

Spatio-temporal analysis with short- and long-memory dependence: a state-space approach

Guillermo Ferreira¹  · Jorge Mateu² · Emilio Porcu³

Received: 29 July 2016 / Accepted: 27 April 2017 / Published online: 10 May 2017
© Sociedad de Estadística e Investigación Operativa 2017

Abstract This paper deals with the estimation and prediction problems of spatio-temporal processes by using state-space methodology. The spatio-temporal process is represented through an infinite moving average decomposition. This expansion is well known in time series analysis and can be extended straightforwardly in space–time. Such an approach allows easy implementation of the Kalman filter procedure for estimation and prediction of linear time processes exhibiting both short- and long-range dependence and a spatial dependence structure given on the locations. Furthermore, we consider a truncated state-space equation, which allows to calculate an approximate likelihood for large data sets. The performance of the proposed Kalman filter approach is evaluated by means of several Monte Carlo experiments implemented under different scenarios, and it is illustrated with two applications.

Keywords Kalman filter algorithm · Second-order stationary · Space–time geostatistics · Time series models

Mathematics Subject Classification 62M10 · 62M20 · 62M30

✉ Guillermo Ferreira
gferreir@udec.cl

¹ Department of Statistics, Universidad de Concepción, Concepción, Chile

² Department of Mathematics, University Jaume I, Castellón, Spain

³ Department of Mathematics, University Federico Santa María, Valparaiso, Chile

1 Introduction

In the last decade, there has been an increasing interest in modeling spatio-temporal data that result from dynamic processes in constant evolution in both space and time. Geostatistics has been used to deal with such spatio-temporal processes by providing covariance models to analyze the dependence structure. Since the eighties, a wealth of papers have provided a variety of spatio-temporal models that have been used in many scientific areas, such as hydrology, environmental sciences, geology, astronomy, neuroscience, ecology, atmospheric sciences, oceanography, and economy; this list is indeed virtually endless [see for instance, [Eynon and Switzer \(1983\)](#), [Bilonick \(1985\)](#), [Oehlert \(1993\)](#) among others]. Usually, the proposed models exhibit a hierarchical structure incorporating spatial or spatio-temporal dependencies. Indeed, approaches based on hierarchical spatio-temporal structures have been used to achieve flexible models that capture the dynamic behavior of the data. This has been the case in [Brown et al. \(1994\)](#) who applied a hierarchical model to a relatively low-dimensional space–time air pollution problem. Similarly, [Handcock and Wallis \(1994\)](#) used a Bayesian kriging approach for space–time modeling of meteorological fields. [Waller et al. \(1997\)](#) employed hierarchical Bayesian space–time models for mapping disease rates. [Wikle et al. \(1998\)](#) used this class of models in the analysis of monthly averaged maximum temperature data distributed in space and time, and [Hughes et al. \(1999\)](#) used hidden Markov models with unobserved weather states to model space–time atmospheric precipitation. We can also find applications in environmental pollution problems, where, for example, [Ippoliti \(2001\)](#) analyzed the levels of sulfur dioxide in Milan (Italy), providing a spatio-temporal state-space representation. [Fasso et al. \(2007\)](#) combined the state-space method with calibration techniques and applied them to fine particulate matter (PM₁₀) data in the spatio-temporal dimension. [Cameletti et al. \(2013\)](#) considered a hierarchical spatio-temporal model for particulate matter concentrations. This proposal involves a Gaussian Field, affected by a measurement error, and a state process characterized by a first-order autoregressive dynamic model and spatially correlated innovations. Another interesting spatio-temporal application for the interpolation of daily rainfall data using state-space models has been proposed by [Militino et al. \(2015\)](#). For an overview of hierarchical dynamical spatio-temporal models, the recent book by [Cressie and Wikle \(2011\)](#) provides an excellent starting point for researchers in this area.

