Multivariate Kalman Filtering for Spatio-Temporal Processes

Guillermo Ferreira (✉ gpuc.ferreira@gmail.com)

Universidad de Concepcion <https://orcid.org/0000-0002-7233-9885>

Jorge Mateu

Jaume I University: Universitat Jaume I

Emilio Porcu

Trinity School

Alfredo Alegría

Universidad Técnica Federico Santa María: Universidad Tecnica Federico Santa Maria

Research Article

Keywords: Cross-covariance, Geostatistics, Kalman Filter, State Space System, Time-Varying Models

Posted Date: December 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1152897/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Multivariate Kalman Filtering for Spatio-Temporal Processes

Guillermo Ferreira* · Jorge Mateu · Emilio Porcu · Alfredo Alegría

Received: date / Accepted: date

Abstract An increasing interest in models for multivariate spatio-temporal processes has been noted in the last years. Some of these models are very flexible and can capture both marginal and cross spatial associations amongst the components of the multivariate process. In order to contribute to the statistical analysis of these models, this paper deals with the estimation and prediction of multivariate spatio-temporal processes by using multivariate state-space models. In this context, a multivariate spatio-temporal process is represented through the well-known Wold decomposition. Such an approach allows for an easy implementation of the Kalman filter to estimate linear temporal processes exhibiting both short and long range dependencies, together with a spatial correlation structure. We illustrate, through simulation experiments, that our method offers a good balance between statistical efficiency and computational complexity. Finally, we apply the method for the analysis of a bivariate dataset on average daily temperatures and maximum daily solar radiations from 21 meteorological stations located in a portion of south-central Chile.

* Dr. Guillermo Ferreira (corresponding author) ,
Department of Statistics,
Universidad de Concepción
Tel.: +56-41-2203162
E-mail: gferreir@udec.cl

Dr. Jorge Mateu ,
Department of Mathematics,
University Jaume I, Castellón, Spain

Dr. Emilio Porcu,
School of Computer Science and Statistics,
Trinity College Dublin, Ireland

Dr. Alfredo Alegría ,
Department of Mathematics,
Universidad Técnica Federico Santa María, Chile

Keywords Cross-covariance · Geostatistics · Kalman Filter · State Space System · Time-Varying Models

PACS 62M10 · 62P20

1 Introduction

The use of state-space models has been quite popular in past literature, and efficient algorithms such as Kalman Filter (*KF* henceforth) have been developed for the modeling of sophisticated time series models, see e.g. [16], [20], [26], [10] and [39]. Once the state-space representation is established, the *KF* algorithm can be used to estimate the state vector, as well as the model parameters, and build the multi-step-ahead predictor of the process and its mean square error matrix. The idea of developing techniques for the estimation of an unobserved state from the observed process goes back to the sixties. A friendly introduction to the general idea of the *KF* is offered in [25], as well as in [30]. More extensive references include [23], [14], [22] and [21]. It is well-known that the *KF* algorithm provides an efficient mean to evaluate the maximum likelihood estimates of Gaussian processes. This outperformance makes it a natural candidate to make statistical inferences for spatio-temporal processes. In this context, the *KF* has been implemented to perform inference and prediction for processes in the physical, ecological, environmental, and biological sciences, among others. Since the nineties, a wealth of papers have provided a variety of methodologies related to the *KF*; for illustrative purposes, just a few representative contributions are cited below.

[28] considered a mixed approach between the *KF* algorithm and Kriging methodology (named as kriged Kalman Filter), in which the state equation incorporates different forms of temporal dynamics to model space-time interactions. [24] and [40] developed empirical Bayesian space-time *KF* models for the analysis of snow water equivalent and monthly precipitation. [41] proposed a spatio-temporal dynamic model formulation with restricted parameter matrices based on prior scientific knowledge, and developed a general expectation maximization (GEM) algorithm to carry out the estimations. [37] applied a dynamic state-space model to a sequence of SeaWiFS satellite images on the Lake Michigan. In this study the authors implemented a comprehensive version of the *KF*, called Ensemble Kalman filter, which allows to deal with problems of nonlinearities and high dimensionality inherent in satellite images. To deal with forecasting on spatio-temporal processes, [42] used the state-space system and a time-varying parameter least squares autoregressive system, with their respective solving algorithms, the *KF*, and autoregressive Adaptive Least Squares (ALS). [4] discussed the methods available for data assimilation in atmospheric models, including Ensemble Kalman filter. More recently, [15] have proposed a state-space methodology to model spatio-temporal processes where the temporal dependency is captured by short or long memory models, such as $\text{ARMA}(p, q)$ or $\text{ARFIMA}(p, d, q)$ through the infinite moving average representation $\text{MA}(\infty)$. Finally, [27] have used the univariate *KF* to make predictions at unobserved locations by using AR-type autoregressive temporal dependence models.

In the multivariate context, there has been a notable interest in incorporating multiple variables in spatio-temporal modeling of real phenomena that are encountered in many branches of science. An important tool to model these kind of data comes in terms of multivariate spatio-temporal covariances that may involve nonstationarity and interactions between different variables, locations, and times. In this context, [18] reviewed the construction of valid cross-covariance models, which include the linear model of coregionalization, convolution methods, the multivariate Matérn, the multivariate Wendland [12] and spatio-temporal extensions. Looking for appropriate methods of estimation has been a major problem that has led to a long discussion by several geostatistical researchers, for example [3], [2], [5] and [1] among others. These methodologies have as a main goal spatio-temporal interpolation and prediction at locations with unobserved data that allow for useful inventory and monitoring purposes. In this spirit, there are authors that have used multivariate models to analyze datasets with spatio-temporal dependencies. This is the case of [13] who proposed a class of conditionally specified models for the analysis of multivariate spatio-temporal processes, where the dependence structure across processes and over space and time is completely specified through a neighborhood structure. [43] studied the ensemble filter algorithm as an alternative to calculate the background error covariances on building a balanced error model for an intermediate coupled model for El Niño-Southern Oscillation (ENSO) predictions. Their approach to build such a model is proposed on the basis of the multivariate empirical orthogonal functions method. [33] considered multivariate autoregressive models for the estimation of the brain activity from electroencephalographic (EEG) time series. They used Kalman filtering to estimate the source dynamics between the EEG and the neural activity into the brain which can be computed using Maxwell equations. [7] introduced a fully Bayesian methodology with multivariate spatio-temporal dependencies to estimate quarterly measures of average monthly income over various geographies of the US. These authors computed the *KF* and Kalman smoothing equations within each MCMC iteration. However, to the best of our knowledge, there are no studies that analyze the finite sample behavior of the *KF* estimator, for both short and long memory multivariate spatio-temporal processes. We note that the extension of the univariate *KF* (as proposed in [15] and/or [27]) to the multivariate context could be considered sort of a simple case from a mathematical point of view. However, there are several reasons that motivate the extension to the multivariate approach. First, in terms of inference, parameter estimation through optimization of the truncated likelihood is not straightforward from the univariate case. This procedure involves the cross-correlation amongst the components of the processes forming the multivariate observations which comes in the temporal covariance. Additionally, it makes some further burden estimating the spatial correlation structure within the multivariate state-space representation. Finally, another aspect that in the univariate case is not direct, is the way the temporal cross-covariance in the multivariate VARFIMA models is added into the *KF* algorithm. This cross-covariance is well-defined under the condition that the coefficients of the $MA(\infty)$ representation are absolutely summable.

In this paper, we indeed propose a simple and flexible modeling strategy for multivariate spatio-temporal models through the well-known Wold decomposition, which under the assumption of absolutely summable coefficients provides an explicit for-

mula for the temporal structure of the covariance. This allows to cover the family of multivariate linear processes such as VARMA and VARFIMA to represent the temporal dependence, and to characterize the spatial association through cross-covariance functions of Matérn type. We use the *KF* algorithm for estimation of spatio-temporal processes as in [15] and find that only a few terms of the truncated representation of $MA(\infty)$ expansions are enough to capture the multivariate spatio-temporal dependencies in the process. Furthermore, by using this truncated representation in state-space models, we achieve great savings in both memory and computation of the likelihood. Finally, taking advantage of the benefits of *KF*, we propose a methodology to predict the process at unobserved locations.

The plan of the paper is the following. Section 2 discusses a class of multivariate spatio-temporal process and their representation in $MA(\infty)$ expansion. Section 3 presents the state-space models and the algorithm for estimating the parameters involved in the spatio-temporal dependence structure. In Section 4 a simulation study reveals the good performance of the estimation method with the *KF* technique. Section 5 applies our proposal to model the space-time variability of the average daily temperature and maximum daily solar radiation. The main conclusions are summarized in Section 6. A supplementary material is also considered.

2 A class of spatio-temporal processes

We consider L -variate spatio-temporal Gaussian processes,

$$\left\{ \mathbf{Y}_t(\mathbf{s}) = \left[Y_t^{(1)}(\mathbf{s}), \dots, Y_t^{(L)}(\mathbf{s}) \right]^\top : \mathbf{s} \in \mathbb{R}^2, t \in \mathbb{R} \right\},$$

where $\mathbf{s}$ is a spatial location and t is a temporal instant. Here, $\top$ is the transpose operator and $Y_t^{(\ell)}(\mathbf{s})$ denotes the ℓ -th component of the process. For a set of available temporal instants $t \in \{1, \dots, T\}$, we adopt a modeling strategy based on the following expansion

$$\mathbf{Y}_t(\mathbf{s}) = \mathbf{M}_t(\mathbf{s})\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t(\mathbf{s}), \quad \text{with} \quad \boldsymbol{\varepsilon}_t(\mathbf{s}) = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\eta}_{t-j}(\mathbf{s}). \quad (1)$$

