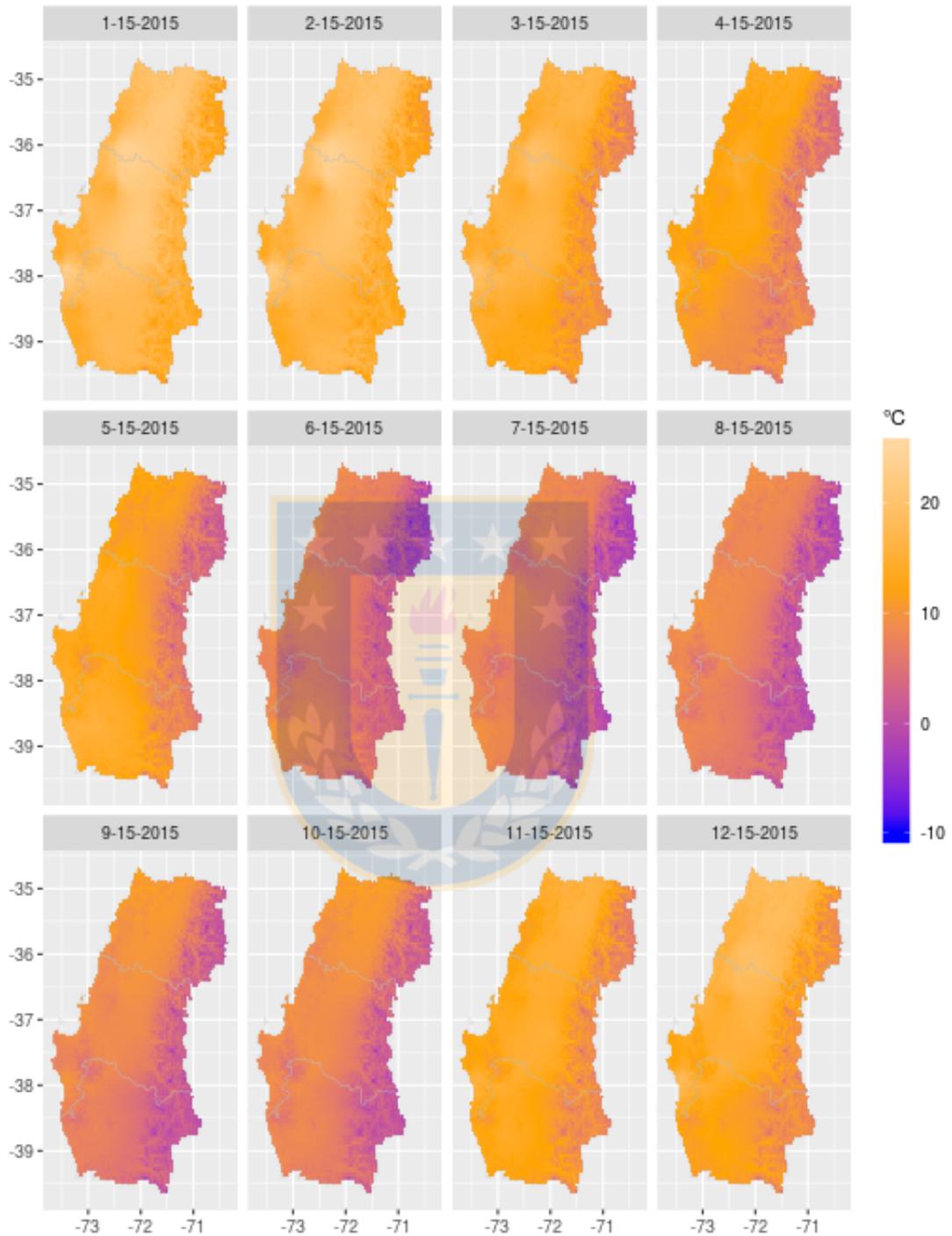


Figura 4.4: Mapa de predicción de temperatura promedio asociada a la quincena de cada mes del año 2015. Predicciones logradas a través del modelo espacio-tiempo AR(1) propuesto. Fuente: Elaboración propia.