In these modeling exercises, the Kalman filter algorithm has proved to be a powerful tool for the statistical treatment of state-space models, providing the estimation of parameters (given in the state vector) and the prediction of unobserved values at a specific location. For instance, [Huang and Cressie \(1996\)](#) and [Wikle \(2003\)](#) developed empirical Bayesian space–time Kalman filter models for the investigation of snow water equivalent and monthly precipitation. [Mardia et al. \(1998\)](#) considered a mixed approach between the Kalman filter 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. [Wikle and Cressie \(1999\)](#) presented an approach to space–time prediction that achieves dimension reduction through a spatio-temporal Kalman filter. [Xu and Wikle \(2007\)](#) proposed a spatio-temporal dynamic model formulation with restricted parameter matrices based on prior scien-

tific knowledge, and developed a general expectation–maximization (GEM) algorithm to carry out the estimations. [Stroud et al. \(2010\)](#) applied a dynamic state-space model to a sequence of SeaWiFS satellite images on the Lake Michigan, where a great amount of sediments was observed after a great storm. In this study, the authors implemented a comprehensive version of the Kalman filter, called ensemble Kalman filter, which allows to deal with problems of nonlinearities and high dimensionality inherent in satellite images. They were able to provide maps of concentrations of sediments with uncertainties in space and time. To deal with forecasting on spatio-temporal processes, [Zes \(2014\)](#) used the state-space system and time-varying parameter least squares autoregressive system, with their respective solving algorithms, the Kalman filter, and autoregressive adaptive least squares (ALS). More recently, [Bocquet et al. \(2015\)](#) discussed the methods available for data assimilation in atmospheric models, including ensemble Kalman filter. A common feature in the above-mentioned papers is that the unobserved state vector is responsible of capturing the temporal dependence through a Markovian temporal evolution, with autoregressive or vector autoregressive type of models. This means that the proposed approaches belong to the class of short-memory models in time series data.

Our approach considers a more general framework. Specifically, we allow to capture the temporal dependence of both short- and long-memory processes, as well as modeling the spatial dependence. We use the Kalman filter algorithm for estimation and prediction of spatio-temporal processes, but on the basis of a new updating scheme of the unobserved state vector, which is different from the above-mentioned proposals for the following reasons:

- We propose the use of an infinite moving average expansion ($MA(\infty)$) as a form of representing linear processes to deal with spatio-temporal models.
- Our proposal includes short- or long-memory models to capture the temporal dependence, such as $ARMA(p, q)$ and $ARFIMA(p, d, q)$ models.
- Instead of directly calculating the likelihood of the spatio-temporal process, we propose an approximation to the likelihood function based on the truncated state-space representation. This, to some extent, reduces the size of memory required and overcomes the computational burdens.
- Finally, a methodology of imputation of missing observations is also proposed.

The plan of the paper is the following. Section 2 discusses a class of spatio-temporal processes and their representation in $MA(\infty)$ expansions. Section 3 presents the state-space models and the Kalman filter algorithm for estimating the parameters involved in the temporal dependence and the spatial structure. In Sect. 4, a simulation study reveals the adequacy and the good behavior of our proposal under a variety of practical scenarios. Section 5 applies our proposal to global total column ozone levels and Irish wind speed data. The paper ends with some final conclusions and a discussion.

2 A class of spatio-temporal processes

Consider the class of spatio-temporal processes given by the infinite moving average expansion,

$$Y_t(\mathbf{s}) = \mathbf{M}_t(\mathbf{s})\boldsymbol{\beta} + \varepsilon_t(\mathbf{s}), \tag{1}$$

$$\varepsilon_t(\mathbf{s}) = \sum_{j=0}^{\infty} \psi_j \eta_{t-j}(\mathbf{s}), \tag{2}$$

for $t = 1, \dots, T$, where \mathbf{s} represents a location in the spatial domain $D \subset \mathbb{R}^2$, $\boldsymbol{\beta} = [\beta_1, \dots, \beta_p]^\top$ is a vector of parameters $\mathbf{M}_t(\mathbf{s}) = [M_t^{(1)}(\mathbf{s}), \dots, M_t^{(p)}(\mathbf{s})]$ is a p -dimensional vector of non-stochastic regressors, $\{\psi_j\}$ is a sequence of coefficients satisfying $\sum_{j=0}^{\infty} \psi_j^2 < \infty$, and $\{\eta_t(\mathbf{s})\}$ a sequence of temporally independent and spatially stationary Gaussian processes, with $\mathbb{E}(\eta_t(\mathbf{s})) = 0 \forall \mathbf{s} \in D$, and covariance function

$$\text{cov}(\eta_t(\mathbf{s}), \eta_t(\mathbf{s}')) = C^\eta(\|\mathbf{s} - \mathbf{s}'\|; \boldsymbol{\theta}), \quad \mathbf{s}, \mathbf{s}' \in D. \tag{3}$$

