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

Estimation of the variance-covariance matrix for the Maximum Likelihood Estimator in Space-Time autoregressive model in the presence of missing data

Tesis para optar al grado de Magíster en Estadística

Salomé Andrea Zaldúa Flores

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

Estimation of the variance-covariance matrix for the Maximum Likelihood Estimator in Space-Time autoregressive model in the presence of missing data

Profesor Guía	:	Bernardo Lagos-Álvarez	Firma	.....
Profesor Colaborador	:	Leonardo Padilla Sepúlveda	Firma	.....
Profesor Colaborador	:	Jorge Figueroa-Zúñiga	Firma	.....
Profesor Colaborador	:	Jorge Mateu	Firma	.....
Nombre Memorante	:	Salomé Zaldúa Flores	Firma	.....
Teléfono	:	+56 9 96171063		
e-mail	:	szaldua@udec.cl		

Concepción, 2025

Dedicatoria

A mis hijos, Cristóbal y Andrés.

Agradecimientos

Quisiera agradecer a Dios por todo lo que me ha dado, mis capacidades y las oportunidades que han hecho posible este camino.

A mi madre, por todo su apoyo y cariño; a mis sobrinos y sobrinas, por ser siempre una alegría en vida; a mi esposo y mis hijos, siempre me han acompañado y apoyado con paciencia en todo los momentos de mi vida.

A la Universidad de Concepción, especialmente al Departamento de Estadística de la Facultad de Ciencias Físicas y Matemáticas y a todos quienes lo conforman. Agradezcol a mis profesores Bernardo Lagos, Leonardo Padilla y Jorge Figueroa, quienes con paciencia y compromiso me guiaron durante estos años, entregandome las herramientas y motivacion para realizar esta investigación. A mis compañeros de magister y a los estudiantes que tuve el privilegio enseñarles mientras yo aun estaba aprendiendo.

A Dirección de Postgrado, por creer nuevamente en mi y apoyar economicamente esta etapa de formación.

Por supuesto, a mis amigos y amigas, en especial a Benito, Maca, Rosario, Stephy y Eduardo, Gracias por estar siempre y por escucharme (aunque a veces no entendieran

nada de mis calculos).

A todos ellos que sin su colaboración, directa o indirecta, no habría sido posible el desarrollo de esta investigación.

Table of Contents

List of Figures

List of Tables

Resumen

Los datos espacio-temporales han cobrado una creciente relevancia en diversas disciplinas como la epidemiología, las ciencias sociales, los estudios ambientales, la meteorología y la agricultura. El análisis preciso de estas bases de datos es fundamental para una contar con modelos e inferencia confiables, especialmente en presencia de datos faltantes, lo que requiere el uso de algoritmos eficientes capaces de manejar escenarios complejos. El objetivo principal de esta investigación es estimar la matriz de varianza-covarianzas del estimador máximo verosimil en un modelo autorregresivo espacio-temporal con datos faltantes. Los errores estándar se obtuvieron invirtiendo la Matriz de Información de Fisher (FIM) observada, calculada utilizando el método de Louis con aproximaciones Monte Carlo para evaluar esperanzas incalculables. Continuando el trabajo de Padilla et al. (2020), quienes utilizaron el filtro de Kalman y un algoritmo de Esperanza-Maximización Generalizado para estimar y predecir bajo datos incompletos, evaluamos la precisión de los estimadores de los parámetros y comparamos los resultados obtenidos mediante el método de Louis con los errores estándar derivados del bootstrap espacio-temporal. Nuestros resultados destacan la precisión, estabilidad y ventajas computacionales del enfoque propuesto, especialmente bajo niveles moderados de datos faltantes. Esta contribución metodológica permite comprender las fortalezas y limitaciones de ambos enfoques, y ofrece una base sólida para futuras investigaciones y aplicaciones en modelos espacio-temporales.

Abstract

Space-time data have become increasingly important in diverse fields, including epidemiology, social sciences, environmental studies, meteorology, and agriculture. Accurate analysis of these types of databases is crucial for reliable modeling and inference, particularly when dealing with missing data, and therefore requires efficient algorithms that can handle complex scenarios. The main objective of this research is to estimate the variance-covariance matrix for the Maximum Likelihood Estimator in a Space-Time Autoregressive model in the presence of missing data. Standard errors were obtained by inverting the observed Fisher Information Matrix (FIM), which is computed via Louis' method using Monte Carlo approximations to evaluate intractable expectations. Building on the research of Padilla et al. (2020), who employed the Kalman filter and the Generalized Expectation-Maximization algorithm for accurate parameter estimation and prediction under incomplete data, we asses the precision of parameter estimates and compare the results obtained via Louis' method with those using space-time bootstrap standard errors. Our findings highlight the accuracy, stability, and computational advantages of the proposed approach, particularly in the presence of moderate levels of missing data. This methodological contribution provides insight into the respective strengths and limitations of these two approaches, offering a foundation for future research and applications in space-time models.

Keywords: standard errors, Louis' method, EM algorithm, Fisher InformationMatrix, Parametric Bootstrap

Chapter 1

Introduction

Space-time data are becoming increasingly important in different disciplines, including epidemiology, social sciences, environmental studies, meteorology, and agriculture, among others. Advancements in technology have enabled the acquisition of increasingly detailed spatial data over time. Analyzing these large and complex space-time databases demands high-performance computing resources and the development of more specific, accurate, and sophisticated statistical models (Wang et al., 2010; Katzfuss and Cressie, 2011).

A common challenge in space-time analysis is dealing with missing or incomplete data, which can significantly impact forecasting accuracy. Additionally, incorporating adequate correlations between spatially and temporally proximate observations adds another layer of complexity to the analysis. Numerous approaches, including space-time covariance functions and dynamic models, have been employed to address these challenges. The statistical correlation between observations can be effectively described through space-time covariance functions (Gneiting et al., 2006), while dynamical models offer a probabilistic mechanism for understanding the temporal evolution of observations and for appropriately managing diverse configurations of incomplete data (Katzfuss and Cressie, 2011; Lagos-

Álvarez et al., 2019; Padilla et al., 2020; Douc et al., 2014). In this context, Space-Time Autoregressive (STAR) models have proved especially useful for predicting environmental and climate variables, such as temperature, air pollution, ozone concentration, and precipitation (Cameletti et al., 2013; Lagos-Álvarez et al., 2019; Padilla et al., 2020; Sahu and Bakar, 2012b,a; Sigrist et al., 2011).

Previous research by Padilla et al. (2020) demonstrated the effectiveness of their stationary Gaussian STAR model, which utilizes the Kalman filter (KF) and the Generalized Expectation-Maximization (GEM) algorithm to provide accurate estimations and reliable predictions in the presence of missing data. The KF, developed by Rudolph Kalman in 1960, is a recursive algorithm that, in conjunction with kriging, computes optimal space-time predictions by integrating past and current data. The Expectation-Maximization (EM) algorithm, developed by Dempster et al. in 1977, is a stable iterative method for Maximum Likelihood Estimation (MLE) with incomplete data that ensures convergence to a local stationary point, guaranteeing the stability and accuracy of the estimates generated by the model. The versatility of the EM algorithm to handle different scenarios of incomplete data, such as missing values, grouped, censored, or truncated data, random effects, hidden Markov models, and finite mixture models, makes it a powerful tool in statistics (Gupta and Chen, 2011; Meilijson, 1989; McLachlan and Krishnan, 2008). The GEM algorithm, an extension of the EM algorithm, is particularly useful when closed-form solutions that globally maximize the auxiliary $\mathcal{Q}$-function (in the M-step) are unavailable. In such instances, the parameters vector is partitioned into subvectors, some of which are solved using the Newton-Raphson algorithm, facilitating a more manageable estimation process.

Building on the model developed by Padilla et al. (2020), the main objective of this research is to estimate the variance-covariance matrix for the MLE within a STAR Gaussian model in the presence of missing data. This objective will be accomplished by calculating the inverse of the observed Fisher Information Matrix (FIM) for parameter inference.

Under regularity conditions, MLE converges asymptotically to a normal distribution, with its variance-covariance matrix given by the inverse of the FIM. The FIM quantifies the information contained in the data concerning unknown parameters, aiding in the analysis of estimation error through confidence intervals and hypothesis tests. A common limitation of the EM algorithm is its inability to estimate the variance-covariance matrix of the MLE automatically. Various methods have been developed to estimate the FIM in the presence of missing data; however, these methods are often model-specific (Baker, 1992; Delyon et al., 1999; Gong et al., 2021; Jamshidian and Jennrich, 2000; Louis, 1982; Meng and Rubin, 1991; Oakes, 2002; Surya, 2024; Walsh, 2006; Wei and Tanner, 1990). In the case of nonindependent data, the exact computation of the FIM can be complex, analytically and computationally, and certain approaches may be susceptible to the stability of the empirical Hessian matrix and cannot guarantee that it will remain a positive definite matrix (Duan and Fulop, 2011).

Louis' method (Louis, 1982) is the most popular and stable theoretical approach for calculating the observed FIM using the EM algorithm. By applying the "Missing Information Principle" proposed by Orchard and Woodbury (University, 1972), Louis demonstrated that the observed FIM can be expressed in terms of the complete data log-likelihood function. However, calculating the FIM analytically can be quite challenging, especially in dynamic models, due to the need to address the conditional variance of the score function and

the conditional expectation of the Hessian matrix. Although the conditional expectation of the Hessian matrix coincides with the second derivative of the auxiliary $\mathcal{Q}$-function, the main difficulty lies in computing the conditional variance of the complete-data log-likelihood function score, which involves complex cross-products of the state process. Numerous researchers have highlighted these challenges (Baker, 1992; Holmes, 2014; Lystig and Hughes, 2002; Turner et al., 1998) and proposed different solutions

Numerical approximations through Monte Carlo (MC) techniques provide an effective solution for the intricate analytical expressions in our model. To compare the results produced by Louis' method, we will implement a Space-Time Parametric Bootstrap (STPB) (Fassò and Cameletti, 2009), to estimate the standard error of parameter estimates.

The performance of Louis' method will be assessed through a simulation study and with real data on average daily air temperature from the National Agroclimatic Network of Chile, specifically covering the regions of Maule, Ñuble, Biobío, and La Araucanía from $2015 - 2022$. The results generated provide insights into the respective strengths and limitations of these two approaches.

Chapter 2

Theoretical Framework

2.1 Space-time processes and modeling definition

A Gaussian space-time random field (GSTRF) $\mathcal{Z}_{\mathcal{D}\times\mathcal{T}}$ indexed in space by $\mathbf{s} \in \mathbb{R}^d$ and in time by $t \in \mathbb{Z}^+$ is defined to

$$\mathcal{Z}_{\mathcal{D}\times\mathcal{T}} := \{Z_t(\mathbf{s}) : \mathbf{s} \in \mathcal{D}, t \in \mathcal{T}\}$$

where $Z_t(\mathbf{s})$ is an interesting random variable, and the realization of this stochastic process is

$$\mathbf{Z}_{n\times T} = (Z_{t_1}(\mathbf{s}_1), \ldots, Z_{t_1}(\mathbf{s}_n), Z_{t_2}(\mathbf{s}_1), \ldots, Z_{t_2}(\mathbf{s}_n), \ldots, Z_{t_T}(\mathbf{s}_1), \ldots, Z_{t_T}(\mathbf{s}_n))^\top$$

at known locations $\mathbf{s}_i$ and time t_k with $i = 1, \ldots, n$ and $k = 1, 2, \ldots, T$, where n is the number of spatial locations or measurement stations and T is the total number of potential temporal observations in location $\mathbf{s}_i$. For notational convenience, in some cases, we will use $Z(\mathbf{s}, t)$, but always considering that it is a space-time process with a discrete-time

index.

Our modelling strategies for the interesting random variable were as follows (for more detail, see Lagos-Álvarez et al. (2019); Padilla et al. (2020)):

$$Z_t(\mathbf{s}) = \mu_t(\mathbf{s}) + \varepsilon_t(\mathbf{s}) + \omega_t(\mathbf{s}), \tag{2.1a}$$

$$\varepsilon_t(\mathbf{s}) = \phi_1\varepsilon_{t-1}(\mathbf{s}) + \phi_2\varepsilon_{t-2}(\mathbf{s}) + \ldots + \phi_p\varepsilon_{t-p}(\mathbf{s}) + \eta_t(\mathbf{s}) \tag{2.1b}$$

where $\mu_t(\mathbf{s})$ is a systematic component that explains most of the variation of $Z_t(\mathbf{s})$, $\boldsymbol{\omega}_{\mathcal{D}\times\mathcal{T}} := \{\omega_t(\mathbf{s}) : \mathbf{s} \in \mathcal{D},\ t \in \mathcal{T}\}$ is a Gaussian White Noise with mean zero and variance σ^2_ω, known as nugget effect $\left(\boldsymbol{\omega}_{\mathcal{D}\times\mathcal{T}} \sim \mathcal{GWN}(0, \sigma^2_\omega)\right)$. $\varepsilon_t(\mathbf{s})$ is the unobserved state processes, we assume that $\boldsymbol{\varepsilon}_{\mathcal{D}\times\mathcal{T}} := \{\varepsilon_t(\mathbf{s}) : \mathbf{s} \in \mathcal{D},\ t \in \mathcal{T}\}$ is a stationary Gaussian space-time autoregressive process of p-order with mean zero where the coefficients $\phi_1, \ldots, \phi_p$ are chosen such that the absolute values of all roots, possibly complex, of $\lambda^p - \sum_{m=1}^{p} \phi_m \lambda^{p-m} = 0$ are less than 1, to achieve temporal stationarity; this is fully detailed in Huang and Cressie (1996). Finally $\boldsymbol{\eta}_{\mathcal{D}\times\mathcal{T}} := \{\eta_t(\mathbf{s}) : \mathbf{s} \in \mathcal{D},\ t \in \mathcal{T}\}$ is a stationary Gaussian space-time process, with mean zero and space-time covariance function, independent of $\boldsymbol{\omega}_{\mathcal{D}\times\mathcal{T}}$.

The covariance function of $\mathcal{Z}_{\mathcal{D}\times\mathcal{T}}$ is defined as:

$$\begin{aligned} \mathcal{C}^Z(\mathbf{s}_i, \mathbf{s}_j; t_k, t_l) &:= \text{Cov}\left[Z(\mathbf{s}_i, t_k), Z(\mathbf{s}_j, t_l)\right] \\ &= E\big[\left(Z(\mathbf{s}_i, t_k) - \mu(\mathbf{s}_i, t_k)\right)\left(Z(\mathbf{s}_j, t_l) - \mu(\mathbf{s}_j, t_l)\right)\big] \end{aligned}$$

where $\mathbf{s}_i, \mathbf{s}_j \in \mathcal{D}$ and $t_k, t_l \in \mathcal{T}$.

Optimal least-squares prediction, or kriging, relies on the appropriate specification of the

space-time covariance structure. In practice, estimation and modeling call for simplifying assumptions, such as stationarity, separability, and full symmetry (Gneiting et al., 2006):

- A STRF $\mathcal{Z}_{\mathcal{D}\times\mathcal{T}}$ is strictly stationary if its probability distribution is translation invariant. Specifically, for any two given vectors, $\mathbf{s}_i - \mathbf{s}_j = \mathbf{h} \in \mathbb{R}$ for spatial separation and $t_k - t_l = \Delta t \in \mathbb{N}$ for temporal lag, the following hold:

$$Z(\mathbf{s}_1, t_1), \ldots, Z(\mathbf{s}_i, t_k), \ldots, Z(\mathbf{s}_n, t_T)$$

and

$$Z(\mathbf{s}_1 + \mathbf{h}, t_1 + \Delta t), \ldots, Z(\mathbf{s}_i + \mathbf{h}, t_k + \Delta t), \ldots, Z(\mathbf{s}_n + \mathbf{h}, t_T + \Delta t)$$

have the same multivariate distribution function for all $\mathbf{s}_1, \ldots, \mathbf{s}_n, t_1, \ldots, t_T, \mathbf{h}, \Delta t$. In practice, it is impossible to determine this assumption, and we can only estimate the first few moments of the distribution.

- A STRF $\mathcal{Z}_{\mathcal{D}\times\mathcal{T}}$ is second-order stationary if:

$$E\left[Z(\mathbf{s}, t)\right] := \mu(\mathbf{s}, t) := \mu, \quad \forall (\mathbf{s}, t) \in \mathbb{R}^d \times \mathbb{Z}^+$$

$$\text{Cov}\left[Z(\mathbf{s}_i, t_k), Z(\mathbf{s}_j, t_l)\right] := \mathcal{C}^Z(||\mathbf{s}_i - \mathbf{s}_j||, |t_k - t_l|) := \mathcal{C}^Z(\mathbf{h}, \Delta t) \quad \forall \mathbf{s}_i, \mathbf{s}_j, t_k, t_l$$

that is, if it has a constant mean and the covariance function depends on the observation sites through $\mathbf{h}$ and on the observation times through Δt. Where $||\cdot||$ indicates the Euclidean distance.

If a space-time process has both, spatially and temporally stationary covariance, we say that the process has stationary covariance. For stationarity tests, see the review mentioned in Fuentes (2005).

- A STRF $\mathcal{Z}_{\mathcal{D}\times\mathcal{T}}$ is said to have separable covariance if there exist purely spatial and purely temporal covariance functions, such that

$$\begin{aligned}\text{Cov}\left[Z(\mathbf{s}_i,t_k),Z(\mathbf{s}_j,t_l)\right] &= \text{Cov}\left[Z(\mathbf{s}_i),Z(\mathbf{s}_j)\right]\cdot\text{Cov}\left[Z(t_k),Z(t_l)\right]\\ &= \mathcal{C}^S(\mathbf{s}_i,\mathbf{s}_j)\cdot\mathcal{C}^T(t_k,t_l)\end{aligned}$$

where $\mathcal{C}^S(\cdot,\cdot)$ is the purely spatial covariance function and $\mathcal{C}^T(\cdot,\cdot)$ is the purely temporal covariance function. This structure corresponds to a product model of a separable covariance function. Mitchell et al. (2005) characterize this model using the Kronecker matrix product. Let C^S and C^T be the matrices of variances and covariances coming from the covariance function with dimensions $n\times n$ and $T\times T$ for space and time, respectively, and let C^Z be the matrix of variances and covariances of dimension $nT\times nT$ of the joint-process. The covariance is said to be separable if, and only if, $C^Z = C^S\otimes C^T$, where $\otimes$ represents the Kronecker product of both matrices. Several statistical tests for separability have been proposed based on parametric models (Shitan and Brockwell, 1995; Mitchell et al., 2005, 2006; Fuentes, 2006).

A stationary space-time covariance function is separable if stationary exists, purely spatial, and purely temporal covariance functions $\mathcal{C}^S(\mathbf{h})$ and $\mathcal{C}^T(\Delta t)$, respectively.

We can factor the space-time covariance function as:

$$\mathcal{C}^Z(\mathbf{h}, \Delta t) = \frac{\mathcal{C}(\mathbf{h}, 0) \cdot \mathcal{C}(\mathbf{0}, \Delta t)}{\mathcal{C}(\mathbf{0}, 0)}$$

for all $(\mathbf{h}, \Delta t) \in \mathbb{R}^d \times \mathbb{R}$ (Mitchell et al. (2006)).

- For a STRF $\mathcal{Z}_{\mathcal{D}\times\mathcal{T}}$ to have fully symmetric covariance, the following must hold

$$\mathcal{C}^Z((\mathbf{s}_i, \mathbf{s}_j; t_k, t_l) = \mathcal{C}^Z((\mathbf{s}_i, \mathbf{s}_j; t_l, t_k)$$
$$\text{Cov}\left[Z(\mathbf{s}_i, t_k), Z(\mathbf{s}_j, t_l)\right] = \text{Cov}\left[Z(\mathbf{s}_i, t_l), Z(\mathbf{s}_j, t_k)\right]$$

A stationary space-time covariance function is fully symmetric if

$$\mathcal{C}^Z(\mathbf{h}, \Delta t) = \mathcal{C}^Z(\mathbf{h}, -\Delta t) = \mathcal{C}^Z(-\mathbf{h}, \Delta t) = \mathcal{C}^Z(-\mathbf{h}, -\Delta t)$$

for all $(\mathbf{h}, \Delta t) \in \mathbb{R}^d \times \mathbb{R}$.

Tests for full symmetry can be used to reject separability (Lu and Zimmerman, 2005), then covariance structures that are not fully symmetric are non-separable (Gneiting, 2002).

Gneiting et al. (2006) proposed a summary of the relationships between the various notions in terms of classes of space-time covariance functions (Figure 2.1). The largest class is that of general, stationary, or non-stationary covariance functions. A separable covariance can be stationary or non-stationary and, similarly, for fully symmetric covariances. However, a separable covariance function is always fully symmetric, but not vice versa,

which has implications in testing and model fitting. In particular, to reject separability, it suffices to reject full symmetry.

Cressie and Wikle (2015) established that a necessary and sufficient condition for a space-

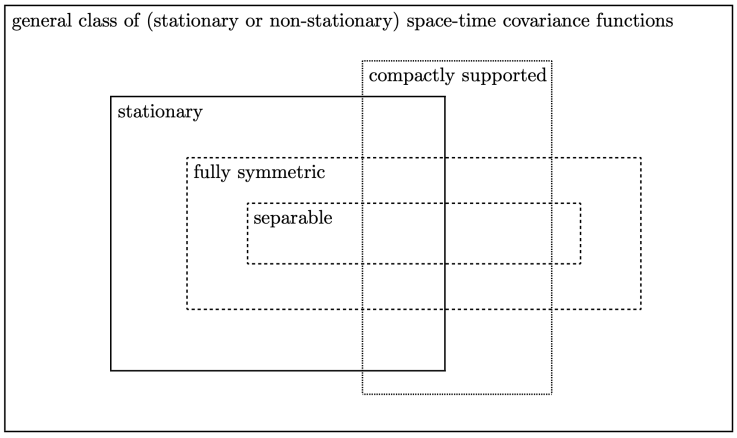

Figure 2.1: Schematic illustration of the relationships between separable, fully symmetric, stationary, and compactly supported covariances within the general class of (stationary or non-stationary) space-time covariance functions.

time covariance function on $\in \mathbb{R}^d \times \mathbb{R}$ is to be symmetric, and positive-definite, that is,

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j \mathcal{C}^Z((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) \geq 0$$

which, in the case of a stationary space-time covariance function

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j \mathcal{C}^Z((\mathbf{s}_i - \mathbf{s}_j, t_i - t_j) \geq 0$$

for any m, any $\{a_i\}$, and any $(\mathbf{s}_i, t_i)$.

It can be quite difficult to check whether a function is positive definite, which forms one of the key difficulties in constructing parametric space-time covariance models. A convenient

choice is a standard isotropic model that guarantees positive definiteness.

A stationary STRF $\mathcal{Z}_{\mathcal{D}\times\mathcal{T}}$ has spatially isotropic covariance function if

$$\text{Cov}\left[Z(\mathbf{s}_i, t_k), Z(\mathbf{s}_j, t_l)\right] = \mathcal{C}^Z(||\mathbf{h}||; t_k, t_l), \quad \forall(\mathbf{s}, t) \in \mathbb{R}^d \times \mathbb{R}$$

This means that the covariate function only depends on $||\mathbf{h}||$. The isotropy property can be thought of as an invariance property under rotations.

Different authors propose standard admissible models for spatial covariance function. Here are some of those proposed by Matérn (1986):

$$\mathcal{C}^S(\mathbf{s}_i, \mathbf{s}_j) = \sigma_S^2\, \rho^S(\mathbf{s}_i, \mathbf{s}_j) = \frac{\sigma_S^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{||\mathbf{s}_i - \mathbf{s}_j||}{\alpha}\right)^{\nu} \mathcal{K}_\nu \left(\frac{||\mathbf{s}_i - \mathbf{s}_j||}{\alpha}\right) \tag{2.2}$$

where $\sigma_S^2 > 0, \alpha > 0, \nu > 0$, $\mathcal{K}$ is the modified Bessel function of the second kind of order ν and $\Gamma(\cdot)$ is the gamma function. The range parameter α controls the decay rate with distance, with larger values of α corresponding to more highly correlated observations.

A popular special case of the Matérn family is the exponential model

$$\mathcal{C}^S(\mathbf{s}_i, \mathbf{s}_j) = \sigma_S^2\, \rho^S(\mathbf{s}_i, \mathbf{s}_j) = \sigma_S^2\, \exp\left(-\frac{||\mathbf{s}_i - \mathbf{s}_j||}{\alpha}\right) \tag{2.3}$$

which is obtained when $\nu = 1/2$. Also, when $v \to \infty$ is obtained the gaussian model

$$\mathcal{C}^S(\mathbf{s}_i, \mathbf{s}_j) = \sigma_S^2\, \rho^S(\mathbf{s}_i, \mathbf{s}_j) = \sigma_S^2\, \exp\left(-\left(\frac{||\mathbf{s}_i - \mathbf{s}_j||}{\alpha}\right)^2\right) \tag{2.4}$$

In our model (2.1a), we assume temporal stationarity, which leads to a space-time covari-

ance function becomes:

$$\begin{aligned}\mathcal{C}^{Z}(\mathbf{s}_i, \mathbf{s}_j, t_k, t_l) &:= \mathrm{Cov}\left[Z_{t_k}(\mathbf{s}_i), Z_{t_l}(\mathbf{s}_j)\right] \\ &= \mathrm{Cov}\left[\varepsilon_{t_k}(\mathbf{s}_i), \varepsilon_{t_l}(\mathbf{s}_j)\right] + \mathrm{Cov}\left[\omega_{t_k}(\mathbf{s}_i), \omega_{t_l}(\mathbf{s}_j)\right] \\ &= \mathcal{C}^{\varepsilon}_{\kappa}(\mathbf{s}_i, \mathbf{s}_j) + \sigma^2_{\omega}\delta_{\{\kappa=0;\mathbf{s}_i=\mathbf{s}_j\}}\end{aligned}$$

where $\kappa = |t_k - t_l| = 0, 1, 2, \ldots, p$ and δ is the indicator function. The covariance function $\mathcal{C}^{\varepsilon}_{\kappa}(\mathbf{s}_i, \mathbf{s}_j)$ comes from the multiplication of equation (2.1b) by $\varepsilon_{t-m}(\mathbf{s})$ and taking expectations:

$$\mathcal{C}^{\varepsilon}_{\kappa}(\mathbf{s}_i, \mathbf{s}_j) = \sum_{m=1}^{p} \phi_m \mathcal{C}^{\varepsilon}_{\kappa-m}(\mathbf{s}_i, \mathbf{s}_j) + \mathcal{C}^{\eta}(\mathbf{s}_i, \mathbf{s}_j)\delta_{\{\kappa;0\}} \tag{2.5}$$

when $\delta_{\{\kappa,0\}}$ is 1 if $\kappa = 0$ and 0 in other case. $\mathcal{C}^{\varepsilon}_{\kappa}(\mathbf{s}_i, \mathbf{s}_j)$ can be determined by solving the $p + 1$ Yule-Walker equations. For more details, see Huang and Cressie (1996); Padilla (2018); Lagos-Álvarez et al. (2019). The covariance function, $\mathcal{C}^{\eta}(\mathbf{s}_i, \mathbf{s}_j)$, is temporarlly uncorrelated, then

$$\mathcal{C}^{\eta}(\mathbf{s}_i, \mathbf{s}_j) = \mathrm{Cov}\left[\eta_{t_k}(\mathbf{s}_i), \eta_{t_l}(\mathbf{s}_j)\right] = \mathcal{C}^{S}(\mathbf{s}_i, \mathbf{s}_j) \quad \text{when} \quad t_k = t_l$$

In case we assume that $\varepsilon_t(\mathbf{s})$ is a space-time autoregressive stationary process of order $p = 1$, AR(1), the space-time covariance function can be written as:

$$\begin{aligned}\mathcal{C}^{\varepsilon}_{\kappa}(\mathbf{s}_i, \mathbf{s}_j) &= \frac{\phi^{\kappa}}{(1-\phi^2)}\, \mathcal{C}^{\eta}(\mathbf{s}_i, \mathbf{s}_j) \\ &= \frac{\phi^{\kappa}}{(1-\phi^2)}\, \sigma^2_{\eta}\, \rho^{\eta}(\mathbf{s}_i, \mathbf{s}_j), \qquad \rho^{T}(\kappa) = \phi^{\kappa} \\ &= \frac{\sigma^2_{\eta}}{(1-\phi^2)}\, \rho^{T}(\kappa)\, \rho^{\eta}(\mathbf{s}_i, \mathbf{s}_j), \qquad \sigma^2_{\varepsilon} = \frac{\sigma^2_{\eta}}{(1-\phi^2)}\end{aligned} \tag{2.6}$$

$$\mathcal{C}^{\varepsilon}_{\kappa}(\mathbf{s}_i, \mathbf{s}_j) = \sigma^2_{\varepsilon}\, \rho^{T}(\kappa)\, \rho^{\eta}(\mathbf{h})$$

where $\rho^T(\kappa)$ is the temporal correlation function, $\rho^\eta(\mathbf{h})$ is the spatial correlation function, then $\mathcal{C}_\kappa^\varepsilon(\mathbf{s}_i, \mathbf{s}_j)$ corresponds to a separable space-time covariance function (Gneiting et al., 2006). For more space-time covariance function review Martinez (2008), Chen et al. (2021) and Porcu et al. (2021).

2.2 State-space representation, Kalman Filter and Kalman Smoother

The use of Kalman recursions requires writing the model in a state-space form. The state-space model of $\mathcal{Z}_{\mathcal{D}\times\mathcal{T}}$ is derived from a linear dynamic model described by the observation equation (2.7a), and the latent state equation (2.7b). One of the primary advantages of state-space models is their versatility in handling several missing data configurations (Shumway and Stoffer, 1982). For further details on state-space models, refer to Shumway and Stoffer (2017) and Hamilton (1994). For a dynamical space-time model, check Cressie and Wikle (2015). Additionally, details about our specific model can be found in Padilla (2018); Padilla et al. (2020). A summary is provided below:

$$Z_t(\mathbf{s}) = \mathbf{X}_t(\mathbf{s})\boldsymbol{\beta} + \boldsymbol{\mathcal{H}}^\top \boldsymbol{\xi}_t(\mathbf{s}) + \omega_t(\mathbf{s}) \tag{2.7a}$$

$$\boldsymbol{\xi}_t(\mathbf{s}) = \mathbf{F}\boldsymbol{\xi}_{t-1}(\mathbf{s}) + \mathbf{V}_t(\mathbf{s}) \tag{2.7b}$$

where:

- $Z_t(\mathbf{s})$ represents the sampled observation at time t at location $\mathbf{s}$,
- $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_\ell)^\top$ is an ℓ-dimensional vector of parameters associated to $\mathbf{X}_t(\mathbf{s}) = \left(X_t^{(1)}(\mathbf{s}), \ldots, X_t^{(\ell)}(\mathbf{s})\right)$, an ℓ-dimensional vector of non-stochastic regressors,

- $\boldsymbol{\mathcal{H}}$ is a p-dimensional vector known as the observation operator,
- $\boldsymbol{\xi}_t(\mathbf{s})$ is a p-dimensional state vector, $\boldsymbol{\xi}_t(\mathbf{s}) := \big(\boldsymbol{\varepsilon}_t(\mathbf{s}), \ldots, \boldsymbol{\varepsilon}_{t-p+1}(\mathbf{s})\big)^\top$
- $\omega_t(\mathbf{s})$ represents the nugget effect defined in (2.1a),
- $\mathbf{F}$ is a $p \times p$-dimensional matrix called state transition operator,
- $\mathbf{V}_t(\mathbf{s})$ is a p-dimensional vector for the spatial innovation.

With n spatial observations at time t, the state vector becomes (pointing his dimensions) $\boldsymbol{\xi}_{t_{np\times 1}} = \big(\boldsymbol{\xi}_t(\mathbf{s}_1)^\top, \ldots, \boldsymbol{\xi}_t(\mathbf{s}_n)^\top\big)^\top$, and the state equation (2.7b) is transformed to:

$$\boldsymbol{\xi}_{t_{np\times 1}} = \boldsymbol{\Phi}_{np\times np} \cdot \boldsymbol{\xi}_{t-1_{np\times 1}} + \mathbf{V}_{t_{np\times 1}} \tag{2.8}$$

where $\boldsymbol{\Phi}_{np\times np} = \mathbf{I}_{n\times n} \otimes \mathbf{F}_{p\times p}$ is the extended state transition operator and $\mathbf{V}_{t_{np\times 1}} = (\mathbf{V}_t(\mathbf{s}_1)^\top, \ldots, \mathbf{V}_t(\mathbf{s}_n)^\top)^\top$ is the spatial innovation vector incorporating the terms of innovation in each location.

To properly incorporate the identification of missing data into the model, a missing data identifier observation operator $\mathbf{L}^{obs}$ is used alongside the observation operator $\mathcal{H}$. Consequently, the observation equation (2.7a) for all n spatial observations is:

$$\mathbf{Z}_{t_{n_t\times 1}} = \mathbf{X}_{t_{n_t\times \ell}} \cdot \boldsymbol{\beta}_{\ell\times 1} + \boldsymbol{\Lambda}_{t_{n_t\times np}} \cdot \boldsymbol{\xi}_{t_{np\times 1}} + \mathbf{W}_{t_{n_t\times 1}} \tag{2.9}$$

where $\boldsymbol{\Lambda}_{t_{n_t\times np}} = \mathbf{L}^{obs}_{t_{n_t\times n}} \otimes \mathcal{H}^\top_{1\times p}$ is the extended observation operator that allows mapping correctly the information at time t to the state process. $\mathbf{L}^{obs}_t = \big(\boldsymbol{e}_{t_1}, \boldsymbol{e}_{t_2}, \ldots, \boldsymbol{e}_{t_{n_t}}\big)^\top$ with $\{\boldsymbol{e}_{i_j}, i_j \in \mathbf{I}_t\}$ is a set of canonical vectors of dimension $(1 \times n)$ where $\mathbf{I}_t = \{i_j, j = 1, \ldots, n_t\}$ was the index set associated to the locations $\mathbf{s}_{i_j}$, $\forall j = 1, \ldots, n_t$ where was a

register at the time t. With $\mathbf{W}_t$ being the error vector for n_t locations.

In our model, we assumed that $\boldsymbol{\varepsilon}_{\mathcal{D}\times\mathcal{T}}$ is uncorrelated with any realization of $\mathbf{V}_t$ and $\mathbf{W}_t$, which are independent as defined in (2.1b).

For a complete matrix representation of the state-space model (2.8, 2.9), see the appendix section VII.1.

We model the system using Gaussian distributions, ensuring that the joint, marginal, and conditional distributions remain Gaussian. The state process follows a first-order Hidden Markov model, with unobserved states, $\boldsymbol{\xi}_{1:T}$, which are conditionally independent given the present, characterized by the initial distribution $f(\boldsymbol{\xi}_1)$ and the transition kernel $f(\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1})$, yielding the following distribution function for the state vector:

$$f(\boldsymbol{\xi}_{1:T}) = f(\boldsymbol{\xi}_1)\prod_{t=2}^{T} f(\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1})$$

The observations $\mathbf{Z}_{1:T}$ are considered conditionally independent given $\boldsymbol{\xi}_{1:T}$. Consequently, the conditional density is expressed as follows

$$f(\mathbf{Z}_{1:T}|\boldsymbol{\xi}_{1:T}) = \prod_{t=1}^{T} f(\mathbf{Z}_t|\boldsymbol{\xi}_t)$$

Furthermore, it is possible to derive the posterior distribution function to predict $\boldsymbol{\xi}_{1:T}$ given the observations $\mathbf{Z}_{1:T}$

$$f(\boldsymbol{\xi}_{1:T}|\mathbf{Z}_{1:T}) \propto f(\mathbf{Z}_1|\boldsymbol{\xi}_1)f(\boldsymbol{\xi}_1)\prod_{t=2}^{T} f(\mathbf{Z}_t|\boldsymbol{\xi}_t)f(\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1})$$

The assumptions outlined facilitate the derivation of the filtering distribution $f(\boldsymbol{\xi}_t|\mathbf{Z}_{1:t})$ and the smoothing distribution $f(\boldsymbol{\xi}_t|\mathbf{Z}_{1:T})$ through the Kalman recursions. The KF provides the equations for filtering and forecasting. The filtered estimate relies on present and past observations, while the forecast relies only on past data. This indicates that the KF defines how to update the filter with new observations from time $t-1$ to t without reprocessing the entire dataset of observed data (Shumway and Stoffer, 1982). Conversely, the Kalman Smoother (KS) utilizes estimates from the KF to produce more accurate smooth estimates, as it incorporates all available observations up until time T.