Predicción año 2016

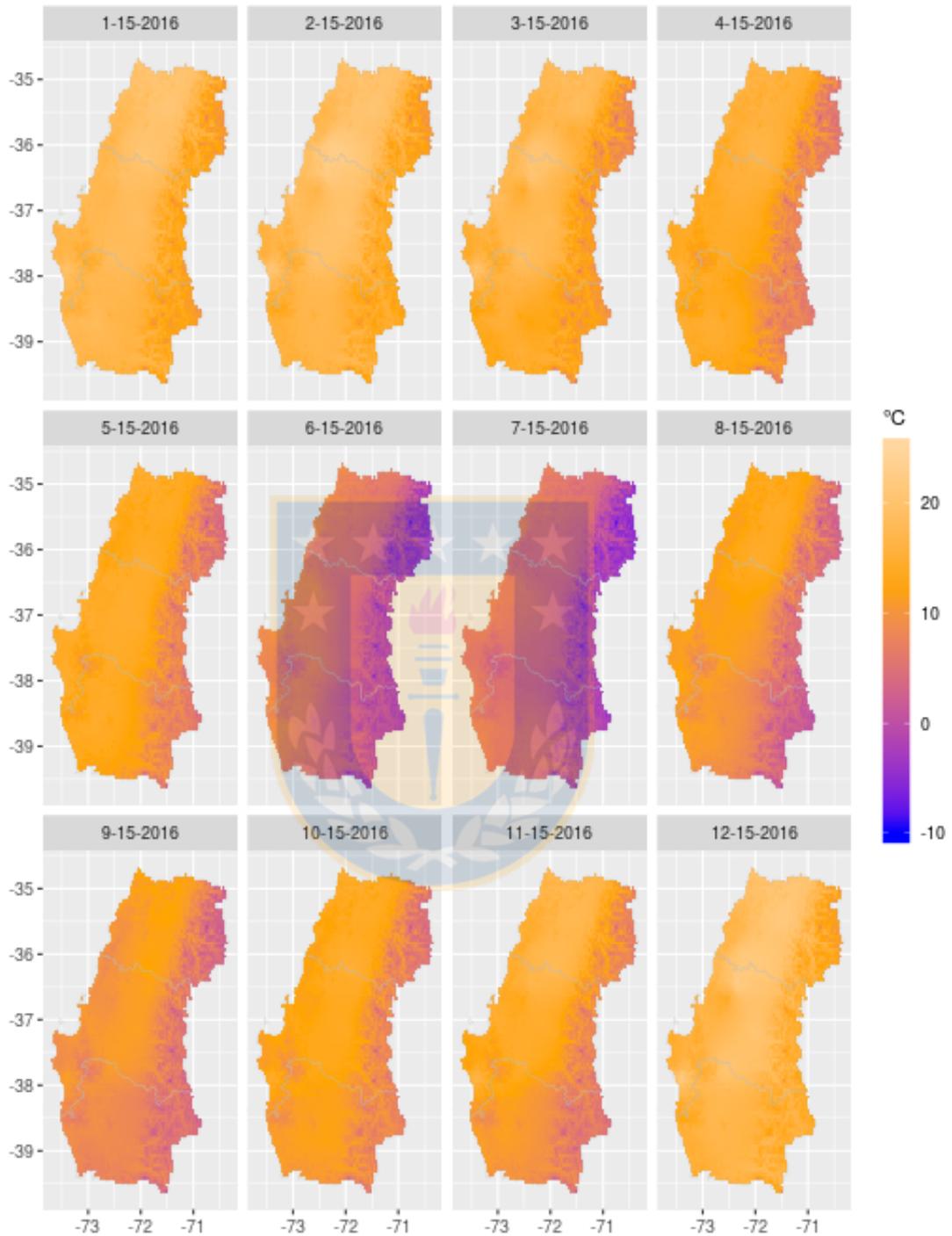


Figura 4.5: Mapa de predicción de temperatura promedio asociada a la quincena de cada mes del año 2016. Predicciones logradas a través del modelo espacio-tiempo AR(1) propuesto. Fuente: Elaboración propia.