Here, $C^\eta : [0, \infty) \rightarrow \mathbb{R}$ such that the composition $C^\eta \circ \|\cdot\| : \mathbb{R}^2 \rightarrow \mathbb{R}$ is an isotropic covariance function (Daley and Porcu 2014), where $\|\cdot\|$ is the Euclidean distance, \circ denotes composition, and $\boldsymbol{\theta}$ is a parameter vector. This representation is similar to the well-known MA(∞) decomposition for the error sequence $\{\varepsilon_t(\mathbf{s})\}$. We say that a stationary process with lag- h temporal covariance $\kappa(h)$ is said to have *short-memory* if $\sum_{h=-\infty}^{\infty} |\kappa(h)| < \infty$, and in this case the process in (2) will be called a *short-memory process*. On the other hand, if $\sum_{h=-\infty}^{\infty} |\kappa(h)| = \infty$, the process in Eq. (2) will be called a *long-memory process*. Another characterization is based directly on the MA(∞) decomposition of the process (2). It is said that the process $\{\varepsilon_t(\mathbf{s})\}$ has short-memory if $\psi_j \sim \exp(-aj)$ for $j \geq 1$, with a a positive constant. On the other hand, the process has long-memory if $\psi_j \sim j^{d-1}$ for some $d \in (0, 1/2)$ [see Palma et al. (2013)]. Here, \sim means that the ratio of both sides tends to one. We now consider the issue of the spatial and temporal dependencies of the process $Y_t(\mathbf{s})$. In the case of the MA(∞) decomposition for the spatio-temporal process defined by (1), the covariance across both space and time is given by

$$\begin{aligned} \text{cov}(Y_t(\mathbf{s}), Y_{t'}(\mathbf{s}')) &= \sum_{i,j=0}^{\infty} \psi_j \psi_i \text{cov}(\eta_{t-j}(\mathbf{s}), \eta_{t'-i}(\mathbf{s}')) \\ &= C^\eta(\xi) \sum_{j=0}^{\infty} \psi_j \psi_{j+|t-t'|} = C^\eta(\xi) \kappa(h), \end{aligned}$$

where $\xi = \|\mathbf{s} - \mathbf{s}'\|$, $\mathbf{s}, \mathbf{s}' \in D$, $h = t - t'$, $t, t' \in \mathbb{Z}_+$ and $\kappa(\cdot)$ is a temporal covariance function. Thus, the spatio-temporal covariance function of the process $Y_t(\mathbf{s})$ can be written as the product of a purely spatial with a purely temporal covariance function. In such a case, we say that $Y_t(\mathbf{s})$ has a separable spatio-temporal covariance function (Gneiting 2002). For a neater and self-contained exposition, we discuss some examples below.

Example 1 As an example of short-memory process, we consider the regression model with autoregressive moving average errors, denoted by ARMA(p, q), and defined as

$$\varepsilon_t(\mathbf{s}) = \sum_{j=1}^p \phi_j \varepsilon_{t-j}(\mathbf{s}) + \sum_{j=1}^q \theta_j \eta_{t-j}(\mathbf{s}) + \eta_t(\mathbf{s}), \quad \mathbf{s} \in D,$$

for $t = 1, 2, \dots, T$. When $p = q = 1$, i.e., $\varepsilon_t(\mathbf{s}) = \phi \varepsilon_{t-1}(\mathbf{s}) + \theta \eta_{t-1}(\mathbf{s}) + \eta_t(\mathbf{s})$, we obtain a special case of the errors in (2) with coefficients $\psi_j = (\phi - \theta) \phi^{j-1}$ for $j \geq 1$ and $\psi_j = 1$ for $j = 0$ in the MA(∞) process (hereafter, we assume that $\psi_0 = 1$, unless specified otherwise). When $q = 0$, we have an autoregressive AR(p) process defined as $\varepsilon_t(\mathbf{s}) = \sum_{j=1}^p \phi_j \varepsilon_{t-j}(\mathbf{s}) + \eta_t(\mathbf{s})$, for $t = 1, 2, \dots, T$. If $p = 1$, then $\varepsilon_t(\mathbf{s}) = \phi \varepsilon_{t-1}(\mathbf{s}) + \eta_t(\mathbf{s})$, this process is a special case of the errors in (2) with coefficients $\psi_j = \phi^j$ in the MA(∞) process. Straightforward calculations show that the temporal covariance function from an AR(1) process is $\kappa(h) = \frac{\phi^h}{1-\phi^2}$ for $h > 0$, see Mikosch et al. (1995) for other MA(∞) representations on ARMA models.