Distributions used in the KF and KS:

- Initial values

$$\boldsymbol{\xi}_1 \sim Gau(\boldsymbol{\xi}_{1|0}, \mathbf{P}_{1|0}) \tag{2.10}$$

$$\boldsymbol{\xi}_{1|0} := \mathrm{E}[\boldsymbol{\xi}_1] = \mathbf{0}$$
$$\mathbf{P}_{1|0} := \mathrm{Var}[\boldsymbol{\xi}_1] = \mathcal{C}_\kappa^\varepsilon(\mathbf{s}_i, \mathbf{s}_j)$$

  where $\mathcal{C}_\kappa^\varepsilon(\mathbf{s}_i, \mathbf{s}_j)$ is from (2.5)

- Measurement Distribution:

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

- State Evolution Distribution:

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

- Prediction Distribution:

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

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

- Filtering Distribution:

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

$$\begin{aligned}
\boldsymbol{\xi}_{t|t} &= \boldsymbol{\xi}_{t|t-1} + \mathbf{G}_{t|t-1}\boldsymbol{\Delta}_{t|t-1}^{-1}(\mathbf{Z}_t - \mathbf{Z}_{t|t-1}) \\
\mathbf{G}_{t|t-1} &:= \mathrm{Cov}[\boldsymbol{\xi}_t, \mathbf{Z}_t|\mathbf{Z}_{1:t-1}] = \mathbf{P}_{t|t-1}\boldsymbol{\Lambda}_t^\top \\
\mathbf{Z}_{t|t-1} &= \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\Lambda}_t\boldsymbol{\xi}_{t|t-1} \\
\boldsymbol{\Delta}_{t|t-1} &= \boldsymbol{\Lambda}_t\mathbf{P}_{t|t-1}\boldsymbol{\Lambda}_t^\top + \sigma_\omega^2\mathbf{I}_{n_t} \\
\mathbf{P}_{t|t} &= \mathbf{P}_{t|t-1} - \mathbf{G}_{t|t-1}\boldsymbol{\Delta}_{t|t-1}^{-1}\mathbf{G}_{t|t-1}^\top
\end{aligned}$$

- Smoothing Distribution:

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

$$\boldsymbol{\xi}_{t|T} := \mathrm{E}[\boldsymbol{\xi}_t|\mathbf{Z}_{1:T}] = \boldsymbol{\xi}_{t|t} + \mathbf{J}_t(\boldsymbol{\xi}_{t+1|T} - \boldsymbol{\xi}_{t+1|t})$$

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

$$\mathbf{P}_{t|T} := \mathrm{Var}[\boldsymbol{\xi}_t|\mathbf{Z}_{1:T}] = \mathbf{P}_{t|t} + \mathbf{J}_t(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t})\mathbf{J}_t^\top$$

The term $SGau$ refers to the Singular Multivariate Gaussian distribution, defined as:

$$f(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-k/2}|\mathbf{D}_{\boldsymbol{\Sigma}}|^{-1/2} \exp\left\{ -\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^-(\mathbf{X} - \boldsymbol{\mu}) \right\}$$

the Moore-Penrose generalized inverse and its determinant are defined by the eigendecomposition of $\boldsymbol{\Sigma}_{q\times q}$ and the reduction of its dimension using the rank k of the singular matrix $\boldsymbol{\Sigma}$ (Holbrook, 2018; Mardia et al., 1979), which in our case corresponds to the covariance matrix $\mathbf{Q}$:

$$\mathbf{Q} = \boldsymbol{E}\boldsymbol{D}_{\mathbf{Q}}\boldsymbol{E}^\top, \quad \text{con } \mathbf{D}_{\mathbf{Q}} = diag\{d_1, d_2, \ldots, d_k\}$$

and $\mathbf{E}$ is a full-rank matrix of dimensions $q \times k$, whose columns are the eigenvectors associated with the k positive eigenvalues.

In our model, the covariance matrix $\mathbf{Q}$ is defined as:

$$\mathbf{Q} = \sigma_\eta^2 \cdot \mathbf{R} = \sigma_\eta^2 \left(\boldsymbol{\rho}^\eta(\alpha) \otimes \mathbf{T}^{AR}\right) = \sigma_\eta^2 \left(\exp\left(-\frac{\mathbf{H}}{\alpha}\right) \otimes \mathbf{T}^{AR}\right) = \mathcal{C}^\eta(\alpha) \otimes \mathbf{T}^{AR}$$

where $\mathbf{H}$ represents the Euclidean distance matrix, calculated from the spatial separation vector $\mathbf{h}_{ij} = ||\mathbf{s}_i - \mathbf{s}_j||$ for all locations, observe

$$\boldsymbol{\rho}^{\eta}(\alpha) = \begin{pmatrix} \rho^{\eta}(\mathbf{s}_1, \mathbf{s}_1) & \rho^{\eta}(\mathbf{s}_1, \mathbf{s}_2) & \cdots & \rho^{\eta}(\mathbf{s}_1, \mathbf{s}_n) \\ \rho^{\eta}(\mathbf{s}_2, \mathbf{s}_1) & \rho^{\eta}(\mathbf{s}_2, \mathbf{s}_2) & \cdots & \rho^{\eta}(\mathbf{s}_2, \mathbf{s}_n) \\ \vdots & \vdots & \cdots & \vdots \\ \rho^{\eta}(\mathbf{s}_n, \mathbf{s}_1) & \rho^{\eta}(\mathbf{s}_n, \mathbf{s}_2) & \cdots & \rho^{\eta}(\mathbf{s}_n, \mathbf{s}_n) \end{pmatrix}_{n \times n} \qquad \mathbf{T}^{AR} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix}_{p \times p}$$

$$\rho^{\eta}(\mathbf{s}_i, \mathbf{s}_j) = \exp\left(-\frac{\mathbf{h}_{ij}}{\alpha}\right)$$

2.3 Maximum Likelihood Estimation via the Generalized Expectation-Maximization algorithm

We defined the complete data, $\mathbf{Z}_{1:T}^{(c)}$, as the union of the observed data, $\mathbf{Z}_{1:T}^{(o)}$, and unobserved or missing data, $\mathbf{Z}_{1:T}^{(m)}$. In our model, the missing data corresponds to the latent state process, $\boldsymbol{\xi}_{1:T}$. The following notation is adopted:

$$\mathbf{Z}_{1:T}^{(c)} := \left(\mathbf{Z}_{1:T}^{(o)},\ \mathbf{Z}_{1:T}^{(m)}\right) := \left(\mathbf{Z}_{1:T}^{(o)},\ \boldsymbol{\xi}_{1:T}\right) := \left(\mathbf{Z}_{1:T},\ \boldsymbol{\xi}_{1:T}\right) \tag{2.11}$$

Then, the formulation of the joint density for the complete-data is:

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

and the complete-data log-likelihood function is

$$\ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}^{(c)}\right) := \log(L(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}))$$
$$2\ell_c(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}) = \ell^{(1)}\left(\boldsymbol{\beta}, \sigma_\omega^2|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}\right) + \ell^{(2)}\left(\boldsymbol{\phi}, \alpha, \sigma_\eta^2|\boldsymbol{\xi}_{1:T}\right) + \ell^{(3)}\left(\boldsymbol{\xi}_1\right)$$

with

$$\begin{aligned}
\ell^{(1)}(\boldsymbol{\beta}, \sigma_\omega^2|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}) &= -\ln\{\sigma_\omega^2\}\sum_{t=1}^{T} n_t \\
&\quad - \frac{1}{\sigma_\omega^2}\sum_{t=1}^{T}\left(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_t\right)^\top\left(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_t\right) \\
\ell^{(2)}(\boldsymbol{\phi}, \alpha, \sigma_\eta^2|\boldsymbol{\xi}_{1:T}) &= -(T-1)(\ln|\mathbf{D}_\mathbf{Q}|) - \sum_{t=2}^{T}\left(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\right)^\top\mathbf{Q}^-\left(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\right) \\
&= -(T-1)(np\ \ln\{\sigma_\eta^2\} + \ln|\mathbf{D}_\mathbf{R}|) \\
&\quad - \frac{1}{\sigma_\eta^2}\sum_{t=2}^{T}\left(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\right)^\top\mathbf{R}^-\left(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\right) \\
\ell^{(3)}(\boldsymbol{\xi}_1) &= -\ln|\mathbf{P}_{1|0}| - \left(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_{1|0}\right)^\top\mathbf{P}_{1|0}^{-1}\left(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_{1|0}\right)
\end{aligned}$$

In the presence of missing data, the most straightforward method for calculating the MLE is through the EM algorithm. This algorithm solved the observed-data log-likelihood function, ℓ_o, by indirectly iterating over the complete-data log-likelihood function, ℓ_c, using the conditional expectation of the latent state given the observed data and the currently estimated parameters.

The EM algorithm is resumed in two main steps:

1. E-step: Compute the conditional expectation of the complete data log-likelihood

given the observed data through the auxiliary $\mathcal{Q}$-function

$$\begin{aligned}\mathcal{Q}\left(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(i-1)}\right) &:= E\left\{2\ell_c(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}^{(c)})\middle|\mathbf{Z}_{1:T}^{(o)},\widehat{\boldsymbol{\Theta}}^{(i-1)}\right\}\\ &= E\left\{2\ell_c(\boldsymbol{\Theta}|\mathbf{Z}_{1:T},\boldsymbol{\xi}_{1:T})\middle|\mathbf{Z}_{1:T},\widehat{\boldsymbol{\Theta}}^{(i-1)}\right\}\\ &= q^{(1)}\left(\boldsymbol{\beta},\sigma_\omega^2|\mathbf{Z}_{1:T},\widehat{\boldsymbol{\Theta}}^{(i-1)}\right)+q^{(2)}\left(\boldsymbol{\phi},\alpha,\sigma_\eta^2|\mathbf{Z}_{1:T},\widehat{\boldsymbol{\Theta}}^{(i-1)}\right)\\ &\qquad\qquad + q^{(3)}\left(\mathbf{Z}_{1:T},\widehat{\boldsymbol{\Theta}}^{(i-1)}\right)\end{aligned}$$

where

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

with

$$\mathbf{S}_{22} := \sum_{t=2}^{T}\left[\mathbf{P}_{t|T}+\boldsymbol{\xi}_{t|T}\boldsymbol{\xi}_{t|T}^\top\right]$$

$$\mathbf{S}_{11} := \sum_{t=2}^{T}\left[\mathbf{P}_{t-1|T}+\boldsymbol{\xi}_{t-1|T}\boldsymbol{\xi}_{t-1|T}^\top\right]$$

$$\mathbf{S}_{21} := \sum_{t=2}^{T}\left[\mathbf{P}_{t,t-1|T}+\boldsymbol{\xi}_{t|T}\boldsymbol{\xi}_{t-1|T}^\top\right]$$

where $\mathbf{P}_{t,t-1|T} := \mathrm{E}\left[(\boldsymbol{\xi}_t-\boldsymbol{\xi}_{t|T})(\boldsymbol{\xi}_{t-1}-\boldsymbol{\xi}_{t-1|T})^\top\middle|\mathbf{Z}_{1:T}\right]$ is the commonly called *lag-one covariance smoother* (Shumway and Stoffer, Property P6.3), which is obtained

by calculating:

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

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

2. M-step: compute the parameters estimates that maximizes $\mathcal{Q}\left(\mathbf{\Theta}|\widehat{\mathbf{\Theta}}^{(i-1)}\right)$

$$\widehat{\mathbf{\Theta}}^{(i)} = \arg\max_{\mathbf{\Theta}} \left\{\mathcal{Q}\left(\mathbf{\Theta}|\widehat{\mathbf{\Theta}}^{(i-1)}\right)\right\}$$

All the maximization calculations are detailed in the appendix section VII.2, obtained the following parameters estimates :

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

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

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

$$\widehat{\boldsymbol{\phi}}_j^{(i)} = \operatorname{tr}\left\{\widehat{\mathbf{Q}}^{-\,(i-1)}\mathbf{S}_{21}^{(i-1)}\left(\boldsymbol{\upsilon}\otimes\{\phi_{,j}\}_{1\le j\le p}\right)\right\}\cdot\operatorname{tr}\left\{\widehat{\mathbf{Q}}^{-\,(i-1)}\mathbf{S}_{11}^{(i-1)}\left(\boldsymbol{\upsilon}^\top\boldsymbol{\upsilon}\right)\right\}^{-1}, \quad \forall\, j = 1,\ldots,p$$

where $\{\phi_{,j}\}_{1\le j\le p}$ correspond to the p unit vectors of dimension $p\times 1$ obtained differentiating by $\boldsymbol{\phi}$ (Bürger, 2014), these mean:

$$\{\phi_{,j}\}_{1\le j\le p} = \begin{Bmatrix}\phi_{,1}\\ \vdots \\ \phi_{,p}\end{Bmatrix} = \begin{Bmatrix}\dfrac{\partial\phi}{\partial\phi_1}\\ \vdots \\ \dfrac{\partial\phi}{\partial\phi_p}\end{Bmatrix}$$

When some of the parameters cannot be solved analytically, it becomes necessary to employ the GEM algorithm, which is based on a single Newton-Raphson step (McLachlan and Krishnan, 2008). Subsequently, the vector parameters can be represented as $(\boldsymbol{\Theta}^{\bullet\top}, \alpha)^\top$, where α denoting a scalar parameter that requires resolution through the Newton-Raphson algorithm, then, $\boldsymbol{\Theta}^\bullet = (\boldsymbol{\beta}^T, \boldsymbol{\phi}^T, \sigma_\eta^2, \sigma_\omega^2)^T$ with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_\ell)^T, \boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^\top$

The update $\widehat{\alpha}^{(i)}$ has the form $\widehat{\alpha}^{(i)} = \widehat{\alpha}^{(i-1)} + w^{(i-1)}\gamma^{(i-1)}$ where $0 < w^{(i-1)} \leq 1$, in practice $w^{(i-1)} = 1$ and

$$\gamma^{(i-1)} = -\frac{\partial^2 \mathcal{Q}\left(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{\bullet(i-1)}, \widehat{\alpha}^{(i-1)}\right)}{\partial\alpha\partial\alpha}^{-1} \frac{\partial \mathcal{Q}\left(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{\bullet(i-1)}, \widehat{\alpha}^{(i-1)}\right)}{\partial\alpha}$$

The update $\widehat{\boldsymbol{\Theta}}^{(i)}$ maximizes the $\mathcal{Q}$-function to increase the likelihood at each iteration, so $\mathcal{Q}(\widehat{\boldsymbol{\Theta}}^{(i)}|\widehat{\boldsymbol{\Theta}}^{(i-1)}) \geq \mathcal{Q}(\widehat{\boldsymbol{\Theta}}^{(i-1)}|\widehat{\boldsymbol{\Theta}}^{(i-1)})$ need to holds to ensure that the algorithm monotonically converges to a local stationary point, the MLE (McLachlan and Krishnan, 2008)

Finally, the EM algorithm converges according to a chosen tolerance level δ:

1. $\dfrac{||\widehat{\boldsymbol{\Theta}}^{(i)} - \widehat{\boldsymbol{\Theta}}^{(i-1)}||}{||\widehat{\boldsymbol{\Theta}}^{(i-1)}||} < \delta_a$

   or

2. $\dfrac{||\ell_o(\widehat{\boldsymbol{\Theta}}^{(i)}) - \ell_o(\widehat{\boldsymbol{\Theta}}^{(i-1)})||}{||\ell_o(\widehat{\boldsymbol{\Theta}}^{(i-1)})||} < \delta_b$

The EM algorithm converges to a local stationary point; therefore, it is recommended to verify the results using different starting values (Xu and Wikle, 2007; Bickel et al., 1998). The number of iterations needed to achieve δ is influenced by the proportion of missing data; when this proportion is large, the convergence may be slow (McLachlan and Krish-

nan, 2008; Tanner, 1993).

A common criticism of the EM algorithm is its inability to estimate the variance-covariance matrix of the MLE automatically. Several studies have addressed this issue for specific models. Since our interest is in calculating the standard errors of MLE for our model, a review of some of these methodologies will be presented in the following chapter.

2.4 Estimation of the variance-covariance matrix of the Maximum Likelihood Estimation in the presence of missing data

Under regularity conditions, the MLE $\widehat{\boldsymbol{\Theta}}_{ML} := \widehat{\boldsymbol{\Theta}}$, is consistent, meaning that it converges in probability to the true parameter value, $\widehat{\boldsymbol{\Theta}} \xrightarrow[n\to\infty]{P} \boldsymbol{\Theta}$. Additionally, the MLE is asymptotically efficient, indicating that the variance of $\widehat{\boldsymbol{\Theta}}$ reaches the Cramer-Rao lower bound, given by $\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\Theta})$. Furthermore, the MLE is asymptotically multivariate normally distributed, i.e.,

$$\widehat{\boldsymbol{\Theta}} \xrightarrow[n\to\infty]{d} \mathcal{N}_{\ell+p+3}(\boldsymbol{\Theta},\, \boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\Theta}))$$

where $\boldsymbol{\mathcal{I}}(\boldsymbol{\Theta})$ denotes the FIM. These properties are also fulfilled in state-space models when the KF is stable, which means the covariance of the filter converges to a stationary value (Shumway and Stoffer, 2017), as well as in Hidden Markov Models (Bickel et al., 1998).

The FIM can be defined as the negative expected value of the second derivative of the log-likelihood function with respect to the parameter vector $\boldsymbol{\Theta}$. Equivalently, it can be expressed as the negative expectation of the Hessian matrix of the log-likelihood with respect to $\boldsymbol{\Theta}$. This concept is also referred to as the expected FIM for any random vector, $\mathbf{Z}$ (Mardia et al., 1979).

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\Theta}) = E\left\{ \left(\frac{\partial \ell(\boldsymbol{\Theta}|\mathbf{Z})}{\partial \boldsymbol{\Theta}} \right) \left(\frac{\partial \ell(\boldsymbol{\Theta}|\mathbf{Z})}{\partial \boldsymbol{\Theta}} \right)^{\top} \right\} = -E\left\{ \frac{\partial^2 \ell(\boldsymbol{\Theta}|\mathbf{Z})}{\partial \boldsymbol{\Theta} \partial \boldsymbol{\Theta}^{\top}} \right\} = -E\Big\{ H(\boldsymbol{\Theta}|\mathbf{Z}) \Big\} \tag{2.12}$$

In contrast, the observed FIM is defined as follows:

$$\boldsymbol{I}_o(\widehat{\boldsymbol{\Theta}}) = -\frac{\partial^2 \ell(\boldsymbol{\Theta}|\mathbf{Z})}{\partial \boldsymbol{\Theta} \partial \boldsymbol{\Theta}^{\top}} \bigg|_{\boldsymbol{\Theta}=\widehat{\boldsymbol{\Theta}}} \tag{2.13}$$

This definition allows us to leaving out the need for expectation calculations (McLachlan and Krishnan, 2008; Walsh, 2006; Tanner, 1993)). As noted by Efron and Hinkley (1978) and Louis (1982), the observed FIM is often significantly easier to compute than the expected FIM and provides a more accurate estimate of the variance of $\widehat{\boldsymbol{\Theta}}$. This advantage is particularly relevant when MLE is asymptotically unbiased and serves as a sufficient estimator of $\boldsymbol{\Theta}$, meaning that under large-sample conditions, we have $E\{\widehat{\boldsymbol{\Theta}}\} = \boldsymbol{\Theta}$.

However, calculating the observed FIM is not that simple in the presence of missing data. We defined the complete data (2.11) as the union of the observed data and the unobserved

(missing) data. By extending this definition to the log-likelihood functions, we could write

$$\begin{aligned} \log f\left(\mathbf{Z}_{1:T}^{(c)}|\boldsymbol{\Theta}\right) &= \log f\left(\mathbf{Z}_{1:T}^{(o)}|\boldsymbol{\Theta}\right) + \log f\left(\boldsymbol{\xi}_{1:T}|\mathbf{Z}_{1:T}^{(o)}, \boldsymbol{\Theta}\right) \\ \ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}^{(c)}\right) &= \ell_o\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}^{(o)}\right) + \ell_m\left(\boldsymbol{\xi}_{1:T}|\mathbf{Z}_{1:T}^{(o)}, \boldsymbol{\Theta}\right) \end{aligned} \tag{2.14}$$

Rearranging the expression (2.14) and differentiating the negative of both sides twice with respect to $\boldsymbol{\Theta}$ gives:

$$-\frac{\partial^2 \ell_o\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}\right)}{\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}^\top} = -\frac{\partial^2 \ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}\right)}{\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}^\top} + \frac{\partial^2 \ell_m\left(\boldsymbol{\xi}_{1:T}|\mathbf{Z}_{1:T}, \boldsymbol{\Theta}\right)}{\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}^\top} \tag{2.15}$$

Using the notation of equation (2.12, 2.13), then taking expectation over the conditional distribution of $\boldsymbol{\xi}_{1:T}|\mathbf{Z}_{1:T}$ yields

$$\begin{aligned} \boldsymbol{I}_o\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}\right) &= E\left\{\boldsymbol{I}_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}\right)\middle|\mathbf{Z}_{1:T}\right\} - E\left\{\boldsymbol{I}_m\left(\boldsymbol{\xi}_{1:T}|\mathbf{Z}_{1:T}, \boldsymbol{\Theta}\right)\middle|\mathbf{Z}_{1:T}\right\} \\ \boldsymbol{I}_o\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}\right) &= \boldsymbol{\mathcal{I}}_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}\right) - \boldsymbol{\mathcal{I}}_m\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}\right) \end{aligned} \tag{2.16}$$

The expression (2.16), known as the "Missing Information Principle", was proposed by University (1972). Intuitively, it asserts that the observed FIM, $\boldsymbol{I}_o$, is equivalent to the expected complete-data FIM, $\boldsymbol{\mathcal{I}}_c$, minus the expected missing-data FIM, $\boldsymbol{\mathcal{I}}_m$, which represents the "missing information" resulting from unobservable data. Louis (1982) demonstrated that the observed FIM can be expressed entirely in terms of conditional expectations derived from the complete-data log-likelihood function:

$$\boldsymbol{I}_o\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}\right) = -E\left\{\frac{\partial^2 \ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}\right)}{\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}^\top}\middle|\mathbf{Z}_{1:T}\right\} - \mathrm{Var}\left\{\frac{\partial \ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}\right)}{\partial\boldsymbol{\Theta}}\middle|\mathbf{Z}_{1:T}\right\} \tag{2.17}$$

This expression arises from the derivations of the definitions of the observed FIM. Using the notes of Holmes (2016), it is possible to demonstrate Louis' results. To simplify the notation for the demonstration, let $\mathbf{Z} := \mathbf{Z}_{1:T}^{(o)}$ represent the observed data, and $\boldsymbol{\xi} := \boldsymbol{\xi}_{1:T}$ represent the missing data. $(\mathbf{Z}, \boldsymbol{\xi})$ constitutes the complete data, while the use of a prime/apostrophe indicates partial derivatives. Thus, the observed FIM is defined as follows:

$$\begin{aligned} \boldsymbol{I}_o(\boldsymbol{\Theta}|\mathbf{Z}) &= -\frac{\partial^2 \ell(\boldsymbol{\Theta}|\mathbf{Z})}{\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}^\top} = -\frac{\partial^2 \log f(\mathbf{Z}|\boldsymbol{\Theta})}{\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}^\top} = -\frac{\partial S(\mathbf{Z}|\boldsymbol{\Theta})}{\partial\boldsymbol{\Theta}^\top} \\ &= \left(\frac{\partial \log f(\mathbf{Z}|\boldsymbol{\Theta})}{\partial\boldsymbol{\Theta}}\right)\left(\frac{\partial \log f(\mathbf{Z}|\boldsymbol{\Theta})}{\partial\boldsymbol{\Theta}}\right)^\top = S(\mathbf{Z}|\boldsymbol{\Theta})\, S(\mathbf{Z}|\boldsymbol{\Theta})^\top \end{aligned} \tag{2.18}$$

where $S(\mathbf{Z}|\boldsymbol{\Theta})$ is the score function for the observed data (the gradient of the observed-data log-likelihood function with respect to $\boldsymbol{\Theta}$). This allows us to reformulate the score function for the observed data

$$S(\mathbf{Z}|\boldsymbol{\Theta}) = \frac{\partial \log \int_{\boldsymbol{\xi}} f_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})\, d\boldsymbol{\xi}}{\partial\boldsymbol{\Theta}} = \frac{\int_{\boldsymbol{\xi}} f'_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})\, d\boldsymbol{\xi}}{\int_{\boldsymbol{\xi}} f_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})\, d\boldsymbol{\xi}} = \frac{\int_{\boldsymbol{\xi}} f'_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})\, d\boldsymbol{\xi}}{f_{\boldsymbol{\xi}}(\boldsymbol{\xi}|\boldsymbol{\Theta})} \tag{2.19}$$

By multiplying and dividing the integrand in the numerator by the joint density for the complete data $f_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})$ and rearranging, we get:

$$\int_{\boldsymbol{\xi}} f'_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})\, d\boldsymbol{\xi} = \int_{\boldsymbol{\xi}} \frac{f'_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})}{f_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})} f_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})\, d\boldsymbol{\xi} = \int_{\boldsymbol{\xi}} S(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})\, f_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})\, d\boldsymbol{\xi} \tag{2.20}$$

where $S(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})$ is the complete-data score function. Substituting (2.20) into (2.19) yields:

$$\begin{aligned} S(\mathbf{Z}|\boldsymbol{\Theta}) &= \frac{\int_{\boldsymbol{\xi}} f'_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})\, d\boldsymbol{\xi}}{f_{\boldsymbol{\xi}}(\boldsymbol{\xi}|\boldsymbol{\Theta})} \\ &= \int_{\boldsymbol{\xi}} S(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta}) \frac{f_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})}{f_{\boldsymbol{\xi}}(\boldsymbol{\xi}|\boldsymbol{\Theta})} d\boldsymbol{\xi} \\ &= \int_{\boldsymbol{\xi}} S(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})\, f_{\boldsymbol{\xi}|\mathbf{Z}}(\boldsymbol{\xi}|\mathbf{Z}, \boldsymbol{\Theta}) d\boldsymbol{\xi} \\ &= E_{\boldsymbol{\xi}|\mathbf{Z}}\Big\{ S(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta}) \Big\} \end{aligned} \tag{2.21}$$

Thus, the observed-data score function is the conditional expectation of the complete-data score function, given the observed data. This result is Equation 3.1 in Louis (1982). To derive Equation 3.2 of Louis' Method, he defines $B(\cdot)$ as the negative second derivative of the log-likelihood (consistent with the definition of the expected FIM). In the context of the complete data, we have

$$\begin{aligned} B(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta}) = -S'(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta}) &= -\frac{\partial f'_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta}) / f'_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}^{\top}} \\ &= -\left(\frac{f''_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})}{f_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})} - \frac{f'_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})\, f'_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})^{\top}}{f_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})^2} \right) \\ &= -\left(\frac{f''_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})}{f_{\mathbf{Z},\boldsymbol{\xi}}(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})} - S(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})\, S(\mathbf{Z}, \boldsymbol{\xi}|\boldsymbol{\Theta})^{\top} \right) \end{aligned} \tag{2.22}$$

From the observed-data FIM (2.18), we obtain:

$$\begin{aligned}
\boldsymbol{I}_o\left(\boldsymbol{\Theta}|\mathbf{Z}\right) &= -\frac{\partial S\left(\mathbf{Z}|\boldsymbol{\Theta}\right)}{\partial \boldsymbol{\Theta}^\top} = -\frac{\partial \int_{\boldsymbol{\xi}} f'_{\mathbf{Z},\boldsymbol{\xi}}\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right) d\boldsymbol{\xi} / f_{\boldsymbol{\xi}}\left(\boldsymbol{\xi}|\boldsymbol{\Theta}\right)}{\partial \boldsymbol{\Theta}^\top} \\
&= -\left(\frac{\int_{\boldsymbol{\xi}} f''_{\mathbf{Z},\boldsymbol{\xi}}\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right) d\boldsymbol{\xi}}{f_{\boldsymbol{\xi}}\left(\boldsymbol{\xi}|\boldsymbol{\Theta}\right)} - \frac{\int_{\boldsymbol{\xi}} f'_{\mathbf{Z},\boldsymbol{\xi}}\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right) \int_{\boldsymbol{\xi}} f'_{\mathbf{Z},\boldsymbol{\xi}}\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right)^\top}{f_{\boldsymbol{\xi}}\left(\boldsymbol{\xi}|\boldsymbol{\Theta}\right)^2} \right) \\
&= -\left(\frac{\int_{\boldsymbol{\xi}} f''_{\mathbf{Z},\boldsymbol{\xi}}\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right) d\boldsymbol{\xi}}{f_{\boldsymbol{\xi}}\left(\boldsymbol{\xi}|\boldsymbol{\Theta}\right)} - S\left(\mathbf{Z}|\boldsymbol{\Theta}\right) S\left(\mathbf{Z}|\boldsymbol{\Theta}\right)^\top \right)
\end{aligned} \tag{2.23}$$

Taking the first term of (2.23), multiplying and dividing the numerator by the joint density for the complete data $f_{\mathbf{Z},\boldsymbol{\xi}}\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right)$, as in (2.20), we gets:

$$\begin{aligned}
\frac{\int_{\boldsymbol{\xi}} f''_{\mathbf{Z},\boldsymbol{\xi}}\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right) d\boldsymbol{\xi}}{f_{\boldsymbol{\xi}}\left(\boldsymbol{\xi}|\boldsymbol{\Theta}\right)} &= \frac{\int_{\boldsymbol{\xi}} f''_{\mathbf{Z},\boldsymbol{\xi}}\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right) d\boldsymbol{\xi}}{f_{\mathbf{Z},\boldsymbol{\xi}}\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right)} \frac{f_{\mathbf{Z},\boldsymbol{\xi}}\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right)}{f_{\boldsymbol{\xi}}\left(\boldsymbol{\xi}|\boldsymbol{\Theta}\right)} \\
&= \int_{\boldsymbol{\xi}} \frac{f''_{\mathbf{Z},\boldsymbol{\xi}}\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right)}{f_{\mathbf{Z},\boldsymbol{\xi}}\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right)} f_{\boldsymbol{\xi}|\mathbf{Z}}(\boldsymbol{\xi}|\mathbf{Z},\boldsymbol{\Theta}) d\boldsymbol{\xi}
\end{aligned} \tag{2.24}$$

This is the conditional expetation $E_{\boldsymbol{\xi}|\mathbf{Z}}\left\{ f''_{\mathbf{Z},\boldsymbol{\xi}}\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right) / f_{\mathbf{Z},\boldsymbol{\xi}}\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right) \right\}$. That expression is in the (2.22), then

$$\frac{f''_{\mathbf{Z},\boldsymbol{\xi}}\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right)}{f_{\mathbf{Z},\boldsymbol{\xi}}\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right)} = -B\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right) + S\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right) S\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right)^\top \tag{2.25}$$

Combining these results leads to Equation 3.2 in Louis (1982) and, finally, to (2.17)

$$\begin{aligned}
\boldsymbol{I}_o\left(\boldsymbol{\Theta}|\mathbf{Z}\right) &= E_{\boldsymbol{\xi}|\mathbf{Z}}\left\{ B\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right) - S\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right) S\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right)^\top \right\} + S\left(\mathbf{Z}|\boldsymbol{\Theta}\right) S\left(\mathbf{Z}|\boldsymbol{\Theta}\right)^\top \\
&= E_{\boldsymbol{\xi}|\mathbf{Z}}\left\{ B\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right) \right\} - E_{\boldsymbol{\xi}|\mathbf{Z}}\left\{ S\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right) S\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right)^\top \right\} \\
&\qquad + E_{\boldsymbol{\xi}|\mathbf{Z}}\left\{ S\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right) \right\} E_{\boldsymbol{\xi}|\mathbf{Z}}\left\{ S\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right) \right\}^\top \\
\boldsymbol{I}_o\left(\boldsymbol{\Theta}|\mathbf{Z}\right) &= E_{\boldsymbol{\xi}|\mathbf{Z}}\left\{ -\frac{\partial^2 \log f\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right)}{\partial \boldsymbol{\Theta} \partial \boldsymbol{\Theta}^\top} \right\} - \mathrm{Var}_{\boldsymbol{\xi}|\mathbf{Z}}\left\{ \frac{\partial \log f\left(\mathbf{Z},\boldsymbol{\xi}|\boldsymbol{\Theta}\right)}{\partial \boldsymbol{\Theta}} \right\}
\end{aligned} \tag{2.26}$$

McLachlan and Krishnan (2008) published a demonstration of Louis' method using an alternative approach, arriving at the same results. Hence, the first term in (2.17) and in (2.26) corresponds to the expected complete-data FIM, $\boldsymbol{\mathcal{I}}_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}^{(c)}\right) \equiv \boldsymbol{\mathcal{I}}_c\left(\boldsymbol{\Theta}|\mathbf{Z}, \boldsymbol{\xi}\right)$, while the second term refers to the expected missing-data FIM. It is possible to demonstrate that, under regularity conditions, $\boldsymbol{\mathcal{I}}_c$ is equivalent to the second derivative of the auxiliary $\mathcal{Q}$-function utilized in the EM-algorithm, specifically,

$$\begin{aligned}
\boldsymbol{\mathcal{I}}_c\left(\widehat{\boldsymbol{\Theta}}|\mathbf{Z}_{1:T}\right) &= -E\left\{\frac{\partial^2 \ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}\right)}{\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}^\top}\middle|\mathbf{Z}_{1:T}\right\}\Bigg|_{\boldsymbol{\Theta}=\widehat{\boldsymbol{\Theta}}} \\
&= -\frac{\partial^2 E\left\{\ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}\right)|\mathbf{Z}_{1:T}\right\}}{\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}^\top}\Bigg|_{\boldsymbol{\Theta}=\widehat{\boldsymbol{\Theta}}} \\
&= -\frac{\partial^2 \mathcal{Q}(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}})}{\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}^\top}\Bigg|_{\boldsymbol{\Theta}=\widehat{\boldsymbol{\Theta}}}
\end{aligned}$$

where $\mathcal{Q}(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}})$ represents the conditional expectation of the complete-data log-likelihood given the observed data and the parameter estimates $\widehat{\boldsymbol{\Theta}}$.

The last term in (2.17) and (2.26) corresponds to the expected missing-data FIM. This is defined as the conditional variance of the first derivative of the complete data log-likelihood function

$$\begin{aligned}
\boldsymbol{\mathcal{I}}_m\left(\boldsymbol{\Theta}, \mathbf{Z}_{1:T}^{(c)}|\mathbf{Z}_{1:T}^{(o)}\right) &= \mathrm{Var}\left\{\frac{\partial \ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}^{(c)}\right)}{\partial\boldsymbol{\Theta}}\middle|\mathbf{Z}_{1:T}^{(o)}\right\} \\
&= E\left\{\left(\frac{\partial \ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}\right)}{\partial\boldsymbol{\Theta}}\right)\left(\frac{\partial \ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}\right)}{\partial\boldsymbol{\Theta}}\right)^\top\middle|\mathbf{Z}_{1:T}\right\} \\
&\quad - E\left\{\frac{\partial \ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}\right)}{\partial\boldsymbol{\Theta}}\middle|\mathbf{Z}_{1:T}\right\} E\left\{\frac{\partial \ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}\right)}{\partial\boldsymbol{\Theta}}\middle|\mathbf{Z}_{1:T}\right\}^\top
\end{aligned} \tag{2.27}$$

At the MLE calculated by the EM algorithm, the gradient of the observed-data log-likelihood is zero (McLachlan and Krishnan, 2008), resulting in the vanishing of the second term in (2.27). Finally, the observed FIM at $\widehat{\boldsymbol{\Theta}}$ can be expressed as:

$$\begin{aligned}\boldsymbol{I}_o\left(\widehat{\boldsymbol{\Theta}}|\mathbf{Z}_{1:T}\right) = &-E\left\{\frac{\partial^2 \ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}\right)}{\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}^\top}\middle|\mathbf{Z}_{1:T}\right\} \\ &- E\left\{\left(\frac{\partial \ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}\right)}{\partial\boldsymbol{\Theta}}\right)\left(\frac{\partial \ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}\right)}{\partial\boldsymbol{\Theta}}\right)^\top\middle|\mathbf{Z}_{1:T}\right\}\Bigg|_{\boldsymbol{\Theta}=\widehat{\boldsymbol{\Theta}}}\end{aligned} \tag{2.28}$$

In the case of nonindependent data, the exact computation of the observed FIM can be complex, due to the need for some impossible and tedious expectations that involve quadratic, cubic, quartic, and bilinear forms of the state process (Holmes, 2014; Turner et al., 1998; Tanner, 1993). As a result, several methods have been proposed to approximate this matrix numerically. However, most of these approaches primarily focus on facilitating the EM algorithm by substituting analytical expectation calculations with techniques such as Monte Carlo methods, stochastic approximation algorithms, or data augmentation procedures, while the computation of the FIM remains on the side. This section offers an overview of the strategies proposed by Delyon et al. (1999); Tanner and Wong (1987); Wei and Tanner (1990) and also includes the Supplemented EM (SEM) algorithm.

Meng and Rubin (1991) proposed the SEM algorithm to automatically compute the variance-covariance matrix of the MLE by relying solely on the expected complete-data FIM. They demonstrated that

$$\boldsymbol{I}_o\left(\widehat{\boldsymbol{\Theta}}|\mathbf{Z}_{1:T}^{(o)}\right)^{-1} = \boldsymbol{\mathcal{I}}_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}^{(c)}\right)^{-1}(\mathbf{I} - DM)^{-1}$$

where $\boldsymbol{\mathcal{I}}_c\, DM$ represents the variance due to the missing data. The main challenge of this methodology lies in computing the DM matrix. DM is the Jacobian matrix of $\mathbf{M}(\boldsymbol{\Theta}^{(i)}) = \boldsymbol{\Theta}^{(i+1)}$. To approximate DM, the authors used numerical differentiation. They illustrated their approach with univariate, multinomial, and bivariate data examples. Furthermore, the SEM algorithm seems to be susceptible to numerical inaccuracies and instability, particularly in high-dimensional models (McLachlan and Krishnan, 2008; Jamshidian and Jennrich, 2000).