This representation is parenthetical to the well-known $MA(\infty)$ decomposition for the errors with $\boldsymbol{\Psi}_j$ a sequence of $L \times L$ matrices with absolutely summable components, i.e., $\sum_{j=0}^{\infty} |\psi_{\ell\ell'}^{(j)}| < \infty$, where $\psi_{\ell\ell'}^{(j)}$ denotes the row ℓ , column ℓ' element of the moving average parameter matrix $\boldsymbol{\Psi}_j$ associated with lag j . The main advantage of this representation is that it provides an explicit formula to obtain the temporal covariance matrix, which is responsible for the characterization of the temporal dependence, given by

$$\Gamma(u; \boldsymbol{\Sigma}) = \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\Sigma} \boldsymbol{\Psi}_{k+u}^\top, \quad u = 0, \pm 1, \pm 2, \dots, \quad (2)$$

where Σ is an $L \times L$ matrix whose (ℓ, ℓ') -entry is given by $\rho_{\ell\ell'}\sigma_\ell\sigma_{\ell'}$, and $\ell, \ell' \in \{1, \dots, L\}$. Here $\sigma_\ell^2 > 0$ is the marginal variance of the process $\eta_t^{(\ell)}(\mathbf{s})$, and $\rho_{\ell\ell'}$ is the colocated correlation coefficient between processes $\eta_t^{(\ell)}(\mathbf{s})$ and $\eta_t^{(\ell')}(\mathbf{s})$, for $\ell, \ell' \in \{1, \dots, L\}$, with $\rho_{\ell\ell} = 1$. Additionally, the vector MA(∞) representation used in this paper satisfies the absolute summability condition, which ensures that the vector process is ergodic for the mean and for the second moment (see [23], page 263 for more details). In (1), $\mathbf{M}_t(\mathbf{s}) = \text{diag}(\mathbf{M}_{t,1}^\top(\mathbf{s}), \dots, \mathbf{M}_{t,L}^\top(\mathbf{s}))$, where each $\mathbf{M}_{t,\ell}^\top(\mathbf{s})$ is a $p_\ell \times 1$ vector of non-stochastic regressors, for $\ell = 1, \dots, L$. Thus, $\mathbf{M}_t(\mathbf{s})$ is an $L \times \sum_{\ell=1}^L p_\ell$ matrix, $\beta = (\beta_1^\top, \dots, \beta_L^\top)^\top$ is a $\sum_{\ell=1}^L p_\ell \times 1$ vector of parameters, with each $\beta_\ell^\top$ being a $p_\ell \times 1$ vector. Finally, $\boldsymbol{\eta}_t(\mathbf{s}) = [\eta_t^{(1)}(\mathbf{s}), \dots, \eta_t^{(L)}(\mathbf{s})]^\top$ is an L -variate spatio-temporal stationary Gaussian process with zero mean and covariance function given by

$$\text{Cov} \left\{ \eta_t^{(\ell)}(\mathbf{s}), \eta_{t'}^{(\ell')}(\mathbf{s}') \right\} = \begin{cases} \Sigma C_{\ell\ell'}(\mathbf{s}, \mathbf{s}') & \text{if } |t - t'| = 0 \\ 0 & \text{if } |t - t'| \neq 0, \end{cases}$$

for all $t, t' \in \mathbb{R}$, $\mathbf{s}, \mathbf{s}' \in \mathbb{R}^2$. Throughout, we assume that $C_{\ell\ell'}(\mathbf{s}, \mathbf{s}') = C(\mathbf{s}, \mathbf{s}'; \psi)$ is a univariate parametric correlation model, which depends on a vector of unknown parameters ψ . The spatio-temporal covariance structure of the process defined in (1) is

$$\begin{aligned} \text{Cov}(\mathbf{Y}_t(\mathbf{s}), \mathbf{Y}_{t'}(\mathbf{s}')) &= \mathbb{E}(\mathbf{Y}_t(\mathbf{s})\mathbf{Y}_{t'}(\mathbf{s}')^\top) \\ &= \sum_{k,l=0}^{\infty} \Psi_k \text{Cov}(\boldsymbol{\eta}_{t-k}(\mathbf{s}), \boldsymbol{\eta}_{t'-l}(\mathbf{s}')) \Psi_l^\top \\ &= C(\mathbf{s}, \mathbf{s}'; \psi) \left(\sum_{k=0}^{\infty} \Psi_k \Sigma \Psi_{k+t-t'}^\top \right) \\ &= C(\mathbf{s}, \mathbf{s}'; \psi) \Gamma(t, t'; \Sigma). \end{aligned} \quad (3)$$

We conclude that $\text{Cov}(\mathbf{Y}_t(\mathbf{s}), \mathbf{Y}_{t'}(\mathbf{s}'))$ is a separable spatio-temporal covariance function [19], where $\Gamma(u; \Sigma)$ is an $L \times L$ matrix that regulates the temporal dependency of the process given in (2). An alternative way of writing the functional form of the space-time covariance matrix of this process is the following. Define the $LN \times 1$ vector $\mathbf{Y}_t = [\mathbf{Y}_t(\mathbf{s}_1), \dots, \mathbf{Y}_t(\mathbf{s}_N)]^\top$ containing the L -dimensional data values at N spatial locations, at each time instant t , then the spatio-temporal covariance structure (3) is equivalent to

$$\mathbf{C}^{\mathbf{Y}} = \mathbf{C} \otimes \mathbf{\Gamma} = \left[C(\mathbf{s}_i, \mathbf{s}_j; \psi) \right]_{i,j=1}^N \otimes \left[\mathbf{\Gamma}(t, t'; \Sigma) \right]_{t,t'=1}^T, \quad (4)$$

where $\mathbf{C}$ and $\mathbf{\Gamma}$ are the spatial and temporal covariances, respectively, and $\otimes$ is the Kronecker product. In this paper, the spatial covariance function $C(\mathbf{s}_i, \mathbf{s}_j; \psi)$ is specified by an admissible covariance. In particular, we consider the general class of Matérn covariance models [29] given by

$$C(\mathbf{s}_i, \mathbf{s}_j; \psi) = \frac{1}{2^{\nu-1} \Gamma(\nu)} \left(\frac{\|\mathbf{s}_i - \mathbf{s}_j\|}{\alpha} \right)^\nu K_\nu \left(\frac{\|\mathbf{s}_i - \mathbf{s}_j\|}{\alpha} \right), \quad \text{for } i, j = 1, \dots, N \quad (5)$$

where $\alpha > 0$, $\nu \geq 0$, $\sigma^2 > 0$ and K_ν is the modified Bessel function of the second kind of order ν . A popular special case of the Matérn family is the exponential model $C(\mathbf{s}, \mathbf{s}'; \psi) = \exp\left(-\frac{\|\mathbf{s}-\mathbf{s}'\|}{\alpha}\right)$ which is obtained when $\nu = 1/2$. The temporal covariance matrix $\Gamma(u; \Sigma)$ is specified through multivariate ARFIMA linear processes [6].

We note that in general *KF* methods work under the assumption of separability of the covariance structures in space and time, as in a non-separable case the autoregressive component would be lost; in particular the ARFIMA structure would have no sense (we refer to contributions [15] and [32], among other authors, to argue about the separability assumption). However, the separability in space and time does not mean independence among the processes that form the multivariate process observation. These processes are clearly dependent due mainly to the temporal correlation structure. We come back to this at the end of this section.

For a neater and self-contained exposition, the following examples discuss some stationary multivariate spatio-temporal ARFIMA processes with their structure of temporal covariance.

Example 21 *An example of the regression model (1) is the L -variate ARMA process of orders p and q that satisfies the equation*

$$\Phi(B)\varepsilon_t(\mathbf{s}) = \Theta(B)\eta_t(\mathbf{s}), \quad t = 1, \dots, T,$$

where $\Phi(z) = \mathbf{I}_L - \Phi_1 z - \dots - \Phi_p z^p$ and $\Theta(z) = \mathbf{I}_L + \Theta_1 z + \dots + \Theta_q z^q$ are matrix-valued polynomials having full rank, $\mathbf{I}_r$ denotes the $r \times r$ identity matrix, and B denotes the backward shift operator. Each component of the matrices $\Phi(z)$ and $\Theta(z)$ is a polynomial with real coefficients and degree less than or equal to p and q , respectively. A multivariate process of this nature is commonly described as a VARMA process (the initial letter denoting “vector”). The MA(∞) representation of the VARMA process is $\varepsilon_t(\mathbf{s}) = \sum_{j=0}^{\infty} \Psi_j \eta_{t-j}(\mathbf{s})$ provided that $\det \Phi(z) \neq 0$ for all $z \in \mathbb{C}$ such that $|z| \leq 1$. The matrices Ψ_j are obtained recursively from the equations

$$\Psi_j = \Theta_j + \sum_{k=0}^{\infty} \Phi_k \Psi_{j-k}, \quad j = 0, 1, \dots \quad (6)$$

where $\Theta_0 = \mathbf{I}_L$, $\Theta_j = \mathbf{0}$, for $j > q$, and $\Phi_j = \mathbf{0}$, for $j > p$. The temporal covariance matrix $\Gamma(u; \Sigma)$ can be expressed as

$$\Gamma(u; \Sigma) = \sum_{j=0}^{\infty} \Psi_j \Sigma \Psi_{j+u}^\top,$$

where the matrices Ψ_j are found from (6). A particular case of this model is the autoregressive vector of order one, VAR (1), defined by

$$\varepsilon_t(\mathbf{s}) = \Phi \varepsilon_{t-1}(\mathbf{s}) + \eta_t(\mathbf{s}), \quad \text{with } \Psi_j = \Phi^j, \quad j = 1, 2, \dots,$$

where $\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ \sigma_1 \sigma_2 \rho & \sigma_2^2 \end{bmatrix}$.

Example 22 A more general class of stationary processes are the vector autoregressive-fractionally integrated moving-average (VARFIMA) processes, given by

$$\Phi(B) \text{diag}(\nabla^d) \varepsilon_t(\mathbf{s}) = \Theta(B) \eta_t(\mathbf{s}), \quad (7)$$

where

$$\text{diag}(\nabla^d) = \begin{bmatrix} \nabla^{d_1} & 0 & \dots & 0 \\ 0 & \nabla^{d_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \nabla^{d_L} \end{bmatrix}.$$

Here, $\nabla = 1 - B$, and $0 < d_\ell < 1/2$, for all $\ell = 1, \dots, L$. Following [38], the model defined in (7) can be represented as

$$\text{diag}(\nabla^d) \Phi(B) \varepsilon_t(\mathbf{s}) = \Theta(B) \eta_t(\mathbf{s}).$$

These representations are identical if either $\Phi(B)$ is diagonal or the values of the differencing parameters remain the same across $\ell = 1, \dots, L$ (see [38] for details). In the univariate case ($L = 1$) the distinction between both models is irrelevant.