Predicción año 2017

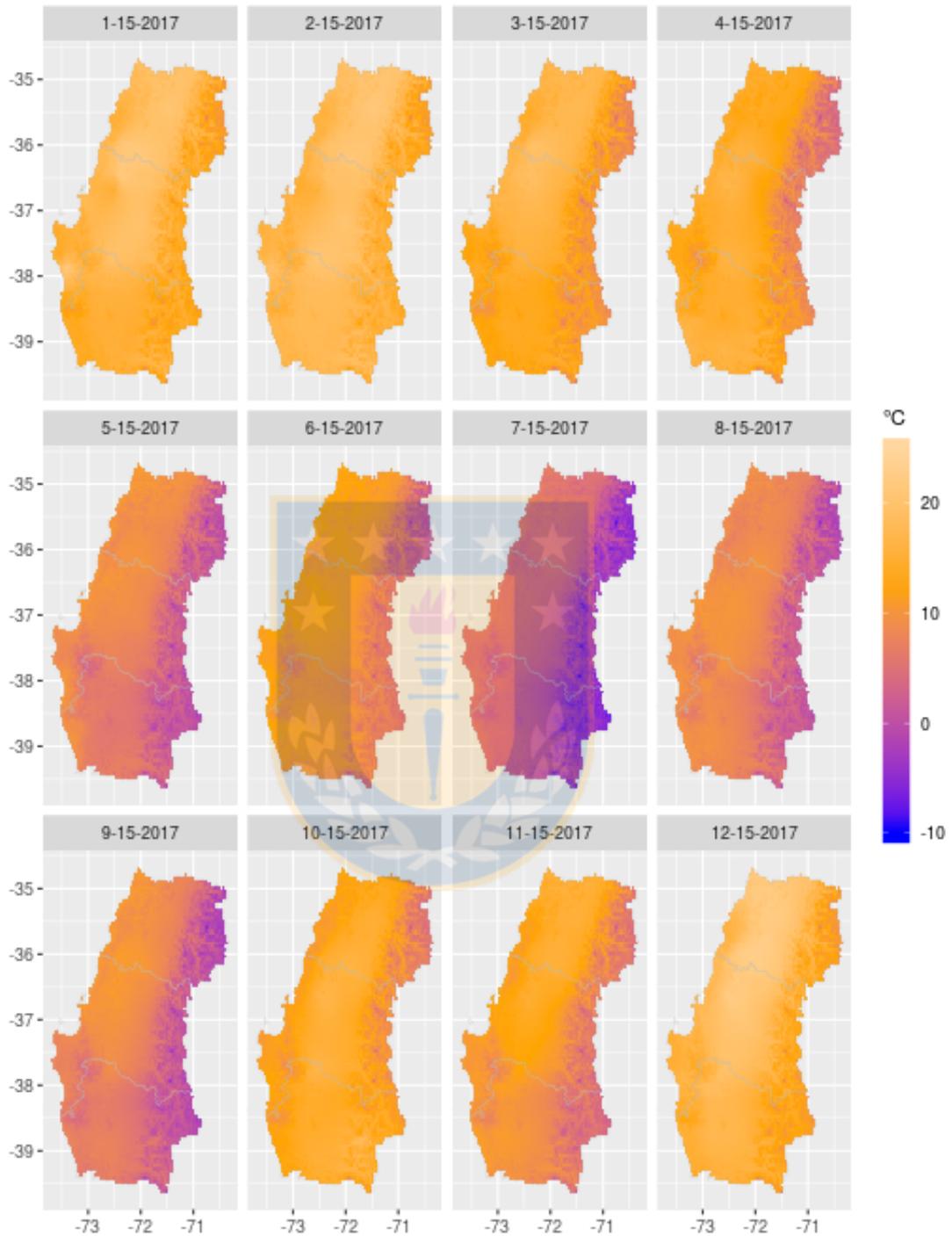


Figura 4.6: Mapa de predicción de temperatura promedio asociada a la quincena de cada mes del año 2017. Predicciones logradas a través del modelo espacio-tiempo $AR(1)$ propuesto. Fuente: Elaboración propia.