A straightforward extension of the EM algorithm is the Monte Carlo EM (MCEM) algorithm.Wei and Tanner (1990) proposed to simulate M realizations of the missing data, denoted $\boldsymbol{\xi}_{1:T}^{(j)} = \left\{\boldsymbol{\xi}_{1:T}^{(1)}, \ldots, \boldsymbol{\xi}_{1:T}^{(M)}\right\}$, from the conditional distribution, $f\left(\boldsymbol{\xi}_t|\widehat{\boldsymbol{\Theta}}^{(i-1)}, \mathbf{Z}_{1:T}\right)$. This approach replaces the analytical expectation in the E-step with an empirical average over the simulations and then continues with the usual M-step.

$$\widetilde{\mathcal{Q}}\left(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(i-1)}\right) = \frac{1}{M}\sum_{j=1}^{M} \ell_c\left(\boldsymbol{\Theta}^{(i-1)}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}^{(j)}\right)$$

The authors note that when $M = 1$, the missing data summarizes the conditional distribution, such as its expected value. This leads to the initial EM algorithm, which utilized an expected value of the missing data state, $\boldsymbol{\xi}_{t|T} = \mathrm{E}\left[f\left(\boldsymbol{\xi}_t|\widehat{\boldsymbol{\Theta}}^{(i-1)}, \mathbf{Z}_{1:T}\right)\right]$. Following the same idea, they propose an approximation of the observed FIM via Monte Carlo ap-

proach:

$$\widetilde{\boldsymbol{I}}_o\left(\widehat{\boldsymbol{\Theta}}|\mathbf{Z}_{1:T}\right) = \frac{1}{M}\sum_{j=1}^{M}\frac{\partial^2\ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T},\boldsymbol{\xi}_{1:T}^{(j)}\right)}{\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}^\top} + \frac{1}{M}\sum_{j=1}^{M}\left(\frac{\partial\ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T},\boldsymbol{\xi}_{1:T}^{(j)}\right)}{\partial\boldsymbol{\Theta}}\right)\left(\frac{\partial\ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T},\boldsymbol{\xi}_{1:T}^{(j)}\right)}{\partial\boldsymbol{\Theta}}\right)^\top - \frac{1}{M}\sum_{j=1}^{M}\left(\frac{\partial\ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T},\boldsymbol{\xi}_{1:T}^{(j)}\right)}{\partial\boldsymbol{\Theta}}\right)\frac{1}{M}\sum_{j=1}^{M}\left(\frac{\partial\ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T},\boldsymbol{\xi}_{1:T}^{(j)}\right)}{\partial\boldsymbol{\Theta}}\right)^\top\Bigg|_{\boldsymbol{\Theta}=\widehat{\boldsymbol{\Theta}}} \tag{2.29}$$

Data augmentation (DA) algorithms refer to methods that augment observed data using the conditional distribution of the latent state and parameters, thereby computing the entire posterior density rather than just the maximizer (Tanner and Wong, 1987). These algorithms are especially common in Bayesian statistics, where several approaches implement the DA principle. Wei and Tanner (1990) proposed two so-called "Poor Man's DA (PMDA) algorithms: PDMA-1 and PDMA-2. The primary difference between them is how multiple imputations of the latent state are made. The PDMA-1 is a non-iterative algorithm for obtaining a refinement approximation of the observed posterior distribution. It is a simple modification to the MCEM algorithm, except that, after obtaining $\widehat{\boldsymbol{\Theta}}$ in the final iteration, multiple imputations of the latent state are drawn from the conditional predictive distribution to approximate the entire (complete-data) posterior density. In PDMA-2, handles scenarios where the exact predictive distribution is available by using importance sampling. Otherwise, an approximation of the predictive distribution is required. PDMA-2 is an improvement of PDMA-1 by enhancing the efficiency of the algorithm. Nonetheless, both serve as starting points for more advanced DA algorithm (Tanner, 1993).

Delyon et al. (1999) proposed the Stochastic Approximation EM (SAEM) algorithm as

an extension of the standard EM procedure. Its basic idea is to split the E-step into a simulation step (S-step) and an averaging step (A-step). In each S-step, the algorithm generates M realizations of the missing data vector $\boldsymbol{\xi}_{1:T}^{(j)}$ from the conditional distribution $f\left(\boldsymbol{\xi}_{1:T}|\widehat{\boldsymbol{\Theta}}^{(i-1)}, \mathbf{Z}_{1:T}\right)$ and the Monte Carlo integration is substituted by a stochastic averaging procedure during the A-step. The M-step then proceeds as in the classical EM algorithm. Formally, the stochastic approximation for the $\mathcal{Q}$-function is given by

$$\widehat{\mathcal{Q}}\left(\boldsymbol{\Theta}\right)^{(i)} \approx \widehat{\mathcal{Q}}\left(\boldsymbol{\Theta}\right)^{(i-1)} + \gamma^{(i)}\left(\frac{1}{M}\sum_{j=1}^{M}\ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}^{(j)}\right) - \widehat{\mathcal{Q}}\left(\boldsymbol{\Theta}\right)^{(i-1)}\right)$$

where $\gamma^{(i)}$ is a sequence of positive step size that typically decreases over iterations.

The authors also compute the observed FIM through stochastic approximation of the Hessian of the complete-data log-likelihood and the variance of the corresponding score function. Specifically, they define:

$$\begin{aligned} H\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}^{(c)}\right)^{(i)} \approx H\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}^{(c)}\right)^{(i-1)} + \gamma^{(i)}\Bigg(\frac{1}{M}\sum_{j=1}^{M}\frac{\partial^2\ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}^{(j)}\right)}{\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}^\top} \\ + \frac{\partial\ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}^{(j)}\right)}{\partial\boldsymbol{\Theta}}\frac{\partial\ell_c\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}^{(j)}\right)}{\partial\boldsymbol{\Theta}}^\top - H\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}^{(c)}\right)^{(i-1)}\Bigg) \\ S\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}^{(c)}\right)^{(i)} \approx S\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}^{(c)}\right)^{(i-1)} + \gamma^{(i)}\Bigg(\frac{1}{M}\sum_{l=1}^{M}\left(\frac{\partial\ell(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}^{(j)})}{\partial\boldsymbol{\Theta}}\right)\left(\frac{\partial\ell(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}^{(j)})}{\partial\boldsymbol{\Theta}}\right)^\top \\ - S\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}^{(c)}\right)^{(i-1)}\Bigg) \end{aligned}$$

Once the iterative procedure has converged to the final estimate $\widehat{\boldsymbol{\Theta}}$, the observed FIM is approximated by

$$\widetilde{\boldsymbol{I}}_o\left(\widehat{\boldsymbol{\Theta}}|\mathbf{Z}_{1:T}^{(o)}\right) = H\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}^{(c)}\right)^{(i)} - S\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}^{(c)}\right)^{(i)} S\left(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}^{(c)}\right)^{(i)^\top}\Big|_{\boldsymbol{\Theta}=\widehat{\boldsymbol{\Theta}}} \tag{2.30}$$

where the superscript (i) indicates values obtained at iteration i. This expression combines the Hessian-based component with the score-based variance to yield a stochastic approximation of the observed FIM.

SAEM is often very useful for models in which the traditional EM algorithm cannot analytically solve the E-step. However, adding sampling and averaging iterations at each step requires additional coding effort and computational cost.

Tanner (1993) and Turner et al. (1998) proposed the simplest methodology, which is more applicable to our model. Their approach involves simulating M realizations of the unobservable (latent) data from the filtering distribution of the KF, that is, the conditional distribution for the latent state given the observed data, and then explicitly computing the missing-data FIM. They note that this approach constitutes a MC approximation: the sample mean of these missing-data FIMs will converge to the expected value. In our case:

$$\begin{aligned}\boldsymbol{\mathcal{I}}_m(\boldsymbol{\Theta}, \mathbf{Z}_{1:T}^{(c)}|\mathbf{Z}_{1:T}^{(o)}) &= \mathrm{Var}\left\{\frac{\partial \ell_c(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}^{(c)})}{\partial \boldsymbol{\Theta}}\middle|\mathbf{Z}_{1:T}\right\} \\ &= E\left\{\left(\frac{\partial \ell_c(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T})}{\partial \boldsymbol{\Theta}}\right)\left(\frac{\partial \ell_c(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T})}{\partial \boldsymbol{\Theta}}\right)^\top\middle|\mathbf{Z}_{1:T}\right\} \\ &\quad - E\left\{\frac{\partial \ell_c(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T})}{\partial \boldsymbol{\Theta}}\middle|\mathbf{Z}_{1:T}^{(o)}\right\} E\left\{\frac{\partial \ell_c(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T})}{\partial \boldsymbol{\Theta}}\middle|\mathbf{Z}_{1:T}\right\}^\top\end{aligned} \tag{2.31}$$

$$\widetilde{\mathcal{I}}_m(\boldsymbol{\Theta}, \mathbf{Z}_{1:T}^{(c)}|\mathbf{Z}_{1:T}^{(o)}) = \frac{1}{M}\sum_{j=1}^{M}\left(\frac{\partial \ell_c(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}^{(j)})}{\partial \boldsymbol{\Theta}}\right)\left(\frac{\partial \ell_c(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}^{(j)})}{\partial \boldsymbol{\Theta}}\right)^\top - \frac{1}{M}\sum_{j=1}^{M}\left(\frac{\partial \ell_c(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}^{(j)})}{\partial \boldsymbol{\Theta}}\right)\frac{1}{M}\sum_{j=1}^{M}\left(\frac{\partial \ell_c(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}^{(j)})}{\partial \boldsymbol{\Theta}}\right)^\top$$

$$\widetilde{\mathcal{I}}_m(\boldsymbol{\Theta}, \mathbf{Z}_{1:T}^{(c)}|\mathbf{Z}_{1:T}^{(o)}) = \frac{1}{M}\sum_{j=1}^{M}\left(\frac{\partial \ell_c(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}^{(j)})}{\partial \boldsymbol{\Theta}}\right)\left(\frac{\partial \ell_c(\boldsymbol{\Theta}|\mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}^{(j)})}{\partial \boldsymbol{\Theta}}\right)^\top \tag{2.32}$$

Reconstructing the latent state $\boldsymbol{\xi}_t$ at time t given observed data $\mathbf{Z}_{1:T}$ is a common problem in state-space models. The most challenging case is in the presence of nonlinear models because the smoothing distribution is often intractable. Consequently, numerous methods have been proposed as an alternative to approximate the smoothing distributions, facilitating the generation of more accurate simulated samples of the latent state for parameter estimation, such as ensemble KS, Particle filters, Conditional Particle filters, among others (Chau et al., 2023). The model is a Gaussian State-Space model, so the Smoothing Distribution from KS directly provides a closed-form smoothing distribution. Hence, simulating the latent state is straightforward:

$$\boldsymbol{\xi}_{1:T}^{(j)} = [\boldsymbol{\xi}_t|\mathbf{Z}_{1:T}] \sim Gau(\boldsymbol{\xi}_{t|T}, \mathbf{P}_{t|T}) \tag{2.33}$$

When Monte Carlo simulations are used, monitoring the convergence of these simulations is essential for determining an appropriate number M (Tanner, 1993)

Another way to estimate the standard errors of our model is through resampling methods, such as Bootstrap. Bootstrap is a powerful, relatively easy-to-implement technique that is useful for a wide variety of problems, especially when parametric inference is highly

complex. However, it can be computationally intensive. The basic idea of bootstrap is to approximate the sampling distribution of an estimator by the empirical distribution of the observed data by resampling B times from the original data with replacement (maintaining the same sample size). From each resampling data set, bootstrap sample, we re-estimate the parameters, yielding $\left\{\widehat{\boldsymbol{\Theta}}^*\right\}_{b=1}^{B} =: \left\{\widehat{\boldsymbol{\Theta}}_1^*, \widehat{\boldsymbol{\Theta}}_2^*, \ldots, \widehat{\boldsymbol{\Theta}}_B^*\right\}$. Then, we use the empirical distribution of these B estimates to approximate the true distribution of $\boldsymbol{\Theta}$ to finally estimate the standard error for $\widehat{\boldsymbol{\Theta}}$.

Most bootstrap techniques have been implemented for purely spatial or purely temporal models, and relatively few works have been proven in the space-time context. Cressie and Wikle (2015) recommend the Nonparametric Monte Carlo Bootstrap (NMCB) introduced by 1991 for state-space models because the methodology proposed by them is appropriate for the space-time models used by us. Specifically, Stoffer and Wall applied the bootstrap to the residuals, $\boldsymbol{\epsilon}_t$, to obtain asymptotically consistent standard errors. The residuals are defined as the sequence of errors in the linear prediction of $\mathbf{Z}_t$ given the data $\{\mathbf{Z}_1, \ldots, \mathbf{Z}_{t-1}\}$:

$$\boldsymbol{\epsilon}_t = \mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_{t|t-1}$$

the residuals are supposed to be normally distributed with mean zero and covariance matrix, $\boldsymbol{\Sigma}_t$.

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Lambda}_t\mathbf{P}_{t|t-1}\boldsymbol{\Lambda}_t^\top + \sigma_\omega^2\mathbf{I}_{n_t}$$

Writing the state space model (2.8, 2.9) including the residuals, leads to

$$\begin{aligned} \boldsymbol{\xi}_{t+1|t} &= \boldsymbol{\Phi}\boldsymbol{\xi}_{t|t-1} + \boldsymbol{\Phi}\mathbf{K}_t\boldsymbol{\epsilon}_t \\ \mathbf{Z}_t &= \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\Lambda}_t\boldsymbol{\xi}_{t|t-1} + \boldsymbol{\epsilon}_t \end{aligned} \tag{2.34}$$

where $\mathbf{K}_t$ is the Kalman gain matrix.

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\boldsymbol{\Lambda}_t^\top \boldsymbol{\Sigma}_t^{-1}$$

this representation simplifies the bootstrap methodology by consolidating all randomness into a single error term, the residuals $\boldsymbol{\epsilon}_t$, which can then be resampled to generate bootstrap replicates. Then, the NMCB is a straightforward algorithm:

1. Calculate the residuals sequence, $\{\widehat{\boldsymbol{\epsilon}}_t\}_{t=1}^T$, the residuals covariance matrix, $\widehat{\boldsymbol{\Sigma}}_t$, and the Kalman gain matrix, $\widehat{\mathbf{K}}_t$.

2. Obtain standardized residuals, $\widehat{\mathbf{E}}_t = \widehat{\boldsymbol{\Sigma}}_t^{-1/2}\widehat{\boldsymbol{\epsilon}}_t$, then build a bootstrap sample with replacement of standardized residuals $\left\{\widehat{\mathbf{E}}_t^*\right\}_{t=1}^T$

3. Construct a bootstrap data set $\{\mathbf{Z}_t^*\}_{t=1}^T$ using the residuals representation form of the model (2.34)

$$\begin{aligned} \widehat{\boldsymbol{\xi}}_{t+1|t}^* &= \widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\xi}}_{t|t-1}^* + \widehat{\boldsymbol{\Phi}}\widehat{\mathbf{K}}_t\widehat{\mathbf{E}}_t^* \\ \widehat{\mathbf{Z}}_t^* &= \mathbf{X}_t\widehat{\boldsymbol{\beta}} + \boldsymbol{\Lambda}_t\widehat{\boldsymbol{\xi}}_{t|t-1}^* + \widehat{\mathbf{E}}_t^* \end{aligned} \tag{2.35}$$

4. Repeat B times steps 2 to 3 to obtain a set of replications of $\left\{\widehat{\boldsymbol{\Theta}}^*\right\}_{b=1}^B$ to estimate the distribution of $\widehat{\boldsymbol{\Theta}}$

Another approach was presented by Fassò and Cameletti (2009), who implemented a Space-Time Parametric Bootstrap (STPB) for a hierarchical model closely resembling our state-space representation. They generate bootstrap samples directly from the state-space equations using all the assumptions of the model and the $\widehat{\boldsymbol{\Theta}}$ obtained from the EM algorithm, and fixing the covariates for all the simulations. The STPB algorithm is as follows:

1. Simulate the initial latent state vector $\widetilde{\boldsymbol{\xi}}_0 \sim Gau\left(\boldsymbol{\xi}_{1|0}, \mathbf{P}_{1|0}\right)$ in accordance with (2.10)

2. For $t = 1, \ldots, T$ repeat from (a) to (d) to generate a bootstrap sample $\{\mathbf{Z}^*\}_{t=1}^{T}$.

   (a) Simulate a Gaussian random vector $\widetilde{\mathbf{V}}_t \sim Gau\left(\mathbf{0}, \mathcal{C}^{\eta}(\widehat{\alpha}, \widehat{\sigma}_{\eta}^2)\right)$ using (2.3)

   (b) Use equation (2.8) to update the latent state

$$\boldsymbol{\xi}_t^* = \widehat{\boldsymbol{\Phi}}\, \boldsymbol{\xi}_{t-1}^* + \widetilde{\mathbf{V}}_t$$

   (c) Simulate a Gaussian random vector $\widetilde{\mathbf{W}}_t \sim Gau\left(\mathbf{0}, \widehat{\sigma}_{\omega}^2 \mathbf{I}_n\right)$

   (d) Obtain the bootstrap observation vector at time t using the observation equation (2.9) with the real covariates matrix $\mathbf{X}_t$

$$\mathbf{Z}_t^* = \mathbf{X}_t \widehat{\boldsymbol{\beta}} + \boldsymbol{\Lambda}_t \boldsymbol{\xi}_t^* + \widetilde{\mathbf{W}}_t$$

3. Having generated B bootstrap sample $\{\mathbf{Z}_t^*\}_{b=1}^{B}$, recompute the parameters for each sample using the GEM algorithm.

By following these steps, STPB utilizes the assumed model structure to generate synthetic datasets and assess the sampling variability of the parameter estimates under realistic space-time dynamics.

In both methodologies, the bootstrap estimate of the standard error of MLE is given by the

sample standard deviation of the bootstrap replications:

$$\widehat{se}_{B,i} = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{\Theta}_i^{*b} - \frac{1}{B}\sum_{c=1}^{B}\widehat{\Theta}_i^{*c}\right)^2}, \quad i = 1, \ldots, \ell + p + 3.$$

Determining the optimal number of bootstrap replications B generally depends on empirical testing for each specific scenario. However, certain standard recommendations and methods are referenced. Efron and Tibshirani (1994) suggested that 50 to 100 bootstrap replications are often sufficient for standard error estimation, but for confidence interval construction using percentiles, $B = 1000$ is needed at least to achieve acceptable accuracy. Andrews and Buchinsky (2000) proposed a three-step method to choose B to achieve a desired level of accuracy to calculate standard errors, confidence intervals, p-values, and bias correction. On the other hand, Fassò and Cameletti (2009) assesses the accuracy of bootstrap estimates by examining the length, PI_χ, of the confidence intervals for the true standard error of the parameter estimate:

$$PI_{\chi,\widehat{\Theta}_i} = \widehat{se}^2_{B,i}(\widehat{\Theta}_i)\sqrt{(B-1)}\left(\sqrt{\frac{1}{\chi^2_{\alpha/2;B-1}}} - \sqrt{\frac{1}{\chi^2_{1-\alpha/2;B-1}}}\right)$$

When the bootstrap distribution of $\widehat{\Theta}_i^*$ is approximately normal, the standard normal and percentile confidence intervals will be nearly identical. Additionally, when $B \to \infty$, the bootstrap histogram will become normally shaped. In practice, however, this normal approximation may fail in certain cases, making it critical to check the bootstrap distributions of $\widehat{\Theta}_i^*$ even when B is big enough or to rely on a percentile interval instead. Although various improvements and corrections exist for percentile intervals, the primary objective is to compare these results with those obtained with the observed FIM.

Chapter 3

Methodology

3.1 Parameter estimation and Statistical inference

Building on the research presented in Padilla et al. (2020), we utilized the GEM algorithm to fit a first-order autoregressive space-time model to our database. The model is specified as follows:

$$
\begin{aligned}
Z_t(\mathbf{s}) &= \mu_t(\mathbf{s}) + \varepsilon_t(\mathbf{s}) + \omega_t(\mathbf{s}), \quad \omega_{\mathcal{D}\times\mathcal{T}} \sim \mathcal{GWN}(0, \sigma_\omega^2) \\
\mu_t(\mathbf{s}) &= \beta_0 + \beta_1 \sin\left(\frac{2\pi t}{365.25}\right) + \beta_2 \cos\left(\frac{2\pi t}{365.25}\right) + \beta_3\, x + \beta_4\, y + \beta_5\, alt \\
\varepsilon_t(\mathbf{s}) &= \phi\varepsilon_{t-1}(\mathbf{s}) + \eta_t(\mathbf{s})
\end{aligned}
\tag{3.1}
$$

where the systematic component, $\mu_t(\mathbf{s})$, includes the covariates for each location (see section 3.3 for more details). Also, a harmonic component, β_1 and β_2, is included to capture the seasonal effect. The spatial innovation process, $\eta_{\mathcal{D}\times\mathcal{T}}$ is a stationary Gaussian space-time process with mean zero and covariance function associated with an exponential model defined in (2.3), independent of $\omega_{\mathcal{D}\times\mathcal{T}}$.

The estimation of the variance-covariance matrix for the MLE was computed by inverting the observed FIM using Louis' method (Louis, 1982), as described in equation (2.28). Details of the differentiation of the complete-data log-likelihood can be found in the appendix section VII.2; these results were also validated through numerical methods. To asses the cubic, quartic and complex cross-products expectations necessary for calculating the missing-data FIM, we followed the approach of Tanner (1993) and Turner et al. (1998) using the equation (2.32) and the KS distribution (2.33). For the MC simulations, Wei and Tanner (1990) recommended $M = 5000$, monitoring the convergence of the state process. Some convergence plots are presented in the appendix section VII.4.

To compare our results using Louis' method, we implemented the NMCB and STPB procedures with $B \approx 500, 1000,$ and 1500. To assess the accuracy of the standard error estimates, $\widehat{se}_B$, we used the PI_χ method. Subsequently, we applied the Lilliefors-Kolmogorov-Smirnov Normality Test to the Bootstrap parameter estimations, because if the bootstrap distribution of $\widehat{\Theta}_i^*$ is approximately normal, then the standard normal and percentile intervals should nearly agree (Efron and Tibshirani, 1994). Additionally, Figures VII.5 - VII.8 in appendix section VII.4 present comparative plots that illustrate the behavior of the bootstrap samples.

To construct the $(1-\alpha)100\%$ Confidence Interval, CI($\widehat{\Theta}_i$), $i = 1, \ldots, \ell, p+3$, we used $\alpha = 0.05$. For the estimates obtained via GEM - Louis' method, we employed standard normal CI $\left[\widehat{\Theta}_i \mp z_{1-\alpha/2} \sqrt{\left[I_o^{-1}(\widehat{\Theta})\right]_{ii}}\right]$. Whereas, for the Bootstrap methods, we utilized the Percentile CI (PI) $\left[\widehat{\Theta}^*_{i,\alpha/2};\ \widehat{\Theta}^*_{i,1-\alpha/2}\right]$, since the percentile approach relies on the empirical distribution of $\widehat{\Theta}^*$ without assuming any underlying distribution. Additionally, we calculated the relative standard error (RSE $= \widehat{se}_i/\widehat{\Theta}_i * 100$) and the length of the CI to

facilitate comparison between the results obtained from these two methodologies.

3.2 Simulation data analysis

A simulation study was conducted to evaluate the behavior of the standard errors estimated using Louis' method under different proportions of missing data. Five schemes of data randomly removed (NA) were examined, with proportions of 0, 0.1, 0.2, 0.3, 0.35, and 0.4 to the total amount of space-time data ($n \times T = 10.000$). These schemes were tested in two scenarios, following the simulation study by Padilla et al. (2020).

For each scenario, $n = 25$ spatial locations were utilized, distributed regularly on the unit square over $T = 400$ temporal observations. No covariates were included in the data simulations, and β_0 was fixed at 0. The nugget effect was set to 10% of the total variance of the process, resulting in $\sigma^2_\omega = 0.1$. An autoregressive coefficient $\phi = 0.7$ was chosen, as values too close to 1 can increase the divergence rate of the algorithm. The spatial innovation variance, σ^2_η was calculated in such a way that $\sigma^2_\eta = 0.9 \cdot (1 - \phi^2) = 0.459$. Two scenarios were defined for the reach, α: scenario1 was characterized by strong spatial dependence ($\alpha = 0.8$) and scenario 2 by weaker spatial dependence ($\alpha = 0.4$). The reach was the only parameter tested at different levels across the missing data schemes, given its difficulty in estimation and significant impact on the behavior of the model.

Computation time was recorded using the system.time function for the MC simulation, which is required for reconstructing the latent state necessary to calculate the missing-data FIM, as well as for the parameter estimates across the $B = 500$ bootstrap samples in each scheme of missing data. All simulations were executed on the Department of

Statistics Linux server: Ubuntu 24.04.2 LTS, kernel 6.8.0-56-generic, equipped with an Intel ® Xeon ® E5-2620 v4 CPU (32 cores, 2.10 GHz) and 132 GB of RAM.

3.3 Real data analysis

The autoregressive space-time model was initially evaluated using average daily air temperature data by (Padilla, 2018) utilizing databases from the National Agroclimatic Network accessible through the AGROMET portal (AGROMET) for the Maule, Ñuble, Biobío, and La Araucanía regions. Their research yielded encouraging results that highlight the efficacy of the model in data prediction. The AGROMET portal was created in 2013 through a collaboration between the Chilean Ministry of Agriculture and the National Agroclimatic Network Technical Consortium. This consortium integrates several key organizations: the Instituto de Investigaciones Agropecuarias (INIA), the Asociación de Exportadores de Frutas de Chile A.G. (ASOEX), the Fundación para el Desarrollo Frutícola (FDF), the Centro Cooperativo para el Desarrollo Vitivinícola S.A. via the Red Meteorológica de Vinos de Chile (METEOVID), Dirección Meteorológica de Chile (DMC), Universidad Austral de Chile (UACH) y Centro de Estudios Avanzados en Zonas Áridas (CEAZA). Collectively, these entities provide essential climate information and insights to the agriculture, forestry, and livestock sectors, emphasizing their impact on productive processes. Currently, AGROMET collects data from 424 meteorological stations across Chile, covering variables such as air temperature, precipitation, solar radiation intensity, and relative humidity, among others.

This research analyzed the space-time behavior of average daily air temperature for the same regions previously studied by Padilla et al. (2020), extending the period from January 1, 2015, to December 31, 2022, which is a total number of temporal observations,

$T = 2922$. New EMAs are added to the portal annually; by 2022, $n = 89$ EMAs were operating within the study area. The selection of the study area, which represents a portion of south-central Chile, was justified by its significant economic contribution to the national Gross Domestic Product (GDP), accounting for 41.6% in the agriculture, forestry, and livestock sector in 2021 (CORFO, 2022).

Figure 3.1 illustrates their geographic distribution, and in Table VII.3, detailed information about their UTM coordinates is provided. The stations in the Maule and Biobío regions are mainly concentrated between the intermediate depression and the coast. In contrast, in Ñuble and La Araucanía regions, they are distributed across almost the entire surface of the region. The minimum and maximum distances between EMAs in the Maule Region are 5.07 km and 166.79 km, respectively. In the Ñuble Region, these distances range from 10.22 km to 122.51 km; in the Biobío Region, from 3.78 km to 207.29 km; and in the La Araucanía Region, from 5.97 km to 206.23 km.

Table 3.1 presents the number of EMAs with available data for each region by year. The years 2013 and 2014 were excluded from the analysis due to the limited data availability; including these years would have substantially increased the overall percentage of missing data (nearly 40%). Figure 3.2 illustrates the distribution of missing data by month for each EMA, revealing that most of the missing data can be attributed to the increasing number of EMAs over the years. Likewise, there are short (days) and extended (months or even years) periods with missing data, likely resulting from technical failures with the EMA sensors. Overall, the complete database used in this research contains approximately 29% missing data. Desglose by region indicates a missing data rate of 35.3% for Maule Region, 23.8% for Ñuble Region, 35.7% for the Biobío Region, and 25.2% for La Araucanía

Region. The proportion of missing data in the Maule Region was particularly high because of the increase in the number of stations during the last year, some of which were discarded, reducing the missing data proportion from 35.3% to 31.3%, ultimately leaving $n_{\text{Maule}} = 17$ and $n_{\text{total}} = 85$.

An exploratory data analysis was performed, comparing the 89 time-series, which revealed periods with clearly anomalous behavior in some EMAs, presumably due to sensor faults. These intervals were subsequently excluded. Furthermore, a multicollinearity analysis was conducted to validate the use of covariates to model the deterministic component. The covariates selected for each EMA were related to the UTM coordinates represented as $\mathbf{s} = (x, y)$ along with *alt* indicating the elevation above sea level.

Table 3.1: Changes in the number of meteorological stations (EMAs) across the study area in Chile, 2013 - 2022.

Region	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Maule	6	8	9	9	13	16	16	16	16	21
Ñuble	6	6	7	7	8	11	11	11	11	12
Biobío	5	5	5	7	7	9	11	11	13	14
Araucanía	9	9	19	21	27	37	40	40	41	42
Total	26	28	40	44	55	73	78	78	81	89

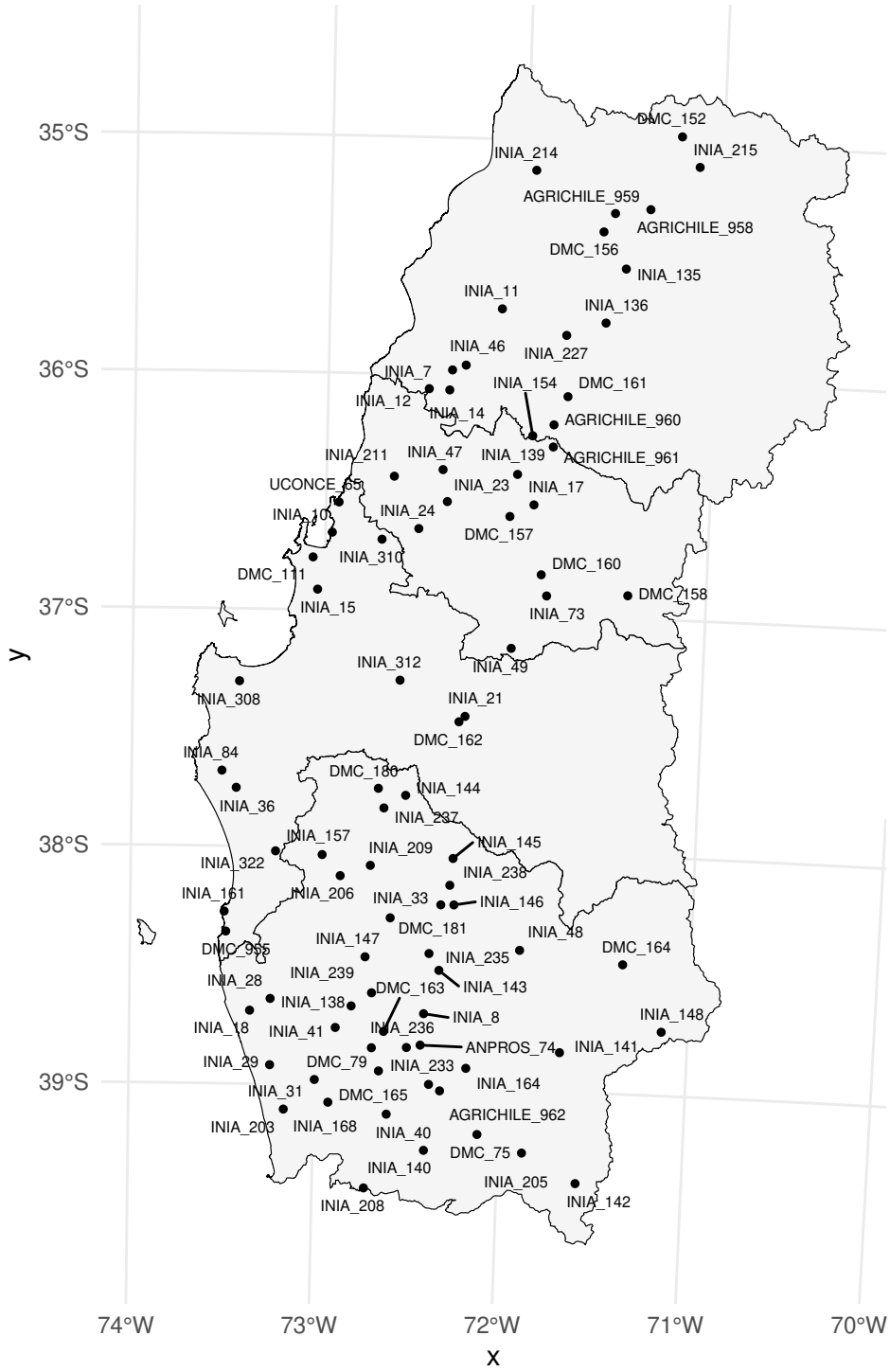

Figure 3.1: Geographical distribution of the 89 meteorological stations (EMAs) operating in 2022 across the Maule, Ñuble, Biobío, and La Araucanía regions.

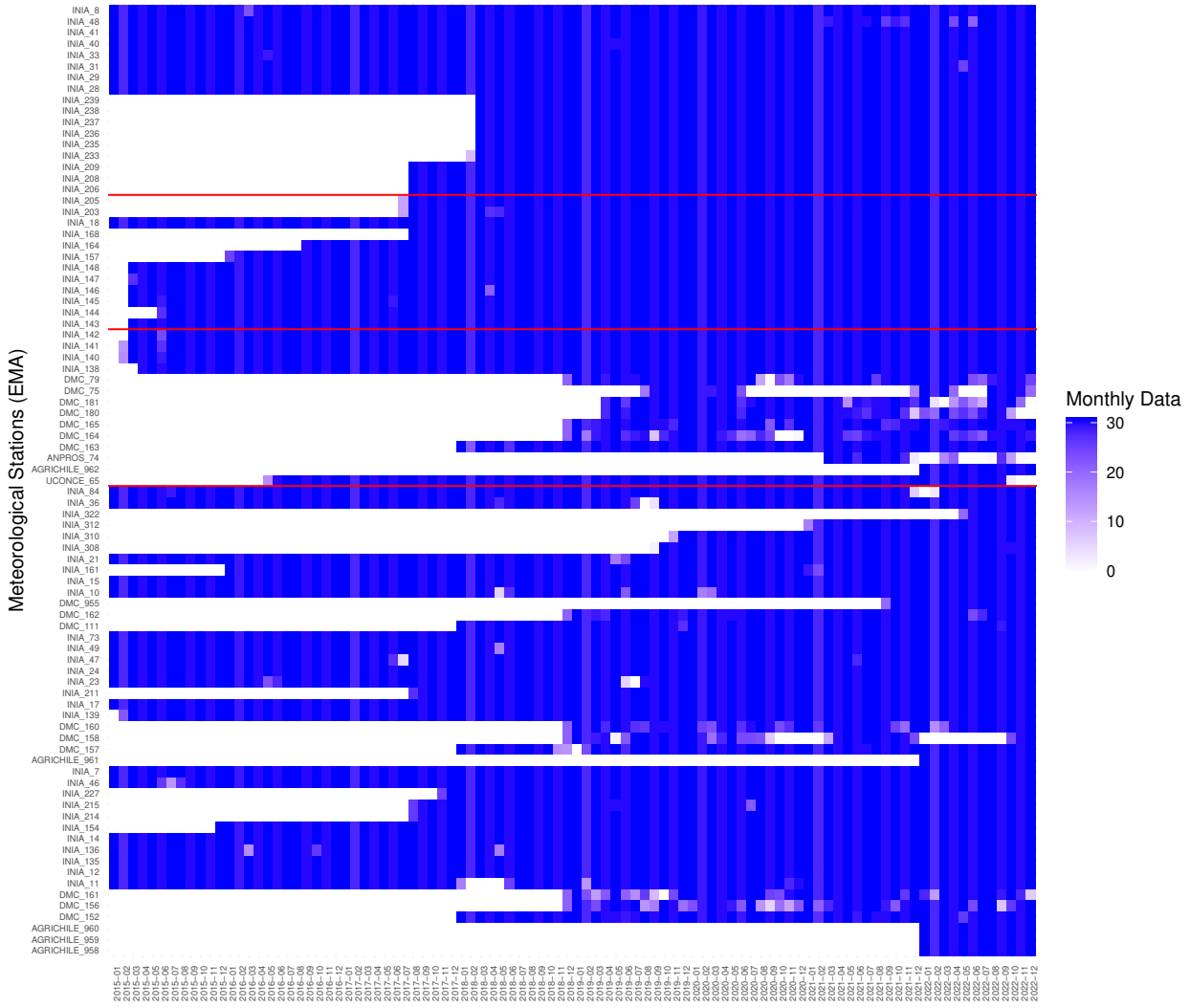

Figure 3.2: Distribution of monthly missing data per meteorological stations (EMAs). The red line indicates the regional division for the study region, from top to bottom: Maule, Ñuble, Biobío, and La Araucanía.