The simplest case is the VARFIMA model with $p = q = 0$, where we have the dynamic $\varepsilon_t(\mathbf{s}) = \text{diag}(\nabla^{-d}) \eta_t(\mathbf{s})$, or equivalently, the ℓ -th component can be written as

$$\varepsilon_t^{(\ell)}(\mathbf{s}) = \sum_{j=0}^{\infty} \psi_{j,\ell} \eta_{t-j}^{(\ell)}(\mathbf{s}), \quad \ell = 1, \dots, L, \quad (8)$$

where $\psi_{j,\ell} = \frac{\Gamma(j+d_\ell)}{\Gamma(j+1)\Gamma(d_\ell)}$, and $\Gamma(\cdot)$ is the gamma function. The (ℓ, ℓ') -th element of the temporal covariance function for model (8) is

$$\Gamma_{\ell\ell'}(u; \Sigma) = \sum_{j=0}^{\infty} \psi_{j,\ell} \Sigma_{\ell\ell'} \psi_{j+u,\ell} = \Sigma_{\ell\ell'} \frac{\Gamma(1-d_\ell-d_{\ell'})}{\Gamma(d_\ell)\Gamma(1-d_{\ell'})} \frac{\Gamma(u+d_{\ell'})}{\Gamma(u+1-d_\ell)},$$

for every $\ell, \ell' = 1, \dots, L$. Other variants of the VARFIMA model can be found in [38]. For computational efficiency, it is well known in the literature that $\Gamma_{\ell\ell'}(u; \Sigma)$ can be calculated as follows

$$\sum_{j=0}^{\infty} \psi_{j,\ell} \Sigma_{\ell\ell'} \psi_{j+u,\ell} = \Sigma_{\ell\ell'} \frac{\Gamma(1-d_\ell-d_{\ell'})}{\Gamma(1-d_\ell)\Gamma(1-d_{\ell'})} \prod_{0 \leq k \leq u} \frac{k-1+d_\ell}{k-d_{\ell'}}.$$

In relation to the separability condition, some comments are in order. Although separable covariance functions can not take into account the interaction between the spatial and temporal components, such an interaction can be injected through the space-time trend $\mathbf{M}_t(\mathbf{s})\beta$, see [17]. On the other hand, separable structures allow for considerable computational gains as, for a given set of space-time locations and observations, the space-time covariance matrix factors into the product of a purely spatial with a purely temporal covariance matrix. Another advantage of the separable space-time covariance structure is the construction of a valid (positive definite) parametric model for the space-time correlation structure, which is not straightforward if we consider nonseparable space-time covariance models. For details, the reader is referred to [34] with the references therein.

3 State-Space representation

In this section we will discuss the state-space (SS) representation of the model defined in equation (1) with a space-time covariance structure given by (3) or (4). Once the SS system has been established, inferences are made about the state equation using all the information available until time t through the KF algorithm. Then if $\mathbf{Y}_t(\mathbf{s})$ is an observation available at an instant of time t for some location $\mathbf{s}$, the SS system for spatio-temporal processes is defined as follows

$$\begin{aligned} \mathbf{Y}_t(\mathbf{s}) &= [\mathbf{G}_t(\mathbf{s}) \quad \mathbf{M}_t(\mathbf{s})] \begin{bmatrix} \mathbf{X}_t(\mathbf{s}) \\ \boldsymbol{\beta}_t \end{bmatrix} + \mathbf{W}_t(\mathbf{s}), \\ \begin{bmatrix} \mathbf{X}_{t+1}(\mathbf{s}) \\ \boldsymbol{\beta}_{t+1} \end{bmatrix} &= \begin{bmatrix} \mathbf{F}_t(\mathbf{s}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix} \begin{bmatrix} \mathbf{X}_t(\mathbf{s}) \\ \boldsymbol{\beta}_t \end{bmatrix} + \begin{bmatrix} \mathbf{H} \\ \mathbf{0} \end{bmatrix} \mathbf{V}_t(\mathbf{s}), \end{aligned} \quad (9)$$

where $\mathbf{G}_t(\mathbf{s})$ is the observation operator vector, $\mathbf{M}_t(\mathbf{s})$ is a vector of exogenous or predetermined variables, $[\mathbf{X}_t(\mathbf{s}) \quad \boldsymbol{\beta}_t]^\top$ is a state vector, and $\mathbf{W}_t(\mathbf{s})$ is an observation noise with variance $\mathbf{R}$. In addition, $\mathbf{F}_t(\cdot)$ is a state transition operator, $\mathbf{H}$ is a linear operator, $\mathbf{V}_t(\mathbf{s})$ is spatially colored, temporally white and Gaussian with mean zero and a common covariance matrix $\mathbf{Q}$, and $\mathbf{V}_t(\mathbf{s})$ and $\mathbf{W}_t(\mathbf{s})$ are uncorrelated, that is, $E(\mathbf{W}_t(\mathbf{s})\mathbf{V}_t(\mathbf{s})^\top) = \mathbf{0}$, for all $\mathbf{s}$ and t . The process (1) can be represented by an SS system as above by generalizing the infinite-dimensional equations given by [15] to the spatio-temporal multivariate case. This can be achieved by assignation of the $(j+1)$ -th component of the state vector as the lag in j steps of the L -dimensional Gaussian process $\boldsymbol{\eta}_t(\mathbf{s})$, i.e. $\mathbf{X}_t^{(j+1)}(\mathbf{s}) = \boldsymbol{\eta}_{t-j}(\mathbf{s})$, for $j = 0, 1, \dots$. In this case, the process specified by (1) can be represented by the following infinite-dimensional state-space system

$$\begin{aligned} \mathbf{Y}_t(\mathbf{s}) &= [(\mathbf{I}_L \quad \boldsymbol{\Psi}_1 \quad \boldsymbol{\Psi}_2 \quad \boldsymbol{\Psi}_3 \quad \dots), \mathbf{M}_t(\mathbf{s})] \begin{bmatrix} \mathbf{X}_t(\mathbf{s}) \\ \boldsymbol{\beta}_t \end{bmatrix}, \\ \begin{bmatrix} \mathbf{X}_{t+1}(\mathbf{s}) \\ \boldsymbol{\beta}_{t+1} \end{bmatrix} &= \begin{bmatrix} \mathbf{F}_t^\top & \mathbf{0} \\ \mathbf{0}^\top & \mathbf{I}_p \end{bmatrix} \begin{bmatrix} \mathbf{X}_t(\mathbf{s}) \\ \boldsymbol{\beta}_t \end{bmatrix} + \begin{bmatrix} [\boldsymbol{\eta}_{t+1}(\mathbf{s}) \quad 0 \quad 0 \quad 0 \quad \dots]^\top \\ \mathbf{0} \end{bmatrix}, \end{aligned}$$

where $\mathbf{G} = [\mathbf{I}_L, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \boldsymbol{\Psi}_3, \dots]$, $\mathbf{F} = \begin{bmatrix} \mathbf{0}^\top & \mathbf{I}_L \\ \mathbf{I}_\infty & \mathbf{0} \end{bmatrix}$, $\mathbf{I}_\infty = \text{diag}\{1, 1, \dots\}$, and $\mathbf{W}_t = \mathbf{0}$. The computational burdens for estimating the VARFIMA model are well-known in terms of dimensionality and span of data series ([38], [31]). We thus truncate the expansion in (1) after some positive integer m , so that

$$\mathbf{Y}_t(\mathbf{s}) = \mathbf{M}_t(\mathbf{s})\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}_t(\mathbf{s}), \quad (10)$$

where $\tilde{\varepsilon}_t(\mathbf{s}) = \sum_{j=0}^m \Psi_j \boldsymbol{\eta}_{t-j}(\mathbf{s})$. Thus, the SS representation of the model (10) is considered, with observation and state equations given by

$$\begin{aligned} \mathbf{Y}_t(\mathbf{s}) &= [(\mathbf{I}_L \ \Psi_1 \ \Psi_2 \ \cdots \ \Psi_m), \mathbf{M}_t(\mathbf{s})] \begin{bmatrix} \mathbf{X}_t(\mathbf{s}) \\ \boldsymbol{\beta} \end{bmatrix}, \\ \begin{bmatrix} \mathbf{X}_{t+1}(\mathbf{s}) \\ \boldsymbol{\beta} \end{bmatrix} &= \begin{bmatrix} - & - & \mathbf{F} & - & - & \mathbf{0}_{1 \times p} \\ \mathbf{0}_{1 \times L(m+1)} & \mathbf{I}_p & & & & \end{bmatrix} \begin{bmatrix} \mathbf{X}_t(\mathbf{s}) \\ \boldsymbol{\beta} \end{bmatrix} + \begin{bmatrix} \mathbf{V}_{t+1}(\mathbf{s}) \\ \mathbf{0}_{p \times 1} \end{bmatrix} \end{aligned} \quad (11)$$

where $\mathbf{F} = \begin{bmatrix} \mathbf{0}_{L \times Lm} & \mathbf{0}_{L \times L} \\ \mathbf{I}_{Lm} & \mathbf{0}_{Lm \times L} \end{bmatrix}$ and $\mathbf{V}_{t+1}(\mathbf{s}) = [\boldsymbol{\eta}_{t+1}(\mathbf{s}) \ 0 \ 0 \ \cdots \ 0]^\top$. Let us denote by $\mathbb{M}_{p \times q}$ the space of $p \times q$ matrices with real elements. Thus, we have that $\mathbf{G} \in \mathbb{M}_{L \times L(m+1)}$, $\mathbf{X}_t(\mathbf{s}) \in \mathbb{M}_{L(m+1) \times 1}$, $\mathbf{F} \in \mathbb{M}_{L(m+1) \times L(m+1)}$ and $\mathbf{V}_{t+1}(\mathbf{s}) \in \mathbb{M}_{L(m+1) \times 1}$. A study about the truncation level m will be presented later through a similarity study for spatio-temporal data following a truncated MA representation as in (10).

3.1 Derivation of the *KF* algorithm

The *KF* is a powerful tool to make inferences about the state vector which allows to calculate the conditional mean and covariance matrix of the state vector $[X_t(\mathbf{s}), \boldsymbol{\beta}(\mathbf{s})]^\top$. For simplicity, we assume that the trend function in (10) is identically equal to zero. Thus, (11) becomes the usual state equation. The *KF* recursion equations are well-known, but we present them here to introduce notation and for completeness. First, define the $NL(m+1)$ -dimensional vector $\mathbf{X}_t = [\mathbf{X}_t(\mathbf{s}_1), \dots, \mathbf{X}_t(\mathbf{s}_N)]^\top$ as an unobservable spatio-temporal state process. In addition, we define the best linear unbiased predictor (BLUP) of the unobserved state $\mathbf{X}_t(\mathbf{s})$ and its error variance-covariance matrix as follows

$$\begin{aligned} \hat{\mathbf{X}}_t(\mathbf{s}) &= \mathbb{E}(\mathbf{X}_t(\mathbf{s}) | \mathbf{X}_1, \dots, \mathbf{X}_t), \\ \boldsymbol{\Omega}_t(\mathbf{s}, \mathbf{s}') &= \text{Cov}(\mathbf{X}_t(\mathbf{s}) - \hat{\mathbf{X}}_t(\mathbf{s}), \mathbf{X}_t(\mathbf{s}') - \hat{\mathbf{X}}_t(\mathbf{s}')). \end{aligned}$$