Example 2 Other example of the regression model (1) is the stationary fractional noise (FN) errors (2) with infinite moving average coefficients $\psi_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}$, where $\Gamma(\cdot)$ is the Gamma function and d is the long-memory coefficient. The expected value is given by $\mathbb{E}(Y_t(\mathbf{s})) = \mathbf{M}_t(\mathbf{s})\boldsymbol{\beta}$, while the temporal covariance function is $\kappa(h) = \sigma^2 \frac{\Gamma(1-2d)\Gamma(h+d)}{\Gamma(1-d)\Gamma(d)\Gamma(h+1-d)}$, for $h > 0$, with $d \in (0, 1/2)$. A natural extension of the FN model is the stationary autoregressive fractionally integrated moving average ARFIMA(p, d, q) process, defined by $\Phi(B)\varepsilon_t(\mathbf{s}) = \Theta(B)(1-B)^{-d}\eta_t(\mathbf{s})$, for $t = 1, 2, \dots, T$, where B is the backward shift operator, $\Phi(B) = 1 + \phi_1 B + \dots + \phi_p B^p$ is an autoregressive polynomial, and $\Theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ is a moving average polynomial. The infinite moving average coefficients ψ_j satisfy the following relation

$$\psi_0 = 1, \quad \psi_j = \pi_j(d) + \sum_{i=1}^{\min(j,p)} \phi_i \psi_{j-i} + \sum_{i=1}^{\min(j,q)} \theta_i \pi_{j-i}(d), \quad j \geq 1,$$

where the weights $\pi_j(d)$ are given by $\pi_j(d) = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}$, for $0 < d < 1/2$; see Kokoszka and Taqqu (1995) for more details.

For the spatial covariance structure defined in (3), we consider the general class of Matérn covariance models (Matérn 1986) given by

$$C^\eta(\boldsymbol{\xi}; \boldsymbol{\theta}) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (2\sqrt{\nu}\rho\xi)^\nu K_\nu(2\sqrt{\nu}\rho\xi), \quad \xi \geq 0, \quad \boldsymbol{\theta} = (\sigma^2, \rho, \nu)^\top, \quad (4)$$

where $\rho > 0$, $\nu \geq 0$, $\sigma^2 > 0$ and K_ν is the modified Bessel function of the second kind of order ν . Known special cases will be shown in detail in Sect. 4.

3 State-space representation

Before starting with the state-space (SS throughout) representation defined through Eq. (1), a general version of the SS system for spatio-temporal processes is given by

$$\begin{aligned}
 Y_t(\mathbf{s}) &= [G_t(\mathbf{s}) \quad \mathbf{M}_t(\mathbf{s})] \begin{bmatrix} X_t(\mathbf{s}) \\ \boldsymbol{\beta}_t(\mathbf{s}) \end{bmatrix} + W_t(\mathbf{s}), \\
 \begin{bmatrix} X_{t+1}(\mathbf{s}) \\ \boldsymbol{\beta}_{t+1}(\mathbf{s}) \end{bmatrix} &= \begin{bmatrix} F_t(\mathbf{s}) & \mathbf{0} \\ \mathbf{0} & I_p \end{bmatrix} \begin{bmatrix} X_t(\mathbf{s}) \\ \boldsymbol{\beta}_t(\mathbf{s}) \end{bmatrix} + \begin{bmatrix} H \\ \mathbf{0} \end{bmatrix} V_t(\mathbf{s}), \tag{5}
 \end{aligned}$$

where $Y_t(\cdot)$ is the observation for time t at location $\mathbf{s} \in D$, $G_t(\mathbf{s})$ is an observation operator, $\mathbf{M}_t(\mathbf{s})$ is a vector of exogenous or predetermined variables, $[X_t(\mathbf{s}) \quad \boldsymbol{\beta}_t(\mathbf{s})]^\top$ is a state vector, and $W_t(\mathbf{s})$ is an observation noise with variance R_W . In addition, $F_t(\cdot)$ is a state transition operator, I_r denotes the $r \times r$ identity matrix hereafter, H is a linear operator, $V_t(\mathbf{s})$ is a spatially colored, temporally white and Gaussian with mean zero and covariance function $\text{cov}(V_t(\mathbf{s}), V_t(\mathbf{s}')) = C^V(\xi; \boldsymbol{\theta})$, and $V_t(\mathbf{s})$ is uncorrelated with $W_t(\mathbf{s})$, i.e., $\mathbb{E}(W_t(\mathbf{s})V_{t'}(\mathbf{s}')) = 0$ for all $\mathbf{s} \in D$ and for all $t, t' \in \mathbb{Z}_+$.