Varianza de predicción, año 2015

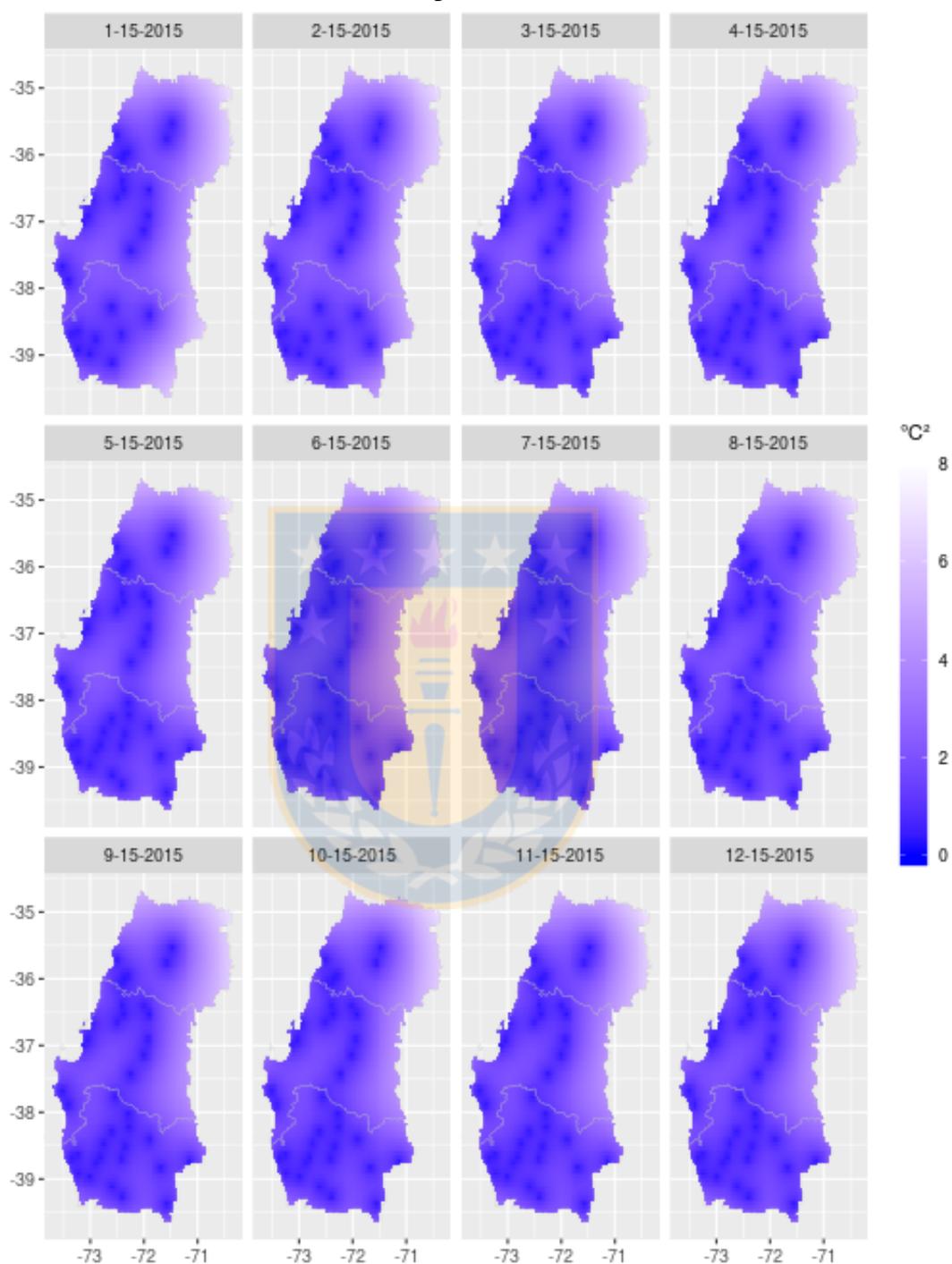


Figura 4.7: Mapa de varianza de predicción de temperatura promedio asociada a la quincena de cada mes del año 2015. Predicciones logradas a través del modelo espacio-tiempo AR(1) propuesto. Fuente: Elaboración propia.