The Kalman recursive equations are defined as follows for the initial conditions $\hat{\mathbf{X}}_1(\mathbf{s}) = \mathbb{E}(\boldsymbol{\eta}_{1-j}(\mathbf{s})) = \mathbf{0}$, for $j = 0, 1, \dots, m$, and

$$\begin{aligned} \boldsymbol{\Omega}_1(\mathbf{s}, \mathbf{s}') &= \mathbb{E}(\mathbf{X}_1(\mathbf{s})\mathbf{X}_1(\mathbf{s}')^\top) = \begin{bmatrix} C(\mathbf{s}, \mathbf{s}'; \boldsymbol{\psi})\boldsymbol{\Gamma}(0; \boldsymbol{\Sigma}) & \cdots & \mathbf{0}_{L \times L} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{L \times L} & \cdots & C(\mathbf{s}, \mathbf{s}'; \boldsymbol{\psi})\boldsymbol{\Gamma}(0; \boldsymbol{\Sigma}) \end{bmatrix} \\ &= \mathbf{I}_{(m+1)} \otimes C(\mathbf{s}, \mathbf{s}'; \boldsymbol{\psi})\boldsymbol{\Gamma}(0; \boldsymbol{\Sigma})_{(L \times L)}. \end{aligned}$$

The *KF* allows to estimate the state vector $\mathbf{X}_{t+1}(\mathbf{s})$ and its prediction variance based on the information available at time t . These estimators are given by

$$\begin{aligned} \hat{\mathbf{X}}_{t+1}(\mathbf{s}) &= \mathbf{F}\hat{\mathbf{X}}_t(\mathbf{s}) + \boldsymbol{\Theta}_t(\mathbf{s})\boldsymbol{\Delta}_t^{-1}(\mathbf{Y}_t - \hat{\mathbf{Y}}_t), \\ \boldsymbol{\Omega}_{t+1}(\mathbf{s}, \mathbf{s}') &= \mathbf{F}\boldsymbol{\Omega}_t(\mathbf{s}, \mathbf{s}')\mathbf{F}^\top + \mathbf{Q}_t^\eta(\mathbf{s}, \mathbf{s}') - \boldsymbol{\Theta}_t(\mathbf{s})\boldsymbol{\Delta}_t^{-1}\boldsymbol{\Theta}_t^\top(\mathbf{s}'), \end{aligned} \quad (12)$$

where

$$\begin{aligned}
\Delta_t &= \text{Var}(\mathbf{Y}_t - \hat{\mathbf{Y}}_t | \mathbf{Y}_{t-1}) = \text{Var}(G(\mathbf{X}_t - \hat{\mathbf{X}}_t) | \mathbf{Y}_{t-1}) \\
&= \begin{bmatrix} \mathbf{G}\boldsymbol{\Omega}_t(\mathbf{s}_1, \mathbf{s}_1)\mathbf{G}^\top & \cdots & \mathbf{G}\boldsymbol{\Omega}_t(\mathbf{s}_1, \mathbf{s}_N)\mathbf{G}^\top \\ \vdots & \ddots & \vdots \\ \mathbf{G}\boldsymbol{\Omega}_t(\mathbf{s}_1, \mathbf{s}_N)\mathbf{G}^\top & \cdots & \mathbf{G}\boldsymbol{\Omega}_t(\mathbf{s}_N, \mathbf{s}_N)\mathbf{G}^\top \end{bmatrix}, \\
\boldsymbol{\Theta}_t(\mathbf{s}) &= \text{Cov}(\mathbf{X}_{t+1}(\mathbf{s}), \mathbf{Y}_t - \hat{\mathbf{Y}}_t) \\
&= \begin{bmatrix} \text{Cov}(\mathbf{X}_{t+1}(\mathbf{s}), \mathbf{Y}_t(\mathbf{s}_1) - \hat{\mathbf{Y}}_t(\mathbf{s}_1)) \\ \vdots \\ \text{Cov}(\mathbf{X}_{t+1}(\mathbf{s}), \mathbf{Y}_t(\mathbf{s}_N) - \hat{\mathbf{Y}}_t(\mathbf{s}_N)) \end{bmatrix} = \begin{bmatrix} \mathbf{F}\boldsymbol{\Omega}_t(\mathbf{s}, \mathbf{s}_1)\mathbf{G}^\top \\ \vdots \\ \mathbf{F}\boldsymbol{\Omega}_t(\mathbf{s}, \mathbf{s}_N)\mathbf{G}^\top \end{bmatrix}, \\
\hat{\mathbf{Y}}_t &= \mathbb{E}(\mathbf{Y}_t | \mathbf{Y}_1, \dots, \mathbf{Y}_{t-1}) = \begin{bmatrix} \mathbf{G}\mathbf{X}_t(\mathbf{s}_1) \\ \vdots \\ \mathbf{G}\mathbf{X}_t(\mathbf{s}_N) \end{bmatrix}, \\
\mathbf{Q}_t^\eta(\mathbf{s}, \mathbf{s}') &= [\text{Cov}(\mathbf{V}_{t+1}(\mathbf{s}), \mathbf{V}_{t+1}(\mathbf{s}'))] = \mathbf{T}_{(m+1)} \otimes C(\mathbf{s}, \mathbf{s}'; \boldsymbol{\psi})\boldsymbol{\Gamma}(0; \boldsymbol{\Sigma}),
\end{aligned}$$

where $\mathbf{T}_{(m+1)}$ has dimension $(m+1) \times (m+1)$ and is given by $\mathbf{T}_{(m+1)} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$.

Let $\boldsymbol{\theta}$ be the vector that contains all the parameters involved in the model, then the log-likelihood function (omitting a constant) can be obtained from (12),

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{t=1}^T \log|\Delta_t(\boldsymbol{\theta})| + \boldsymbol{\epsilon}_t(\boldsymbol{\theta})^\top \Delta_t(\boldsymbol{\theta})^{-1} \boldsymbol{\epsilon}_t(\boldsymbol{\theta}),$$

where $\boldsymbol{\epsilon}_t(\boldsymbol{\theta}) = \mathbf{Y}_t - \hat{\mathbf{Y}}_t$ is the innovation vector and $\Delta_t(\boldsymbol{\theta})$ is the innovation covariance matrix at time t . Hence the exact maximum likelihood estimate (MLE) provided by the Kalman equations (12) is given by $\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})$. Note that the Kalman equations (12) can be applied directly to the general state-space representation (9) or to the truncated representation (11), yielding in the latter case an approximated MLE.

It is worth noting that for time series models with a long memory structure, [8] proved that the exact likelihood function can be computed recursively in a finite number of steps. On the other hand, an approximation to the likelihood function based on the truncated SS equation is also possible obtaining a considerable reduction in the number of iterations of the algorithm. Indeed, the order of computation of the algorithm is $m^2 \times T$ instead of T^3 . In the context of multivariate spatio-temporal processes the number of iterations increases considerably, for example, if N is the number of spatial locations, we also need to solve matrices of order $LN \times LN$ for each iteration, i.e. the algorithm has an order of computation of $(LN)^3 \times m^2 \times T$.

Here, and in order to reduce computational time, we have implemented the *KF* algorithm in C source code connected to R free software [35] through the interface called `.C` [9]; in particular, the QR decomposition method for the inverse of matrices was implemented into pure C. On the other hand, numerical optimisation of the Gaussian log-likelihood function to obtain the QML estimates was carried out using the `optim` command of R. This method makes use of the subroutine “L-BFGS-B” corresponding to a quasi-Newton method [44], [11]. A numerical analysis about the computational time to estimates a VAR(1) model is available in the supplementary material, Section A.1.

3.2 Spatial interpolation through the *KF*

Following the proposal of [27] for the univariate case, we propose spatial interpolation in a multivariate context by means of the *KF* algorithm, at locations where we do not have information. Denote by s_0 a location where there is no record of $\mathbf{Y}_t(\cdot)$. The equations that allow us to achieve predictions for $\mathbf{Y}_t(s_0)$ given the information $\mathbf{Y}_t$, for $t = 1, \dots, T$, are given by

$$\begin{aligned}\widehat{\mathbf{X}}_{t+1}(s_0) &= \mathbf{F}\widehat{\mathbf{X}}_t(s_0) + \boldsymbol{\Theta}_t(s_0)\boldsymbol{\Delta}_t^{-1}(\mathbf{Y}_t - \widehat{\mathbf{Y}}_t), \\ \boldsymbol{\Omega}_{t+1}(s_0) &= \mathbf{F}\boldsymbol{\Omega}_t(s_0)\mathbf{F}^\top + \mathbf{Q}_t^\eta(s_0) - \boldsymbol{\Theta}_t(s_0)\boldsymbol{\Delta}_t^{-1}\boldsymbol{\Theta}_t(s)^\top,\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{\Theta}_t(s_0) &= \text{Cov}(\mathbf{X}_{t+1}(s_0), \mathbf{Y}_t | \mathbf{Y}_{t-1}) = \begin{bmatrix} \text{Cov}(\mathbf{X}_{t+1}(s_0), \mathbf{Y}_t(s_1)) \\ \vdots \\ \text{Cov}(\mathbf{X}_{t+1}(s_0), \mathbf{Y}_t(s_N)) \end{bmatrix} = \begin{bmatrix} \mathbf{F}\boldsymbol{\Omega}_t(s_0, s_1)\mathbf{G}^\top \\ \vdots \\ \mathbf{F}\boldsymbol{\Omega}_t(s_0, s_N)\mathbf{G}^\top \end{bmatrix}, \\ \boldsymbol{\Omega}_t(s_0) &= \text{Cov}(\mathbf{X}_t(s_0), \mathbf{X}_t | \mathbf{X}_{t-1}), \mathbf{Q}_t^\eta(s_0) = \text{Cov}(\mathbf{V}_{t+1}(s_0), \mathbf{V}_{t+1}), \text{ and} \\ \boldsymbol{\Theta}_t(s) &= \text{Cov}(\mathbf{X}_{t+1}(s), \mathbf{Y}_t | \mathbf{Y}_{t-1}).\end{aligned}$$

Then, the updated equation through the *KF* leads

$$\boldsymbol{\Omega}_{t+1}(s_0, s_0) = \mathbf{F}\boldsymbol{\Omega}_t(s_0, s_0)\mathbf{F}^\top + \mathbf{Q}_t^\eta(s_0, s_0) - \boldsymbol{\Theta}_t(s_0)\boldsymbol{\Delta}_t^{-1}\boldsymbol{\Theta}_t(s_0)^\top,$$

and thus the prediction variance at location s_0 is given by

$$\boldsymbol{\Delta}_t(s_0) = \text{Var}(\mathbf{Y}_t(s_0) - \widehat{\mathbf{Y}}_t(s_0)) = \mathbf{G}\boldsymbol{\Omega}_t(s_0, s_0)\mathbf{G}^\top.$$

Note that according to the state vector update in (12), it requires that all observations in time are available. Then, if the data have missing values, they must be previously imputed, a methodology in this line is proposed by [32] for the space-time autoregressive univariate model.

4 Simulation studies

In this section, several Monte Carlo experiments are carried out to analyze the finite sample behavior of the KF estimator for short and long memory spatio-temporal processes. We work under the bivariate ($L = 2$) and trivariate ($L = 3$) cases. We consider spatial sites being uniformly distributed on the square $[0, 1]^2$, with $N \times T = 25 \times 200$ or $N \times T = 100 \times 300$. Thus, when $L = 2$, we have 10×10^3 and 60×10^3 observations, respectively.