The process (1) can be represented by a SS system as above by generalizing the infinite-dimensional equations given by Hannan and Deistler (1988) and Ferreira et al. (2013) to the spatio-temporal case. This can be achieved by assignation of the $(j + 1)$ component of the state vector as the lag in j steps of the process $\{\eta_t(\mathbf{s})\}$ defined in the model error $\{\varepsilon_t(\mathbf{s})\}$, i.e., $X_t^{(j+1)}(\mathbf{s}) = \eta_{t-j}(\mathbf{s})$ for $j = 0, 1, \dots$. In such a case, the process specified by (1) can be represented by the following infinite-dimensional state-space system

$$\begin{aligned}
 Y_t(\mathbf{s}) &= [G \quad \mathbf{M}_t(\mathbf{s})] \begin{bmatrix} X_t(\mathbf{s}) \\ \boldsymbol{\beta}(\mathbf{s}) \end{bmatrix}, \\
 \begin{bmatrix} X_{t+1}(\mathbf{s}) \\ \boldsymbol{\beta}(\mathbf{s}) \end{bmatrix} &= \begin{bmatrix} F_t & \mathbf{0} \\ \mathbf{0} & I_p \end{bmatrix} \begin{bmatrix} X_t(\mathbf{s}) \\ \boldsymbol{\beta}(\mathbf{s}) \end{bmatrix} + \begin{bmatrix} [1 \quad 0 \quad 0 \quad 0 \quad \dots]^\top \\ \mathbf{0} \end{bmatrix} \eta_{t+1}(\mathbf{s}), \tag{6}
 \end{aligned}$$

for $t = 1, \dots, T$, where $G = [1 \quad \psi_1 \quad \psi_2 \quad \dots]$, $F_t = \begin{bmatrix} \mathbf{0}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ I_\infty & \mathbf{0} \end{bmatrix}$, $I_\infty = \text{diag}\{1, 1, \dots\}$, and $R_W = 0$, i.e., we assume that our observed data are measured without an additive error. From these equations, observe that $H = [1, 0, 0, \dots]^\top$ and $V_t(\mathbf{s}) = \eta_{t+1}(\mathbf{s})$. Estimation techniques related to (6) have a notorious computational burden; for this reason, we truncate the expansion in (2) after some $m \in \mathbb{Z}_+$ components, so that an approximation for $\{Y_t(\mathbf{s})\}$ can be written as

$$Y_t(\mathbf{s}) = \mathbf{M}_t(\mathbf{s})\boldsymbol{\beta} + \sum_{j=0}^m \psi_j \eta_{t-j}(\mathbf{s}), \quad \mathbf{s} \in D, \tag{7}$$

for some positive integer m . Thus, the finite-dimensional SS representation of model (7) is considered, with observation and state equations given by

$$Y_t(\mathbf{s}) = \left[[1 \quad \psi_1 \quad \psi_2 \quad \dots \quad \psi_m], \quad \mathbf{M}_t(\mathbf{s}) \right] \begin{bmatrix} X_t(\mathbf{s}) \\ \boldsymbol{\beta}(\mathbf{s}) \end{bmatrix},$$

$$\begin{bmatrix} X_{t+1}(\mathbf{s}) \\ \boldsymbol{\beta}(\mathbf{s}) \end{bmatrix} = \begin{bmatrix} - & - & F_t & - & - & \mathbf{0}^\top \\ - & - & - & - & - & - \\ \mathbf{0}_{p \times (m+1)} & - & - & - & - & I_p \end{bmatrix} \begin{bmatrix} X_t(\mathbf{s}) \\ \boldsymbol{\beta}(\mathbf{s}) \end{bmatrix} + \begin{bmatrix} [1 & 0 & 0 & \dots & 0]^\top \\ \mathbf{0}_{p \times 1} \end{bmatrix} \eta_{t+1}(\mathbf{s}), \tag{8}$$