Varianza de predicción, año 2016

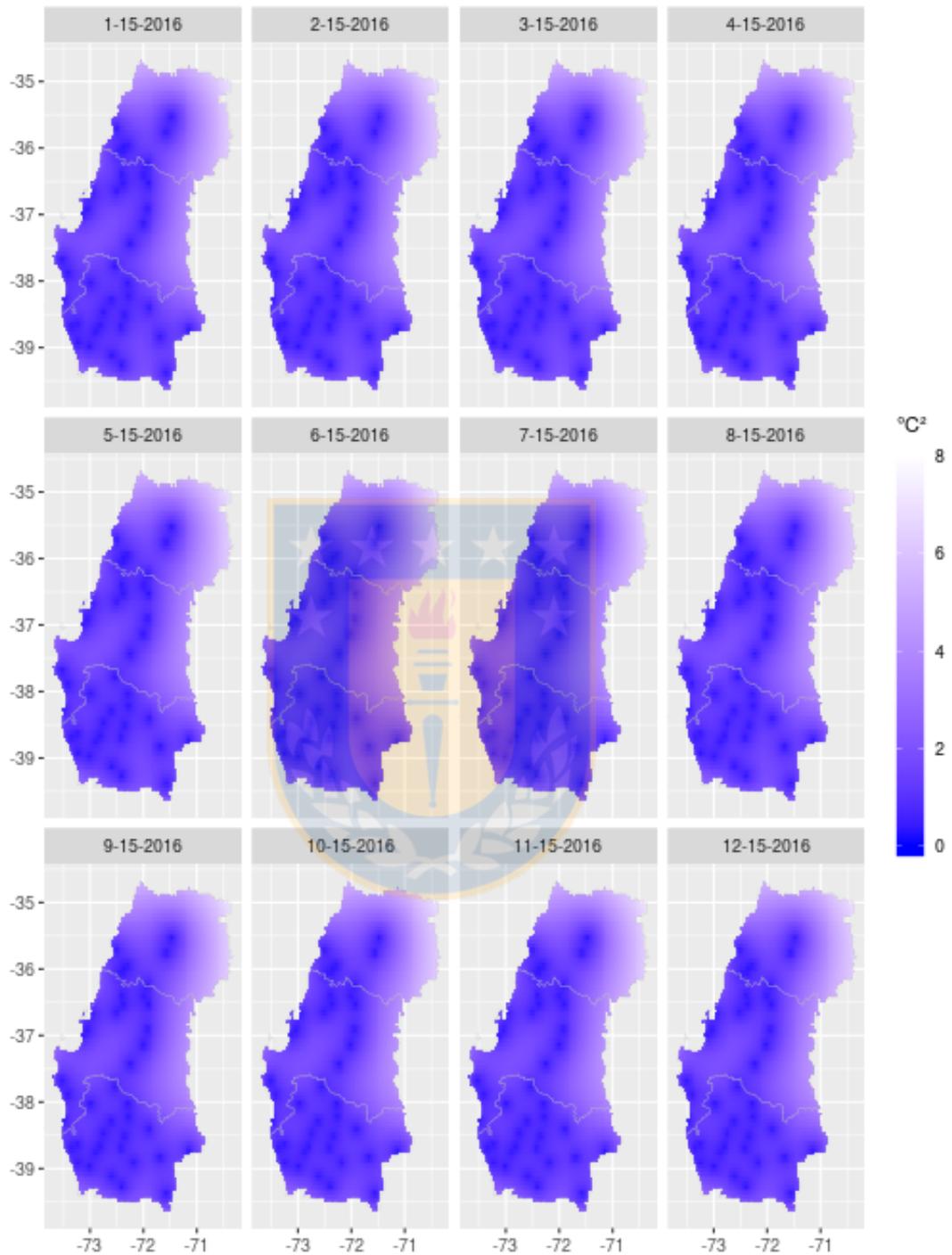


Figura 4.8: Mapa de varianza de predicción de temperatura promedio asociada a la quincena de cada mes del año 2016. Predicciones logradas a través del modelo espacio-tiempo AR(1) propuesto. Fuente: Elaboración propia.