We consider a correlation structure $C(\mathbf{s}, \mathbf{s}'; \alpha)$ of Matérn type given in (5), which is the most popular correlation structure used throughout the spatial statistics literature. In particular, we use two special cases of the Matérn model:

- **Model 1.** Matérn with smoothness parameter $\nu = \frac{1}{2}$, corresponding to the exponential model

$$C(\mathbf{h}; \alpha) = \exp \left\{ -\frac{\|\mathbf{h}\|}{\alpha} \right\} \quad \text{with} \quad \mathbf{h} = \mathbf{s} - \mathbf{s}'. \quad (13)$$

- **Model 2.** Matérn with smoothness parameter $\nu = \frac{3}{2}$, which leads to

$$C(\mathbf{h}; \alpha) = \left(1 + \frac{\|\mathbf{h}\|}{\alpha} \right) \exp \left\{ -\frac{\|\mathbf{h}\|}{\alpha} \right\}.$$

Here, $\alpha > 0$ is a scale parameter that controls the rate of the spatial dependence decay; a large value of α corresponds to a high spatial correlation. For the temporal dependence, we consider two models, a VAR(1) in a bivariate short memory case, and a VARFIMA(0, d , 1) in a trivariate long memory case. The source codes used in the paper in R and C are available upon request to the corresponding author.

4.1 Short memory case

Consider the VAR(1) model with the errors defined by (1), where

$$\begin{bmatrix} \varepsilon_t^{(1)}(\mathbf{s}) \\ \varepsilon_t^{(2)}(\mathbf{s}) \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_{t-1}^{(1)}(\mathbf{s}) \\ \varepsilon_{t-1}^{(2)}(\mathbf{s}) \end{bmatrix} + \begin{bmatrix} \eta_t^{(1)}(\mathbf{s}) \\ \eta_t^{(2)}(\mathbf{s}) \end{bmatrix}. \quad (14)$$

The sequence of matrices in (1) satisfies $\Psi_j = \Phi^j = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}^j$. The unknown parameters are Φ and ϑ , where $\vartheta = (\sigma_1^2, \sigma_2^2, \rho_{12}, \alpha)^\top$. We study the following scenarios:

Case 1. $\theta = (\Phi, \vartheta)^\top = (0.45, 0.2, 0.15, 0.65, 1, 1, 0.5, 0.5)^\top$

Case 2. $\theta = (\Phi, \vartheta)^\top = (0.15, 0, 0, 0.35, 1, 0.5, -0.3, 1.5)^\top$.

We assume that we observe a spatio-temporal process $\varepsilon_t(\mathbf{s})$ at $t = 1, \dots, T$ and $\mathbf{s} = (s_1, \dots, s_N)$, on a regular rectangular grid of N spatial locations in $[0, 1]^2$ and at equidistant time points. For the data generation scheme, the process is generated

recursively from (14) with initial values $\varepsilon_1^{(\ell)}(\mathbf{s}) \sim N(0, \sigma_\ell^2)$ for $\ell = 1, 2$. Finally, we simulate each process 100 times, and for each simulation the *KF* estimates are evaluated by the relative bias (RelBias) and the mean square errors (MSE) defined as

$$\text{RelBias}(\theta) = \frac{1}{100} \sum_{i=1}^{100} (\hat{\theta}_i / \theta - 1) \quad \text{and} \quad \text{MSE}(\theta) = \frac{1}{100} \sum_{i=1}^{100} (\hat{\theta}_i - \theta)^2,$$

where $\hat{\theta}_i$ is the *KF* estimate of θ_i for the i th realization. Table 1 shows the estimates of the parameters. Note that the estimates are very close to their theoretical counterparts. Furthermore, we highlight that the goodness-of-fit criteria, such as the standard deviation (sd), bias and $\sqrt{\text{MSE}}$ are very similar in comparison to the level of truncation. Indeed, if the truncation parameter is $m = 5$, the truncated *KF* works extremely well for both sample sizes. In order to give a neater picture about the finite sample performance of the *KF* algorithm, we show a set of boxplots displaying the behavior of the estimators of the parameters obtained from the previous Monte Carlo experiments (see the supplementary material in Section A.2). We next consider Model 2 and parameter vector as in Case 2. Due to space constraints, only a subset of the results are presented, in particular, we consider a sample size of $T = 200, 300$ and $m = 5, 10$. The same evidence is obtained in this case, see Table 2.

4.2 Long memory case

This subsection considers the impact of VARFIMA(0, d , 1) parameters on the performance of the *KF* algorithm. We focus on the following 3-dimensional model

$$\begin{bmatrix} \nabla^{d_1} & 0 & 0 \\ 0 & \nabla^{d_2} & 0 \\ 0 & 0 & \nabla^{d_3} \end{bmatrix} \begin{bmatrix} \varepsilon_t^{(1)}(\mathbf{s}) \\ \varepsilon_t^{(2)}(\mathbf{s}) \\ \varepsilon_t^{(3)}(\mathbf{s}) \end{bmatrix} = \begin{bmatrix} 1 + \theta_{11}B & 0 & 0 \\ 0 & 1 + \theta_{22}B & 0 \\ 0 & 0 & 1 + \theta_{33}B \end{bmatrix} \begin{bmatrix} \eta_t^{(1)}(\mathbf{s}) \\ \eta_t^{(2)}(\mathbf{s}) \\ \eta_t^{(3)}(\mathbf{s}) \end{bmatrix}, \text{ or} \\ \varepsilon_t^{(\ell)}(\mathbf{s}) = \sum_{k=0}^{\infty} \psi_{k,\ell} \eta_{t-k}^{(\ell)}(\mathbf{s}), \quad \ell = 1, 2, 3, \quad (15)$$

with $\psi_{k,\ell} = \frac{\Gamma(k+d_\ell)}{\Gamma(k+1)\Gamma(d_\ell)} + \theta_{\ell\ell} \frac{\Gamma(k+d_\ell-1)}{\Gamma(k)\Gamma(d_\ell)}$, where d_ℓ is the long memory coefficient such that $0 < d < 1/2$ and $\theta_{\ell\ell}$ is the moving average coefficient satisfying $|\theta_{\ell\ell}| < 1$. Concerning the innovations $\{\eta_t^{(\ell)}(\mathbf{s})\}$ in MA(∞) representation (15), these were generated by using the innovation algorithm (see [23]). In this implementation, the temporal covariance of the process $\{\varepsilon_t^{(\ell)}(\mathbf{s})\}$ is given by

$$\Gamma_{\ell\ell'}(u; \Sigma) = \left[\Sigma_{\ell\ell'} \Gamma_{\ell\ell'}(u) \right]_{\ell,\ell'=1}^3,$$

Table 1 Summary results for the KF estimates with N spatial locations on the square $[0, 1]^2$ when Model 1 is considered.

Case		Estimation							
		$\hat{\phi}_{11}$	$\hat{\phi}_{21}$	$\hat{\phi}_{12}$	$\hat{\phi}_{22}$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\rho}$	$\hat{\alpha}$
1	$N = 25$								
$m = 5$									
$T = 200$	Mean	0.4618	0.1498	0.1075	0.6497	0.9570	0.9717	0.4802	0.4665
	sd	0.0162	0.0256	0.0189	0.0126	0.0374	0.0322	0.0194	0.0167
	RelBias	0.0262	-0.2510	-0.2831	-0.0004	-0.0430	-0.0283	-0.0397	-0.0670
	$\sqrt{\text{MSE}}$	0.0199	0.0563	0.0465	0.0125	0.0569	0.0428	0.0277	0.0374
$T = 300$	Mean	0.4661	0.1531	0.1072	0.6472	0.9541	0.9674	0.4778	0.4669
	sd	0.0144	0.0209	0.0196	0.0099	0.0347	0.0317	0.0183	0.0167
	RelBias	0.0358	-0.2347	-0.2854	-0.0044	-0.0459	-0.0326	-0.0445	-0.0663
	$\sqrt{\text{MSE}}$	0.0215	0.0514	0.0471	0.0102	0.0574	0.0454	0.0287	0.0371
$m = 10$									
$T = 200$	Mean	0.4711	0.1320	0.1052	0.6714	0.9706	0.9754	0.4844	0.4750
	sd	0.0175	0.0203	0.0228	0.0125	0.0299	0.0266	0.0154	0.0167
	RelBias	0.0469	-0.3398	-0.2989	0.0329	-0.0294	-0.0246	-0.0312	-0.0500
	$\sqrt{\text{MSE}}$	0.0282	0.0709	0.0503	0.0247	0.0418	0.0361	0.0219	0.0300
$T = 300$	Mean	0.4757	0.1356	0.1053	0.6720	0.9673	0.9745	0.4822	0.4695
	sd	0.0159	0.0181	0.0176	0.0113	0.0313	0.0257	0.0141	0.0146
	RelBias	0.0570	-0.3219	-0.2982	0.0339	-0.0327	-0.0255	-0.0356	-0.0610
	$\sqrt{\text{MSE}}$	0.0302	0.0669	0.0480	0.0247	0.0451	0.0361	0.0227	0.0338
1	$N = 100$								
$m = 5$									
$T = 200$	Mean	0.4641	0.1502	0.1087	0.6482	0.9572	0.9722	0.4865	0.4639
	sd	0.0150	0.0222	0.0195	0.0110	0.0362	0.0324	0.0161	0.0180
	RelBias	0.0314	-0.2488	-0.2752	-0.0028	-0.0428	-0.0278	-0.0271	-0.0721
	$\sqrt{\text{MSE}}$	0.0205	0.0544	0.0456	0.0111	0.0559	0.0426	0.0209	0.0403
$T = 300$	Mean	0.4626	0.1531	0.1098	0.6476	0.9481	0.9614	0.4781	0.4621
	sd	0.0127	0.0217	0.0157	0.0094	0.0362	0.0356	0.0157	0.0195
	RelBias	0.0281	-0.2346	-0.2678	-0.0037	-0.0519	-0.0386	-0.0439	-0.0759
	$\sqrt{\text{MSE}}$	0.0177	0.0515	0.0430	0.0095	0.0629	0.0521	0.0268	0.0425
$m = 10$									
$T = 200$	Mean	0.4724	0.1303	0.1102	0.6720	0.9644	0.9681	0.4827	0.4695
	sd	0.0174	0.0167	0.0180	0.0125	0.0356	0.0299	0.0124	0.0187
	RelBias	0.0497	-0.3483	-0.2653	0.0338	-0.0356	-0.0319	-0.0346	-0.0610
	$\sqrt{\text{MSE}}$	0.0282	0.0716	0.0436	0.0252	0.0502	0.0435	0.0213	0.0357
$T = 300$	Mean	0.4758	0.1334	0.1029	0.6751	0.9529	0.9501	0.4799	0.4678
	sd	0.0149	0.0188	0.0171	0.0145	0.0509	0.0564	0.0289	0.0174
	RelBias	0.0573	-0.3330	-0.3141	0.0386	-0.0471	-0.0499	-0.0401	-0.0644
	$\sqrt{\text{MSE}}$	0.0297	0.0692	0.0501	0.0290	0.0691	0.0751	0.0350	0.0365

where

$$\Gamma_{\ell\ell'}(u) = \frac{\Gamma(1 - d_\ell - d_{\ell'})\Gamma(u + d_{\ell'})}{\Gamma(1 - d_{\ell'})\Gamma(d_{\ell'})\Gamma(u + 1 - d_{\ell'})} \times \left[1 + \theta_{\ell\ell}\theta_{\ell'\ell'} + \theta_{\ell\ell}\frac{u - d_\ell}{u - 1 + d_{\ell'}} + \theta_{\ell'\ell'}\frac{u + d_{\ell'}}{u + 1 - d_\ell} \right],$$

Table 2 Summary results for the KF estimates with N spatial locations on the square $[0, 1]^2$ when Model 2 is considered.