Chapter 4

Results and Discussion

4.1 Simulation data analysis

Across both scenarios, MLEs obtained using the GEM algorithm were practically unaffected by the proportion of missing data (Tables 4.2 and 4.4).In scenario 1 (strong spatial dependence), the difference between the estimated parameter and the true value of the parameter gradually increases when NA exceeds 0.3. Conversely, in scenario 2 (weak spatial dependence), the largest differences appear at $NA = 0.4$, although a small difference is visible in almost all the estimations. These findings are clearly illustrated in Figures 4.1 and 4.2.

Standard errors computed in this study demonstrated remarkable stability and consistency: not all the CI($\widehat{\Theta}_i$) increase in conjunction with the proportion of missing data. Only $\widehat{\alpha}$ and $\widehat{\sigma}^2_\eta$ showed a clear trend across both scenarios when employing Louis' method. The STPB standard errors displayed no systematic pattern. This contrasts with Padilla et al. (2020) (Table 4.1), who reported that empirical standard error systematically increases with a higher proportion of missing data. Although the proportion of missing data directly af-

fects the quality of statistical inference, the literature does not establish an acceptable limit for the amount of missing data; its impact must be reviewed for each combination of model and database (Little and Rubin, 2002). According to McLachlan and Krishnan (2008), when the proportion of missing data is low, the calculation of the complete-data expected FIM remains highly accurate, thereby constraining the variance increase due to the missing values.

In both scenarios, the standard errors calculated using Louis' methods are narrower than those obtained through STPB across all parameters except for the reach. For $\widehat{\alpha}$. In scenario 1, this relationship remains valid up to $NA = 0.3$; for values exceeding $NA = 0.35$, the PI from STPB becomes narrower. In scenario 2, the CI from Louis' method was longer than the PI from STPB across all missing data schemes, with the longer intervals observed for NA above 0.35. It is noteworthy that the CI($\widehat{\alpha}$) in scenario 2 was more precise than in scenario 1 until $NA = 0.2$. These findings align with those reported by De Oliveira and Han (2022), who showed that the observed FIM about the reach decreases as the spatial dependence increases, leading to wider CIs in scenarios characterized by strong spatial dependence. However, for $NA > 0.3$, the length of CI($\widehat{\alpha}$) in scenario 2 is initially 31% greater than in scenario 1 and continues to widen thereafter, showing the impact of missing data on the standard error estimation.

The nugget effect, $\widehat{\sigma}^2_\omega$ displayed values of RSE that ranged from 5.2 % to 9.3 % in scenario 1 and from 5.9 % to 21.4 % in scenario 2. In scenario 1, at $NA = 0.35$, the CI calculated using Louis' method excludes the true value of the parameter. Similarly, in scenario 2, at $NA = 0.4$, both methods fail to include it. When comparing our results with those reported from Padilla et al. (2020) (Table 4.1), we note that in scenarios char-

acterized by strong spatial dependence, specifically at $NA = 0.2$ and $NA = 0.4$, nearly all estimated parameters, with the exception of the nugget, exhibit smaller standard errors when utilizing the Louis' method. In contrast, in scenarios with weak spatial dependence at $NA = 0.2$ and $NA = 0.4$, the nugget and reach obtained smaller empirical standard errors at the same proportion of missing data.

The spatial innovation variance, $\widehat{\sigma}^2_\eta$, exhibited a similar behavior to the nugget in scenario 1. In contrast, in scenario 2, the CI calculated using Louis' methods with $NA = 0.2$ and $NA = 0.35$ encompassed the true value of the parameter, while the STPB failed at $NA = 0.4$. These findings suggest that when spatial dependence is weak, it substantially increases the probability of Type I error for this parameter.

The autoregressive coefficient, $\widehat{\phi}$, was the most accurately estimated parameter, with RSE remaining below 1.2 % in all cases, with the exception of the interval at $NA = 0.4$ in scenario 2, where the intervals did not encompass the true value of the parameter. On the other hand, the intercept, $\widehat{\beta}_0$, showed the highest RSE, primarily because the true value of the parameters is 0. Nevertheless, the length of CI was consistent across both scenarios, regardless of the method employed, and all the intervals included the true value of the parameter.

Reconstructing the latent state through MC simulation using the KS distribution required 2.3 to 2.4 hours per dataset, and the computing time is not influenced by the proportion of missing data. In contrast, computing the bootstrap estimates per scheme of missing data took approximately 10 to 11.4 hours. All the bootstrap samples converged under the GEM algorithm and attained good accuracy in their estimates for $B = 500$ (maximum

$PI_{\chi} = 0.00016$); however, not all the bootstrap estimate sets were normally distributed (Table VII.1).

The results demonstrate that the standard error estimates derived from Louis' method are stable, consistent, and computationally efficient, offering robust estimations for our model in scenarios with up to 30 - 35% missing data, although special attention should be paid to $\widehat{\alpha}$, regardless of strong or weak spatial dependence. While the STPB serves as a valuable alternative for validation,it entails a substantial computational cost, increasing runtime by nearly an order of magnitude.

Table 4.1: Parameter Estimates and Empirical Standard Errors for 1000 simulations of STRF $\mathcal{Z}_{\mathcal{D}\times\mathcal{T}}$), $\widehat{se}_E$, for Scenarios 1 and 2 with $NA = 0$, $NA = 0.2$ and $NA = 0.4$ reported by Padilla et al. (2020)

	Scenario 1					
	$NA = 0$		$NA = 0.2$		$NA = 0.4$	
Paremeter	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\widehat{se}_E$	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\widehat{se}_E$	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\widehat{se}_E$
$\widehat{\beta}_0$	0.00064	0.03560	0.00102	0.03620	-0.00082	0.04628
$\widehat{\sigma}^2_\omega$	0.10003	0.00419	0.09925	0.00511	0.09819	0.00678
$\widehat{\phi}$	0.69956	0.01068	0.69831	0.01154	0.69579	0.01303
$\widehat{\alpha}$	0.80190	0.04182	0.79643	0.04253	0.78652	0.04851
$\widehat{\sigma}^2_\eta$	0.45923	0.01789	0.45969	0.01784	0.46042	0.01921
	Scenario 2					
	$NA = 0$		$NA = 0.2$		$NA = 0.4$	
Paremeter	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\widehat{se}_E$	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\widehat{se}_E$	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\widehat{se}_E$
$\widehat{\beta}_0$	-0.00018	0.03393	-0.00262	0.04097	0.00047	0.04352
$\widehat{\sigma}^2_\omega$	0.09969	0.00613	0.09891	0.00779	0.09722	0.01045
$\widehat{\phi}$	0.69939	0.01040	0.69820	0.01156	0.69609	0.01239
$\widehat{\alpha}$	0.39964	0.01614	0.39830	0.01797	0.39672	0.01997
$\widehat{\sigma}^2_\eta$	0.45927	0.01366	0.46001	0.01560	0.46321	0.01839

Table 4.2: Summary of Parameter Estimates, Standard Errors, Relative Standard Error (RSE), Confidence Intervals (CI or PI), and Interval Lengths for simulate data analysis Scenario 1 ($\beta_0 = 0, \sigma^2_\omega = 0.1, \phi = 0.7, \sigma^2_\eta = 0.459$) with a strong spatial dependence ($\alpha = 0.8$) with different proportion of missing data (NA_i) using GEM-Louis' method and STPB.

Scenario	$NA_1 = 0$											
1	GEM-Louis' method						STPB					
Parameter	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\sqrt{\left[\boldsymbol{I}_o^{-1}(\widehat{\boldsymbol{\Theta}})\right]_{ii}}$	RSE	CI-L	CI-U	CI Length	$\widehat{se}_B$	RSE	PI-L	PI-U	PI length	PI_χ
$\widehat{\beta}_0$	-0.00174	0.01007	-578.884	-0.02148	0.01800	0.03948	0.02630	4781.818	-0.05760	0.05677	0.11437	0.00009
$\widehat{\sigma}^2_\omega$	0.09724	0.00502	5.165	0.08740	0.10708	0.01968	0.00832	8.542	0.07945	0.11476	0.03531	0.00001
$\widehat{\phi}$	0.69898	0.00626	0.896	0.68671	0.71125	0.02454	0.00962	1.376	0.67958	0.71823	0.03865	0.00001
$\widehat{\alpha}$	0.79691	0.02327	2.921	0.75129	0.84253	0.09124	0.02776	3.485	0.74295	0.85113	0.10818	0.00010
$\widehat{\sigma}^2_\eta$	0.46078	0.00759	1.648	0.44590	0.47566	0.02976	0.01442	3.132	0.43442	0.48706	0.05264	0.00003
Scenario	$NA_2 = 0.1$											
1	GEM-Louis' method						STPB					
Parameter	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\sqrt{\left[\boldsymbol{I}_o^{-1}(\widehat{\boldsymbol{\Theta}})\right]_{ii}}$	RSE	CI-L	CI-U	CI Length	$\widehat{se}_B$	RSE	PI-L	PI-U	PI length	PI_χ
$\widehat{\beta}_0$	-0.00021	0.01040	-4951.349	-0.02059	0.02017	0.04076	0.02638	1055.200	-0.05760	0.05677	0.11437	0.00009
$\widehat{\sigma}^2_\omega$	0.09996	0.00523	5.230	0.08971	0.11021	0.02050	0.00923	9.250	0.07945	0.11476	0.03531	0.00001
$\widehat{\phi}$	0.70047	0.00633	0.903	0.68807	0.71287	0.02480	0.01055	1.506	0.67958	0.71823	0.03865	0.00001
$\widehat{\alpha}$	0.79982	0.02530	3.163	0.75023	0.84941	0.09918	0.03047	3.808	0.74295	0.85113	0.10818	0.00012
$\widehat{\sigma}^2_\eta$	0.45669	0.00808	1.770	0.44084	0.47254	0.03170	0.01530	3.350	0.43442	0.48706	0.05264	0.00003
Scenario	$NA_3 = 0.2$											
1	GEM-Louis' method						STPB					
Parameter	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\sqrt{\left[\boldsymbol{I}_o^{-1}(\widehat{\boldsymbol{\Theta}})\right]_{ii}}$	RSE	CI-L	CI-U	CI Length	$\widehat{se}_B$	RSE	PI-L	PI-U	PI length	PI_χ
$\widehat{\beta}_0$	-0.00119	0.01476	-1240.506	-0.03012	0.02774	0.05786	0.03096	-1433.333	-0.06548	0.06763	0.13311	0.00012
$\widehat{\sigma}^2_\omega$	0.09861	0.00939	9.525	0.08020	0.11702	0.03682	0.01003	10.273	0.07811	0.11951	0.04140	0.00001
$\widehat{\phi}$	0.69972	0.00645	0.922	0.68708	0.71236	0.02528	0.01126	1.613	0.67356	0.72020	0.04664	0.00002
$\widehat{\alpha}$	0.81656	0.02989	3.660	0.75798	0.87514	0.11716	0.03200	3.933	0.75261	0.87720	0.12459	0.00013
$\widehat{\sigma}^2_\eta$	0.45774	0.00940	2.054	0.43931	0.47617	0.03686	0.01639	3.571	0.42311	0.49060	0.06749	0.00003

Table 4.3: (Continuation) Summary of Parameter Estimates, Standard Errors, Relative Standard Error (RSE), Confidence Intervals (CI or PI), and Interval Lengths for simulate data analysis Scenario 1 ($\beta_0 = 0, \sigma^2_\omega = 0.1, \phi = 0.7, \sigma^2_\eta = 0.459$) with a strong spatial dependence ($\alpha = 0.8$) with different proportion of missing data (NA_i) using GEM-Louis' method and STPB.

Scenario	$NA_4 = 0.3$											
1	GEM-Louis' method						STPB					
Parameter	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\sqrt{\left[\boldsymbol{I}_o^{-1}(\widehat{\boldsymbol{\Theta}})\right]_{ii}}$	RSE	CI-L	CI-U	CI Length	$\widehat{se}_B$	RSE	PI-L	PI-U	PI length	PI_χ
$\widehat{\beta}_0$	-0.01000	0.01114	-111.426	-0.03184	0.01184	0.04368	0.02910	-280.077	-0.07103	0.05044	0.12147	0.00011
$\widehat{\sigma}^2_\omega$	0.09059	0.00700	7.728	0.07687	0.10431	0.02744	0.01039	11.669	0.07091	0.11308	0.04217	0.00001
$\widehat{\phi}$	0.69652	0.00652	0.936	0.68375	0.70929	0.02554	0.01139	1.639	0.67269	0.71734	0.04465	0.00002
$\widehat{\alpha}$	0.77318	0.03253	4.207	0.70943	0.83693	0.12750	0.03179	4.122	0.71048	0.83805	0.12757	0.00013
$\widehat{\sigma}^2_\eta$	0.46765	0.00989	2.114	0.44827	0.48703	0.03876	0.01637	3.485	0.43757	0.50203	0.06446	0.00003
Scenario	$NA_5 = 0.35$											
1	GEM-Louis' method						STPB					
Parameter	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\sqrt{\left[\boldsymbol{I}_o^{-1}(\widehat{\boldsymbol{\Theta}})\right]_{ii}}$	RSE	CI-L	CI-U	CI Length	$\widehat{se}_B$	RSE	PI-L	PI-U	PI length	PI_χ
$\widehat{\beta}_0$	-0.00884	0.01145	-129.570	-0.03129	0.01361	0.04490	0.03269	-402.586	-0.07607	0.05723	0.13330	0.00013
$\widehat{\sigma}^2_\omega$	0.12335	0.00762	6.178	0.10841	0.13829	0.02988	0.01329	10.851	0.09620	0.14889	0.05269	0.00002
$\widehat{\phi}$	0.71128	0.00671	0.944	0.69812	0.72444	0.02632	0.01211	1.705	0.68734	0.73340	0.04606	0.00002
$\widehat{\alpha}$	0.80059	0.04763	5.950	0.70723	0.89395	0.18672	0.03563	4.447	0.73410	0.87262	0.13852	0.00016
$\widehat{\sigma}^2_\eta$	0.43095	0.01245	2.889	0.40655	0.45535	0.04880	0.01952	4.512	0.39487	0.47097	0.07610	0.00005
Scenario	$NA_6 = 0.4$											
1	GEM-Louis' method						STPB					
Parameter	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\sqrt{\left[\boldsymbol{I}_o^{-1}(\widehat{\boldsymbol{\Theta}})\right]_{ii}}$	RSE	CI-L	CI-U	CI Length	$\widehat{se}_B$	RSE	PI-L	PI-U	PI length	PI_χ
$\widehat{\beta}_0$	-0.00651	0.01337	-205.445	-0.03272	0.01970	0.05242	0.03018	-463.594	-0.06457	0.05495	0.11952	0.00011
$\widehat{\sigma}^2_\omega$	0.08898	0.00825	9.273	0.07281	0.10515	0.03234	0.01203	13.708	0.06508	0.11583	0.05075	0.00002
$\widehat{\phi}$	0.69179	0.00689	0.996	0.67829	0.70529	0.02700	0.01059	1.534	0.66945	0.71181	0.04236	0.00001
$\widehat{\alpha}$	0.76717	0.04859	6.334	0.67194	0.86240	0.19046	0.03123	4.083	0.70728	0.82751	0.12023	0.00012
$\widehat{\sigma}^2_\eta$	0.47444	0.01373	2.895	0.44752	0.50136	0.05384	0.01823	3.823	0.44195	0.51286	0.07091	0.00004

Table 4.4: Summary of Parameter Estimates, Standard Errors, Relative Standard Error (RSE), Confidence Intervals (CI or PI), and Interval Lengths for simulate data analysis Scenario 2 ($\beta_0 = 0, \sigma^2_\omega = 0.1, \phi = 0.7, \sigma^2_\eta = 0.459$) with a weaker spatial dependence ($\alpha = 0.4$) with different proportion of missing data (NA_i) using GEM-Louis' method and STPB.

Scenario	$NA_1 = 0$											
2	GEM-Louis' method						STPB					
Parameter	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\sqrt{\left[\boldsymbol{I}_o^{-1}(\widehat{\boldsymbol{\Theta}})\right]_{ii}}$	RSE	CI-L	CI-U	CI Length	$\widehat{se}_B$	RSE	PI-L	PI-U	PI length	PI_χ
$\widehat{\beta}_0$	-0.00796	0.00794	-99.733	-0.02352	0.00759	0.03111	0.02159	-289.410	-0.05346	0.03797	0.09143	0.00006
$\widehat{\sigma}^2_\omega$	0.09626	0.00567	5.890	0.08515	0.10738	0.02223	0.00927	9.639	0.07772	0.11687	0.03915	0.00001
$\widehat{\phi}$	0.69337	0.00608	0.877	0.68145	0.70530	0.02385	0.00977	1.409	0.67487	0.71311	0.03824	0.00001
$\widehat{\alpha}$	0.40648	0.01962	4.827	0.36803	0.44493	0.07690	0.01995	4.928	0.36472	0.44148	0.07676	0.00005
$\widehat{\sigma}^2_\eta$	0.46806	0.00409	0.874	0.46005	0.47607	0.01602	0.01520	3.251	0.43494	0.49570	0.06076	0.00003
Scenario	$NA_2 = 0.1$											
2	GEM-Louis' method						STPB					
Parameter	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\sqrt{\left[\boldsymbol{I}_o^{-1}(\widehat{\boldsymbol{\Theta}})\right]_{ii}}$	RSE	CI-L	CI-U	CI Length	$\widehat{se}_B$	RSE	PI-L	PI-U	PI length	PI_χ
$\widehat{\beta}_0$	-0.00986	0.00795	-80.584	-0.02544	0.00572	0.03116	0.02317	-216.138	-0.05735	0.03743	0.09478	0.00007
$\widehat{\sigma}^2_\omega$	0.09357	0.00929	9.924	0.07537	0.11177	0.03640	0.00980	10.388	0.07594	0.11615	0.04021	0.00001
$\widehat{\phi}$	0.69486	0.00641	0.923	0.68229	0.70743	0.02514	0.01032	1.484	0.67523	0.71522	0.03999	0.00001
$\widehat{\alpha}$	0.40404	0.02160	5.347	0.36170	0.44638	0.08468	0.02150	5.325	0.36371	0.44596	0.08225	0.00006
$\widehat{\sigma}^2_\eta$	0.47017	0.00474	1.009	0.46087	0.47946	0.01859	0.01515	3.232	0.43643	0.49653	0.06010	0.00003
Scenario	$NA_3 = 0.2$											
2	GEM-Louis' method						STPB					
Parameter	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\sqrt{\left[\boldsymbol{I}_o^{-1}(\widehat{\boldsymbol{\Theta}})\right]_{ii}}$	RSE	CI-L	CI-U	CI Length	$\widehat{se}_B$	RSE	PI-L	PI-U	PI length	PI_χ
$\widehat{\beta}_0$	-0.00241	0.00812	-337.079	-0.01833	0.01351	0.03184	0.02449	-1309.626	-0.04868	0.04664	0.09532	0.00007
$\widehat{\sigma}^2_\omega$	0.10010	0.01158	11.568	0.07740	0.12279	0.04539	0.01103	11.021	0.07630	0.12276	0.04646	0.00002
$\widehat{\phi}$	0.69332	0.00577	0.833	0.68200	0.70463	0.02263	0.01103	1.591	0.67167	0.71448	0.04281	0.00002
$\widehat{\alpha}$	0.40917	0.02755	6.732	0.35518	0.46316	0.10798	0.02361	5.772	0.36235	0.45172	0.08937	0.00007
$\widehat{\sigma}^2_\eta$	0.46817	0.00513	1.095	0.45812	0.47822	0.02010	0.01689	3.605	0.43383	0.50019	0.06636	0.00004

Table 4.5: (Continuation) Summary of Parameter Estimates, Standard Errors, Relative Standard Error (RSE), Confidence Intervals (CI or PI), and Interval Lengths for simulated data analysis Scenario 2 ($\beta_0 = 0, \sigma^2_\omega = 0.1, \phi = 0.7, \sigma^2_\eta = 0.459$) with a weaker spatial dependence ($\alpha = 0.4$) with different proportion of missing data (NA_i) using GEM-Louis' method and STPB.

Scenario	$NA_4 = 0.3$											
2	GEM-Louis' method						STPB					
Parameter	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\sqrt{\left[\boldsymbol{I}_o^{-1}(\widehat{\boldsymbol{\Theta}})\right]_{ii}}$	RSE	CI-L	CI-U	CI Length	$\widehat{se}_B$	RSE	PI-L	PI-U	PI length	PI_χ
$\widehat{\beta}_0$	-0.00593	0.00910	-153.453	-0.02376	0.01190	0.03566	0.02443	-552.715	-0.05052	0.04655	0.09707	0.00007
$\widehat{\sigma}^2_\omega$	0.09999	0.00718	7.177	0.08593	0.11406	0.02813	0.01265	12.836	0.07136	0.12534	0.05398	0.00002
$\widehat{\phi}$	0.69236	0.00658	0.951	0.67945	0.70526	0.02581	0.01156	1.673	0.66751	0.71304	0.04553	0.00002
$\widehat{\alpha}$	0.42319	0.04255	10.055	0.33979	0.50659	0.16680	0.02382	5.622	0.37652	0.46769	0.09117	0.00007
$\widehat{\sigma}^2_\eta$	0.47628	0.00673	1.412	0.46310	0.48946	0.02636	0.01894	3.960	0.43613	0.51564	0.07951	0.00004
Scenario	$NA_5 = 0.35$											
2	GEM-Louis' method						STPB					
Parameter	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\sqrt{\left[\boldsymbol{I}_o^{-1}(\widehat{\boldsymbol{\Theta}})\right]_{ii}}$	RSE	CI-L	CI-U	CI Length	$\widehat{se}_B$	RSE	PI-L	PI-U	PI length	PI_χ
$\widehat{\beta}_0$	-0.01026	0.009263	-90.260	-0.02842	0.00789	0.03631	0.02456	-249.340	-0.05968	0.03894	0.09862	0.00008
$\widehat{\sigma}^2_\omega$	0.09421	0.01070	11.361	0.07323	0.11519	0.04196	0.01243	13.328	0.06944	0.12431	0.05487	0.00002
$\widehat{\phi}$	0.70124	0.00496	0.707	0.69153	0.71095	0.01942	0.01086	1.550	0.67971	0.72140	0.04169	0.00001
$\widehat{\alpha}$	0.39459	0.11595	29.385	0.16733	0.62186	0.45453	0.02742	6.998	0.34062	0.44601	0.10539	0.00009
$\widehat{\sigma}^2_\eta$	0.46193	0.01072	2.321	0.44091	0.48294	0.04203	0.01818	3.935	0.42310	0.49437	0.07127	0.00004
Scenario	$NA_6 = 0.4$											
2	GEM-Louis' method						STPB					
Parameter	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\sqrt{\left[\boldsymbol{I}_o^{-1}(\widehat{\boldsymbol{\Theta}})\right]_{ii}}$	RSE	CI-L	CI-U	CI Length	$\widehat{se}_B$	RSE	PI-L	PI-U	PI length	PI_χ
$\widehat{\beta}_0$	-0.02024	0.01297	-64.070	-0.04565	0.00518	0.05083	0.02231	-114.235	-0.06539	0.02833	0.09372	0.00006
$\widehat{\sigma}^2_\omega$	0.05778	0.01234	21.362	0.03359	0.08198	0.04839	0.00886	14.949	0.04657	0.08966	0.04309	0.00001
$\widehat{\phi}$	0.67431	0.00753	1.117	0.65955	0.68908	0.02953	0.01053	1.561	0.65451	0.69582	0.04131	0.00001
$\widehat{\alpha}$	0.39001	0.09605	24.627	0.20176	0.57826	0.37650	0.02594	6.704	0.33448	0.44209	0.10761	0.00008
$\widehat{\sigma}^2_\eta$	0.51619	0.01115	2.159	0.49434	0.53804	0.04370	0.01469	2.851	0.48068	0.53973	0.05905	0.00003

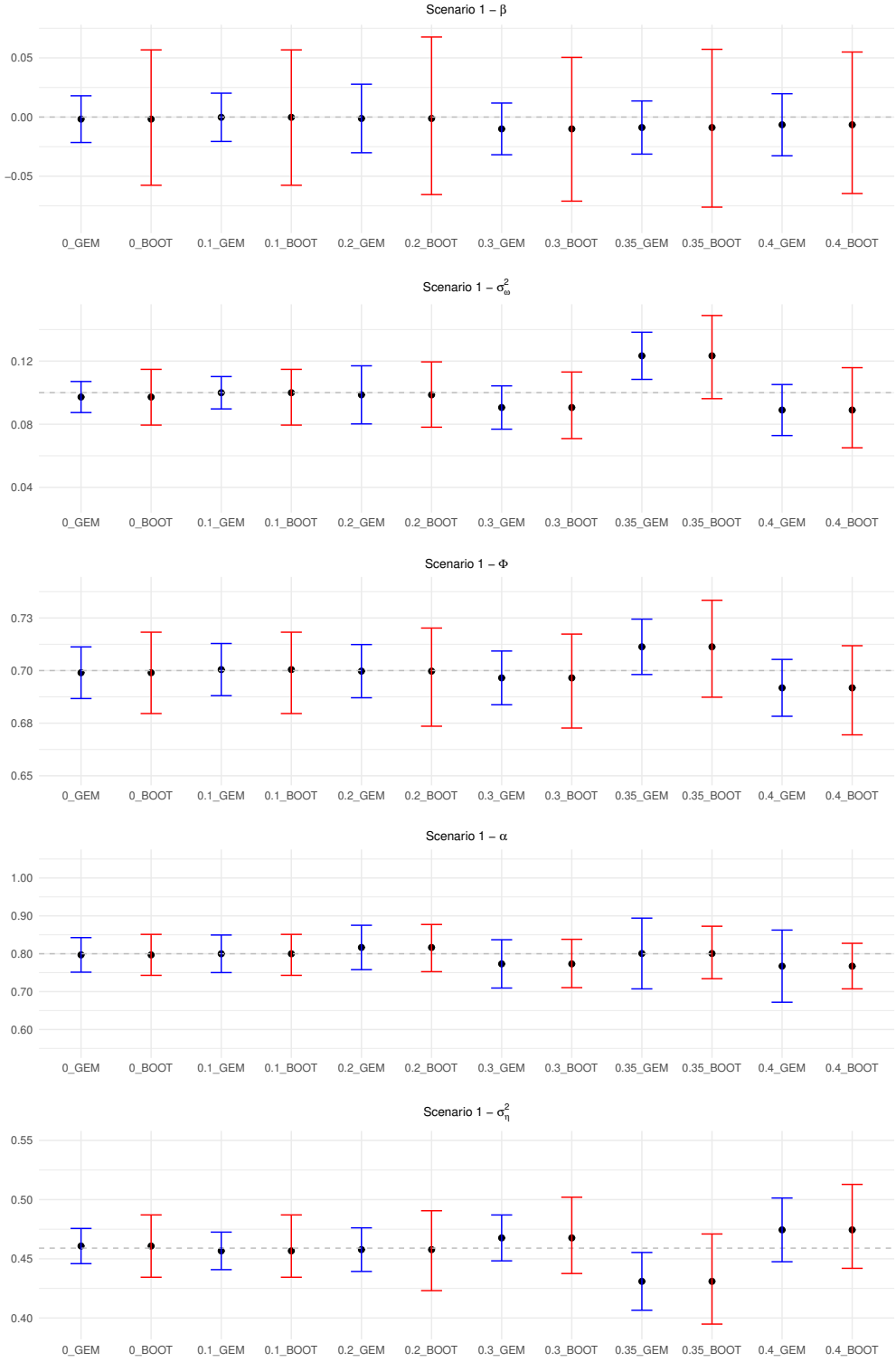

Figure 4.1: Comparison of Confidence Intervals for the parameter estimates from the simulated data in Scenario 1 (strong spatial dependence, $\alpha = 0.8$). The grey dashed line denotes the true parameter values used in the simulation ($\beta_0 = 0, \sigma^2_\omega = 0.1, \phi = 0.7, \sigma^2_\eta = 0.459$). Confidence intervals derived using Louis' method are shown in blue (GEM), while those obtained with the STPB are depicted in red (BOOT). The x-axis indicates the different missing data schemes.

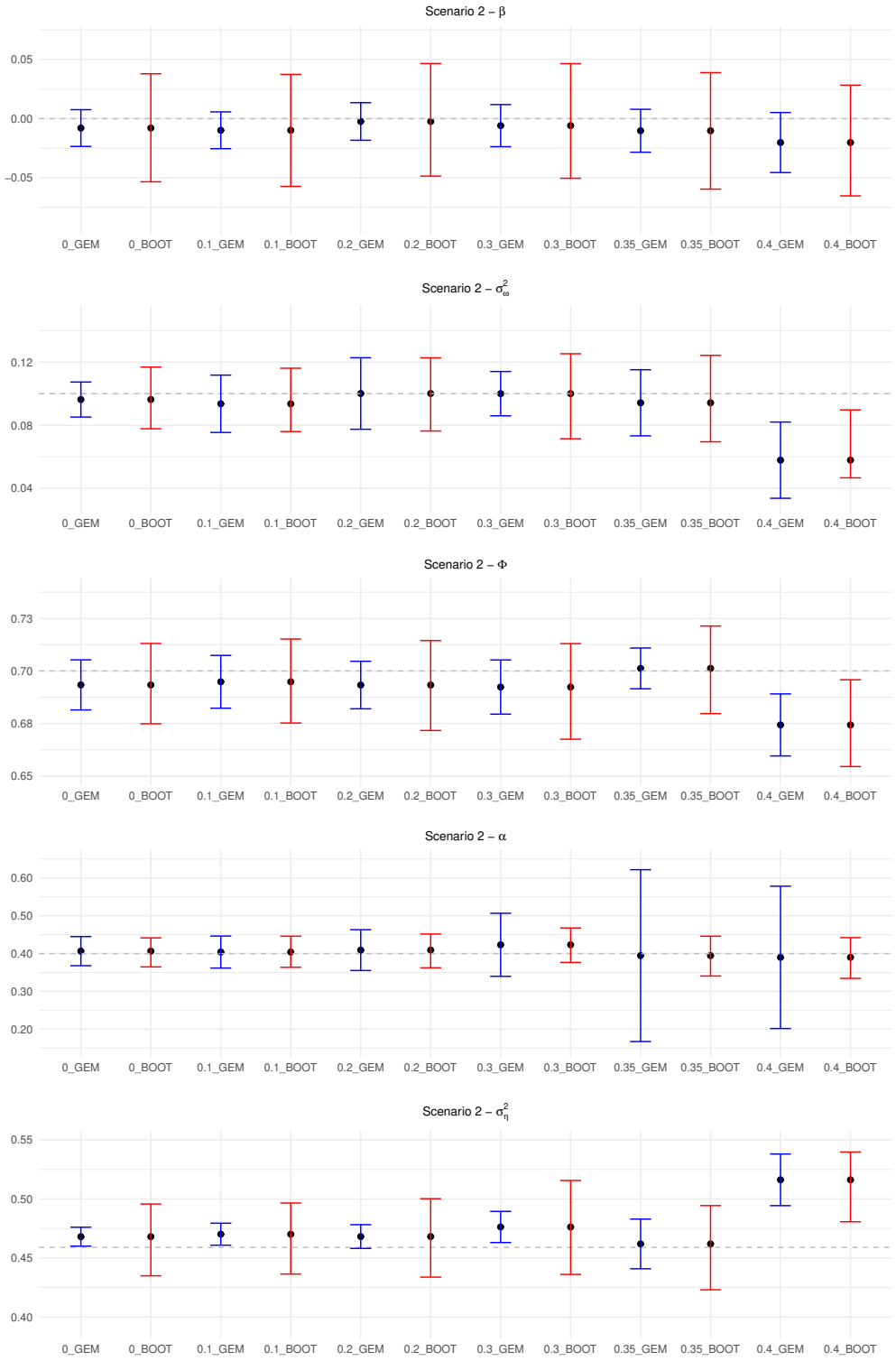

Figure 4.2: Comparison of Confidence Intervals for the parameter estimates from the simulated data in Scenario 2 (weak spatial dependence, $\alpha = 0.4$). The grey dashed line denotes the true parameter values used in the simulation ($\beta_0 = 0, \sigma^2_\omega = 0.1, \phi = 0.7, \sigma^2_\eta = 0.459$). Confidence intervals derived using Louis' method are shown in blue (GEM), while those obtained with the STPB are depicted in red (BOOT). The x-axis indicates the different missing data schemes.

4.2 Exploratory real data analysis

The relations between the average daily air temperature versus the covariates: x (longitude), y (latitude), and *alt* (elevation) are illustrated in Figure 4.3. Usually, latitude (Subfigure 4.3 b) and elevation (Subfigure 4.3 c) exhibit a more pronounced influence on temperature: higher temperatures are recorded farther north, while lower temperatures are recorded at higher elevations. Although one might anticipate longitude (Subfigure 4.3 a) to behave similarly to elevation, the a-priori correlation analyses and multicollinearity analyses show neither strong correlation nor strong multicollinearity, indicating that none of these covariates need to be eliminated (Table 4.6).

Table 4.6: Variance Inflation Factors (VIF) and Spearman correlation coefficients among the covariates considered in the real data analysis

Covariate	$\sin\left(\frac{2\pi t}{365.25}\right)$	$\cos\left(\frac{2\pi t}{365.25}\right)$	x	y	*alt*
VIF - Entire polygon	1.00006	1.00006	2.03346	1.51331	1.70132

	Spearman - Entire polygon		
Covariate	x	y	*alt*
x	1	-	-
y	0.38	1	-
alt	0.58	-0.06	1

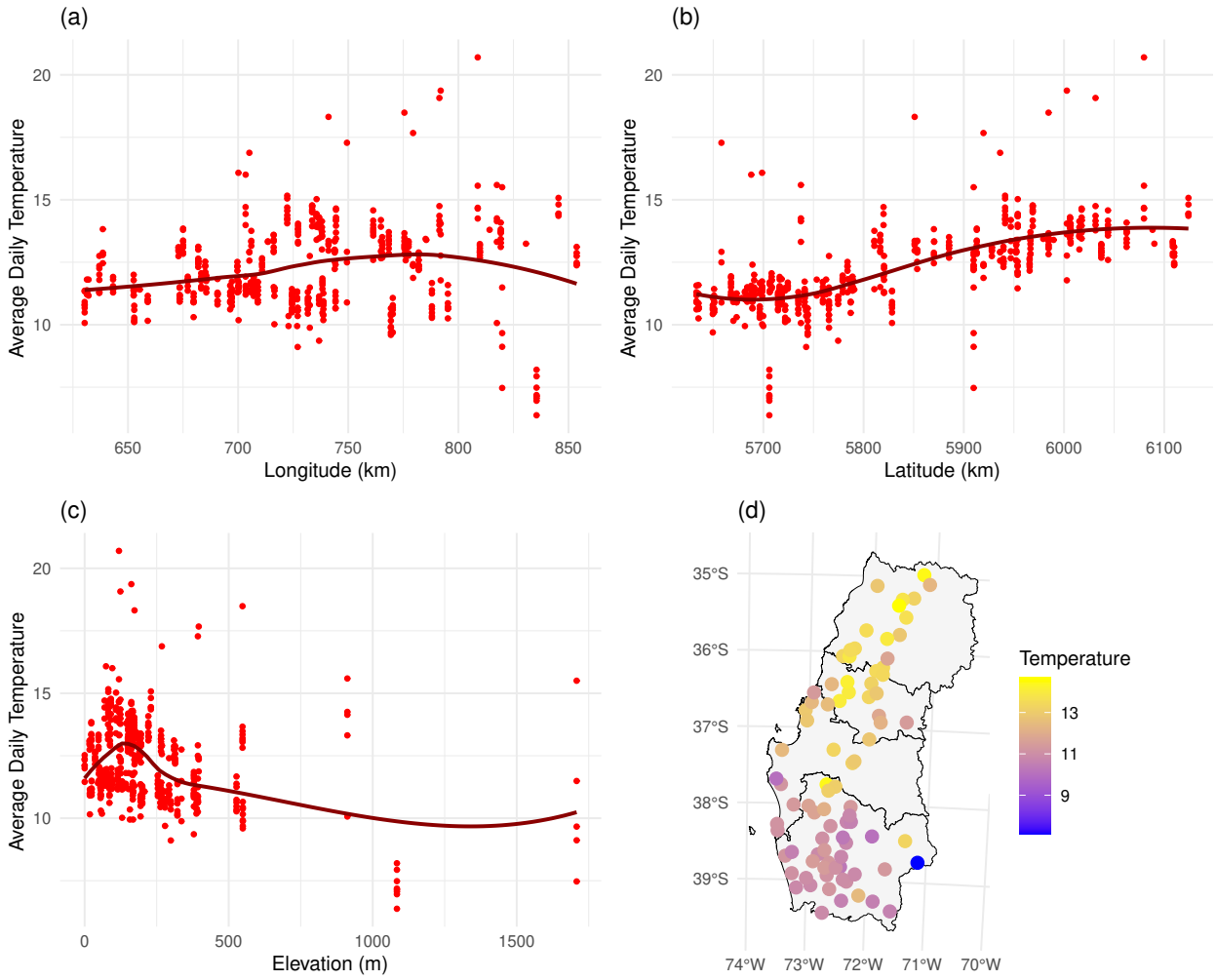

Figure 4.3: Temperature versus covariates with polynomial regression fit curves: (a-b) spatial coordinates x (longitude) and y (latitude) and (c) elevation. (d) Spatial distribution of average temperatures in 2022

4.3 Real data analysis

To fit the model, we used a dataset with $n \times T = 85 \cdot 2922 = 248.370$ space-time observations, which represents an extensive volume of data. The computational requirements for data simulations at this scale are highly intensive, combined with the spatial irregularity in the distribution of the EMAs across regions (Figure 3.1) and the variability in temperature ranges (Figure 4.3), motivated us to conduct the analysis separately for each region. Moreover, preliminary analyses showed that generating bootstrap samples for two or more regions simultaneously failed to reproduce adequately the original data structure, further justifying a regional-level approach. Two bootstrap methodologies were implemented to compare with Louis' method: a Parametric bootstrap (STPB) and a Non-Parametric one (NMCB). The NMCB approach generated bootstrap samples with erratic and unrealistic behavior, resulting in values that fell outside the range of the observed temperatures in the original data. Reported a similar situatioin Fernández-Casal et al. (2024), who encountered problems using a Cholesky decomposition-based resampling algorithm (as used in NMCB), resulting in samples that fail to reflect the observed data patterns. Consequently, this unconditional bootstrap algorithm is unsuitable for simulating data that exhibit spatial dependence (or are conditioned on a latent state). Conversely, STPB produced bootstrap samples consistent with the empirical dataset (see Figures VII.6, VII.5, VII.7, VII.8).