Varianza de predicción, año 2017

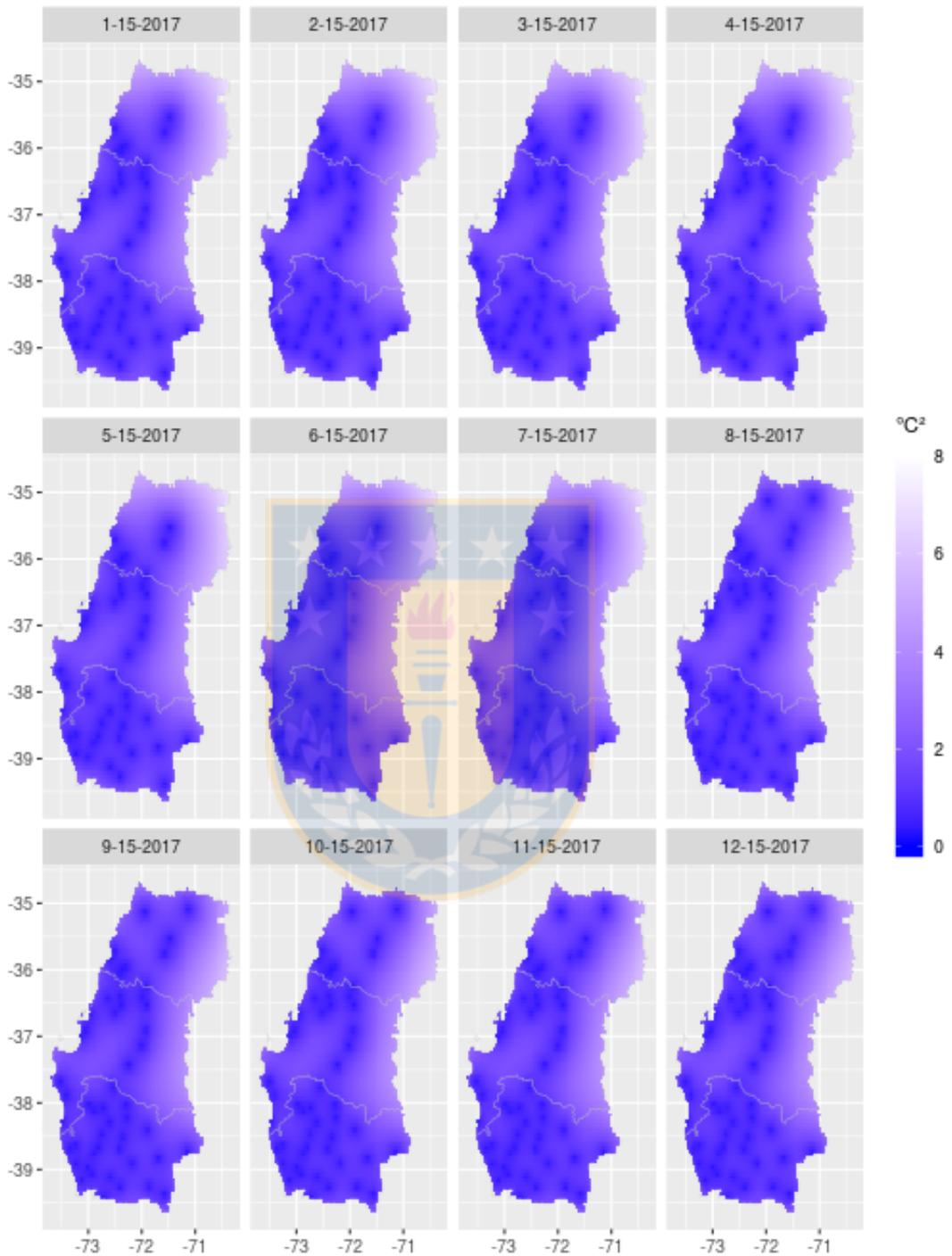


Figura 4.9: Mapa de varianza de predicción de temperatura promedio asociada a la quincena de cada mes del año 2017. Predicciones logradas a través del modelo espacio-tiempo AR(1) propuesto. Fuente: Elaboración propia.