Case		Estimation							
		$\hat{\phi}_{11}$	$\hat{\phi}_{21}$	$\hat{\phi}_{12}$	$\hat{\phi}_{22}$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\rho}$	$\hat{\alpha}$
2	$N = 25$								
$T = 200$									
$m = 5$	Mean	0.1750	0.0240	0.0399	0.3384	0.8111	0.4482	-0.2674	1.2759
	sd	0.0190	0.0174	0.0334	0.0271	0.0577	0.0238	0.0198	0.0474
	RelBias	0.1668	0.0064	-0.0214	-0.0332	-0.1889	-0.1036	-0.1086	-0.1494
	$\sqrt{\text{MSE}}$	0.0314	0.0296	0.0520	0.0294	0.1975	0.0570	0.0381	0.2290
$m = 10$	Mean	0.1712	0.0198	0.0435	0.3443	0.8172	0.4465	-0.2704	1.2696
	sd	0.0157	0.0149	0.0332	0.0235	0.0495	0.0249	0.0208	0.0423
	RelBias	0.1523	-0.1326	-0.0976	-0.0142	-0.1794	-0.0995	-0.0990	-0.1547
	$\sqrt{\text{MSE}}$	0.0260	0.0228	0.0546	0.0190	0.1860	0.0554	0.0356	0.2363
$T = 300$									
$m = 5$	Mean	0.1729	0.0177	0.0451	0.3450	0.8206	0.4502	-0.2703	1.2680
	sd	0.0125	0.0145	0.0309	0.0185	0.0492	0.0245	0.0198	0.0449
	RelBias	0.1523	-0.1001	-0.0645	-0.0142	-0.1794	-0.0995	-0.0990	-0.1547
	$\sqrt{\text{MSE}}$	0.0260	0.0228	0.0546	0.0190	0.1860	0.0554	0.0356	0.2363
$m = 10$	Mean	0.1706	0.0182	0.0438	0.3440	0.8210	0.4513	-0.2704	1.2675
	sd	0.0158	0.0145	0.0313	0.0204	0.0508	0.0252	0.0192	0.0398
	RelBias	0.1373	-0.0231	-0.0045	-0.0172	-0.1790	-0.0973	-0.0988	-0.1550
	$\sqrt{\text{MSE}}$	0.0259	0.0233	0.0537	0.0212	0.1860	0.0547	0.0353	0.2358

and $\Sigma_{\ell\ell'} = [\rho_{\ell\ell'} \sigma_{\ell} \sigma_{\ell'}]_{\ell, \ell'=1}^3$. The parameters considered are the following: $(d_1, d_2, d_3) = (0.4, 0.3, 0.2)$, and $(\theta_{11}, \theta_{22}, \theta_{33}) = (-0.7, -0.5, -0.3)$. To capture the spatial dependence, we consider a trivariate covariance exponential model, defined in equation (13). The parameters of Model 1 together with the variance and correlation parameters are given by $\boldsymbol{\vartheta} = (1, 1, 1, 0.5, 0.5)^\top$. Note that we have assumed $\rho_{\ell\ell'} = 0.5$ for $\ell, \ell' \in \{1, 2, 3\}$ in this simulation scenario. Table 3 reports the results of the Monte Carlo experiment. It is clearly noted that the observed means for the estimates are close to their expected values. Again, with such a small m the KF technique provides satisfactory results.

4.3 Truncation level m

An interesting question that arises from our proposal is which is the optimal truncation level to obtain estimates close to the true value of the parameter. In this section, a study is carried out consisting of data coming from Model 1 for both a VAR(1) and a VARFIMA trivariate model defined in (14) and (15), respectively, and where we consider different values of the truncation level m . In each case, we have used $N = 25$ and 250 observations over time. Figures 1-2 show the estimates for both models. Each horizontal line represents the true value of the parameter. From both Figures we underline that for $m = 5$ the estimates are close to the horizontal line, and this is improved as this value increases. For $m = 30, 40$ the estimates stabilize.

Table 3 Summary results for the KF estimates with N spatial locations on the square $[0, 1]^2$ for the trivariate covariance model in 13.

Model		Estimation										
		$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_3$	$\hat{\theta}_{11}$	$\hat{\theta}_{22}$	$\hat{\theta}_{33}$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_3^2$	$\hat{\rho}$	$\hat{\alpha}$
1	$N = 25$											
$m = 5$												
$T = 200$	Mean	0.4161	0.2566	0.1727	-0.6467	-0.4223	-0.2244	0.9313	0.9402	0.9755	0.5322	0.4651
	sd	0.0170	0.0121	0.0132	0.0179	0.0124	0.0128	0.0308	0.0257	0.0239	0.0124	0.0313
	RelBias	0.0402	-0.1447	-0.1365	-0.0761	-0.1554	-0.2520	-0.0687	-0.0598	-0.0245	0.0644	-0.0698
	$\sqrt{\text{MSE}}$	0.0233	0.0450	0.0303	0.0562	0.0787	0.0767	0.0753	0.0651	0.0341	0.0345	0.0468
$T = 300$	Mean	0.4111	0.2545	0.1688	-0.6535	-0.4242	-0.2293	0.9271	0.9318	0.9624	0.5302	0.4703
	sd	0.0139	0.0107	0.0113	0.0163	0.0107	0.0113	0.0280	0.0260	0.0254	0.0122	0.0237
	RelBias	0.0278	-0.1515	-0.1562	-0.0665	-0.1517	-0.2356	-0.0729	-0.0682	-0.0376	0.0603	-0.0594
	$\sqrt{\text{MSE}}$	0.0177	0.0467	0.0332	0.0493	0.0766	0.0716	0.0780	0.0729	0.0453	0.0325	0.0379
$m = 10$												
$T = 200$	Mean	0.4023	0.2494	0.1689	-0.6623	-0.4268	-0.2334	0.9143	0.9373	0.9768	0.5344	0.4748
	sd	0.0160	0.0147	0.0131	0.0144	0.0156	0.0165	0.0259	0.0268	0.0238	0.0119	0.0240
	RelBias	0.0058	-0.1688	-0.1554	-0.0538	-0.1465	-0.2220	-0.0857	-0.0627	-0.0232	0.0688	-0.0503
	$\sqrt{\text{MSE}}$	0.0161	0.0527	0.0337	0.0403	0.0749	0.0686	0.0895	0.0681	0.0331	0.0364	0.0347
$T = 300$	Mean	0.4032	0.2489	0.1677	-0.6628	-0.4285	-0.2337	0.9183	0.9346	0.9690	0.5326	0.4738
	sd	0.0142	0.0116	0.0115	0.0137	0.0110	0.0116	0.0221	0.0259	0.0246	0.0099	0.0180
	RelBias	0.0079	-0.1703	-0.1615	-0.0532	-0.1431	-0.2210	-0.0817	-0.0654	-0.0310	0.0652	-0.0525
	$\sqrt{\text{MSE}}$	0.0144	0.0524	0.0343	0.0396	0.0724	0.0673	0.0846	0.0703	0.0395	0.0341	0.0318

This suggests that the estimates obtained by the approximated MLE require a small number of m .

5 Real data illustration

This section analyzes part of the integrated Agromet network which contains more than 100 meteorological stations throughout Chile, including daily temperature data, soil temperature, rainfall, humidity, solar radiation, wind speed and direction, among others. This dataset is reported and updated by the Institute of Agricultural Research (INIA) and can be available from the website <http://www.agromet.cl>.

We focus our interest in three regions, namely Maule, Biobío and Araucanía which represent a portion of south-central Chile. This area is surrounded by mountains (mountain of the coast to the east and mountain of the Andes to the west). According to the 2007 agriculture national census, this area has a surface of 99,206 km^2 . The Maule Region concentrates 17,2% of the crop national area, its main use corresponds to forest plantations, followed by cereals, fruit trees, forage plants and vineyards and parronals, groups that together respond to 94% of the area of crops in the region. The BíoBío region accounts for 28,1% of the crop national area. The main use, with 79,0% of the total, is for forest plantations, with cereals and forage plants, but with a smaller participation. Finally, Araucanía region covers 20,6% of the crop national area, its main use corresponding to forest plantations with 64,3% of that total, followed by grains with 18,5% and forage plants with 9,8%. These

Fig. 1 Estimates of parameters as a function of m for the MA approximation of a VAR(1) model defined in (14) with Model 1 and Case 1 as simulation schemes. Panel (a): Estimation of vector Φ . Panel (b): Estimation of vector θ . In all panels the continuous line represents the true value of the parameter.

Fig. 2 Estimates of parameters as a function of m for the MA approximation of a VARFIMA trivariate model defined in (15) with covariance function following an exponential model. Panel (a): Temporal structure parameters (d , Θ). Panel (b): Estimation of vector θ . In all panels the continuous line represents the true value of the parameter.

three uses account for 92,6% of the region soil. As can be noted, the study area is

mostly devoted to agriculture, and therefore both air temperature and solar radiation are predominant factors in crop growth; note that extreme values of temperature and radiation affect the production and quality of agricultural and fruit products. The frost causes a deterioration of production, reducing the activity of the agricultural industry in the south-central zones of our country, generating losses of thousands of dollars, besides to the paralyzation and the low activity in the exports to the external market. In this way, it is of great interest to study the spatio-temporal variability of these meteorological processes, and therefore to generate proposals that help to explain such variability in a coherent and appropriate way.

In particular, we study the behavior of the space-time variability of the average daily temperature and maximum daily solar radiation, information obtained from the 21 meteorological stations located between Maule, Biobío and Araucanía Regions. Figure 3 displays the spatial distribution of these meteorological stations.

Fig. 3 Locations of the selected 21 meteorological stations in the Maule, Biobío and Araucanía regions, Chile.