The initial number of bootstrap samples, $B \approx 500$, followed the recommendation of Fassò and Cameletti (2009), who attained good estimation accuracy with $PI_{\chi} \leq 0.15$. However, this amount of bootstrap samples is mainly sufficient for simulation scenarios; after evaluating the precision for each estimated parameter, in each region, it became evident that $B \approx 500$ was insufficient to compute reliable CIs for $\widehat{\alpha}$ (except in the La Araucanía Region); for most other parameter estimates, this value of B was adequate. Therefore, the

number of bootstrap samples was increased to around 1500 in line with Efron and Tibshirani (1994). The final number of valid bootstrap samples varies by region, as some bootstrap samples were discarded due to non-convergence of the GEM algorithm. Fassò and Cameletti (2009) attributed these convergence failures to the Newton-Raphson algorithm used to update the reach parameter.

Detailed parameter estimates, standard errors, and intervals by region are provided in Tables 4.7 - 4.10. The analysis of the confidence intervals obtained using GEM-Louis' method indicated that all estimated parameters were statistically significant at the $\alpha = 0.05$ (Figure 4.4), except for one: the coefficient associated with the y coordinate in the Ñuble Region, which was not statistically significant. The STPB yielded the same result.

Concerning all covariates, the harmonic component, $\widehat{\beta}_1$ and $\widehat{\beta}_2$, as well as the elevation coefficient, $\widehat{\beta}_5$, were estimated with the highest precision across regions, with the exception of $\widehat{\beta}_5$ in the Maule region. Using Louis' method, their RSE remained below 2.66 % and under 5 % with STPB. While harmonic coefficients were quite similar among regions, $\widehat{\beta}_2$ exhibited a slightly weaker effect in the Biobío Region. $\widehat{\beta}_5$ ranged from -0.002 to -0.004, indicating that for every 250 to 500 m increase in elevation above sea level, average temperature decreases by approximately $1°C$. While in the Maule Region $\widehat{\beta}_5 = -0.0008$, the same decrease is achieved at approximately 1250 m.

The coefficients $\widehat{\beta}_3$ and $\widehat{\beta}_4$ associated with the spatial covariates $\mathbf{s} = (x, y)$, respectively, varied noticeably across regions, reflecting the geographic distribution of the EMAs. From the exploratory analysis, we identified two expected general spatial patterns: (i) temperatures increase when moving to the north, and (ii) temperatures decrease when transition-

ing from the coast toward the Andes, reaching their minimum at higher elevations. While these patterns are clearly visible when considering all four regions simultaneously (Subfigure 4.3 d), the regional level subset often deviates from this general pattern due to its limited spatial coverage. For example, temperatures in the La Araucanía Region are typically cooler than those in the Maule Region. In particular, the coefficient associated with the y coordinate in the Ñuble Region was not statistically significant. This result may be explained by the relatively short north-south extension of the region (approximately 160 km), which limits the capture of latitudinal temperature variation described in (i). This coefficient also achieved the highest RSE in nearly all regions, reaching 99% with Louis' method and 112% with STPB in the Ñuble Region. For the x coordinate coefficient, Maule and Ñuble regions exhibited the expected behavior mentioned in (ii), probably due to their absence of coastal stations. Along the Chilean coastline, the Humboldt Current tends to lower temperatures, which may explain the opposite behavior observed in the Biobío and La Araucania regions, where several EMAs are located near the shoreline. This is partially illustrated in Subfigure 4.3 c, where the polynomial fit curve suggests lower temperatures at sea level (elevation 0 m). In the Biobío Region, most EMAs are located close to the coastline, with only three are in the intermediate depression, which could account for the different sign in $\widehat{\beta}_3$. Meanwhile, the La Araucanía Region exhibited highly variable behavior, with several EMAs registering unexpected values, most likely due to the topography of the region, which gives rise to a variety of microclimate zones.

The estimated propagation factor was high, $\widehat{\phi} \approx 0.73$ to 0.85, indicating that the average daily temperature has strong temporal correlation with the average daily temperature of the previous day. As in the simulation study, $\widehat{\phi}$ was the most precisely estimated parameter, with RSE under 0.5% using Louis' method and 0.65 % with STPB. The reach parameter,

$\widehat{\alpha}$, suggested a process with strong spatial dependence, because the CI in each region included the maximum distance between EMAs. Despite achieving one of the highest RSEs in almost all regions, RSE remained below 4%. The variance of the spatial process, $\widehat{\sigma}^2_\eta$, fluctuated between 2.1 and 3.6, indicating high spatial variability. A strong correlation was observed between $\widehat{\alpha}$ was and $\widehat{\sigma}^2_\eta$ with $\rho_{\widehat{\alpha},\widehat{\sigma}^2_\eta} \approx 0.99$ across all regions. In contrast, the nugget effect was low, $\widehat{\sigma}^2_\omega \approx 0.06$ to 0.08, reaching its highest value in the Biobío Region ($\widehat{\sigma}^2_\omega \approx 0.13$), meaning that measurement error explained no more than 1.3% of the total variability. The RSE $\widehat{\sigma}^2_\omega$ and $\widehat{\sigma}^2_\eta$ were similar and indicated precise estimations in most cases. However, an exception was observed in the Ñuble region where $\widehat{\sigma}^2_\omega$ exhibited the highest RSE (21%). These results are consistent with those reported by Padilla et al. (2020).

Analyzing the distribution of the bootstrap samples (via normality tests, QQ-plots, and histograms in Table VII.1) revealed that only La Araucanía Region presented a few outlier estimates, which were removed. Outlier estimates may relate to its largest temperature range $[-12.2; 35.5]$ ($\Delta T = 47.7°C$). Furthermore, their estimated values of $\widehat{\alpha}$ and $\widehat{\sigma}^2_\eta$ was the highest, impacting the covariance matrix used to simulate a Gaussian random vector, $\widetilde{\mathbf{V}}_t$, in step 2 of the STPB algorithm.

In the La Araucanía Region ($n = 42$), with just $B = 596$ bootstrap samples, yielded $PI_{\chi,\widehat{\alpha}} = 0.122$, despite being the region to have the highest rate of discarded samples due to non-convergence of the GEM algorithm, which requires generating a greater number of bootstrap samples, increasing substantially STPB computational time. Although McLachlan and Krishnan (2008) and Tanner (1993) note that EM algorithm convergence speed is related to the proportion of missing data, our results suggest it was mainly influenced by

sample size, n, and data variability. Meanwhile, Maule region ($n = 17$), Ñuble Region ($n = 12$), and Biobío Region ($n = 14$) required over 1500 bootstrap samples to achieve less precise values: $PI_{\chi,\widehat{\alpha}} = 2.62$, $PI_{\chi,\widehat{\alpha}} = 1.31$, and $PI_{\chi,\widehat{\alpha}} = 2.03$, respectively. Bootstrap accuracy depends primarily on the information contained in the original data, which depends mainly on n. Even if B is increased and the resampling-related error is reduced, the final precision of the estimate depends on the original sample size (Efron and Tibshirani, 1994).

Standard errors obtained via STPB fall into ranges similar to those found with Louis' method for all the regions. The CI length was examined to compare the two approaches, showing non-uniform behavior for all estimated parameters. For $\widehat{\beta}_1$, $\widehat{\beta}_2$, $\widehat{\sigma}^2_\eta$ and $\widehat{\alpha}$ STPB yielded shorter intervals than Louis' method. In the La Araucanía Region, STPB consistently produces shorter intervals than Louis' method (except for $\widehat{\phi}$), further reflecting the effect of sample size on interval precision. Louis' method yielded a shorter CI for $\widehat{\beta}_5$, $\widehat{\sigma}^2_\omega$ and $\widehat{\phi}$. Significantly, for $\widehat{\phi}$ in all regions, Louis' method led to shorter CI lengths than STPB. Excluding La Araucanía Region, we see that for $\widehat{\beta}_5$, Louis' method provides shorter CIs than STPB, while for $\widehat{\sigma}^2_\omega$, the CI lengths barely differ between the two methods.

Each methodology has particular strengths and weaknesses. There is no universally optimal approach; the decision on which methodology to use should be guided by the specific characteristics of each dataset and the goals of the study. STPB is a valuable option for quantifying uncertainty in space-time models where a well-defined parametric framework exists, even given potential convergence problems, high computational cost, and sensitivity to extreme parameters, as well as to the size of the dataset. Meanwhile, despite

requiring intricate and meticulous derivative calculations, regularity assumptions, and EM convergence to the true MLE, Louis' method leverages model structure to provide robust, efficient, and reliable estimates of parameter uncertainty.

Table 4.7: Summary of Parameter Estimates, Standard Errors, Relative Standard Error (RSE), Confidence Intervals (CI or PI), and Interval Lengths for the Maule Region using GEM-Louis' method and STPB

	Maule Region - GEM-Louis' method						Maule Region - STPB $B = 1526$					
	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\sqrt{\left[\boldsymbol{I}_o^{-1}(\widehat{\boldsymbol{\Theta}})\right]_{ii}}$	RSE	CI-L	CI-U	CI length	$\widehat{se}_B$	RSE	PI-L	PI-U	PI length	δ
$\widehat{\beta}_0$	-6.47037	1.14591	17.710	-8.71636	-4.22438	4.49198	1.45489	22.485	-9.34989	-3.52208	5.82781	0.15042
$\widehat{\beta}_1$	1.57427	0.01728	1.098	1.54040	1.60814	0.06774	0.00134	0.085	1.57149	1.57670	0.00521	< 0.00001
$\widehat{\beta}_2$	5.99659	0.01241	0.207	5.97226	6.02092	0.04866	0.00226	0.038	5.99405	5.99927	0.00522	< 0.00001
$\widehat{\beta}_3$	-0.00373	0.00020	5.362	-0.00412	-0.00334	0.00078	0.00027	7.239	-0.00428	-0.00319	0.00109	< 0.00001
$\widehat{\beta}_4$	0.00381	0.00020	5.249	0.00341	0.00421	0.00080	0.00027	7.087	0.00325	0.00435	0.00110	< 0.00001
$\widehat{\beta}_5$	-0.00082	0.00004	4.878	-0.00091	-0.00073	0.00018	0.00007	8.537	-0.00097	-0.00067	0.00030	< 0.00001
$\widehat{\sigma}^2_\omega$	0.07762	0.00230	2.963	0.07311	0.08213	0.00902	0.00183	2.358	0.07389	0.08130	0.00741	< 0.00001
$\widehat{\phi}$	0.81142	0.00248	0.306	0.80657	0.81627	0.00970	0.00342	0.421	0.80432	0.81754	0.01322	< 0.00001
$\widehat{\alpha}$	259.43299	10.70898	4.128	238.44338	280.42260	41.97922	6.07279	2.341	246.76339	271.37512	24.61173	2.62068
$\widehat{\sigma}^2_\eta$	2.96738	0.11835	3.988	2.73541	3.19935	0.46394	0.05726	1.930	2.85398	3.08695	0.23296	0.00023

Table 4.8: Summary of Parameter Estimates, Standard Errors, Relative Standard Error (RSE), Confidence Intervals (CI or PI), and Interval Lengths for the Ñuble Region Using GEM and STPB Methods

	Ñuble Region - GEM-Louis' method						Ñuble Region - STPB $B = 1532$					
	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\sqrt{\left[\boldsymbol{I}_o^{-1}(\widehat{\boldsymbol{\Theta}})\right]_{ii}}$	RSE	CI-L	CI-U	CI length	$\widehat{se}_B$	RSE	PI-L	PI-U	PI length	δ
$\widehat{\beta}_0$	24.25788	3.66591	15.112	17.07270	31.44305	14.37035	4.29105	17.689	13.66701	33.15308	19.48607	1.30590
$\widehat{\beta}_1$	1.80750	0.04801	2.656	1.71340	1.90160	0.18820	0.00569	0.315	1.79494	1.82004	0.02510	< 0.00001
$\widehat{\beta}_2$	5.99299	0.00647	0.108	5.98030	6.00568	0.02538	0.00568	0.095	5.98147	6.00642	0.02495	< 0.00001
$\widehat{\beta}_3$	-0.00907	0.00033	3.639	-0.00972	-0.00843	0.00129	0.00060	6.613	-0.01041	-0.00769	0.00271	< 0.00001
$\widehat{\beta}_4$	-0.00061	0.00060	99.231	-0.00179	0.00057	0.00236	0.00068	111.678	-0.00203	0.00104	0.00307	< 0.00001
$\widehat{\beta}_5$	-0.00229	0.00003	1.380	-0.00235	-0.00223	0.00012	0.00005	2.182	-0.00240	-0.00219	0.00022	< 0.00001
$\widehat{\sigma}^2_\omega$	0.08115	0.01697	20.912	0.04789	0.11441	0.06652	0.00352	4.333	0.07412	0.08853	0.01441	< 0.00001
$\widehat{\phi}$	0.73094	0.00374	0.512	0.72361	0.73827	0.01466	0.00477	0.653	0.72092	0.73979	0.01887	< 0.00001
$\widehat{\alpha}$	170.27987	4.88892	2.871	160.69759	179.86216	19.16457	4.29594	2.523	162.12704	178.63465	16.50761	1.30888
$\widehat{\sigma}^2_\eta$	3.38785	0.08584	2.534	3.21960	3.55609	0.33649	0.06665	1.967	3.24963	3.52019	0.27056	0.00032

Table 4.9: Summary of Parameter Estimates, Standard Errors, Relative Standard Error (RSE), Confidence Intervals (CI or PI), and Interval Lengths for the Biobío Region Using GEM-Louis' method and STPB

	Biobío Region - GEM-Louis' method						Biobío Region - STPB $B = 1591$					
	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\sqrt{\left[\boldsymbol{I}_o^{-1}(\widehat{\boldsymbol{\Theta}})\right]_{ii}}$	RSE	CI-L	CI-U	CI length	$\widehat{se}_B$	RSE	PI-L	PI-U	PI length	δ
$\widehat{\beta}_0$	-10.01683	1.05738	10.556	-12.08930	-7.94436	4.14494	0.63454	6.335	-11.26560	-8.68824	2.57735	0.02802
$\widehat{\beta}_1$	1.55086	0.02086	1.345	1.50997	1.59174	0.08177	0.00215	0.139	1.54640	1.55540	0.00900	< 0.00001
$\widehat{\beta}_2$	3.90664	0.01835	0.470	3.87067	3.94261	0.07194	0.00230	0.053	3.90232	3.91067	0.00835	< 0.00001
$\widehat{\beta}_3$	0.00989	0.00018	1.792	0.00955	0.01024	0.00070	0.00206	1.613	0.00957	0.01023	0.00066	< 0.00001
$\widehat{\beta}_4$	0.00258	0.00018	7.068	0.00223	0.00294	0.00072	0.00012	4.636	0.00233	0.00282	0.00049	< 0.00001
$\widehat{\beta}_5$	-0.00256	0.00005	1.935	-0.00265	-0.00247	0.00020	0.00013	5.038	-0.00284	-0.00231	0.00053	< 0.00001
$\widehat{\sigma}^2_\omega$	0.13377	0.00283	2.117	0.12822	0.13932	0.01110	0.00286	2.135	0.12789	0.13933	0.01144	< 0.00001
$\widehat{\phi}$	0.84459	0.00255	0.301	0.83960	0.84958	0.00998	0.00348	0.412	0.83710	0.85093	0.01383	< 0.00001
$\widehat{\alpha}$	223.16450	5.93383	2.659	211.53420	234.79479	23.26059	5.40057	2.420	212.16580	233.77950	21.61369	2.02970
$\widehat{\sigma}^2_\eta$	2.10023	0.05353	2.549	1.99531	2.20514	0.20983	0.03748	1.785	2.02793	2.17889	0.15097	0.00010

Table 4.10: Summary of Parameter Estimates, Standard Errors, Relative Standard Error (RSE), Confidence Intervals (CI or PI), and Interval Lengths for the La Araucanía Region using GEM-Louis' method and STPB

	La Araucanía Region - GEM-Louis' method						La Araucanía Region - STPB $B = 1546$					
	$\widehat{\boldsymbol{\Theta}}_{\text{GEM}}$	$\sqrt{\left[\boldsymbol{I}_o^{-1}(\widehat{\boldsymbol{\Theta}})\right]_{ii}}$	RSE	CI-L	CI-U	CI length	$\widehat{se}_B$	RSE	PI-L	PI-U	PI length	δ
$\widehat{\beta}_0$	-45.29522	0.61478	1.357	-46.50020	-44.09025	2.40995	0.04900	0.108	-45.40615	-45.19752	0.20863	0.00017
$\widehat{\beta}_1$	1.65406	0.00856	0.518	1.63728	1.67083	0.03355	0.00014	0.008	1.65378	1.65429	0.00051	< 0.00001
$\widehat{\beta}_2$	4.64276	0.01080	0.233	4.62159	4.66392	0.04233	0.00017	0.004	4.64247	4.64302	0.00055	< 0.00001
$\widehat{\beta}_3$	0.00873	0.00015	1.746	0.00844	0.00903	0.00059	0.00003	0.349	0.00868	0.00879	0.00011	< 0.00001
$\widehat{\beta}_4$	0.00897	0.00018	2.040	0.00861	0.00932	0.00071	0.00001	0.079	0.00895	0.00898	0.00003	< 0.00001
$\widehat{\beta}_5$	-0.00427	0.00004	0.900	-0.00435	-0.00420	0.00015	0.00001	0.159	-0.00428	-0.00426	0.00003	< 0.00001
$\widehat{\sigma}^2_\omega$	0.07720	0.00141	1.824	0.07443	0.07996	0.00553	0.00065	0.846	0.07611	0.07822	0.00211	< 0.00001
$\widehat{\phi}$	0.76313	0.00167	0.219	0.75986	0.76640	0.00654	0.00223	0.292	0.75880	0.76726	0.00846	< 0.00001
$\widehat{\alpha}$	232.71480	4.59858	1.976	223.70161	241.72803	18.02642	1.10109	0.473	230.67560	234.69010	4.01454	0.08560
$\widehat{\sigma}^2_\eta$	3.60579	0.07059	1.958	3.46743	3.74414	0.27671	0.01701	0.472	3.57281	3.63725	0.06444	0.00002

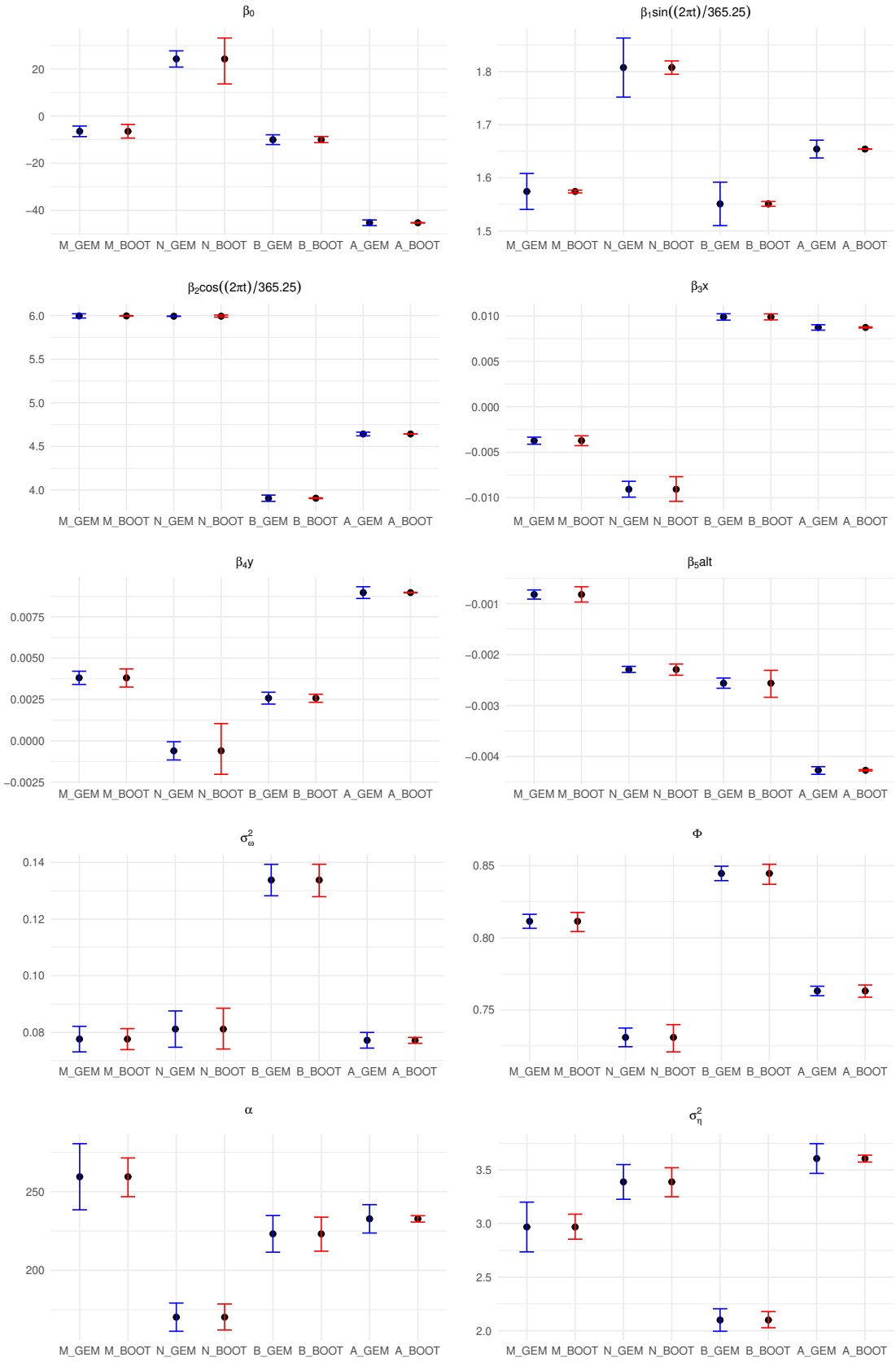

Figure 4.4: Comparison of Confidence Intervals for each region analyzed in the real data study. The capital letters on the x-axis indicate M for Maule Region, N for Ñuble Region, B for Biobío Region, and A for La Araucanía Region. Confidence intervals derived using Louis' method are shown in blue (GEM), while those obtained through the STPB are represented in red (BOOT).

Chapter 5

Conclusions

In this research, the estimated variance-covariance matrix of the maximum likelihood estimator for a space-time autoregressive model in the presence of missing data was successfully approximated analytically and numerically. By inverting the observed Fisher Information Matrix, we calculate standard errors for each parameter, allowing a detailed assessment of the model precision.

The performance of Louis' method was evaluated by comparing it with standard error estimates based on the parametric bootstrap method, in simulation scenarios, and with real observed data. The results from both approaches were largely consistent, indicating that Louis' method provides a robust and computationally efficient solution for complex inference databases involving missing data.

Our proposal for inference in a space-time model in the presence of missing data represents a significant advancement in the field of space-time analysis. Furthermore, it provides a foundation for future developments in both theoretical and applied contexts.

Future works:

- Expanding the model to other climatological variables and higher-order autoregressive models AR(p).
- Finish the development of the CRAN package ”Tools For Statisticians,” including optimized codes for the observed Fisher Information Matrix and Space-Time Parametric Bootstrap.

Chapter 6

Conclusiones

En esta investigación, la matriz de varianza-covarianza del estimador de máxima verosimilitud para un modelo autorregresivo espacio-temporal con datos faltantes fue aproximada exitosamente, tanto de forma analítica como numérica. Al invertir la Matriz de Información de Fisher observada, se calcularon los errores estándar para cada parámetro estimado, lo que permitió evaluar con detalle la precisión del modelo.

El desempeño del método de Louis fue evaluado mediante su comparación con las estimaciones de error estándar obtenidas a partir del método Bootstrap parámetrico, tanto en escenarios de simulación como con datos reales observados. Los resultados obtenidos con ambos enfoques fueron ampliamente consistentes, lo que indica que el metodo de Louis ofrece una solución robusta y computacionalmente eficiente para problemas de inferencia complejos con datos perdidos.

Nuestra propuesta para la inferencia en un modelo espacio-temporal en presencia de datos incompletos representa un avance significativo en el campo del analisis de modelos espacio-temporal. Además, proporciona una base sólida para futuros desarrollos en contextos tanto

teóricos como aplicados.

Trabajos futuros:

- Ampliar el modelo a otras variables climatológicas y a modelos autorregresivos de orden superior AR(p).
- Finalizar el desarrollo del paquete de CRAN ” Tools For Statisticians ”, que incluirá código optimizado para la Matriz de Información de Fisher observada y el Bootstrap Paramétrico espacio-temporal.

Bibliography

AGROMET (n.d.). Red agroclimática nacional. *https://www.agromet.cl*.

Andrews, D. W. K. and Buchinsky, M. (2000). A Three–Step Method for Choosing the Number of Bootstrap Repetitions. *Econometrica*, 68(1):23–51.

Baker, S. G. (1992). A Simple Method for Computing the Observed Information Matrix When Using the EM Algorithm with Categorical Data. *Journal of Computational and Graphical Statistics*, 1(1):63–76.

Bickel, P. J., Ritov, Y., and Rydén, T. (1998). Asymptotic Normality of the Maximum-Likelihood Estimator for General Hidden Markov Models. *The Annals of Statistics*, 26(4):1614–1635.

Brookes, M. (2020). *The Matrix Reference Manual*. Imperial College (online).

Bürger, R. (2014). *Apuntes curso: Mecanica del Medio Continuo*. Centro de Investigacion en Ingenieria y Departamento de Ingenieria Matematica. Facultad de Ciencias Fisicas y Matematicas. Universidad de Concepcion, Concepcion.

Cameletti, M., Lindgren, F., Simpson, D., and Rue, H. (2013). Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis*, 97(2):109–131.

Chau, T. T. T., Ailliot, P., Monbet, V., and Tandeo, P. (2023). Comparison of simulation-based algorithms for parameter estimation and state reconstruction in nonlinear state-space models. *Discrete and Continuous Dynamical Systems – S*, 16(2):240–264.

Chen, W., Genton, M. G., and Sun, Y. (2021). Space-Time Covariance Structures and Models. *Annual Review of Statistics and Its Application*, 8:191–215.

CORFO (2022). *Chile y sus Regiones en datos económicos – Informe Economico para la Descentralizacion*. Unidad de Analisis Territorial. Gerencia de Redes y Territorios CORFO – Gobierno de Chile, https://wapp4.corfo.cl/archivos/wcsconti/ip/grc/chileedition.

Cressie, N. and Wikle, C. K. (2015). *Statistics for spatio–temporal data*. Wiley Series in Probability and Statistics. John Wiley & Sons, New Jersey.

De Oliveira, V. and Han, Z. (2022). On Information About Covariance Parameters in Gaussian Matérn Random Fields. *Journal of Agricultural, Biological and Environmental Statistics*, 27(4):690–712.

Delyon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a Stochastic Approximation Version of the EM Algorithm. *The Annals of Statistics*, 27(1):94–128.

Douc, R., Moulines, E., and Stoffer, D. (2014). *Nonlinear time series. Theory, methods and applications with R examples*. Texts in Statistical Science. Chapman & Hall/CRC, New York.

Duan, J.-C. and Fulop, A. (2011). A stable estimator of the information matrix under EM for dependent data. *Statistics and Computing*, 21(1):83–91.

Efron, B. and Hinkley, D. V. (1978). Assessing the Accuracy of the Maximum Likelihood Estimator: Observed Versus Expected Fisher Information. *Biometrika*, 65(3):457–482.

Efron, B. and Tibshirani, R. (1994). *An Introduction to the Bootstrap*. Monographs on Statistics & Applied Probability 57. Chapman and Hall/CRC, New York, 1st edition edition.

Fassò, A. and Cameletti, M. (2009). The EM algorithm in a distributed computing environment for modelling environmental space–time data. *Environmental Modelling & Software*, 24(9):1027–1035.

Fernández-Casal, R., Castillo-Páez, S., and Francisco-Fernández, M. (2024). Nonparametric Conditional Risk Mapping Under Heteroscedasticity. *Journal of Agricultural, Biological and Environmental Statistics*, 29:56 – 72.

Fuentes, M. (2005). A formal test for nonstationarity of spatial stochastic processes. *Journal of Multivariate Analysis*, 96(1):30–54.

Fuentes, M. (2006). Testing for separability of spatial–temporal covariance functions. *Journal of Statistical Planning and Inference*, 136(2):447–466.

Gneiting, T. (2002). Nonseparable, Stationary Covariance Functions for Space–Time Data. *Journal of the American Statistical Association*, 97(458):590–600.

Gneiting, T., Genton, M., and Guttorp, P. (2006). Geostatistical Space-Time Models, Stationarity, Separability and Full Symmetry. *Statistical Methods for Spatio-temporal Systems (Monographs on Statistics and Applied Probability)*, 107.

Gong, M., Miller, C., Scott, M., O'Donnell, R., Simis, S., Groom, S., Tyler, A., Hunter, P., and Spyrakos, E. (2021). State space functional principal component analysis to identify spatiotemporal patterns in remote sensing lake water quality. *Stochastic Environmental Research and Risk Assessment*, 35(12):2521–2536.

Gupta, M. R. and Chen, Y. (2011). Theory and Use of the EM Algorithm. *Foundations and Trends in Signal Processing*, 4(3):223–296.

Hamilton, J. D. (1994). *Time Series Analysis*. Book collections on Project MUSE. Princeton University Press, New Jersey.

Holbrook, A. (2018). Differentiating the pseudo determinant. *Linear Algebra and its Applications*, 548:293–304.

Holmes, E. (2014). Computation of standardized residuals for (MARSS) models. Technical Report 1411.0045, Northwest Fisheries Science Center, NOAA Fisheries, Seattle, Washington.

Holmes, E. (2016). Notes on computing the Fisher Information matrix for MARSS models. Part II Louis 1982. Technical report, National Marine Fisheries Service & University of Washington.

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

Jamshidian, M. and Jennrich, R. I. (2000). Standard errors for EM estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2):257–270.

Katzfuss, M. and Cressie, N. (2011). Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. *Journal of Time Series Analysis*, 32(4):430–446.

Lagos-Álvarez, B., Padilla, L., Mateu, J., and Ferreira, G. (2019). A Kalman filter method for estimation and prediction of space–time data with an autoregressive structure. *Journal of Statistical Planning and Inference*, 203:117–130.

Little, R. and Rubin, D. (2002). *Statistical Analysis with Missing Data*. Wiley Series in Probability and Statistics. Wiley & Sons, Inc., New Jersey.

Louis, T. A. (1982). Finding the Observed Information Matrix when Using the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(2):226–233.

Lu, N. and Zimmerman, D. L. (2005). The likelihood ratio test for a separable covariance matrix. *Statistics & Probability Letters*, 73(4):449–457.

Lystig, T. C. and Hughes, J. P. (2002). Exact Computation of the Observed Information Matrix for Hidden Markov Models. *Journal of Computational and Graphical Statistics*, 11(3):678–689.

Magnus, J. R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, volume Wiley Series in Probability and Statistics. John Wiley & Sons., New York, second edition.

Mardia, K. V., Kent, J., and Bibby, J. (1979). *Multivariate Analysis*. Probability and Mathematical Statistics: a series of monographs and textbooks. Academic Press, New York, 1 edition.

Martinez, F. (2008). *Modelización de la función de covarianza en procesos espacio–temporales: análisis y aplicaciones*. PhD thesis, Departament d' Estadística i Investigació Operativa. Universitat de Valencia, Valencia.

Matérn, B. (1986). *Spatial Variation*, volume 36 of *Lecture Notes in Statistics*. Springer, New York, 2 edition.

McLachlan, G. and Krishnan, T. (2008). *The EM Algorithm and Extensions*. Wiley Series in Probability and Statistics. Wiley & Sons, Inc., New Jersey, second edition edition.

Meilijson, I. (1989). A Fast Improvement to the EM Algorithm on its Own Terms. *Journal of the Royal Statistical Society. Series B (Methodological)*, 51(1):127–138.

Meng, X.-L. and Rubin, D. B. (1991). Using EM to Obtain Asymptotic Variance-Covariance Matrices: The SEM Algorithm. *Journal of the American Statistical Association*, 86(416):899–909.

Mitchell, M. W., Genton, M. G., and Gumpertz, M. L. (2005). Testing for separability of space–time covariances. *Environmetrics*, 16(8):819–831.

Mitchell, M. W., Genton, M. G., and Gumpertz, M. L. (2006). A likelihood ratio test for separability of covariances. *Journal of Multivariate Analysis*, 97(5):1025–1043.

Oakes, D. (2002). Direct Calculation of the Information Matrix via the EM Algorithm. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 61(2):479–482.

Padilla, L. (2018). Implementacion de un metodo de estimacion y prediccion espacio–tiempo con datos faltantes basado en filtro de Kalman. Master's thesis, Facultad de Ciencias Fisicas y Matematicas. Universidad de Concepcion, Concepcion.

Padilla, L., Lagos-Álvarez, B., Mateu, J., and Porcu, E. (2020). Space–time autoregressive estimation and prediction with missing data based on Kalman filtering. *Environmetrics*, 31(7):e2627.

Petersen, K. B. and Pedersen, M. S. (2012). *The Matrix Cookbook*. Technical University of Denmark, Copenhagen.

Porcu, E., Furrer, R., and Nychka, D. (2021). 30 Years of space–time covariance functions. *WIREs Computational Statistics*, 13(2):e1512.

Sahu, S. and Bakar, K. (2012a). A comparison of Bayesian models for daily ozone concentration levels. *Statistical Methodology*, 9(1):144–157. Special Issue on Astrostatistics + Special Issue on Spatial Statistics.

Sahu, S. K. and Bakar, K. S. (2012b). Hierarchical Bayesian autoregressive models for large space–time data with applications to ozone concentration modelling. *Applied Stochastic Models in Business and Industry*, 28(5):395–415.

Shitan, M. and Brockwell, P. J. (1995). An asymptotic test for separability of a spatial autoregressive model. *Communications in Statistics–theory and Methods*, 24(8):2027–2040.

Shumway, R. H. and Stoffer, D. S. (1982). AN APPROACH TO TIME SERIES SMOOTHING AND FORECASTING USING THE EM ALGORITHM. *Journal of Time Series Analysis*, 3(4):253–264.

Shumway, R. H. and Stoffer, D. S. (2017). *Time Series Analysis and its Applications – with R Examples*. Springer Texts in Statistics. Springer Nature, Switzerland, fourth edition.

Sigrist, F., Künsch, H. R., and Stahel, W. A. (2011). An autoregressive spatio-temporal precipitation model. *Procedia Environmental Sciences*, 3:2–7. 1st Conference on Spatial Statistics 2011 – Mapping Global Change.

Stoffer, D. S. and Wall, K. D. (1991). Bootstrapping State-Space Models: Gaussian Maximum Likelihood Estimation and the Kalman Filter. *Journal of the American Statistical Association*, 86(416):1024–1033.

Surya, B. A. (2024). Maximum likelihood recursive state estimation: An incomplete-information based approach. *Automatica*, 168:111820.

Tanner, M. A. (1993). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Springer Series in Statistics. Springer-Verlag, New York, second edition.

Tanner, M. A. and Wong, W. H. (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association*, 82(398):528–540.

Turner, T. R., Cameron, M. A., and Thomson, P. J. (1998). Hidden Markov Chains in Generalized Linear Models. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 26(1):107–125.

University, C., editor (1972). *A missing information principle: theory and applications*, volume 1, Berkeley. University of California Press.

Walsh, M. J. (2006). Computing the observed information matrix for dynamic mixture models. NUWC-NPT Technical Report 11,768 ADA457101, Naval Undersea Warfare Center Division, Newport, Rhode Island.

Wang, H., Zhang, T., Tang, X., and Liu, Y. (2010). Building a Dynamic, Large-Scale Spatio-temporal Vector Database to Support a National Spatial Data Infrastructure in China. *GIScience & Remote Sensing*, 47(1):135–162.

Wei, G. C. G. and Tanner, M. A. (1990). A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms. *Journal of the American Statistical Association*, 85(411):699–704.