where $F_t = \begin{bmatrix} \mathbf{0}^\top & - & - & - & - & \mathbf{0} \\ - & - & - & - & - & - \\ I_m & - & - & - & - & \mathbf{0}_{m \times 1} \end{bmatrix}$. It is worth noting that the matrices involved in the truncated Kalman equations have the following dimensions. For $\mathbb{M}_{\mathbb{R}(p \times q)}$ being the space of real-value matrices of dimension $p \times q$, we have $G \in \mathbb{M}_{\mathbb{R}(1 \times (m+1))}$, $X_t(\mathbf{s}) \in \mathbb{M}_{\mathbb{R}((m+1) \times 1)}$, $F \in \mathbb{M}_{\mathbb{R}((m+1) \times (m+1))}$ and $H \in \mathbb{M}_{\mathbb{R}((m+1) \times 1)}$. The following result establishes the asymptotic magnitude of the truncation error when approximating (1) by (7).

Proposition 1 *For given $m \in \mathbb{Z}_+$, $t \in \mathbb{Z}$ and $\mathbf{s} = (\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n)^\top$. Let $\boldsymbol{\eta}_t(\mathbf{s}) = (\eta_t(\mathbf{s}_1), \dots, \eta_t(\mathbf{s}_n))^\top$ and $r_m(\mathbf{s}) : D \rightarrow \mathbb{M}_{\mathbb{R}(n \times n)}$ be the error variance matrix, defined as $r_m(\mathbf{s}) = \text{Var}(\sum_{j=0}^\infty \psi_j \boldsymbol{\eta}_{t-j}(\mathbf{s}) - \sum_{j=0}^m \psi_j \boldsymbol{\eta}_{t-j}(\mathbf{s}))$. Then, for large m, n , $a > 0$ and $0 < d < 1/2$, we have the following:*

$$\|r_m(\mathbf{s})\| \sim \begin{cases} \mathcal{O}(nm^{2d-1}), & \text{for a long – memory process} \\ \mathcal{O}(n \exp(-am)), & \text{for a short – memory process,} \end{cases}$$

where $\|A\| = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|$ is the infinity norm of A .

Proof For an infinite moving average process $\{Y_t(\mathbf{s})\}$, the error variance matrix for the truncated series is

$$\begin{aligned} r_m(\mathbf{s}) &= \text{Var} \left(\sum_{j=0}^\infty \psi_j \boldsymbol{\eta}_{t-j}(\mathbf{s}) - \sum_{j=0}^m \psi_j \boldsymbol{\eta}_{t-j}(\mathbf{s}) \right) = \text{Var} \left(\sum_{k=m+1}^\infty \psi_k \boldsymbol{\eta}_{t-k}(\mathbf{s}) \right) \\ &= [C^\eta(\|\mathbf{s}_i - \mathbf{s}_j\|)]_{i,j=1}^n \sum_{k=m+1}^\infty \psi_k^2 = \mathbf{c}_n b_m, \end{aligned}$$

where the coefficients satisfy $\psi_m \sim \frac{\Theta(1)}{\Phi(1)} \frac{m^{d-1}}{\Gamma(d)}$, as $m \rightarrow \infty$ and where d is the long-memory parameter, see Corollary 3.1 of [Kokoszka and Taquq \(1995\)](#). In particular, if $\{Y_t(\mathbf{s})\}$ is a FN(d) process with $\psi_m = \frac{\Gamma(m+d)}{\Gamma(m+1)\Gamma(d)}$, by applying the Stirling’s approximation we have $\psi_m \sim \frac{m^{d-1}}{\Gamma(d)}$, $m \rightarrow \infty$. Furthermore, using Lemma 3.3 of [Palma \(2007\)](#) we get

$$\sum_{k=m+1}^\infty \psi_k^2 \sim m^{2d-1} \Gamma(d)^{-1} (2d-1)^{-1}, \quad 0 < d < 1/2, \tag{9}$$

so that (9) implies $b_m = \mathcal{O}(m^{2d-1})$ when $m \rightarrow \infty$. On the other hand, $\|\mathbf{c}_n\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |C^\eta(\|\mathbf{s}_i - \mathbf{s}_j\|)| \leq n\sigma^2$ where $\sigma^2 = C^\eta(0; \boldsymbol{\theta})$. Thus, $\mathbf{c}_n = \mathcal{O}(n)$ for

large n , proving the result. The proof for the case of a short-memory process is similar to the previous case, using $\psi_m \sim \exp(-am)$ when $m \rightarrow \infty$.