In section A.3 of the supplementary material we can find some exploratory analysis. In particular, Panel (a) of Figure A3.4 shows a plot of the average daily temperature time series observed during the year 2016, and panel (b) displays the maximum daily solar radiation time series. Also, Table A3.1 summarizes some descriptive

statistics of the average daily temperature $Y_t^{(1)}$ and the maximum daily radiation $Y_t^{(2)}$ for some selected meteorological stations.

5.1 The model

We consider the following bivariate spatio-temporal model for the Agromet data

$$\begin{bmatrix} Y_t^{(1)}(\mathbf{s}) \\ Y_t^{(2)}(\mathbf{s}) \end{bmatrix} = \begin{bmatrix} M_{t,1}^\top & 0 \\ 0 & M_{t,2}^\top \end{bmatrix} \begin{bmatrix} \beta_1^\top \\ \beta_2^\top \end{bmatrix} + \begin{bmatrix} \varepsilon_t^{(1)}(\mathbf{s}) \\ \varepsilon_t^{(2)}(\mathbf{s}) \end{bmatrix}. \quad (16)$$

The time series shown in Figure A2.2 seems to present a seasonal component. In particular, a joint analysis of the periodogram of the detrended data and the autocorrelation function (ACF) revealed the presence of plausible seasonal frequencies at $\omega_1 = \frac{2\pi}{366.25}$ and $\omega_2 = \frac{2\pi}{183.25}$. Thus, the following non-stochastic regressors are considered

$$M_{t,1}^\top = M_{t,2}^\top = \beta_0 + \sum_{j=1}^2 \beta_j \sin(t\omega_j) + \alpha_j \cos(t\omega_j) + \beta_3 h(\mathbf{s}) + \beta_4 \text{lat},$$

where $h(\mathbf{s})$ is the elevation and the covariate lat stands for the latitude; note that both affect the temperature and radiation. In order to obtain information on the temporal correlation structure of the data, we study the residuals $\mathbf{e}_T = \{\mathbf{Y}_t(\mathbf{s}) - \mathbf{M}_t(\mathbf{s})\beta | t = 1, \dots, T\}$.

The marginal sample ACF and cross-ACF of the residuals are shown in Figure A3.5 of supplementary material. Both plots suggest that the time series exhibit short-range dependence. In particular, the disturbances $\{\varepsilon_t^{(2)}(\mathbf{s})\}$ in the linear regression model given in (16) have a VAR(1) structure, i.e.,

$$\begin{bmatrix} \varepsilon_t^{(1)}(\mathbf{s}) \\ \varepsilon_t^{(2)}(\mathbf{s}) \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} \varepsilon_{t-1}^{(1)}(\mathbf{s}) \\ \varepsilon_{t-1}^{(2)}(\mathbf{s}) \end{bmatrix} + \begin{bmatrix} \eta_t^{(1)}(\mathbf{s}) \\ \eta_t^{(2)}(\mathbf{s}) \end{bmatrix},$$

where $\{\eta_t(\cdot)\}$ are independent over time and follow a stationary Gaussian spatial process with mean zero and spatial covariance structure defined in (13) with $\ell = 2$. Table 4 reports the parameter estimates using the *KF* with truncation level $m = 5$. We can observe that the estimated autoregressive coefficients are high, $\phi_{11} = 0.5845$ and $\phi_{22} = 0.5033$, which suggest that the temperature and solar radiation data are highly correlated to the temperature of the previous day. On the other hand, the temporal dependence of $Y_t^{(1)}$ on $Y_t^{(2)}$ is strong with $\phi_{12} = 0.2959$, the relationship between $Y_t^{(2)}$ and $Y_t^{(1)}$ is weak with $\phi_{21} = 0.0053$. Additionally, the estimated range parameter is high, $\alpha = 1.4378$, in relation to the spatial sampling scheme, with a maximum distance between the stations of $h = 381.8$ kms. This informs that the spatial process has a high spatial continuity. Finally, the correlation coefficient between $Y_t^{(1)}(\mathbf{s}_i)$ and $Y_t^{(1)}(\mathbf{s}_j)$ for $i, j = 1, \dots, 21$ is positive with value $\rho = 0.1011$.

Table 4 Parameter estimates for average daily temperature and daily solar radiation data.

		Parameter						
$M_t(\mathbf{s})$	β_0	β_1	α_1	β_2	α_2	β_3	β_4	
	35.9932	2.3869	-0.0372	-0.0034	0.1793	0.2545	-0.0085	
$\varepsilon_t(\mathbf{s})$	ϕ_{11}	ϕ_{12}	ϕ_{21}	ϕ_{22}	σ_1^2	σ_2^2	ρ	α
	0.5845	0.2959	0.0053	0.5033	1.0681	0.9994	0.1011	1.4378

5.2 Performance of the prediction procedure

Before making predictions in areas where there are no observations, we conduct a cross-validation study to see the performance of the model represented by equation (16) (denoted here by Model 1). For this, two additional models will be defined, which are described below:

- Model 2: A VAR(1) model as that in (16) is considered. However, the setting of both the SS representation and the KF differs from the proposal of this paper, that is, here we will not use the truncated MA representation as in (10). Here we use the classical development of the KF (see [23]) with $\mathbf{G}_t(\mathbf{s}) = \mathbf{I}_2$, $\mathbf{X}_t(\mathbf{s}) = \mathbf{Y}_t(\mathbf{s})$, $\mathbf{F}_t(\mathbf{s}) = \Phi$, and $\mathbf{V}_t(\mathbf{s}) = [\eta_t^{(1)}(\mathbf{s}), \eta_t^{(2)}(\mathbf{s})]^\top$ in (9).
- Model 3: A more general version of model (16) is considered by using location-dependent parameters that describe spatial non-stationarity in the VAR(1) model.

In this case, the sequence of matrices in (10) satisfies $\Psi_j(\mathbf{s}) = \Phi^j(\mathbf{s}) = \begin{bmatrix} \phi_{11}(\mathbf{s}) & \phi_{12}(\mathbf{s}) \\ \phi_{21}(\mathbf{s}) & \phi_{22}(\mathbf{s}) \end{bmatrix}^j$, with $\mathbf{s} = (\mathbf{s}_1, \dots, \mathbf{s}_{21})$ and $j = 1, \dots, m$.

The estimates of $\Phi(\mathbf{s})$ for Model 3 are carried out using non-parametric techniques based on cubic splines. In Figure 4 the continuous line represents the estimates (through cubic splines) of the spatial movement of the parameters of the VAR(1) model. The dots represent the heuristic approach (i.e., for each location, the parameters determining a stationary VAR(1) process are estimated), and the horizontal dashed line indicates the estimates of stationary VAR(1) model with KF . To compare the prediction performance of Models 1-3, a cross-validation method was proposed. Towards this end, we have removed the data $\{\mathbf{Y}_t(\mathbf{s}_i) : t = 1, \dots, 366\}$ for each $i = 1, \dots, 21$, and predicted $\mathbf{Y}_t(\mathbf{s}_i)$ from the remaining data. The predicted value of $\mathbf{Y}_t(\mathbf{s}_i)$ is denoted by $\mathbf{Y}_t^{(j)}(\mathbf{s}_i)$, where $j = 1, 2, 3$ represents Models 1-3, respectively. Finally, the criterion used to quantify the predictive power was the mean squared error of prediction (MSEP) given by

$$\text{MSEP}(\mathbf{Y}_t^{(\cdot)}(\mathbf{s}_i)) = \frac{1}{366} \sum_{t=1}^{366} [\mathbf{Y}_t(\mathbf{s}_i) - \mathbf{Y}_t^{(\cdot)}(\mathbf{s}_i)]^2,$$

Fig. 4 Agromet data: Top-Left, estimates of $\phi_{11}(s)$. Top-Right, estimates of $\phi_{12}(s)$. Bottom-Left, estimates of $\phi_{21}(s)$. Bottom-Right, estimates of $\phi_{22}(s)$. In all panels the continuous line represents the curves estimated through cubic splines, the dots represent the heuristic approach, and the horizontal dashed line indicates the estimates of stationary VAR(1) model with KF .

and the gain, to quantify how better Model 1 is, is given by

$$\text{Gain}_j = \left(\frac{\text{MSEP}(\mathbf{Y}_t^{(j)}(\mathbf{s}_i))}{\text{MSEP}(\mathbf{Y}_t^{(1)}(\mathbf{s}_i))} - 1 \right) \times 100\% \quad \text{for } j = 2, 3.$$

Table 5 shows the gain results for the MSEP through a cross-validation method using the KF algorithm for Models 1- 3. The gain of Model 1 with respect to Model 2 is similar in both variables, on average the profit is 0.8% and 0.7% respectively. While the gain of Model 1 with respect to Model 3 is significant only in variable $Y_t^{(2)}(\mathbf{s}_i)$, with an average gain of 3.6%. However, the first variable shows a slight gain in favor of Model 3. From this information we can say that our proposal provides a potentially powerful prediction tool compared to Models 2-3.

5.3 Prediction

The KF algorithm has proven to be a powerful tool for interpolation, see [27]. Taking advantage of the benefits of this algorithm, we use the procedure shown in Section 3.2 to predict the average temperature and radiation variables at unobserved locations for the three regions under study, that is, Maule, Biobío and Araucanía regions. For this purpose we considered a grid of 5955 nodes within the study area.

In each of the nodes, we made predictions for the two variables, therefore we interpolated 11910 records. Additionally, the time domain considered was from January 01 to December 31, 2016, that is 365 days, with a total of 4347150 observations.

Table 5 Gain obtained through a cross-validation method using Models 1-3

Station	ID	Gain ₂ %		Gain ₃ %	
		$Y_t^{(1)}(s_i)$	$Y_t^{(2)}(s_i)$	$Y_t^{(1)}(s_i)$	$Y_t^{(2)}(s_i)$
Coronel del Maule	1	0.2149	-0.5302	-0.7299	1.9439
Los despachos	2	1.5979	-0.7680	-0.8654	2.8899
Chanco	3	0.5868	0.4406	-0.1610	3.4481
Santa Sofía	4	-0.9589	1.8506	-0.6698	2.1915
San Clemente	5	0.8052	0.7750	-0.1333	0.8885
Coronel	6	1.0312	7.5340	-0.2641	4.5875
Chiguayante	7	1.6519	2.0511	-0.4154	4.9166
Human	8	1.9147	2.6118	-0.3967	3.6875
Cañete	9	-1.0350	-3.6344	-0.1074	3.3202
Nueva Aldea	10	0.4047	-0.0339	-0.3177	1.4633
Ninhue	11	2.7436	-2.3497	-0.5329	1.4357
Navidad	12	-0.7507	1.1819	-0.5580	3.6217
Punta Parra	13	-1.4067	0.8627	-0.2837	5.6725
Sta Rosa	14	1.5834	1.4270	-0.6897	3.5880
Dominguez	15	0.2926	-1.8261	-0.1475	6.4187
C. Llollinco	16	1.7383	-2.2445	-0.0523	4.2171
Cuarta Faja	17	1.6171	-4.5170	-0.0487	2.8786
Quiripio	18	-1.5453	1.0951	-0.1922	8.7234
San Luis	19	0.4883	0.4539	-0.1199	2.1234
Tranapuate	20	1.1039	1.3299	-0.3594	5.4774
Sta. Adela	21	-1.8847	0.2858	-0.2652	3.1700
Average		0.4854	0.3360	-0.3481	3.6506

Due to space constraints, only a subset of the results are presented, in particular we report the predictions for the fortnights of each month. Figures 5 and 6 display the predictions every fifteen days for all the months of the year. In addition, Figures 7 and 8 show the prediction variance maps.