Xu, K. and Wikle, C. K. (2007). Estimation of parameterized spatio–temporal dynamic models. *Journal of Statistical Planning and Inference*, 137(2):567–588.

Chapter VII

Appendix

VII.1 Arrays for the state-space representation

For one $Z_t(\mathbf{s})$ we have:

State equation:

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

$$\begin{pmatrix} \boldsymbol{\varepsilon}_t(\mathbf{s}) \\ \vdots \\ \boldsymbol{\varepsilon}_{t-p+2}(\mathbf{s}) \\ \boldsymbol{\varepsilon}_{t-p+1}(\mathbf{s}) \end{pmatrix}_{p\times 1} = \begin{pmatrix} \phi_1 & \dots & \phi_{p-1} & \phi_p \\ 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}_{p\times p} \begin{pmatrix} \boldsymbol{\varepsilon}_{t-1}(\mathbf{s}) \\ \vdots \\ \boldsymbol{\varepsilon}_{t-p+1}(\mathbf{s}) \\ \boldsymbol{\varepsilon}_{t-p}(\mathbf{s}) \end{pmatrix}_{p\times 1} + \begin{pmatrix} \eta_t(\mathbf{s}) \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{p\times 1}$$

Observation equation:

$$Z_t(\mathbf{s}) = \mathbf{X}_t(\mathbf{s}) \cdot \boldsymbol{\beta} + \boldsymbol{\mathcal{H}}^\top \cdot \boldsymbol{\xi}_t(\mathbf{s}) + \omega_t(\mathbf{s})$$

$$\Big(Z_t(\mathbf{s})\Big)_{1\times 1} = \Big(X_t^{(1)}(\mathbf{s}) \quad \cdots \quad X_t^{(\ell)}(\mathbf{s})\Big)_{1\times \ell} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_\ell \end{pmatrix}_{\ell\times 1} + \Big(1 \quad 0 \quad \dots \quad 0\Big)_{1\times p} \begin{pmatrix} \boldsymbol{\varepsilon}_t(\mathbf{s}) \\ \vdots \\ \boldsymbol{\varepsilon}_{t-p+2}(\mathbf{s}) \\ \boldsymbol{\varepsilon}_{t-p+1}(\mathbf{s}) \end{pmatrix}_{p\times 1} + \Big(\sigma_\omega^2\Big)_{1\times 1}$$

For all locations at time t, the state equation is defined as follows:

$$\boldsymbol{\xi}_{t_{np\times 1}} = \boldsymbol{\Phi}_{np\times np} \cdot \boldsymbol{\xi}_{t-1_{np\times 1}} + \mathbf{V}_{t_{np\times 1}}$$

$$\boldsymbol{\xi}_t = \begin{pmatrix} \boldsymbol{\xi}_t(\mathbf{s}_1) \\ \vdots \\ \boldsymbol{\xi}_t(\mathbf{s}_n) \end{pmatrix}_{np\times 1}, \qquad \boldsymbol{\xi}_{t-1} = \begin{pmatrix} \boldsymbol{\xi}_{t-1}(\mathbf{s}_1) \\ \vdots \\ \boldsymbol{\xi}_{t-1}(\mathbf{s}_n) \end{pmatrix}_{np\times 1},$$

$$\boldsymbol{\Phi} = \mathbf{I} \otimes \mathbf{F} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}_{n\times n} \otimes \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}_{p\times p} = \begin{pmatrix} \mathbf{F} & 0 & \cdots & 0 \\ 0 & \mathbf{F} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{F} \end{pmatrix}_{np\times np}$$

$$\mathbf{V}_t(\mathbf{s}) = \begin{pmatrix} \eta_t(\mathbf{s}) \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{p\times 1}, \quad \mathbf{V}_t = \begin{pmatrix} \eta_t(\mathbf{s}_1) \\ 0 \\ \vdots \\ 0 \\ \eta_t(\mathbf{s}_2) \\ 0 \\ \vdots \\ 0 \\ \vdots \\ \eta_t(\mathbf{s}_n) \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{np\times 1}, \quad \mathbf{V}_t = \begin{pmatrix} \mathbf{V}_t(\mathbf{s}_1) \\ \vdots \\ \mathbf{V}_t(\mathbf{s}_n) \end{pmatrix}_{np\times 1},$$

Observation equation:

$$\mathbf{Z}_{t_{n_t\times 1}} = \mathbf{X}_{t_{n_t\times \ell}} \cdot \boldsymbol{\beta}_{\ell\times 1} + \boldsymbol{\Lambda}_{t_{n_t\times np}} \cdot \boldsymbol{\xi}_{t_{np\times 1}} + \mathbf{W}_{t_{n_t\times 1}}$$

Those expressions expand naturally to all the location $\mathbf{s}_n$ for all t_T, organized all the data in the vector $\mathbf{Z}_{1:T}$. Where $\boldsymbol{\mu}_t(\mathbf{s})_{(n_T)\times 1} = \mathbf{X}_t(\mathbf{s})\boldsymbol{\beta}$ and $\boldsymbol{\Lambda}_{t_{(n_T)\times np}} = \mathbf{L}_t^{obs} \otimes \boldsymbol{\mathcal{H}}^\top$

$$\mathbf{Z}_{1:T} = \begin{pmatrix} Z_{t_1}(\mathbf{s}_1) \\ Z_{t_1}(\mathbf{s}_2) \\ \vdots \\ Z_{t_1}(\mathbf{s}_n) \\ Z_{t_2}(\mathbf{s}_1) \\ Z_{t_2}(\mathbf{s}_2) \\ \vdots \\ Z_{t_2}(\mathbf{s}_n) \\ \vdots \\ Z_{t_T}(\mathbf{s}_1) \\ Z_{t_T}(\mathbf{s}_2) \\ \vdots \\ Z_{t_T}(\mathbf{s}_n) \end{pmatrix}_{(n_T)\times 1}, \quad \mathbf{X}_t(\mathbf{s}) = \begin{pmatrix} 1 & X_{t_1}^{(1)}(\mathbf{s}_1) & X_{t_1}^{(2)}(\mathbf{s}_1) & \dots & X_{t_1}^{(\ell)}(\mathbf{s}_1) \\ 1 & X_{t_1}^{(1)}(\mathbf{s}_2) & X_{t_1}^{(2)}(\mathbf{s}_2) & \dots & X_{t_1}^{(\ell)}(\mathbf{s}_2) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{t_1}^{(1)}(\mathbf{s}_n) & X_{t_1}^{(2)}(\mathbf{s}_n) & \dots & X_{t_1}^{(\ell)}(\mathbf{s}_n) \\ 1 & X_{t_2}^{(1)}(\mathbf{s}_1) & X_{t_2}^{(2)}(\mathbf{s}_1) & \dots & X_{t_2}^{(\ell)}(\mathbf{s}_1) \\ 1 & X_{t_2}^{(1)}(\mathbf{s}_2) & X_{t_2}^{(2)}(\mathbf{s}_2) & \dots & X_{t_2}^{(\ell)}(\mathbf{s}_2) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{t_2}^{(1)}(\mathbf{s}_n) & X_{t_2}^{(2)}(\mathbf{s}_n) & \dots & X_{t_2}^{(\ell)}(\mathbf{s}_n) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{t_T}^{(1)}(\mathbf{s}_1) & X_{t_T}^{(2)}(\mathbf{s}_1) & \dots & X_{t_T}^{(\ell)}(\mathbf{s}_1) \\ 1 & X_{t_T}^{(1)}(\mathbf{s}_2) & X_{t_T}^{(2)}(\mathbf{s}_2) & \dots & X_{t_T}^{(\ell)}(\mathbf{s}_2) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{t_T}^{(1)}(\mathbf{s}_n) & X_{t_T}^{(2)}(\mathbf{s}_n) & \dots & X_{t_T}^{(\ell)}(\mathbf{s}_n) \end{pmatrix}_{(n_T)\times(\ell+1)}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_\ell \end{pmatrix}_{(\ell+1)\times 1}, \boldsymbol{\Lambda}_{t_{(n_T)\times np}} = \begin{pmatrix} (\boldsymbol{e}_{1_1})_{1\times n} \\ (\boldsymbol{e}_{1_2})_{1\times n} \\ \vdots \\ (\boldsymbol{e}_{1_T})_{1\times n} \\ (\boldsymbol{e}_{2_1})_{1\times n} \\ (\boldsymbol{e}_{2_2})_{1\times n} \\ \vdots \\ (\boldsymbol{e}_{2_T})_{1\times n} \\ \vdots \\ (\boldsymbol{e}_{n_1})_{1\times n} \\ (\boldsymbol{e}_{n_2})_{1\times n} \\ \vdots \\ (\boldsymbol{e}_{n_T})_{1\times n} \end{pmatrix}_{(n_T)\times n} \otimes \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}_{1\times p}, \quad \mathbf{W}_t = \begin{pmatrix} \omega_{t_1}(\mathbf{s}_1) \\ \vdots \\ \omega_{t_1}(\mathbf{s}_n) \\ \omega_{t_2}(\mathbf{s}_1) \\ \vdots \\ \omega_{t_2}(\mathbf{s}_n) \\ \vdots \\ \omega_{t_T}(\mathbf{s}_n) \end{pmatrix}_{(n_T)\times 1}$$

VII.2 Matrix Calculus

All the matrix differential calculus was made using the Theorems and Properties listed in Magnus and Neudecker (1999), Petersen and Pedersen (2012), and Brookes (2020)

1. **First Derivates of auxiliary Q-function from GEM algorithm**

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

$$
\frac{\partial Q(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}})}{\partial\boldsymbol{\Theta}} = \left[\frac{\partial q^{(1)}}{\partial\boldsymbol{\beta}} \quad \frac{\partial q^{(1)}}{\partial\sigma_\omega^2} \quad \frac{\partial q^{(2)}}{\partial\boldsymbol{\phi}} \quad \frac{\partial q^{(2)}}{\partial\alpha} \quad \frac{\partial q^{(2)}}{\partial\sigma_\eta^2}\right]^\top_{\ell+p+3}
$$

$$
\text{with } \frac{\partial q^{(1)}}{\partial\boldsymbol{\beta}} := \left(\frac{\partial q^{(1)}}{\partial\beta_i}\right)_{1\le i\le\ell} \quad \text{and} \quad \frac{\partial q^{(2)}}{\partial\boldsymbol{\phi}} := \left(\frac{\partial q^{(2)}}{\partial\phi_j}\right)_{1\le j\le p}
$$

1.1 $\frac{\partial q^{(1)}}{\partial \boldsymbol{\beta}}$

$$
\begin{aligned}
\frac{\partial\, 2q^{(1)}}{\partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}}\left[-\ln\{\sigma_\omega^2\}\sum_{t=1}^{T} n_t - \frac{1}{\sigma_\omega^2}\,\mathrm{tr}\left\{\sum_{t=1}^{T}\left(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_{t|T}\right)\right.\right. \\
&\qquad\qquad \left.\left.\left(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_{t|T}\right)^\top + \boldsymbol{\Lambda}_t\mathbf{P}_{t|T}\boldsymbol{\Lambda}_t^\top\right\}\right] \\
&= \frac{\partial}{\partial \boldsymbol{\beta}}\left[-\frac{1}{\sigma_\omega^2}\,\mathrm{tr}\left\{\sum_{t=1}^{T} -\mathbf{Z}_t\boldsymbol{\beta}^\top\mathbf{X}_t^\top - \mathbf{X}_t\boldsymbol{\beta}\mathbf{Z}_t^\top + \mathbf{X}_t\boldsymbol{\beta}\boldsymbol{\beta}^\top\mathbf{X}_t^\top + \mathbf{X}_t\boldsymbol{\beta}\boldsymbol{\xi}_{t|T}^\top\boldsymbol{\Lambda}_t^\top + \boldsymbol{\Lambda}_t\boldsymbol{\xi}_{t|T}\boldsymbol{\beta}^\top\mathbf{X}_t^\top\right\}\right] \\
&= -\frac{1}{\sigma_\omega^2}\Bigg(-\sum_{t=1}^{T}\frac{\partial}{\partial \boldsymbol{\beta}}\left[\,\mathrm{tr}\left\{\mathbf{Z}_t\boldsymbol{\beta}^\top\mathbf{X}_t^\top\right\}\right] - \sum_{t=1}^{T}\frac{\partial}{\partial \boldsymbol{\beta}}\left[\,\mathrm{tr}\left\{\mathbf{X}_t\boldsymbol{\beta}\mathbf{Z}_t^\top\right\}\right] + \sum_{t=1}^{T}\frac{\partial}{\partial \boldsymbol{\beta}}\left[\,\mathrm{tr}\left\{\mathbf{X}_t\boldsymbol{\beta}\boldsymbol{\beta}^\top\mathbf{X}_t^\top\right\}\right] \\
&\qquad + \sum_{t=1}^{T}\frac{\partial}{\partial \boldsymbol{\beta}}\left[\,\mathrm{tr}\left\{\mathbf{X}_t\boldsymbol{\beta}\boldsymbol{\xi}_{t|T}^\top\boldsymbol{\Lambda}_t^\top\right\}\right] + \sum_{t=1}^{T}\frac{\partial}{\partial \boldsymbol{\beta}}\left[\,\mathrm{tr}\left\{\boldsymbol{\Lambda}_t\boldsymbol{\xi}_{t|T}\boldsymbol{\beta}^\top\mathbf{X}_t^\top\right\}\right]\Bigg) \\
&= -\frac{1}{\sigma_\omega^2}\Bigg(-\sum_{t=1}^{T}\mathbf{X}_t^\top\mathbf{Z}_t - \sum_{t=1}^{T}\mathbf{X}_t^\top\mathbf{Z}_t + \sum_{t=1}^{T}\mathbf{X}_t^\top\mathbf{X}_t\boldsymbol{\beta} + \sum_{t=1}^{T}\mathbf{X}_t^\top\mathbf{X}_t\boldsymbol{\beta} + \sum_{t=1}^{T}\mathbf{X}_t^\top\boldsymbol{\Lambda}_t\boldsymbol{\xi}_{t|T} + \sum_{t=1}^{T}\mathbf{X}_t^\top\boldsymbol{\Lambda}_t\boldsymbol{\xi}_{t|T}\Bigg) \\
\frac{\partial\, 2q^{(1)}}{\partial \boldsymbol{\beta}} &= -\frac{2}{\sigma_\omega^2}\Bigg(-\sum_{t=1}^{T}\mathbf{X}_t^\top\mathbf{Z}_t + \sum_{t=1}^{T}\mathbf{X}_t^\top\mathbf{X}_t\boldsymbol{\beta} + \sum_{t=1}^{T}\mathbf{X}_t^\top\boldsymbol{\Lambda}_t\boldsymbol{\xi}_{t|T}\Bigg)
\end{aligned}
$$

1.2 $\frac{\partial q^{(1)}}{\partial \sigma_\omega^2}$

$$
\begin{aligned}
\frac{\partial\, 2q^{(1)}}{\partial \sigma_\omega^2} &= \frac{\partial}{\partial \sigma_\omega^2}\left[-\ln\{\sigma_\omega^2\}\sum_{t=1}^{T} n_t - \frac{1}{\sigma_\omega^2}\,\mathrm{tr}\left\{\sum_{t=1}^{T}\left(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_{t|T}\right)\left(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_{t|T}\right)^\top + \boldsymbol{\Lambda}_t\mathbf{P}_{t|T}\boldsymbol{\Lambda}_t^\top\right\}\right] \\
\frac{\partial\, 2q^{(1)}}{\partial \sigma_\omega^2} &= -\sum_{t=1}^{T} n_t\frac{1}{\sigma_\omega^2} + \frac{1}{(\sigma_\omega^2)^2}\,\mathrm{tr}\left\{\sum_{t=1}^{T}\left(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_{t|T}\right)\left(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_{t|T}\right)^\top + \boldsymbol{\Lambda}_t\mathbf{P}_{t|T}\boldsymbol{\Lambda}_t^\top\right\}
\end{aligned}
$$

1.3 $\dfrac{\partial q^{(2)}}{\partial \boldsymbol{\phi}}$

$$\frac{\partial 2q^{(2)}}{\partial \boldsymbol{\phi}} = \frac{\partial}{\partial \boldsymbol{\phi}}\left[-(T-1)\left(\ln|\mathbf{Q}|\right) - \operatorname{tr}\left\{\mathbf{Q}^{-}\left(\mathbf{S}_{22} - \mathbf{S}_{21}\boldsymbol{\Phi}^{\top} - \boldsymbol{\Phi}\mathbf{S}_{21}^{\top} + \boldsymbol{\Phi}\mathbf{S}_{11}\boldsymbol{\Phi}^{\top}\right)\right\}\right]$$

$$\text{Given } \boldsymbol{\Phi}_{np\times np} = \mathbf{I}_{n\times n} \otimes \mathbf{F}_{p\times p} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \otimes \begin{pmatrix} \phi_1 & \phi_2 & \ldots & \phi_{p-1} & \phi_p \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

it is possible to decompose $\boldsymbol{\Phi}$, to isolate the vector $\boldsymbol{\phi}$

$$\boldsymbol{\Phi}_{np\times np} = \left\{\boldsymbol{\upsilon}_{np\times n}^{\top} \otimes \boldsymbol{\phi}_{1\times p}^{\top}\right\} + \mathbf{G}_{np\times np}$$

$$\text{where } \boldsymbol{\upsilon}_{np\times n}^{\top} \otimes \boldsymbol{\phi}_{1\times p}^{\top} = \left\{ \quad \mathbf{I}_{n\times n} \quad \otimes \quad \boldsymbol{\mathcal{H}}_{1\times p}^{\top}\right\}^{\top} \otimes \quad \boldsymbol{\phi}_{1\times p}^{\top}$$

$$= \left\{ \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right\} \otimes \begin{pmatrix} \phi_1 & \phi_2 & \ldots & \phi_{p-1} & \phi_p \end{pmatrix}$$

$$\text{and } \mathbf{G}_{np\times np} = \mathbf{I}_{n\times n} \otimes \mathbf{F}0_{p\times p} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

$$\begin{aligned}
\frac{\partial\, 2q^{(2)}}{\partial \boldsymbol{\phi}} &= \frac{\partial}{\partial \boldsymbol{\phi}}\left[-\operatorname{tr}\left\{ \mathbf{Q}^{-}\mathbf{S}_{22} - \mathbf{Q}^{-}\mathbf{S}_{21}\boldsymbol{\Phi}^{\top} - \mathbf{Q}^{-}\boldsymbol{\Phi}\mathbf{S}_{21}^{\top} + \mathbf{Q}^{-}\boldsymbol{\Phi}\mathbf{S}_{11}\boldsymbol{\Phi}^{\top} \right\}\right] \\
&= \frac{\partial}{\partial \boldsymbol{\phi}}\left[\operatorname{tr}\left\{ \mathbf{Q}^{-}\mathbf{S}_{21}\Big(\Big(\boldsymbol{v}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big) + \mathbf{G}\Big)^{\top}\right\}\right] + \frac{\partial}{\partial \boldsymbol{\phi}}\left[\operatorname{tr}\left\{ \mathbf{Q}^{-}\Big(\Big(\boldsymbol{v}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big) + \mathbf{G}\Big)\mathbf{S}_{21}^{\top}\right\}\right] \\
&\qquad - \frac{\partial}{\partial \boldsymbol{\phi}}\left[\operatorname{tr}\left\{ \mathbf{Q}^{-}\Big(\Big(\boldsymbol{v}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big) + \mathbf{G}\Big)\mathbf{S}_{11}\Big(\Big(\boldsymbol{v}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big) + \mathbf{G}\Big)^{\top}\right\}\right] \\
&= \frac{\partial}{\partial \boldsymbol{\phi}}\left[\operatorname{tr}\left\{ \mathbf{Q}^{-}\mathbf{S}_{21}\left(\boldsymbol{v}\otimes\boldsymbol{\phi}\right)\right\}\right] + \frac{\partial}{\partial \boldsymbol{\phi}}\left[\operatorname{tr}\left\{ \mathbf{Q}^{-}\Big(\boldsymbol{v}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big)\mathbf{S}_{21}^{\top}\right\}\right] \\
&\qquad - \frac{\partial}{\partial \boldsymbol{\phi}}\Bigg[\operatorname{tr}\Big\{ \mathbf{Q}^{-}\Big(\boldsymbol{v}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big)\mathbf{S}_{11}\left(\boldsymbol{v}\otimes\boldsymbol{\phi}\right) + \mathbf{Q}^{-}\Big(\boldsymbol{v}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big)\mathbf{S}_{11}\mathbf{G}^{\top} \\
&\qquad\qquad + \mathbf{Q}^{-}\mathbf{G}\,\mathbf{S}_{11}\left(\boldsymbol{v}\otimes\boldsymbol{\phi}\right) + \mathbf{Q}^{-}\mathbf{G}\,\mathbf{S}_{11}\mathbf{G}^{\top}\Big\}\Bigg] \\
&= 2\frac{\partial}{\partial \boldsymbol{\phi}}\left[\operatorname{tr}\left\{ \mathbf{Q}^{-}\mathbf{S}_{21}\left(\boldsymbol{v}\otimes\boldsymbol{\phi}\right)\right\}\right] - \frac{\partial}{\partial \boldsymbol{\phi}}\left[\operatorname{tr}\left\{ \mathbf{Q}^{-}\Big(\boldsymbol{v}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big)\mathbf{S}_{11}\left(\boldsymbol{v}\otimes\boldsymbol{\phi}\right)\right\}\right] \\
&= 2\operatorname{tr}\left\{ \mathbf{Q}^{-}\mathbf{S}_{21}\frac{\partial}{\partial \boldsymbol{\phi}}\Big[\left(\boldsymbol{v}\otimes\boldsymbol{\phi}\right)\Big]\right\} - \frac{\partial}{\partial \boldsymbol{\phi}}\left[\operatorname{tr}\left\{ \mathbf{Q}^{-}\mathbf{S}_{11}\Big(\boldsymbol{v}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big)\left(\boldsymbol{v}\otimes\boldsymbol{\phi}\right)\right\}\right] \\
&= 2\operatorname{tr}\left\{ \mathbf{Q}^{-}\mathbf{S}_{21}\Big(\boldsymbol{v}\otimes\{\phi_{,j}\}\Big)\right\} - \frac{\partial}{\partial \boldsymbol{\phi}}\Big[\operatorname{tr}\Big\{ \mathbf{Q}^{-}\mathbf{S}_{11}\Big(\boldsymbol{v}^{\top}\boldsymbol{v}\otimes\boldsymbol{\phi}^{\top}\boldsymbol{\phi}\Big)\Big\}\Big] \\
&= 2\operatorname{tr}\left\{ \mathbf{Q}^{-}\mathbf{S}_{21}\Big(\boldsymbol{v}\otimes\{\phi_{,j}\}\Big)\right\} - \operatorname{tr}\left\{ \mathbf{Q}^{-}\mathbf{S}_{11}\Big(\boldsymbol{v}^{\top}\boldsymbol{v}\Big)\right\}\frac{\partial}{\partial \boldsymbol{\phi}}\Big[\operatorname{tr}\Big\{\boldsymbol{\phi}^{\top}\boldsymbol{\phi}\Big\}\Big] \\
\frac{\partial\, 2q^{(2)}}{\partial \phi_j} &= 2\operatorname{tr}\left\{ \mathbf{Q}^{-}\mathbf{S}_{21}\Big(\boldsymbol{v}\otimes\{\phi_{,j}\}\Big)\right\} - 2\operatorname{tr}\left\{ \mathbf{Q}^{-}\mathbf{S}_{11}\Big(\boldsymbol{v}^{\top}\boldsymbol{v}\Big)\right\}\Big(\phi_j\Big), \quad \forall\, j = 1,\dots,p
\end{aligned}$$

where $\{\phi_{,j}\}$ correspond to the p unit vectors of dimension $p \times 1$ obtained by differentiating

$\boldsymbol{\phi}$ (Bürger, 2014)

$$\{\phi_{,j}\} = \frac{\partial \boldsymbol{\phi}}{\partial \phi_j} = \begin{Bmatrix} \frac{\partial \boldsymbol{\phi}}{\partial \phi_1} \\ \vdots \\ \frac{\partial \boldsymbol{\phi}}{\partial \phi_p} \end{Bmatrix} = \begin{Bmatrix} \phi_{,1} \\ \vdots \\ \phi_{,p} \end{Bmatrix} \quad \forall\, j = 1, \ldots, p$$

1.4 $\dfrac{\partial q^{(2)}}{\partial \alpha}$

$$\frac{\partial\, 2q^{(2)}}{\partial \alpha} = \frac{\partial}{\partial \alpha}\left[-(T-1)\left(\ln|\mathbf{Q}|\right) - \operatorname{tr}\left\{ \mathbf{Q}^{-}\left(\mathbf{S}_{22} - \mathbf{S}_{21}\boldsymbol{\Phi}^{\top} - \boldsymbol{\Phi}\mathbf{S}_{21}^{\top} + \boldsymbol{\Phi}\mathbf{S}_{11}\boldsymbol{\Phi}^{\top} \right) \right\} \right]$$

to simplify the expression $\boldsymbol{\Psi}(\boldsymbol{\Phi}) = \mathbf{S}_{22} - \mathbf{S}_{21}\boldsymbol{\Phi}^{\top} - \boldsymbol{\Phi}\mathbf{S}_{21}^{\top} + \boldsymbol{\Phi}\mathbf{S}_{11}\boldsymbol{\Phi}^{\top}$

$$= \frac{\partial}{\partial \alpha}\left[-(T-1)\ln|\mathbf{Q}| \right] - \frac{\partial}{\partial \alpha}\left[\operatorname{tr}\left\{ \mathbf{Q}^{-}\left(\boldsymbol{\Psi}(\boldsymbol{\Phi}) \right) \right] \right\}$$

considering that $\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1} \sim SGau(\Phi\boldsymbol{\xi}_{t-1}; \mathbf{Q})$, where $\mathbf{Q} = \sigma_\eta^2\, \mathbf{R}$,

note that $\mathbf{Q}^{-} = \dfrac{1}{\sigma_\eta^2}\mathbf{R}^{-}$ then $|\mathbf{Q}| = (\sigma_\eta^2)^{np}|\mathbf{R}|$, where $\mathbf{R} = \boldsymbol{\rho}^{\eta}(\alpha) \otimes \mathbf{T}^{AR}$

and $\boldsymbol{\rho}^{\eta}(\alpha) = \exp\left(-\dfrac{\mathbf{H}}{\alpha} \right)$

$$\boldsymbol{\rho}^{\eta}(\alpha) = \begin{pmatrix} \rho^{\eta}(\mathbf{s}_1, \mathbf{s}_1) & \rho^{\eta}(\mathbf{s}_1, \mathbf{s}_2) & \cdots & \rho^{\eta}(\mathbf{s}_1, \mathbf{s}_n) \\ \rho^{\eta}(\mathbf{s}_2, \mathbf{s}_1) & \rho^{\eta}(\mathbf{s}_2, \mathbf{s}_2) & \cdots & \rho^{\eta}(\mathbf{s}_2, \mathbf{s}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{\eta}(\mathbf{s}_n, \mathbf{s}_1) & \rho^{\eta}(\mathbf{s}_n, \mathbf{s}_2) & \cdots & \rho^{\eta}(\mathbf{s}_n, \mathbf{s}_n) \end{pmatrix}_{n \times n} \qquad \mathbf{T}^{AR} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix}_{p \times p}$$

$$= \frac{\partial}{\partial \alpha}\left[-(T-1)\left(np\ \ln\{\sigma_\eta^2\} + \ln|\mathbf{R}| \right) \right] - \frac{\partial}{\partial \alpha}\left[\frac{1}{\sigma_\eta^2} \operatorname{tr}\left\{ \mathbf{R}^{-}\left(\boldsymbol{\Psi}(\boldsymbol{\Phi}) \right) \right\} \right]$$

$$= -(T-1)\ \frac{\partial}{\partial \alpha}\left[\ln|\mathbf{R}| \right] - \frac{1}{\sigma_\eta^2} \operatorname{tr}\left\{ \frac{\partial\, \mathbf{R}^{-}}{\partial \alpha}\left(\boldsymbol{\Psi}(\boldsymbol{\Phi}) \right) \right\}$$

using the results in Holbrook (2018) and the Theorem for the differential of the Moore-Penrose Inverse matrix (Golub, 1973)

$$\frac{\partial\, 2q^{(2)}}{\partial\alpha} = -(T-1)\ \mathrm{tr}\left\{\mathbf{R}^{-}\ \frac{\partial\,\mathbf{R}}{\partial\alpha}\right\} - \frac{1}{\sigma_{\eta}^{2}}\,\mathrm{tr}\left\{\left(-\mathbf{R}^{-}\ \frac{\partial\,\mathbf{R}}{\partial\alpha}\,\mathbf{R}^{-} + \mathbf{R}^{-}\mathbf{R}^{-\top}\ \frac{\partial\,\mathbf{R}^{\top}}{\partial\alpha}\,(\mathbf{I}-\mathbf{R}\mathbf{R}^{-}) + (\mathbf{I}-\mathbf{R}^{-}\mathbf{R})\ \frac{\partial\,\mathbf{R}^{\top}}{\partial\alpha}\,\mathbf{R}^{-\top}\mathbf{R}^{-}\right)\left(\boldsymbol{\Psi}(\boldsymbol{\Phi})\right)\right\}$$

$$\text{where}\ \ \frac{\partial\,\mathbf{R}}{\partial\alpha} = \frac{\partial}{\partial\alpha}\left[\boldsymbol{\rho}^{\eta}(\alpha)\otimes\mathbf{T}^{AR}\right] = \frac{\partial}{\partial\alpha}\left[\exp\left(-\frac{\mathbf{H}}{\alpha}\right)\right]\otimes\mathbf{T}^{AR}$$
$$= \left(\exp\left(-\frac{\mathbf{H}}{\alpha}\right)\circ\frac{\mathbf{H}}{\alpha^{2}}\right)\otimes\mathbf{T}^{AR} = \left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^{2}}\right)\otimes\mathbf{T}^{AR}$$

and the " $\circ$ " represent the Hadamard product.

$$\frac{\partial\ \boldsymbol{\rho}^{\eta}(\alpha)}{\partial\alpha} = \begin{pmatrix} \rho^{\eta}(\mathbf{s}_1,\mathbf{s}_1)\dfrac{||\mathbf{s}_1-\mathbf{s}_1||}{\alpha^{2}} & \rho^{\eta}(\mathbf{s}_1,\mathbf{s}_2)\dfrac{||\mathbf{s}_1-\mathbf{s}_2||}{\alpha^{2}} & \cdots & \rho^{\eta}(\mathbf{s}_1,\mathbf{s}_n)\dfrac{||\mathbf{s}_1-\mathbf{s}_n||}{\alpha^{2}} \\ \rho^{\eta}(\mathbf{s}_2,\mathbf{s}_1)\dfrac{||\mathbf{s}_2-\mathbf{s}_1||}{\alpha^{2}} & \rho^{\eta}(\mathbf{s}_2,\mathbf{s}_2)\dfrac{||\mathbf{s}_2-\mathbf{s}_2||}{\alpha^{2}} & \cdots & \rho^{\eta}(\mathbf{s}_2,\mathbf{s}_n)\dfrac{||\mathbf{s}_2-\mathbf{s}_n||}{\alpha^{2}} \\ \vdots & \vdots & \cdots & \vdots \\ \rho^{\eta}(\mathbf{s}_n,\mathbf{s}_1)\dfrac{||\mathbf{s}_n-\mathbf{s}_1||}{\alpha^{2}} & \rho^{\eta}(\mathbf{s}_n,\mathbf{s}_2)\dfrac{||\mathbf{s}_n-\mathbf{s}_2||}{\alpha^{2}} & \cdots & \rho^{\eta}(\mathbf{s}_n,\mathbf{s}_n)\dfrac{||\mathbf{s}_n-\mathbf{s}_n||}{\alpha^{2}} \end{pmatrix}_{n\times n}$$

Because $\mathbf{R}$ is symmetric, then $\mathbf{R} = \mathbf{R}^{\top}$.

$$\frac{\partial\, 2q^{(2)}}{\partial\alpha} = -(T-1)\ \mathrm{tr}\left\{\mathbf{R}^{-}\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^{2}}\right)\otimes\mathbf{T}^{AR}\right\} - \frac{1}{\sigma_{\eta}^{2}}\,\mathrm{tr}\left\{\left(-\mathbf{R}^{-}\left(\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^{2}}\right)\otimes\mathbf{T}^{AR}\right)\mathbf{R}^{-} + \left(\mathbf{R}^{-}\mathbf{R}^{-}\left(\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^{2}}\right)\otimes\mathbf{T}^{AR}\right)\right)(\mathbf{I}-\mathbf{R}\mathbf{R}^{-}) + (\mathbf{I}-\mathbf{R}^{-}\mathbf{R})\left(\left(\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^{2}}\right)\otimes\mathbf{T}^{AR}\right)\mathbf{R}^{-}\mathbf{R}^{-}\right)\right)\left(\boldsymbol{\Psi}(\boldsymbol{\Phi})\right)\right\}$$

$\mathbf{R}\mathbf{R}^{-}$ and $\mathbf{R}^{-}\mathbf{R}$ are orthogonal projection operators. These means that $\mathbf{R}\mathbf{R}^{-}$ and $\mathbf{R}^{-}\mathbf{R}$ are orthogonal projections onto the range of $\mathbf{R}$ and $\mathbf{R}^{-}$, respectivaly. There-

fore, the two last terms become equal to 0. Finally,

$$\frac{\partial 2q^{(2)}}{\partial \alpha} = -(T-1)\ \mathrm{tr}\left\{\mathbf{R}^{-}\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\right)\otimes\mathbf{T}^{AR}\right\} + \frac{1}{\sigma_\eta^2}\,\mathrm{tr}\left\{\left(\mathbf{R}^{-}\left(\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\right)\otimes\mathbf{T}^{AR}\right)\mathbf{R}^{-}\right)\left(\boldsymbol{\Psi}(\boldsymbol{\Phi})\right)\right\}$$

1.5 $\frac{\partial q^{(2)}}{\partial \sigma_\eta^2}$

$$\begin{aligned}
\frac{\partial q^{(2)}}{\partial\, 2\sigma_\eta^2} =& \frac{\partial}{\partial \sigma_\eta^2}\left[-(T-1)\left(\ln|\mathbf{Q}|\right) - \mathrm{tr}\left\{\mathbf{Q}^{-}\left(\boldsymbol{\Psi}(\boldsymbol{\Phi})\right)\right\}\right] \\
&= -(T-1)\frac{\partial}{\partial \sigma_\eta^2}\left[\ln|\mathbf{Q}|\right] - \mathrm{tr}\left\{\frac{\partial}{\partial \sigma_\eta^2}\left[\mathbf{Q}^{-}\right]\left(\boldsymbol{\Psi}(\boldsymbol{\Phi})\right)\right\} \\
&= -(T-1)\frac{\partial}{\partial \sigma_\eta^2}\left[\left(n\ln(\sigma_\eta^2) + \ln|\mathbf{R}|\right)\right] - \mathrm{tr}\left\{\frac{\partial}{\partial \sigma_\eta^2}\left[\frac{1}{\sigma_\eta^2}\,\mathbf{R}^{-}\right]\left(\boldsymbol{\Psi}(\boldsymbol{\Phi})\right)\right\} \\
&= -(T-1)n\left[\frac{1}{\sigma_\eta^2}\right] - \mathrm{tr}\left\{\frac{\partial}{\partial \sigma_\eta^2}\left[\frac{1}{\sigma_\eta^2}\right]\mathbf{R}^{-}\left(\boldsymbol{\Psi}(\boldsymbol{\Phi})\right)\right\} \\
\frac{\partial 2q^{(2)}}{\partial \sigma_\eta^2} =& -(T-1)\frac{n}{\sigma_\eta^2} + \frac{1}{(\sigma_\eta^2)^2}\,\mathrm{tr}\left\{\mathbf{R}^{-}\left(\boldsymbol{\Psi}(\boldsymbol{\Phi})\right)\right\}
\end{aligned}$$

2. **Second Derivates of auxiliary Q-function from GEM algorithm**

$$\frac{\partial^2 Q(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}})}{\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}^\top} = \begin{pmatrix} \frac{\partial^2 q^{(1)}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} & \frac{\partial^2 q^{(1)}}{\partial\boldsymbol{\beta}\partial\sigma_\omega^2} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \frac{\partial^2 q^{(1)}}{\partial\sigma_\omega^2\partial\boldsymbol{\beta}^\top} & \frac{\partial^2 q^{(1)}}{\partial{\sigma_\omega^2}^2} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{\partial^2 q^{(2)}}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^\top} & \frac{\partial^2 q^{(2)}}{\partial\boldsymbol{\phi}\partial\alpha} & \frac{\partial^2 q^{(2)}}{\partial\boldsymbol{\phi}\partial\sigma_\eta^2} \\ \mathbf{0} & \mathbf{0} & \frac{\partial^2 q^{(2)}}{\partial\alpha\partial\boldsymbol{\phi}^\top} & \frac{\partial^2 q^{(2)}}{\partial\alpha^2} & \frac{\partial^2 q^{(2)}}{\partial\alpha\partial\sigma_\eta^2} \\ \mathbf{0} & \mathbf{0} & \frac{\partial^2 q^{(2)}}{\partial\sigma_\eta^2\partial\boldsymbol{\phi}^\top} & \frac{\partial^2 q^{(2)}}{\partial\sigma_\eta^2\partial\alpha} & \frac{\partial^2 q^{(2)}}{\partial{\sigma_\eta^2}^2} \end{pmatrix}_{(\ell+p+3)\times(\ell+p+3)}$$

with $\frac{\partial^2 q^{(1)}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} := \left(\frac{\partial^2 q^{(1)}}{\partial\beta_i\partial\beta_j}\right)_{1\le i,j\le\ell}$ and $\frac{\partial^2 q^{(2)}}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}^\top} := \left(\frac{\partial^2 q^{(2)}}{\partial\phi_i\partial\phi_j}\right)_{1\le i,j\le p}$

The Schwarz Theorem, also known as the symmetry of second derivatives, states that for a function $f(x, y)$ of two variables x and y that have continuous second partial derivatives, the mixed partial derivatives are equal, $\frac{\partial^2 f}{\partial x\partial y} = \frac{\partial^2 f}{\partial y\partial x}$. This theorem generalizes to functions of more than two variables. It ensures that the order in which partial derivatives are taken does not matter when the second partial derivatives are continuous.