Conclusión

En la práctica, dentro de las ciencias ambientales y geofísicas entre otras, los datos son recopilados en estaciones ubicadas en distintos lugares dentro de una región de interés, es muy habitual la no obtención de dicha información, causando períodos de tiempo en la cual hay datos faltantes. Esto motivó, a través de la representación espacio-estado, implementar un método que nos ayude a enfrentar el análisis bajo esta configuración de los datos, actualizando de manera congruente al momento de incorporarse nueva información en el sistema, dentro de la distribución de filtro.

Metodología de filtraje que no ha sido discutida por ningún autor, dentro de la familia de modelos dinámicos espacio-temporales.

Por otro lado, dada la comparativa en los métodos de estimación de parámetros, se concluyó que los estimadores logrados a través del algoritmo EM presentan mayor poder de eficiencia e insesgadez, siendo éste un método preferencial al momento de modelar.

Bibliografía

Agromet (2017 (accessed February 26, 2017)). Agromet: Red agrometereológica de inia. <http://www.agromet.cl>.

Anderson, B. D. and Moore, J. B. (1979). Optimal filtering, *Englewood Cliffs* **21**: 22–95.

Cressie, N. and Wikle, C. K. (2015). *Statistics for Spatio-temporal Data*, John Wiley & Sons.

Gneiting, T., Genton, M. G. and Guttorp, P. (2006). Geostatistical space-time models, stationarity, separability, and full symmetry, *Monographs On Statistics and Applied Probability* **107**: 151–175.

Huang, H. and Cressie, N. (1996). Spatio-temporal prediction of snow water equivalent using the kalman filter, *Computational Statistics & Data Analysis* **22**: 159–175.

Jazwinski, A. (1970). *Stochastic processes and filtering theory Jazwinski*, AH Academic Press: New York.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems, *Journal of Basic Engineering* **82(1)**: 35–45.

Kwong, K.-S. and Iglewicz, B. (1996). On singular multivariate normal distribution and its applications, *Computational Statistics & Data Analysis* **22(3)**: 271–285.

Matérn, B. (2013). *Spatial Variation*, Vol. **36**, Lecture Notes in Statistics, Springer Science & Business Media.

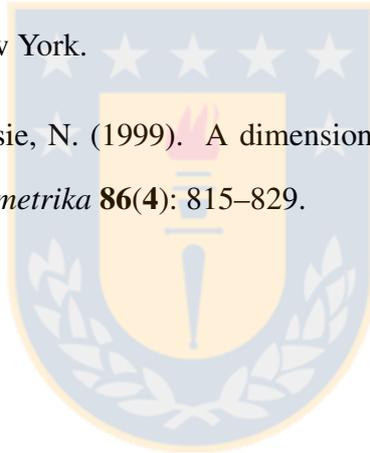
Nadal, I. S. n. and Muñuzuri, V. P. (2006). *Fundamentos de Meteorología*, Vol. **6**, Univ. Santiago de Compostela.

Searle, S. (1979). *R. 1971. Linear Models*, New York: WileySearleLinear models1971.

Shumway, R. and Stoffer, D. (2006). *Time Series Analysis and Its Applications*, Springer. New York.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*, Springer-Verlag, New York.

Wikle, C. K. and Cressie, N. (1999). A dimension-reduced approach to space-time kalman filtering, *Biometrika* **86(4)**: 815–829.