From the images of predictions, a fairly marked general seasonality can be observed. Lower temperatures are associated with the winter season, and higher with the summer season. On the other hand, a very strong systematic component is observed in the Andes mountain in every month. In relation to the prediction variance, we note that in the observation locations a smaller prediction variance is reflected.

6 Conclusions

We have presented a state-space methodology to model multivariate spatio-temporal processes. In particular, we have proposed to model the temporal dependence structure through the infinite moving average representation $MA(\infty)$. In this context, we

Fig. 5 Temperature space-time predictions of the fortnights of each month of the year 2016

Fig. 6 Radiation space-time predictions of the fortnights of each month of the year 2016

have incorporated the ARFIMA models to quantify the temporal correlation and valid Matérn cross-covariance models to characterize the spatial correlation in the spatio-temporal processes. In terms of the estimation procedure, we have proposed an approximation to the likelihood functions via truncation which provides an efficient means to calculate the MLE. Simulation studies evidenced that the proposed

Fig. 7 Temperature prediction variance of the fortnights of each month of the year 2016.

Fig. 8 Radiation prediction variance of the fortnights of each month of the year 2016

approach can be extremely efficient for small truncation levels. Furthermore, this approach allows to overcome the computational burdens while reducing substantially the size of the required memory whenever we deal with large spatio-temporal datasets.

In general, we can mention that one of the major advantages of using multivariate KF , is the use of a single model that explains all the phenomena in question, a remarkable quality in practice, aiming at explaining the phenomenon simply using most of the available information. In contrast to the more classical geostatistical techniques that only incorporate spatial information, where we have to propose a different model associated with a given instant of time.

Our methods can be easily extended to the estimation of multivariate space-time covariance models when considering fully symmetric covariance models. For estimation of asymmetric covariance models ([36]), asymmetry in time should be taken into account in the state-space system following the lines of Section 3. Finally, note that a typical problem, shared in a number of practical applications, is that many multivariate spatio-temporal datasets are affected by missing data. The SS methodology allows to directly tackle this issue only in the univariate time series case. [32] developed space-time estimation and prediction methods in presence of missing data by using an EM algorithm. Thus, that is a promising topic for future research, which will allow to face missing values in multivariate spatio-temporal data.

Acknowledgments

Guillermo Ferreira would like to express his thanks for the support from ENLACE 2018.014.028-1, established by the Universidad de Concepción and Center for the Discovery of Structures in Complex Data (MiDas). Jorge Mateu has been partially funded by grant MTM2016-78917-R from the Ministerio de Economía y Competitividad and by grant UJI-B2018-04 from University Jaume I.

Appendix A. Supplementary data

The following is the Supplementary material related to this article.

References

1. Alegria, A., Porcu, E., Furrer, R.: Asymmetric matrix-valued covariances for multivariate random fields on spheres. *Journal of Statistical Computation and Simulation* **88**(10), 1850–1862 (2018)
2. Bevilacqua, M., Alegria, A., Velandia, D., Porcu, E.: Composite likelihood inference for multivariate gaussian random fields. *Journal of Agricultural, Biological, and Environmental Statistics* **21**(3), 448–469 (2016)
3. Bevilacqua, M., Hering, A.S., Porcu, E.: On the flexibility of multivariate covariance models: Comment on the paper by genton and kleiber. *Statistical Science* **30**(2), 167–169 (2015)
4. Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Zabkar, R., Carmichael, G., Flemming, J., Inness, A., Pagowski, M., Pérez Camacho, J., et al.: Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models. *Atmospheric Chemistry and Physics* **15**(10), 5325–5358 (2015)
5. Bourotte, M., Allard, D., Porcu, E.: A flexible class of non-separable cross-covariance functions for multivariate space-time data. *Spatial Statistics* **18**, 125–146 (2016)
6. Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M.: Time series analysis: forecasting and control. John Wiley & Sons (2015)

7. Bradley, J.R., Holan, S.H., Wikle, C.K.: Multivariate spatio-temporal models for high-dimensional areal data with application to longitudinal employer-household dynamics. *The Annals of Applied Statistics* **9**(4), 1761–1791 (2015)
8. Chan, N.H., Palma, W.: State space modeling of long-memory processes. *Annals of Statistics* pp. 719–740 (1998)
9. Chaudhary, A.: An introduction to the interface between c and r. Tech. rep. (2007)
10. Cheng, C., Sa-Ngasongsong, A., Beyca, O., Le, T., Yang, H., Kong, Z., Bukkapatnam, S.T.: Time series forecasting for nonlinear and non-stationary processes: A review and comparative study. *Iie Transactions* **47**(10), 1053–1071 (2015)
11. Dai, Y.H.: Convergence properties of the bfgs algorithm. *SIAM Journal on Optimization* **13**(3), 693–701 (2002)
12. Daley, D.J., Porcu, E., Bevilacqua, M.: Classes of compactly supported covariance functions for multivariate random fields. *Stochastic Environmental Research and Risk Assessment* **29**(4), 1249–1263 (2015)
13. Daniels, M.J., Zhou, Z., Zou, H.: Conditionally specified space-time models for multivariate processes. *Journal of Computational and Graphical Statistics* **15**(1), 157–177 (2006)
14. Durbin, J., Koopman, S.J.: *Time Series Analysis by State Space Methods*, *Oxford Statistical Science Series*, vol. 24. Oxford University Press, Oxford (2001)
15. Ferreira, G., Mateu, J., Porcu, E.: Spatio-temporal analysis with short- and long-memory dependence: a state-space approach. *TEST* **27**(1), 221–245 (2017)
16. Ferreira, G., Rodríguez, A., Lagos, B.: Kalman filter estimation for a regression model with locally stationary errors. *Computational Statistics & Data Analysis* **62**, 52–69 (2013)
17. Genton, M.G.: Separable approximations of space-time covariance matrices. *Environmetrics: The official journal of the International Environmetrics Society* **18**(7), 681–695 (2007)
18. Genton, M.G., Kleiber, W.: Cross-covariance functions for multivariate geostatistics. *Statistical Science* **30**(2), 147–163 (2015)
19. Gneiting, T.: Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association* **97**(458), 590–600 (2002)
20. Grassi, S., de Magistris, P.S.: When long memory meets the kalman filter: A comparative study. *Computational Statistics & Data Analysis* **76**, 301–319 (2014)
21. Grewal, M.S.: *Kalman filtering*. Springer (2011)
22. Grewal, M.S., Weill, L.R., Andrews, A.P.: *Global positioning system, inertial navigation and integration*. a john wiley and sons. Inc. Publication (2001)
23. Hamilton, J.D.: *Time series analysis*. Princeton university press (2020)
24. Huang, H.C., Cressie, N.: Spatio-temporal prediction of snow water equivalent using the kalman filter. *Computational Statistics & Data Analysis* **22**(2), 159–175 (1996)
25. Kalman, R.E.: A new approach to linear filtering and prediction problems. *Journal of Basic Engineering* **82**(1), 35–45 (1960)
26. Kim, H.M., Ryu, D., Mallick, B.K., Genton, M.G.: Mixtures of skewed kalman filters. *Journal of Multivariate Analysis* **123**, 228 – 251 (2014)
27. Lagos-Álvarez, B., Padilla, L., Mateu, J., Ferreira, G.: 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 (2019)
28. Mardia, K.V., Goodall, C., Redfern, E.J., Alonso, F.J.: The kriged kalman filter. *Test* **7**(2), 217–282 (1998)
29. Matérn, B.: *Spatial variation*. Lecture Notes in Statistics **36** (1986)
30. Maybeck, P.S.: Square root filtering. *Stochastic models, estimation and control* **1**, 368–409 (1979)
31. Morana, C.: Multivariate modelling of long memory processes with common components. *Computational Statistics & Data Analysis* **52**(2), 919 – 934 (2007)
32. Padilla, L., Lagos-Álvarez, B., Mateu, J., Porcu, E.: Space-time autoregressive estimation and prediction with missing data based on kalman filtering. *Environmetrics* p. e2627 (2020)
33. Padilla-Buriticá, J.I., Giraldo, E., Castellanos-Domínguez, G.: Eeg source localization based on multivariate autoregressive models using kalman filtering. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 7151–7154. IEEE (2011)
34. Porcu, E., Furrer, R., Nychka, D.: 30 years of space–time covariance functions. *Wiley Interdisciplinary Reviews: Computational Statistics* p. e1512 (2019)
35. R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2015)

36. Stein, M.L.: Space-time covariance functions. *Journal of the American Statistical Association* **100**(469), 310–321 (2005)
37. Stroud, J.R., Stein, M.L., Lesht, B.M., Schwab, D.J., Beletsky, D.: An ensemble kalman filter and smoother for satellite data assimilation. *Journal of the American Statistical Association* **105**(491), 978–990 (2010)
38. Tsay, W.J.: Maximum likelihood estimation of stationary multivariate arfima processes. *Journal of Statistical Computation and Simulation* **80**(7), 729–745 (2010)
39. Wang, Y., Chaib-draa, B.: Knn-based kalman filter: An efficient and non-stationary method for gaussian process regression. *Knowledge-Based Systems* **114**, 148 – 155 (2016)
40. Wikle, C.K.: Hierarchical models in environmental science. *International Statistical Review* **71**(2), 181–199 (2003)
41. Xu, K., Wikle, C.K.: Estimation of parameterized spatio-temporal dynamic models. *Journal of Statistical Planning and Inference* **137**(2), 567–588 (2007)
42. Zes, D.: Facile spacio-temporal modeling, forecasting with adaptive least squares and the kalman filter. *Journal of Environmental Statistics* **6**(1) (2014)
43. Zheng, F., Zhu, J.: Balanced multivariate model errors of an intermediate coupled model for ensemble kalman filter data assimilation. *Journal of Geophysical Research: Oceans* **113**(C7) (2008)
44. Zhu, C., Byrd, R.H., Lu, P., Nocedal, J.: Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)* **23**(4), 550–560 (1997)