2.1 $\dfrac{\partial^2 q^{(1)}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}$

$$\begin{aligned}
\frac{\partial^2 q^{(1)}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= \frac{\partial}{\partial \boldsymbol{\beta}} \left[-\frac{1}{\sigma_\omega^2} \Big(-\sum_{t=1}^{T} \mathbf{X}_t^\top \mathbf{Z}_t + \sum_{t=1}^{T} \mathbf{X}_t^\top \mathbf{X}_t \boldsymbol{\beta} + \sum_{t=1}^{T} \mathbf{X}_t^\top \boldsymbol{\Lambda}_t \boldsymbol{\xi}_{t|T} \Big) \right] \\
&= -\frac{1}{\sigma_\omega^2} \frac{\partial}{\partial \boldsymbol{\beta}} \left[\sum_{t=1}^{T} \mathbf{X}_t^\top \mathbf{X}_t \boldsymbol{\beta} \right] \\
\frac{\partial^2 q^{(1)}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= -\frac{1}{\sigma_\omega^2} \sum_{t=1}^{T} \mathbf{X}_t^\top \mathbf{X}_t
\end{aligned}$$

2.2 $\dfrac{\partial^2 q^{(1)}}{\partial \boldsymbol{\beta} \partial \sigma_\omega^2}$

$$\begin{aligned}
\frac{\partial^2 q^{(1)}}{\partial \boldsymbol{\beta} \partial \sigma_\omega^2} &= \frac{\partial}{\partial \sigma_\omega^2} \left[-\frac{1}{\sigma_\omega^2} \left(-\sum_{t=1}^{T} \mathbf{X}_t^\top \mathbf{Z}_t + \sum_{t=1}^{T} \mathbf{X}_t^\top \mathbf{X}_t \boldsymbol{\beta} + \sum_{t=1}^{T} \mathbf{X}_t^\top \boldsymbol{\Lambda}_t \boldsymbol{\xi}_{t|T} \right) \right] \\
\frac{\partial^2 q^{(1)}}{\partial \boldsymbol{\beta} \partial \sigma_\omega^2} &= \frac{1}{(\sigma_\omega^2)^2} \Big(-\sum_{t=1}^{T} \mathbf{X}_t^\top \mathbf{Z}_t + \sum_{t=1}^{T} \mathbf{X}_t^\top \mathbf{X}_t \boldsymbol{\beta} + \sum_{t=1}^{T} \mathbf{X}_t^\top \boldsymbol{\Lambda}_t \boldsymbol{\xi}_{t|T} \Big)
\end{aligned}$$

2.3 $\dfrac{\partial^2 q^{(1)}}{\partial {\sigma_\omega^2}^2}$

$$\begin{aligned}
\frac{\partial^2 q^{(1)}}{\partial {\sigma_\omega^2}^2} &= \frac{1}{2} \frac{\partial}{\sigma_\omega^2} \left[-\sum_{t=1}^{T} n_t \frac{1}{\sigma_\omega^2} + \frac{1}{(\sigma_\omega^2)^2} \operatorname{tr} \left\{ \sum_{t=1}^{T} \Big(\Big(\mathbf{Z}_t - \mathbf{X}_t \boldsymbol{\beta} - \boldsymbol{\Lambda}_t \boldsymbol{\xi}_{t|T} \Big) \right. \right. \\
&\qquad \left. \left. \Big(\mathbf{Z}_t - \mathbf{X}_t \boldsymbol{\beta} - \boldsymbol{\Lambda}_t \boldsymbol{\xi}_{t|T} \Big)^\top + \boldsymbol{\Lambda}_t \mathbf{P}_{t|T} \boldsymbol{\Lambda}_t^\top \Big) \right\} \right] \\
\frac{\partial^2 q^{(1)}}{\partial {\sigma_\omega^2}^2} &= \frac{1}{2} \Big(\sum_{t=1}^{T} n_t \frac{1}{(\sigma_\omega^2)^2} - \frac{2}{(\sigma_\omega^2)^3} \operatorname{tr} \left\{ \sum_{t=1}^{T} \Big(\Big(\mathbf{Z}_t - \mathbf{X}_t \boldsymbol{\beta} - \boldsymbol{\Lambda}_t \boldsymbol{\xi}_{t|T} \Big) \right. \right. \\
&\qquad \left. \left. \Big(\mathbf{Z}_t - \mathbf{X}_t \boldsymbol{\beta} - \boldsymbol{\Lambda}_t \boldsymbol{\xi}_{t|T} \Big)^\top + \boldsymbol{\Lambda}_t \mathbf{P}_{t|T} \boldsymbol{\Lambda}_t^\top \Big) \right\} \Big)
\end{aligned}$$

2.4 $\dfrac{\partial^2 q^{(2)}}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}^\top}$

$$\begin{aligned}
\frac{\partial^2 q^{(2)}}{\partial \boldsymbol{\phi}_j \partial \boldsymbol{\phi}^\top} &= \frac{\partial}{\partial \boldsymbol{\phi}} \left[\operatorname{tr} \left\{ \mathbf{Q}^- \mathbf{S}_{21} \Big(\boldsymbol{\upsilon} \otimes \{\phi_{,j}\} \Big) \right\} - \operatorname{tr} \left\{ \mathbf{Q}^- \mathbf{S}_{11} \left(\boldsymbol{\upsilon}^\top \boldsymbol{\upsilon} \right) \right\} \Big(\phi_j \Big) \right] \\
&= - \operatorname{tr} \left\{ \mathbf{Q}^- \mathbf{S}_{11} \left(\boldsymbol{\upsilon}^\top \boldsymbol{\upsilon} \right) \right\} \frac{\partial}{\partial \boldsymbol{\phi}} \left[\Big(\phi_j \Big) \right] \\
\frac{\partial^2 q^{(2)}}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}_j^\top} &= - \operatorname{tr} \left\{ \mathbf{Q}^- \mathbf{S}_{11} \left(\boldsymbol{\upsilon}^\top \boldsymbol{\upsilon} \right) \right\} \{\phi_{,j}\}^\top, \quad \forall\, j = 1, \ldots, p
\end{aligned}$$

2.5 $\dfrac{\partial^2 q^{(2)}}{\partial \alpha^2}$

$$\begin{aligned}
\frac{\partial^2 q^{(2)}}{\partial \alpha^2} &= \frac{1}{2} \frac{\partial}{\partial \alpha} \Bigg[- (T-1)\ \operatorname{tr} \left\{ \mathbf{R}^- \left(\boldsymbol{\rho}^\eta(\alpha) \circ \frac{\mathbf{H}}{\alpha^2} \right) \otimes \mathbf{T}^{AR} \right\} \\
&\qquad + \frac{1}{\sigma_\eta^2} \operatorname{tr} \left\{ \left(\mathbf{R}^- \left(\left(\boldsymbol{\rho}^\eta(\alpha) \circ \frac{\mathbf{H}}{\alpha^2} \right) \otimes \mathbf{T}^{AR} \right) \mathbf{R}^- \right) \Big(\boldsymbol{\Psi}(\boldsymbol{\Phi}) \Big) \right\} \Bigg] \\
&= \frac{1}{2} \Bigg(- (T-1) \operatorname{tr} \Bigg\{ \underbrace{\frac{\partial}{\partial \alpha} \left[\mathbf{R}^- \left(\boldsymbol{\rho}^\eta(\alpha) \circ \frac{\mathbf{H}}{\alpha^2} \right) \otimes \mathbf{T}^{AR} \right]}_{\text{(a)}} \Bigg\} \\
&\qquad + \frac{1}{\sigma_\eta^2} \operatorname{tr} \Bigg\{ \underbrace{\frac{\partial}{\partial \alpha} \left[\mathbf{R}^- \left(\left(\boldsymbol{\rho}^\eta(\alpha) \circ \frac{\mathbf{H}}{\alpha^2} \right) \otimes \mathbf{T}^{AR} \right) \mathbf{R}^- \right]}_{\text{(b)}} \Big(\boldsymbol{\Psi}(\boldsymbol{\Phi}) \Big) \Bigg\} \Bigg)
\end{aligned}$$

Considering that

$$\begin{aligned}
\frac{\partial^2 \boldsymbol{\rho}^\eta(\alpha)}{\partial \alpha^2} &= \frac{\partial}{\partial \alpha} \left[\boldsymbol{\rho}^\eta(\alpha) \circ \frac{\mathbf{H}}{\alpha^2} \right] = \frac{\partial\, \boldsymbol{\rho}^\eta(\alpha)}{\partial \alpha} \circ \frac{\mathbf{H}}{\alpha^2} + \boldsymbol{\rho}^\eta(\alpha) \circ \frac{\partial}{\partial \alpha} \left[\frac{\mathbf{H}}{\alpha^2} \right] \\
&= \boldsymbol{\rho}^\eta(\alpha) \circ \left(\frac{\mathbf{H}}{\alpha^2} \right)^2 - \boldsymbol{\rho}^\eta(\alpha) \circ 2 \cdot \frac{\mathbf{H}}{\alpha^3} = \boldsymbol{\rho}^\eta(\alpha) \circ \left(\frac{\mathbf{H}^2}{\alpha^4} - 2 \cdot \frac{\mathbf{H}}{\alpha^3} \right)
\end{aligned}$$

$$= \begin{pmatrix} \rho^{\eta}(\mathbf{s}_1,\mathbf{s}_1)\left(\frac{||\mathbf{s}_1-\mathbf{s}_1||^2}{\alpha^4}-2\cdot\frac{||\mathbf{s}_1-\mathbf{s}_1||}{\alpha^3}\right) & \cdots & \rho^{\eta}(\mathbf{s}_1,\mathbf{s}_n)\left(\frac{||\mathbf{s}_1-\mathbf{s}_n||^2}{\alpha^4}-2\cdot\frac{||\mathbf{s}_1-\mathbf{s}_n||}{\alpha^3}\right) \\ \rho^{\eta}(\mathbf{s}_2,\mathbf{s}_1)\left(\frac{||\mathbf{s}_2-\mathbf{s}_1||^2}{\alpha^4}-2\cdot\frac{||\mathbf{s}_2-\mathbf{s}_1||}{\alpha^3}\right) & \cdots & \rho^{\eta}(\mathbf{s}_2,\mathbf{s}_n)\left(\frac{||\mathbf{s}_2-\mathbf{s}_n||^2}{\alpha^4}-2\cdot\frac{||\mathbf{s}_2-\mathbf{s}_n||}{\alpha^3}\right) \\ \vdots & \ddots & \vdots \\ \rho^{\eta}(\mathbf{s}_n,\mathbf{s}_1)\left(\frac{||\mathbf{s}_n-\mathbf{s}_1||^2}{\alpha^4}-2\cdot\frac{||\mathbf{s}_n-\mathbf{s}_1||}{\alpha^3}\right) & \cdots & \rho^{\eta}(\mathbf{s}_n,\mathbf{s}_n)\left(\frac{||\mathbf{s}_n-\mathbf{s}_n||^2}{\alpha^4}-2\cdot\frac{||\mathbf{s}_n-\mathbf{s}_n||}{\alpha^3}\right) \end{pmatrix}$$

developing by terms

Term (a)

$$\begin{aligned} \frac{\partial}{\partial\alpha}\left[\mathbf{R}^{-}\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\right)\otimes\mathbf{T}^{AR}\right] &= \frac{\partial\,\mathbf{R}^{-}}{\partial\alpha}\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\right)\otimes\mathbf{T}^{AR} \\ &\quad+\mathbf{R}^{-}\,\frac{\partial}{\partial\alpha}\left[\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\right)\otimes\mathbf{T}^{AR}\right] \\ &= \left(-\mathbf{R}^{-}\left(\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\right)\otimes\mathbf{T}^{AR}\right)\mathbf{R}^{-}\right) \\ &\quad\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\right)\otimes\mathbf{T}^{AR}+\mathbf{R}^{-}\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\left(\frac{\mathbf{H}^2}{\alpha^4}-2\frac{\mathbf{H}}{\alpha^3}\right)\right)\otimes\mathbf{T}^{AR} \end{aligned}$$

Term (b)

$$\begin{aligned} \frac{\partial}{\partial\alpha}\left[\mathbf{R}^{-}\left(\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\right)\otimes\mathbf{T}^{AR}\right)\mathbf{R}^{-}\right] &= \frac{\partial\,\mathbf{R}^{-}}{\partial\alpha}\left(\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\right)\otimes\mathbf{T}^{AR}\right)\mathbf{R}^{-} \\ &\quad+\mathbf{R}^{-}\,\frac{\partial}{\partial\alpha}\left[\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\right)\otimes\mathbf{T}^{AR}\right]\mathbf{R}^{-} \\ &\quad+\mathbf{R}^{-}\left(\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\right)\otimes\mathbf{T}^{AR}\right)\frac{\partial\,\mathbf{R}^{-}}{\partial\alpha} \\ &= \left(-\mathbf{R}^{-}\left(\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\right)\otimes\mathbf{T}^{AR}\right)\mathbf{R}^{-}\right)\left(\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\right)\otimes\mathbf{T}^{AR}\right)\mathbf{R}^{-} \\ &\quad+\mathbf{R}^{-}\left(\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\left(\frac{\mathbf{H}^2}{\alpha^4}-2\frac{\mathbf{H}}{\alpha^3}\right)\right)\otimes\mathbf{T}^{AR}\right)\mathbf{R}^{-} \\ &\quad-\mathbf{R}^{-}\left(\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\right)\otimes\mathbf{T}^{AR}\right)\mathbf{R}^{-}\left(\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\right)\otimes\mathbf{T}^{AR}\right)\mathbf{R}^{-} \\ &= -2\,\mathbf{R}^{-}\left(\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\right)\otimes\mathbf{T}^{AR}\right)\mathbf{R}^{-}\left(\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\right)\otimes\mathbf{T}^{AR}\right)\mathbf{R}^{-} \\ &\quad+\mathbf{R}^{-}\left(\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\left(\frac{\mathbf{H}^2}{\alpha^4}-2\frac{\mathbf{H}}{\alpha^3}\right)\right)\otimes\mathbf{T}^{AR}\right)\mathbf{R}^{-} \end{aligned}$$

Finally,

$$\begin{aligned}\frac{\partial^2 q^{(2)}}{\partial\alpha^2} = \frac{1}{2}\Bigg((T-1)\,\mathrm{tr}\Bigg\{\mathbf{R}^-\Bigg(\Bigg(\boldsymbol{\rho}^\eta(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\Bigg)\otimes\mathbf{T}^{AR}\Bigg)\mathbf{R}^-\Bigg(\Bigg(\boldsymbol{\rho}^\eta(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\Bigg)\otimes\mathbf{T}^{AR}\Bigg)\\ -\mathbf{R}^-\Bigg(\boldsymbol{\rho}^\eta(\alpha)\circ\Bigg(\frac{\mathbf{H}^2}{\alpha^4}-2\frac{\mathbf{H}}{\alpha^3}\Bigg)\Bigg)\otimes\mathbf{T}^{AR}\Bigg\}\\ +\frac{1}{\sigma_\eta^2}\,\mathrm{tr}\Bigg\{\Bigg(-2\,\mathbf{R}^-\Bigg(\Bigg(\boldsymbol{\rho}^\eta(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\Bigg)\otimes\mathbf{T}^{AR}\Bigg)\mathbf{R}^-\Bigg(\Bigg(\boldsymbol{\rho}^\eta(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\Bigg)\otimes\mathbf{T}^{AR}\Bigg)\mathbf{R}^-\\ +\mathbf{R}^-\Bigg(\Bigg(\boldsymbol{\rho}^\eta(\alpha)\circ\Bigg(\frac{\mathbf{H}^2}{\alpha^4}-2\frac{\mathbf{H}}{\alpha^3}\Bigg)\Bigg)\otimes\mathbf{T}^{AR}\Bigg)\mathbf{R}^-\Bigg)\Bigg(\boldsymbol{\Psi}(\boldsymbol{\Phi})\Bigg)\Bigg\}\Bigg)\end{aligned}$$

2.6 $\dfrac{\partial^2 q^{(2)}}{\partial\boldsymbol{\phi}\partial\alpha} = \dfrac{\partial^2 q^{(2)}}{\partial\alpha\partial\boldsymbol{\phi}}$

$$\begin{aligned}\frac{\partial^2 q^{(2)}}{\partial\boldsymbol{\phi}\partial\alpha} &= \frac{\partial}{\partial\alpha}\Bigg[\frac{1}{\sigma_\eta^2}\,\mathrm{tr}\Bigg\{\mathbf{R}^-\mathbf{S}_{21}\Bigg(\boldsymbol{v}\otimes\{\phi_{,j}\}\Bigg)\Bigg\}-\frac{1}{\sigma_\eta^2}\,\mathrm{tr}\Bigg\{\mathbf{R}^-\mathbf{S}_{11}\left(\boldsymbol{v}^\top\boldsymbol{v}\right)\Bigg\}\Bigg(\phi_j\Bigg)\Bigg]\\ &= \frac{1}{\sigma_\eta^2}\,\mathrm{tr}\Bigg\{\frac{\partial\,\mathbf{R}^-}{\partial\alpha}\,\mathbf{S}_{21}\Bigg(\boldsymbol{v}\otimes\{\phi_{,j}\}\Bigg)\Bigg\}-\frac{1}{\sigma_\eta^2}\,\mathrm{tr}\Bigg\{\frac{\partial\,\mathbf{R}^-}{\partial\alpha}\,\mathbf{S}_{11}\left(\boldsymbol{v}^\top\boldsymbol{v}\right)\Bigg\}\Bigg(\phi_j\Bigg)\\ \frac{\partial^2 q^{(2)}}{\partial\boldsymbol{\phi}\partial\alpha} &= -\frac{1}{\sigma_\eta^2}\,\mathrm{tr}\Bigg\{\Bigg(\mathbf{R}^-\Bigg(\Bigg(\boldsymbol{\rho}^\eta(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\Bigg)\otimes\mathbf{T}^{AR}\Bigg)\mathbf{R}^-\Bigg)\mathbf{S}_{21}\Bigg(\boldsymbol{v}\otimes\{\phi_{,j}\}\Bigg)\Bigg\}\\ &\qquad+\frac{1}{\sigma_\eta^2}\,\mathrm{tr}\Bigg\{\Bigg(\mathbf{R}^-\Bigg(\Bigg(\boldsymbol{\rho}^\eta(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\Bigg)\otimes\mathbf{T}^{AR}\Bigg)\mathbf{R}^-\Bigg)\;\mathbf{S}_{11}\left(\boldsymbol{v}^\top\boldsymbol{v}\right)\Bigg\}\Bigg(\phi_j\Bigg)\\ &\qquad\forall\, j=1,\ldots,p\end{aligned}$$

2.7 $\dfrac{\partial^2 q^{(2)}}{\partial {\sigma_\eta^2}^2}$

$$\begin{aligned}
\frac{\partial^2 q^{(2)}}{\partial {\sigma_\eta^2}^2} &= \frac{1}{2}\frac{\partial}{\partial \sigma_\eta^2}\Big[-(T-1)\frac{np}{\sigma_\eta^2} + \operatorname{tr}\left\{ \frac{1}{(\sigma_\eta^2)^2}\, \mathbf{R}^{-}\Big(\mathbf{\Psi}(\mathbf{\Phi})\Big)\right\}\Big] \\
&= \frac{1}{2}\left(-(T-1)\, np\, \frac{\partial}{\partial \sigma_\eta^2}\left[\frac{1}{\sigma_\eta^2}\right] + \frac{\partial}{\partial \sigma_\eta^2}\left[\frac{1}{(\sigma_\eta^2)^2}\right] \operatorname{tr}\left\{\mathbf{R}^{-}\Big(\mathbf{\Psi}(\mathbf{\Phi})\Big)\right\}\right) \\
\frac{\partial^2 q^{(2)}}{\partial {\sigma_\eta^2}^2} &= (T-1)\frac{np}{2(\sigma_\eta^2)^2} - \frac{1}{(\sigma_\eta^2)^3}\ \operatorname{tr}\left\{\mathbf{R}^{-}\Big(\mathbf{\Psi}(\mathbf{\Phi})\Big)\right\}
\end{aligned}$$

2.8 $\dfrac{\partial^2 q^{(2)}}{\partial \boldsymbol{\phi} \partial \sigma_\eta^2} = \dfrac{\partial^2 q^{(2)}}{\partial \sigma_\eta^2 \partial \boldsymbol{\phi}}$

$$\begin{aligned}
\frac{\partial^2 q^{(2)}}{\partial \boldsymbol{\phi} \partial \sigma_\eta^2} &= \frac{\partial}{\partial \sigma_\eta^2}\left[\frac{1}{\sigma_\eta^2}\operatorname{tr}\left\{\mathbf{R}^{-}\mathbf{S}_{21}\Big(\boldsymbol{v}\otimes\{\phi_{,j}\}\Big)\right\} - \frac{1}{\sigma_\eta^2}\operatorname{tr}\left\{\mathbf{R}^{-}\mathbf{S}_{11}\left(\boldsymbol{v}^\top\boldsymbol{v}\right)\right\}\Big(\phi_j\Big)\right] \\
&= \frac{\partial}{\partial \sigma_\eta^2}\left[\frac{1}{\sigma_\eta^2}\right]\operatorname{tr}\left\{\mathbf{R}^{-}\mathbf{S}_{21}\Big(\boldsymbol{v}\otimes\{\phi_{,j}\}\Big)\right\} - \frac{\partial}{\partial \sigma_\eta^2}\left[\frac{1}{\sigma_\eta^2}\right]\operatorname{tr}\left\{\mathbf{R}^{-}\mathbf{S}_{11}\left(\boldsymbol{v}^\top\boldsymbol{v}\right)\right\}\Big(\phi_j\Big) \\
\frac{\partial^2 q^{(2)}}{\partial \boldsymbol{\phi} \partial \sigma_\eta^2} &= -\frac{1}{(\sigma_\eta^2)^2}\operatorname{tr}\left\{\mathbf{R}^{-}\mathbf{S}_{21}\Big(\boldsymbol{v}\otimes\{\phi_{,j}\}\Big)\right\} + \frac{1}{(\sigma_\eta^2)^2}\operatorname{tr}\left\{\mathbf{R}^{-}\mathbf{S}_{11}\left(\boldsymbol{v}^\top\boldsymbol{v}\right)\right\}\Big(\phi_j\Big) \\
&\qquad \forall\ j = 1, \ldots, p
\end{aligned}$$

2.9 $\dfrac{\partial^2 q^{(2)}}{\partial \alpha \partial \sigma_\eta^2} = \dfrac{\partial^2 q^{(2)}}{\partial \sigma_\eta^2 \partial \alpha}$

$$\begin{aligned}
\frac{\partial^2 q^{(2)}}{\partial \sigma_\eta^2 \partial \alpha} &= \frac{1}{2}\frac{\partial}{\partial \alpha}\Big[-(T-1)\frac{np}{\sigma_\eta^2} + \operatorname{tr}\left\{ \left(\frac{1}{(\sigma_\eta^2)^2}\, \mathbf{R}^{-}\right)\Big(\mathbf{\Psi}(\mathbf{\Phi})\Big)\right\}\Big] \\
&= \frac{1}{2(\sigma_\eta^2)^2}\operatorname{tr}\left\{\left(\frac{\partial\, \mathbf{R}^{-}}{\partial \alpha}\right)\Big(\mathbf{\Psi}(\mathbf{\Phi})\Big)\right\} \\
\frac{\partial^2 q^{(2)}}{\partial \sigma_\eta^2 \partial \alpha} &= -\frac{1}{2(\sigma_\eta^2)^2}\operatorname{tr}\left\{\left(\mathbf{R}^{-}\left(\left(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\right)\otimes\mathbf{T}^{AR}\right)\mathbf{R}^{-}\right)\Big(\mathbf{\Psi}(\mathbf{\Phi})\Big)\right\}
\end{aligned}$$

3. **First Derivates of the complete log-likelihood**

$$2\ell_c = 2\left(\ell^{(1)}(\boldsymbol{\beta}, \sigma_\omega^2 | \mathbf{Z}_{1:T}, \boldsymbol{\xi}_{1:T}) + \ell^{(2)}(\boldsymbol{\phi}, \alpha, \sigma_\eta^2 | \boldsymbol{\xi}_{1:T}) + \ell^{(3)}(\boldsymbol{\xi}_1)\right)$$

3.1 $\dfrac{\partial \ell^{(1)}}{\partial \boldsymbol{\beta}}$

$$\begin{aligned}
\frac{\partial 2\ell^{(1)}}{\partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}}\Big[- \ln\{\sigma_\omega^2\} \sum_{t=1}^{T} n_t - \frac{1}{\sigma_\omega^2} \sum_{t=1}^{T} \Big(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_t\Big)^\top \Big(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_t\Big)\Big] \\
&= -\frac{1}{\sigma_\omega^2} \sum_{t=1}^{T} \frac{\partial}{\partial \boldsymbol{\beta}}\Big[\Big(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_t\Big)^\top \Big(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_t\Big)\Big] \\
&= -\frac{1}{\sigma_\omega^2} \sum_{t=1}^{T} -2\mathbf{X}_t^\top \Big(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_t\Big) \\
\frac{\partial \ell^{(1)}}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma_\omega^2} \sum_{t=1}^{T} \Big(\mathbf{X}_t^\top\mathbf{Z}_t - \mathbf{X}_t^\top\mathbf{X}_t\boldsymbol{\beta} - \mathbf{X}_t^\top\boldsymbol{\Lambda}_t\boldsymbol{\xi}_t\Big)
\end{aligned}$$

3.2 $\dfrac{\partial \ell^{(1)}}{\partial \sigma_\omega^2}$

$$\begin{aligned}
\frac{\partial 2\ell^{(1)}}{\partial \sigma_\omega^2} &= \frac{\partial}{\partial \sigma_\omega^2}\Big[- \ln\{\sigma_\omega^2\} \sum_{t=1}^{T} n_t - \frac{1}{\sigma_\omega^2} \sum_{t=1}^{T} \Big(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_t\Big)^\top \Big(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_t\Big)\Big] \\
\frac{\partial \ell^{(1)}}{\partial \sigma_\omega^2} &= \frac{1}{2}\left(- \sum_{t=1}^{T} n_t \frac{1}{\sigma_\omega^2} + \frac{1}{(\sigma_\omega^2)^2} \sum_{t=1}^{T} \Big(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_t\Big)^\top \Big(\mathbf{Z}_t - \mathbf{X}_t\boldsymbol{\beta} - \boldsymbol{\Lambda}_t\boldsymbol{\xi}_t\Big)\right)
\end{aligned}$$

3.3 $\dfrac{\partial \ell^{(2)}}{\partial \boldsymbol{\phi}}$

$$
\begin{aligned}
\frac{\partial\, 2\ell^{(2)}}{\partial \boldsymbol{\phi}} &= \frac{\partial}{\partial \boldsymbol{\phi}}\Big[-(T-1)\left(\ln|\mathbf{Q}|\right) - \sum_{t=2}^{T}\Big(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big)^{\top}\mathbf{Q}^{-}\Big(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big)\Big] \\
&= -\sum_{t=2}^{T}\frac{\partial}{\partial \boldsymbol{\phi}}\Big[-\boldsymbol{\xi}_t^{\top}\mathbf{Q}^{-}\boldsymbol{\Phi}\boldsymbol{\xi}_{t-1} - \boldsymbol{\xi}_{t-1}^{\top}\boldsymbol{\Phi}^{\top}\mathbf{Q}^{-}\boldsymbol{\xi}_t + \boldsymbol{\xi}_{t-1}^{\top}\boldsymbol{\Phi}^{\top}\mathbf{Q}^{-}\boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big] \\
&= \sum_{t=2}^{T}\frac{\partial}{\partial \boldsymbol{\phi}}\Big[\boldsymbol{\xi}_t^{\top}\mathbf{Q}^{-}\boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big] + \frac{\partial}{\partial \boldsymbol{\phi}}\Big[\boldsymbol{\xi}_{t-1}^{\top}\boldsymbol{\Phi}^{\top}\mathbf{Q}^{-}\boldsymbol{\xi}_t\Big] - \frac{\partial}{\partial \boldsymbol{\phi}}\Big[\boldsymbol{\xi}_{t-1}^{\top}\boldsymbol{\Phi}^{\top}\mathbf{Q}^{-}\boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big] \\
&\quad \text{decomposing } \boldsymbol{\Phi} = \Big(\boldsymbol{\upsilon}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big) + \mathbf{G} \\
&= \sum_{t=2}^{T}\frac{\partial}{\partial \boldsymbol{\phi}}\Big[\boldsymbol{\xi}_t^{\top}\mathbf{Q}^{-}\Big(\boldsymbol{\upsilon}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big)\boldsymbol{\xi}_{t-1}\Big] + \frac{\partial}{\partial \boldsymbol{\phi}}\Big[\boldsymbol{\xi}_{t-1}^{\top}\Big(\boldsymbol{\upsilon}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big)^{\top}\mathbf{Q}^{-}\boldsymbol{\xi}_t\Big] \\
&\quad - \frac{\partial}{\partial \boldsymbol{\phi}}\Big[\boldsymbol{\xi}_{t-1}^{\top}\Big(\boldsymbol{\upsilon}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big)^{\top}\mathbf{Q}^{-}\Big(\boldsymbol{\upsilon}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big)\boldsymbol{\xi}_{t-1}\Big] \\
&\quad - \frac{\partial}{\partial \boldsymbol{\phi}}\Big[\boldsymbol{\xi}_{t-1}^{\top}\Big(\boldsymbol{\upsilon}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big)^{\top}\mathbf{Q}^{-}\mathbf{G}\,\boldsymbol{\xi}_{t-1}\Big] - \frac{\partial}{\partial \boldsymbol{\phi}}\Big[\boldsymbol{\xi}_{t-1}^{\top}\mathbf{G}^{\top}\mathbf{Q}^{-}\Big(\boldsymbol{\upsilon}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big)\boldsymbol{\xi}_{t-1}\Big] \\
&= \sum_{t=2}^{T}\boldsymbol{\xi}_t^{\top}\mathbf{Q}^{-}\frac{\partial}{\partial \boldsymbol{\phi}}\Big[\Big(\boldsymbol{\upsilon}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big)\Big]\boldsymbol{\xi}_{t-1} + \boldsymbol{\xi}_{t-1}^{\top}\frac{\partial}{\partial \boldsymbol{\phi}}\Big[(\boldsymbol{\upsilon}\otimes\boldsymbol{\phi})\Big]\mathbf{Q}^{-}\boldsymbol{\xi}_t \\
&\quad - \Big(\boldsymbol{\xi}_{t-1}^{\top}\frac{\partial}{\partial \boldsymbol{\phi}}\Big[(\boldsymbol{\upsilon}\otimes\boldsymbol{\phi})\Big]\mathbf{Q}^{-}\Big(\boldsymbol{\upsilon}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big)\boldsymbol{\xi}_{t-1} \\
&\quad + \boldsymbol{\xi}_{t-1}^{\top}(\boldsymbol{\upsilon}\otimes\boldsymbol{\phi})\,\mathbf{Q}^{-}\frac{\partial}{\partial \boldsymbol{\phi}}\Big[\Big(\boldsymbol{\upsilon}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big)\Big]\boldsymbol{\xi}_{t-1}\Big) - \boldsymbol{\xi}_{t-1}^{\top}\frac{\partial}{\partial \boldsymbol{\phi}}\Big[(\boldsymbol{\upsilon}\otimes\boldsymbol{\phi})\Big]\mathbf{Q}^{-}\mathbf{G}\,\boldsymbol{\xi}_{t-1} \\
&\quad - \boldsymbol{\xi}_{t-1}^{\top}\mathbf{G}^{\top}\mathbf{Q}^{-}\frac{\partial}{\partial \boldsymbol{\phi}}\Big[\Big(\boldsymbol{\upsilon}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big)\Big]\boldsymbol{\xi}_{t-1} \\
\frac{\partial \ell^{(2)}}{\partial \boldsymbol{\phi}} &= \frac{1}{2}\Big(\sum_{t=2}^{T}\boldsymbol{\xi}_t^{\top}\mathbf{Q}^{-}\Big(\boldsymbol{\upsilon}^{\top}\otimes\{\phi_{,j}\}^{\top}\Big)\boldsymbol{\xi}_{t-1} \;+\; \boldsymbol{\xi}_{t-1}^{\top}(\boldsymbol{\upsilon}\otimes\{\phi_{,j}\})\,\mathbf{Q}^{-}\boldsymbol{\xi}_t \\
&\quad - \;\boldsymbol{\xi}_{t-1}^{\top}(\boldsymbol{\upsilon}\otimes\{\phi_{,j}\})\,\mathbf{Q}^{-}\Big(\boldsymbol{\upsilon}^{\top}\otimes\boldsymbol{\phi}^{\top}\Big)\boldsymbol{\xi}_{t-1} - \;\boldsymbol{\xi}_{t-1}^{\top}(\boldsymbol{\upsilon}\otimes\boldsymbol{\phi})\,\mathbf{Q}^{-}\Big(\boldsymbol{\upsilon}^{\top}\otimes\{\phi_{,j}\}^{\top}\Big)\boldsymbol{\xi}_{t-1} \\
&\quad - \;\boldsymbol{\xi}_{t-1}^{\top}(\boldsymbol{\upsilon}\otimes\{\phi_{,j}\})\,\mathbf{Q}^{-}\mathbf{G}\,\boldsymbol{\xi}_{t-1} \;-\; \boldsymbol{\xi}_{t-1}^{\top}\mathbf{G}^{\top}\mathbf{Q}^{-}\Big(\boldsymbol{\upsilon}^{\top}\otimes\{\phi_{,j}\}^{\top}\Big)\boldsymbol{\xi}_{t-1}\Big)
\end{aligned}
$$

3.4 $\dfrac{\partial \ell^{(2)}}{\partial \alpha}$

$$\frac{\partial 2\ell^{(2)}}{\partial \alpha} = \frac{\partial}{\partial \alpha}\Big[-(T-1)\left(\ln|\mathbf{Q}|\right) - \sum_{t=2}^{T}\Big(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big)^{\top}\mathbf{Q}^{-}\Big(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big)\Big]$$

$$= -(T-1)\frac{\partial}{\partial \alpha}\Big[\ln|\mathbf{R}|\Big] - \sum_{t=2}^{T}\Big(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big)^{\top}\Big(\frac{1}{\sigma_\eta^2}\frac{\partial\, \mathbf{R}^{-}}{\partial \alpha}\Big)\Big(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big)$$

$$\frac{\partial \ell^{(2)}}{\partial \alpha} = \frac{1}{2}\Big(-(T-1)\ \mathrm{tr}\Big\{\mathbf{R}^{-}\Big(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\Big)\otimes\mathbf{T}^{AR}\Big\} - \frac{1}{\sigma_\eta^2}\sum_{t=2}^{T}\Big(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big)^{\top}$$

$$\Big(-\mathbf{R}^{-}\Big(\Big(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\Big)\otimes\mathbf{T}^{AR}\Big)\mathbf{R}^{-} + \Big(\mathbf{R}^{-}\mathbf{R}^{-}\Big(\Big(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\Big)\otimes\mathbf{T}^{AR}\Big)\Big)$$

$$(\mathbf{I}-\mathbf{R}\mathbf{R}^{-}) + (\mathbf{I}-\mathbf{R}^{-}\mathbf{R})\Big(\Big(\Big(\boldsymbol{\rho}^{\eta}(\alpha)\circ\frac{\mathbf{H}}{\alpha^2}\Big)\otimes\mathbf{T}^{AR}\Big)\mathbf{R}^{-}\mathbf{R}^{-}\Big)\Big)\Big(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big)\Big)$$