Anexos

Función de Bessel

Una ecuación diferencial muy importante en matemáticas aplicadas es la ecuación diferencial de Bessel, (Bowman 1958):

$$x^2 y'' + xy' + (x^2 - \kappa^2)y = 0,$$

para la cual una solución está dada por la llamada función de Bessel de primera clase de orden κ :

$$\mathcal{J}_\kappa(x) = x^\kappa \sum_{m=0}^{\infty} \frac{(-1)^m x^{2m}}{2^{2m+\kappa} m! \Gamma(m + \kappa + 1)}.$$

La función:

$$\mathcal{I}_\kappa(x) = i^\kappa \mathcal{J}_\kappa(ix) = \sum_{m=0}^{\infty} \frac{x^{2m+\kappa}}{2^{2m+\kappa} m! \Gamma(m + \kappa + 1)},$$

es conocida como la función de Bessel modificada de la primera clase de orden κ . Otra solución de la ecuación diferencial está dada por la función de Bessel modificada de la segunda clase de orden κ :

$$\mathcal{K}_\kappa(x) = \frac{\pi}{2 \sin(\kappa\pi)} [\mathcal{I}_{-\kappa(x)} - \mathcal{I}_\kappa(x)].$$

Distribución Gaussiana singular

Sea Σ una matriz simétrica de dimensión $(p \times p)$ y de rango $k < p$, así Σ^{-1} no está definida. Ahora, Σ tiene sólo k valores propios positivos y los restantes $p - k$ son cero. La p.d.f de la distribución Gaussiana singular $\mathbf{X} \sim SGau(\boldsymbol{\mu}, \Sigma)$ es definida en términos de su inversa generalizada y la reducción de dimensión:

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-k/2} |\mathbf{D}|^{-1/2} \exp \{ (\mathbf{x} - \boldsymbol{\mu})^{\top} \Sigma^{-} (\mathbf{x} - \boldsymbol{\mu}) \},$$

donde $\Sigma = \mathbf{E}\mathbf{D}\mathbf{E}^{\top}$, con $\mathbf{D} = \text{diag}\{d_1, d_2, \dots, d_k\}$ y \mathbf{E} una matriz de dimensión $(p \times k)$ de rango completo k y donde sus columnas son los vectores propios asociados a los k valores propios positivos.

Operadores utilizados en el estimador de ϕ

Sean \mathbf{B} y \mathbf{A}_{ij} matrices cuadradas de dimensión $(k \times k)$ y sea \mathbf{A} una matriz constituida por n^2 bloques dadas por

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \cdots & b_{kk} \end{pmatrix}, \quad \mathbf{A}_{ij} = \begin{pmatrix} a_{11}^{(ij)} & a_{12}^{(ij)} & \cdots & a_{1k}^{(ij)} \\ a_{21}^{(ij)} & a_{22}^{(ij)} & \cdots & a_{2k}^{(ij)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1}^{(ij)} & a_{k2}^{(ij)} & \cdots & a_{kk}^{(ij)} \end{pmatrix},$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1n} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{n1} & \mathbf{A}_{n2} & \cdots & \mathbf{A}_{nn} \end{pmatrix},$$

y sean v_1, v_2, v_3 operadores dados por

$$v_1(\mathbf{B}) = \begin{pmatrix} b_{11} & 0 & \cdots & 0 \\ 0 & b_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_{kk} \end{pmatrix}, v_2(\mathbf{B}) = \begin{pmatrix} 0 & b_{12} & \cdots & b_{1k} \\ b_{21} & 0 & \cdots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \cdots & 0 \end{pmatrix},$$

$$v_3(\mathbf{B}) = \begin{pmatrix} b_{11} \\ b_{12} \\ \vdots \\ b_{1k} \end{pmatrix},$$

así, podemos definir las funciones $\mathcal{H}_1(\cdot), \mathcal{H}_2(\cdot), \mathcal{H}_3(\cdot)$ por

$$\mathcal{H}_1(\mathbf{A}) := v_1 \left(\sum_{i=1}^n \mathbf{A}_{ii} \right), \mathcal{H}_2(\mathbf{A}) := v_2 \left(\sum_{i=1}^n \mathbf{A}_{ii} \right), \mathcal{H}_3(\mathbf{A}) := v_3 \left(\sum_{i=1}^n \mathbf{A}_{ii} \right)$$

la cual son empleadas en el estimador de ϕ en el algoritmo EM.