3.5 $\dfrac{\partial \ell^{(2)}}{\partial \sigma_\eta^2}$

$$\frac{\partial 2\ell^{(2)}}{\partial \sigma_\eta^2} = \frac{\partial}{\partial \sigma_\eta^2}\Big[-(T-1)\left(\ln|\mathbf{Q}|\right) - \sum_{t=2}^{T}\Big(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big)^{\top}\mathbf{Q}^{-}\Big(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big)\Big]$$

$$= \frac{\partial}{\partial \sigma_\eta^2}\Big[-(T-1)\left(np\,\ln\{\sigma_\eta^2\} + \ln|\mathbf{R}|\right)\Big]$$

$$-\sum_{t=2}^{T}\Big(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big)^{\top}\Big(\frac{\partial}{\partial \sigma_\eta^2}\Big[\frac{1}{\sigma_\eta^2}\mathbf{R}^{-}\Big]\Big)\Big(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big)$$

$$= -(T-1)\ np\ \frac{\partial}{\partial \sigma_\eta^2}\Big[\ln\{\sigma_\eta^2\}\Big] - \sum_{t=2}^{T}\Big(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big)^{\top}\Big(\frac{\partial}{\partial \sigma_\eta^2}\Big[\frac{1}{\sigma_\eta^2}\Big]\mathbf{R}^{-}\Big)\Big(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big)$$

$$= -(T-1)\ \frac{np}{\sigma_\eta^2} + \sum_{t=2}^{T}\Big(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big)^{\top}\Big(\frac{1}{(\sigma_\eta^2)^2}\mathbf{R}^{-}\Big)\Big(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big)$$

$$\frac{\partial \ell^{(2)}}{\partial \sigma_\eta^2} = \frac{1}{2}\Big(-(T-1)\ \frac{np}{\sigma_\eta^2} + \sum_{t=2}^{T}\Big(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big)^{\top}\Big(\frac{1}{\sigma_\eta^2}\mathbf{Q}^{-}\Big)\Big(\boldsymbol{\xi}_t - \boldsymbol{\Phi}\boldsymbol{\xi}_{t-1}\Big)\Big)$$

VII.3 Covariates of Meteorological Stations (EMAs)

	EMA	Region	x	y	*alt*
1	AGRICHILE_958	Maule	830688.23	6090139.09	223.00
2	AGRICHILE_959	Maule	814169.94	6088351.31	144.00
3	AGRICHILE_960	Maule	785432.66	5989627.28	180.00
4	AGRICHILE_963	Maule	851203.37	6114301.75	273.00
5	DMC_152	Maule	845463.02	6124141.95	230.00
6	DMC_156	Maule	808782.27	6079755.00	119.00
7	DMC_161	Maule	791957.50	6002803.94	162.00
8	DMC_974	Maule	807773.05	6028775.50	160.00
9	INIA_11	Maule	761320.39	6043827.47	162.00
10	INIA_12	Maule	727182.92	6006552.71	172.00
11	INIA_135	Maule	819228.97	6062403.29	221.00
12	INIA_136	Maule	809746.68	6037170.75	150.00
13	INIA_14	Maule	736745.73	6005964.00	167.00
14	INIA_154	Maule	775488.48	5984538.90	548.00
15	INIA_212	Maule	717905.11	6031615.62	172.00
16	INIA_214	Maule	777386.31	6108539.89	48.00
17	INIA_215	Maule	853700.07	6109886.50	310.00
18	INIA_22	Maule	725160.90	6045739.81	60.00
19	INIA_227	Maule	791374.94	6031430.19	124.00
20	INIA_46	Maule	744415.71	6017539.16	162.00
21	INIA_7	Maule	738042.66	6015307.95	151.00
22	AGRICHILE_961	Ñuble	785108.50	5979296.31	180.00
23	DMC_157	Ñuble	764820.36	5946837.69	151.00
24	DMC_158	Ñuble	819885.69	5909798.20	1708.00
25	DMC_160	Ñuble	779433.32	5919619.04	396.00
26	INIA_139	Ñuble	768411.68	5966489.15	162.00
27	INIA_17	Ñuble	776063.68	5952263.07	194.00
28	INIA_211	Ñuble	710857.68	5965670.43	36.00
29	INIA_23	Ñuble	735612.67	5953815.76	109.00
30	INIA_24	Ñuble	722273.18	5941195.15	82.00
31	INIA_47	Ñuble	733602.57	5968742.21	91.00

	EMA	Region	x	y	*alt*
32	INIA_49	Ñuble	765421.54	5885261.90	265.00
33	INIA_73	Ñuble	781973.34	5909724.59	314.00
34	DMC_111	Biobío	672899.69	5927967.08	17.00
35	DMC_162	Biobío	741025.07	5850948.57	173.00
36	DMC_955	Biobío	632073.73	5753221.75	153.00
37	INIA_10	Biobío	681879.44	5939539.20	190.00
38	INIA_15	Biobío	675004.69	5912922.01	23.00
39	INIA_161	Biobío	631337.86	5762574.11	143.00
40	INIA_21	Biobío	743845.69	5853463.62	195.00
41	INIA_308	Biobío	638542.05	5870132.51	120.00
42	INIA_310	Biobío	705092.11	5936232.32	268.00
43	INIA_312	Biobío	713501.97	5870383.01	120.00
44	INIA_322	Biobío	655346.95	5790674.60	46.00
45	INIA_36	Biobío	636976.06	5820357.26	125.00
46	INIA_84	Biobío	630310.80	5828379.33	194.00
47	UCONCE_65	Biobío	685044.39	5953630.86	0.00
48	AGRICHILE_962	La Araucanía	731948.04	5678562.05	208.00
49	ANPROS_74	La Araucanía	722874.19	5699912.45	157.00
50	DMC_163	La Araucanía	705735.82	5706329.57	86.00
51	DMC_164	La Araucanía	817477.12	5737426.79	912.00
52	DMC_165	La Araucanía	703401.74	5687880.17	96.00
53	DMC_180	La Araucanía	703376.60	5819876.38	69.00
54	DMC_181	La Araucanía	732541.40	5765451.22	336.00
55	DMC_75	La Araucanía	749409.28	5658165.08	393.00
56	DMC_79	La Araucanía	700109.20	5698699.26	74.00
57	INIA_138	La Araucanía	690636.81	5718242.99	55.00
58	INIA_140	La Araucanía	724413.00	5650775.30	266.00
59	INIA_141	La Araucanía	787972.19	5696394.70	527.00
60	INIA_142	La Araucanía	795225.63	5635230.84	393.00
61	INIA_143	La Araucanía	731635.19	5734838.26	263.00
62	INIA_144	La Araucanía	716125.41	5816590.23	88.00

	EMA	Region	x	y	*alt*
63	INIA_145	La Araucanía	738261.79	5787099.92	379.00
64	INIA_146	La Araucanía	738673.03	5765396.36	388.00
65	INIA_147	La Araucanía	697140.58	5741177.39	167.00
66	INIA_148	La Araucanía	835486.10	5705924.86	1084.00
67	INIA_157	La Araucanía	677111.72	5788934.42	76.00
68	INIA_164	La Araucanía	744203.20	5689086.52	293.00
69	INIA_168	La Araucanía	679767.24	5673272.82	56.00
70	INIA_18	La Araucanía	643210.57	5716244.46	64.00
71	INIA_203	La Araucanía	658943.13	5670055.03	18.00
72	INIA_205	La Araucanía	770222.13	5649503.85	279.00
73	INIA_206	La Araucanía	685538.04	5779096.60	76.00
74	INIA_208	La Araucanía	696371.31	5633227.26	86.00
75	INIA_209	La Araucanía	699632.07	5783898.76	173.00
76	INIA_233	La Araucanía	726793.37	5681574.34	168.00
77	INIA_235	La Araucanía	727020.94	5742707.53	299.00
78	INIA_236	La Araucanía	716448.06	5698806.35	130.00
79	INIA_237	La Araucanía	705969.01	5810708.90	81.00
80	INIA_238	La Araucanía	736750.55	5774586.41	379.00
81	INIA_239	La Araucanía	700178.88	5724360.47	60.00
82	INIA_28	La Araucanía	652814.00	5721740.28	322.00
83	INIA_29	La Araucanía	652556.94	5690783.18	93.00
84	INIA_31	La Araucanía	673439.62	5683875.35	24.00
85	INIA_33	La Araucanía	708900.91	5759370.73	254.00
86	INIA_40	La Araucanía	707037.89	5667669.40	115.00
87	INIA_41	La Araucanía	683222.53	5708176.36	50.00
88	INIA_48	La Araucanía	769350.81	5744187.18	549.00
89	INIA_8	La Araucanía	724491.41	5714596.20	200.00

VII.4 Simulations Analysis

Below are examples of the convergence plots (Figures VII.1, VII.2, VII.3, VII.4) from the Monte Carlo simulation of the latent state process for selected locations and time points $[n, t]$, using the KS distribution. These simulations were used to obtain a numerical approximation of the expected missing data FIM.

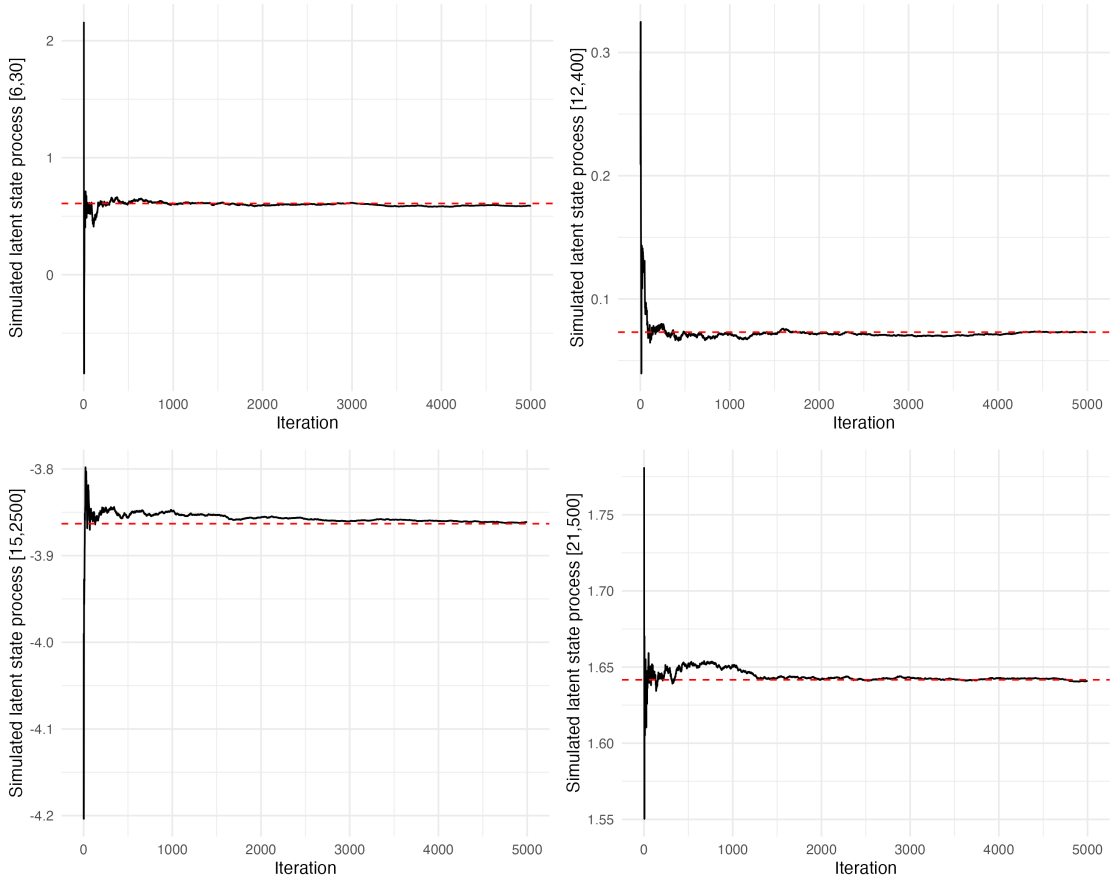

Figure VII.1: Convergence of the simulated latent state process at selected locations and time points $[n, t]$ in the Maule Region. The dashed red line represents $\boldsymbol{\xi}_{t|T}$ which each sample was simulated accordingly.

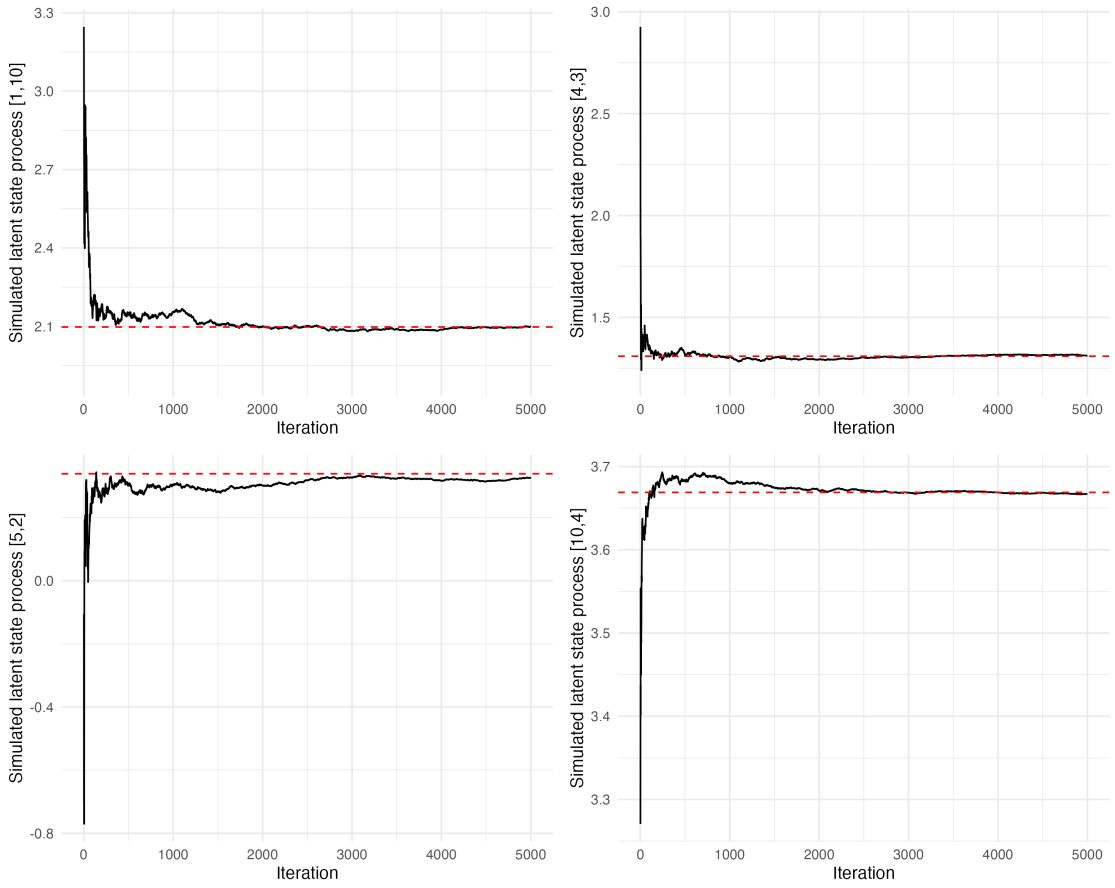

Figure VII.2: Convergence of the simulated latent state process at selected locations and time points $[n, t]$ in the Ñuble Region. The dashed red line represents $\boldsymbol{\xi}_{t|T}$ which each sample was simulated accordingly.

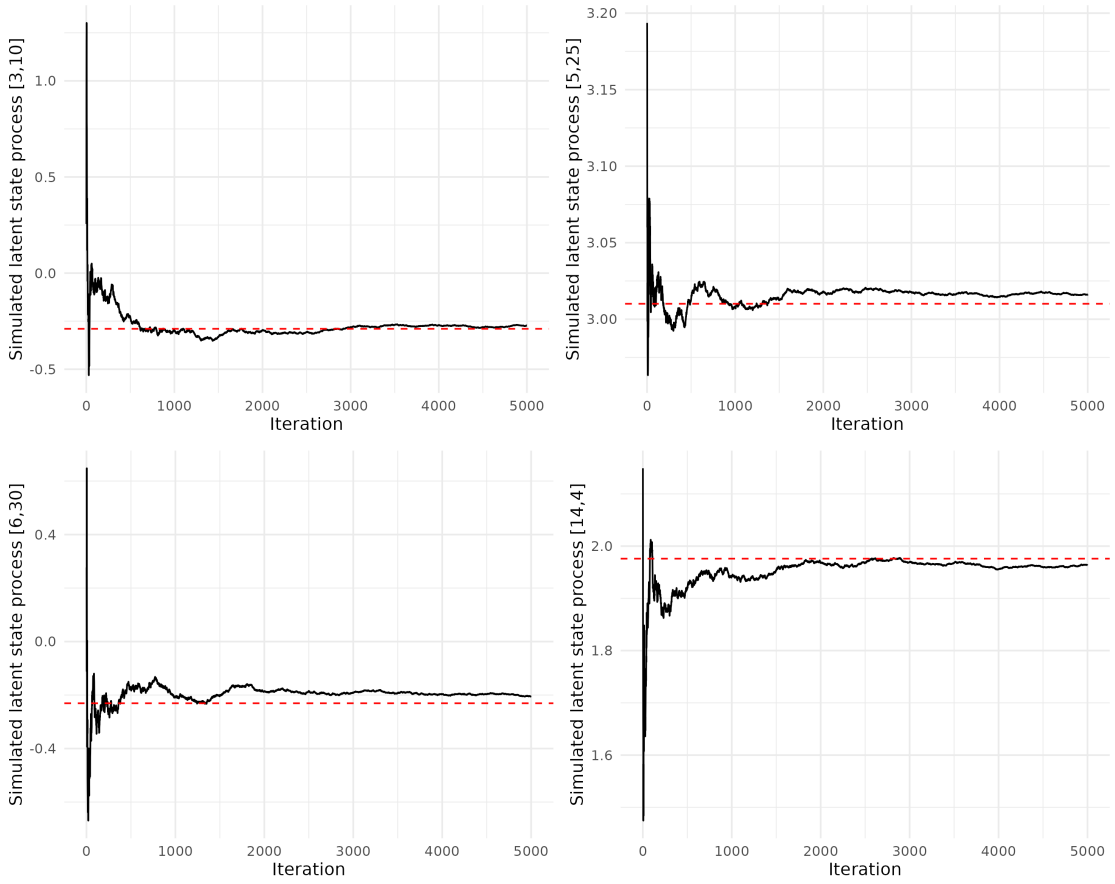

Figure VII.3: Convergence of the simulated latent state process at selected locations and time points $[n, t]$ in the Biobío Region. The dashed red line represents $\boldsymbol{\xi}_{t|T}$ which each sample was simulated accordingly.

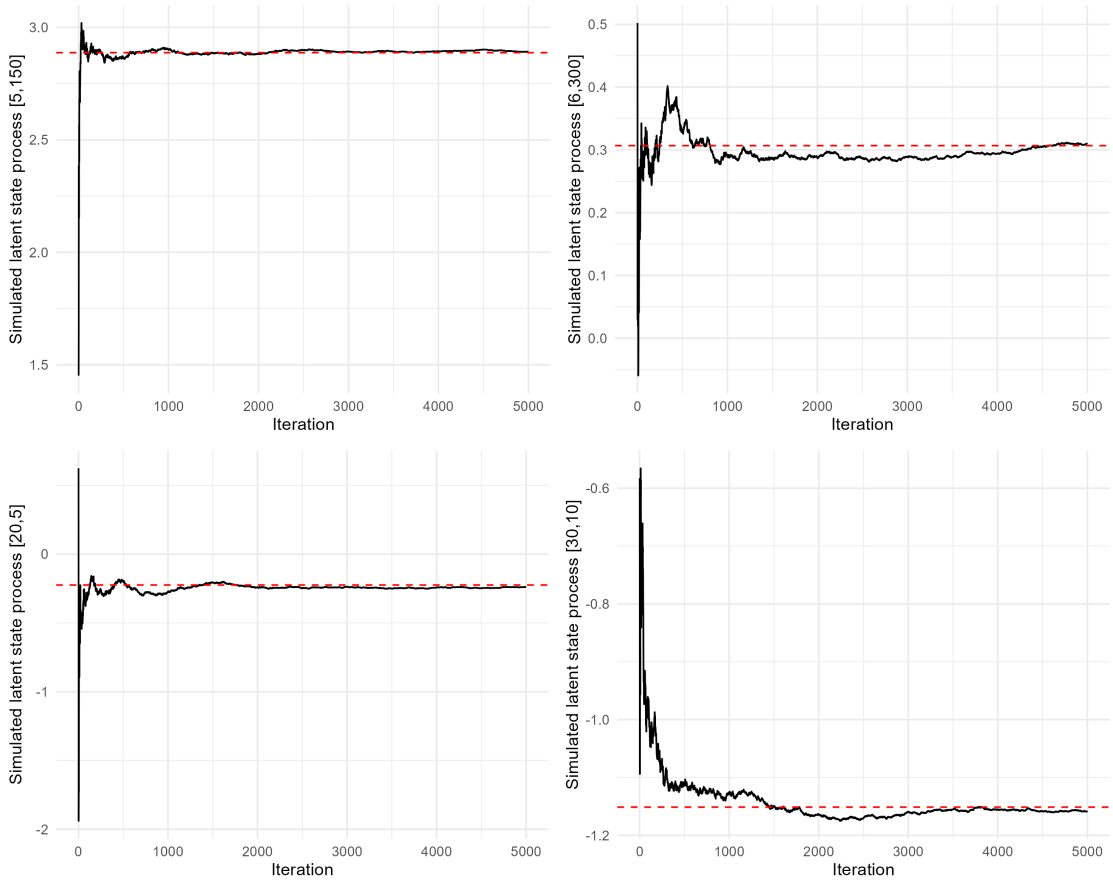

Figure VII.4: Convergence of the simulated latent state process at selected locations and time points $[n, t]$ in the La Araucanía Region. The dashed red line represents $\boldsymbol{\xi}_{t|T}$ which each sample was simulated accordingly.

Normality of the bootstrap samples was evaluated using the Lilliefors (Kolmogorov-Smirnov) test. Additionally, for the real data analysis, a subset of samples from each region was selected and plotted to verify that their behavior was consistent with the original data.(Figures VII.5, VII.6, VII.7, VII.8)

Table VII.1: P-value for Lilliefors Normality Test of the parameter estimation made by STPB

	Scenario 1					
Parameter	$NA_1 = 0$	$NA_2 = 0.1$	$NA_3 = 0.2$	$NA_4 = 0.3$	$NA_5 = 0.35$	$NA_6 = 0.4$
$\widehat{\beta_0}$	$< 0.0001^*$	$< 0.0001^*$	$< 0.0001^*$	$< 0.0001^*$	0.0008^*	0.0009^**
$\widehat{\sigma}^2_{\omega}$	0.0033^*	0.0060^**	0.0005^*	$< 0.0001^*$	0.0903	$< 0.0001^*$
$\widehat{\phi}$	0.6064	0.0056^**	0.7927	0.3353	0.8476	0.4099
$\widehat{\alpha}$	0.4857	0.8059	0.07324	0.0512	0.6146	0.0635
$\widehat{\sigma}^2_{\eta}$	0.1928	0.1708	0.1192	0.8124	0.7369	0.4557
	Scenario 2					
Parameter	$NA_1 = 0$	$NA_2 = 0.1$	$NA_3 = 0.2$	$NA_4 = 0.3$	$NA_5 = 0.35$	$NA_6 = 0.4$
$\widehat{\beta_0}$	$< 0.0001^*$	0.0058^*	0.1315	0.0019^*	0.5313	$< 0.0001^*$
$\widehat{\sigma}^2_{\omega}$	$< 0.0001^*$	$< 0.0001^*$	0.0006^*	$< 0.0001^*$	$< 0.0001^*$	$< 0.0001^*$
$\widehat{\phi}$	0.5350	0.1920	0.7162	0.7162	0.9151	0.4361
$\widehat{\alpha}$	0.1602	0.5478	0.3238	0.4311	0.7988	0.1502
$\widehat{\sigma}^2_{\eta}$	0.1636	0.3974	0.2397	0.0077^*	0.2019	0.0009^*

	Region			
Parameter	Maule	Ñuble	Biobío	La Araucanía
$\widehat{\sigma}^2_{\omega}$	$< 0.0001^*$	0.0001^{**}	0.0097^{**}	$< 0.0001^*$
$\widehat{\phi}$	0.8936	0.5957	0.9901	0.2386
$\widehat{\alpha}$	$< 0.0001^{**}$	0.1689	0.0010^{**}	$< 0.0001^*$
$\widehat{\sigma}^2_{\eta}$	$< 0.0001^{**}$	0.9295	$< 0.0001^{**}$	$< 0.0001^*$

None of the $\widehat{\boldsymbol{\beta}}$ followed a normal distribution. In the Table **??**, * indicates statistical significance, meaning the data do not follow a normal distribution. ** also indicates statistical significance. However, analysis of the histogram, QQ plot, kurtosis (close to 3), and symmetry (close to 0) suggests that the data exhibit a behavior consistent with a normal

distribution.

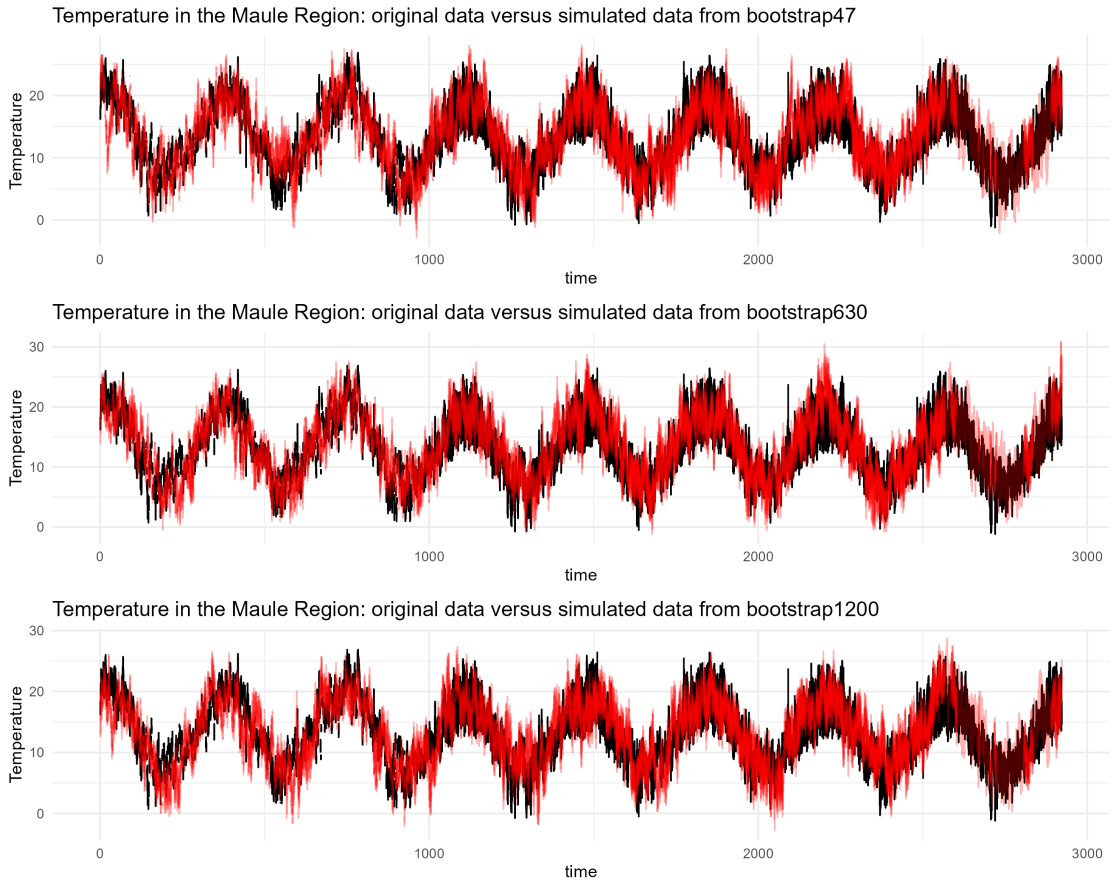

Figure VII.5: Temperature in the Maule Region: original data (black) and simulated data from STPB (red)

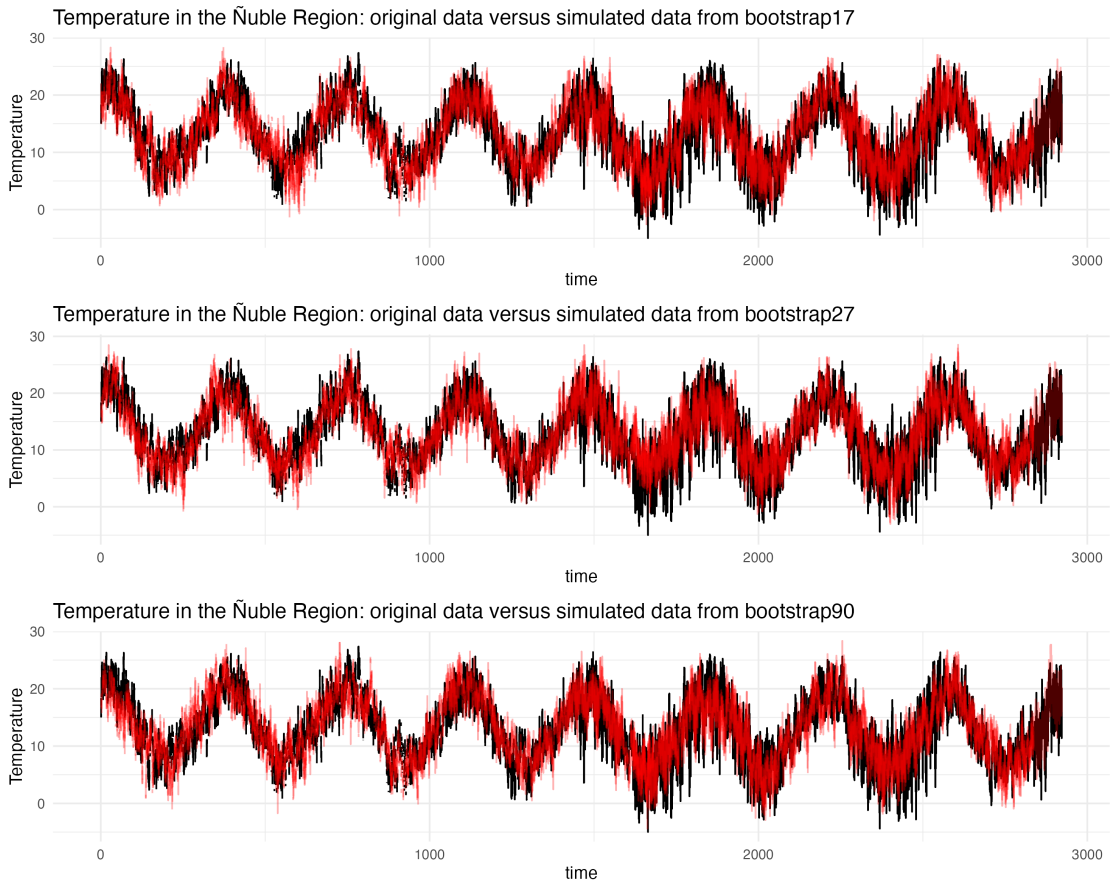

Figure VII.6: Temperature in the Ñuble Region: original data (black) and simulated data from STPB (red)

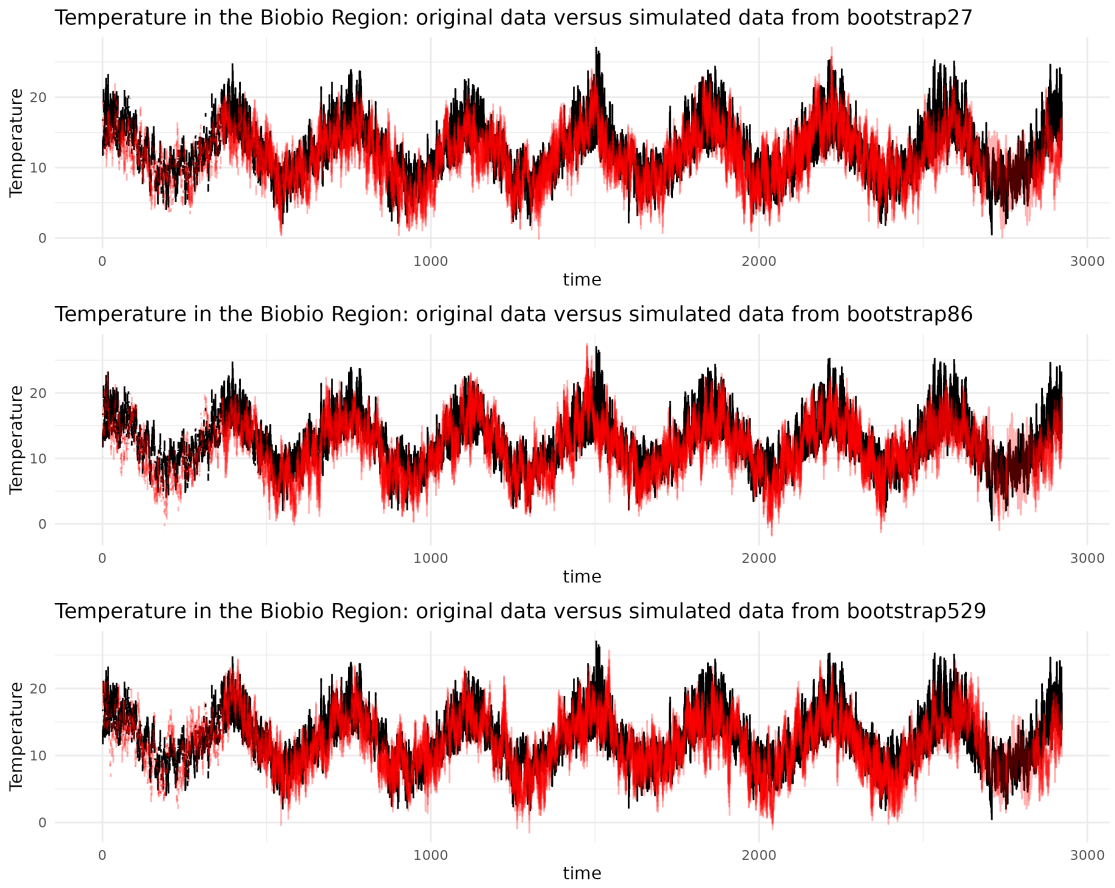

Figure VII.7: Temperature in the Biobío Region: original data (black) and simulated data from STPB (red)

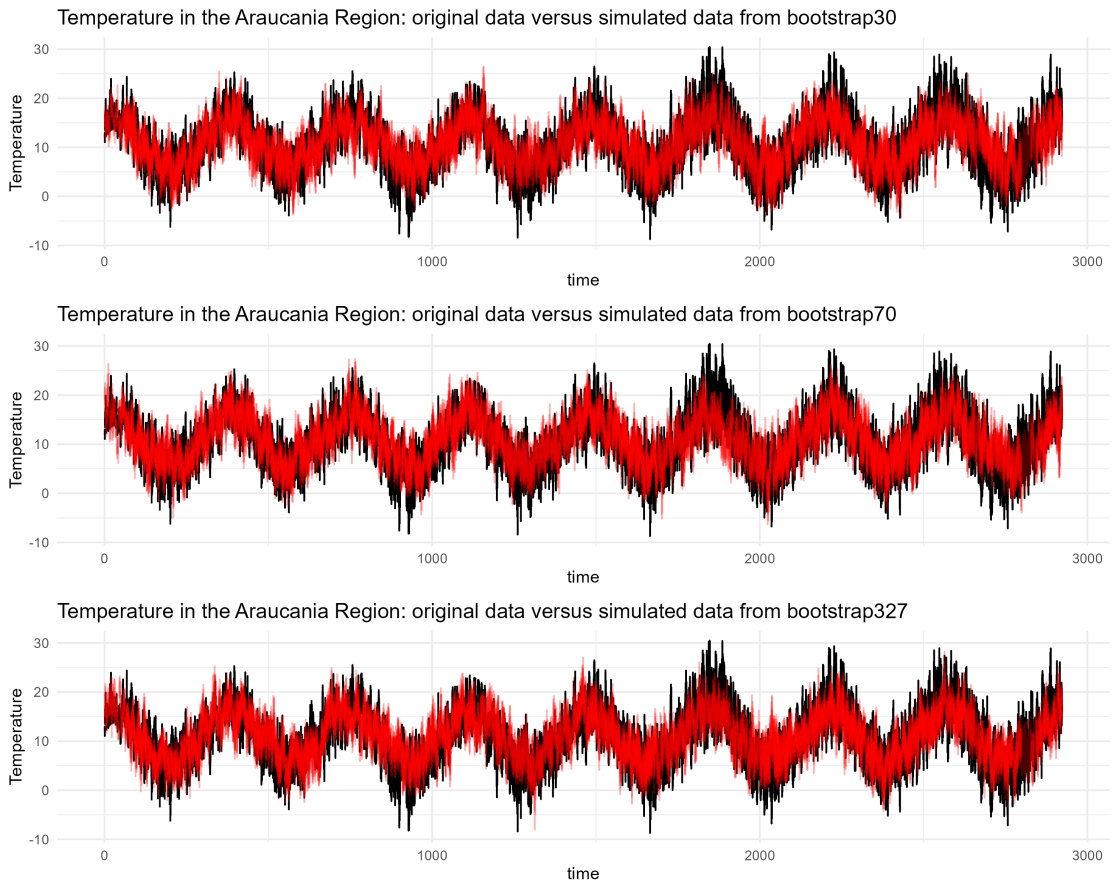

Figure VII.8: Temperature in the Araucanía Region: original data (black) and simulated data from STPB (red)