Some comments are in order. First, space–time asymptotics is still an open problem in the geostatistical literature. Most of the contributions are devoted to either of these two approaches: on the one hand, one might fix the number of spatial sites and let the number of temporal observations tend to infinity. This is the approach followed, for instance, by Li et al. (2008). According to this framework, evaluating the performance of our estimators amounts to apply mutandis the results in Chan and Palma (1998). Another approach might be obtained by increasing domain asymptotics (Guyon 1995). In this case, space–time asymptotics would be covered by the results in Guyon (1995). A very tempting approach might be to consider infill asymptotics in space and increasing domain in time. To the knowledge of the authors, this approach is not available in the literature and represents a major challenge.

3.1 Derivation of the Kalman filter algorithm

We now develop the Kalman filter algorithm for the spatio-temporal process defined in Eq. (7) with the associated SS representation given in (8). The Kalman filter 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 the sake of simplicity, we restrict our attention to studying the behavior of the parameter estimates of the error $\{\varepsilon_t(\cdot)\}$ process in the regression model (1), i.e.,

$$Y_t(\mathbf{s}) = \sum_{j=0}^m \psi_j \eta_{t-j}(\mathbf{s}), \quad \mathbf{s} \in D, \tag{10}$$

in such a case, the state vector is reduced down to $X_t(\mathbf{s})$. The Kalman filter recursive equations are well known, but we present them here to setup notation, and for a self-contained exposition. First, define the $n \times 1$ vector $\mathbf{Y}_t = [Y_t(\mathbf{s}_1), \dots, Y_t(\mathbf{s}_n)]^\top$ containing the data values at n spatial locations, $\{\mathbf{s}_i\}_{i=1}^n$, at time t and let $\mathbf{X}_t = [X_t(\mathbf{s}_1), \dots, X_t(\mathbf{s}_n)]^\top$ be an $n \times 1$ vector for an unobservable spatio-temporal state process for n locations, where each component of this state vector are $(m + 1)$ -dimensional vectors. Let $\widehat{X}_t(\mathbf{s}) = \mathbb{E}(X_t(\mathbf{s}) | \mathbf{X}_1, \dots, \mathbf{X}_t)$ the best linear unbiased predictor (BLUP) of the unobserved state $X_t(\mathbf{s})$ and let $\Omega_t(\mathbf{s}, \mathbf{s}') = \text{cov}(X_t(\mathbf{s}) - \widehat{X}_t(\mathbf{s}), X_t(\mathbf{s}') - \widehat{X}_t(\mathbf{s}'))$ be the state prediction error variance-covariance matrix. Finally, the initial state vector has mean $\widehat{X}_1(\mathbf{s}) = \mathbb{E}([\eta_0(\mathbf{s}), \eta_{-1}(\mathbf{s}), \dots, \eta_{1-m}(\mathbf{s})]^\top) = \mathbf{0}_{((m+1) \times 1)}$ and covariance matrix

$$\begin{aligned} \Omega_1(\mathbf{s}, \mathbf{s}') &= \text{cov}([\eta_0(\mathbf{s}), \eta_{-1}(\mathbf{s}), \dots, \eta_{1-m}(\mathbf{s})]^\top, [\eta_0(\mathbf{s}'), \eta_{-1}(\mathbf{s}'), \dots, \eta_{1-m}(\mathbf{s}')]^\top) \\ &= \text{diag}(\text{cov}(\eta_0(\mathbf{s}), \eta_0(\mathbf{s}')), \dots, \text{cov}(\eta_{1-m}(\mathbf{s}), \eta_{1-m}(\mathbf{s}'))), \end{aligned}$$

which is a $(m + 1) \times (m + 1)$ matrix. The Kalman filter allows to estimate the state vector, $X_{t+1}(\mathbf{s})$ for $\mathbf{s} \in D$ and its prediction error based on the information available at time t . These estimates are given by

$$\widehat{X}_{t+1}(\mathbf{s}) = F\widehat{X}_t(\mathbf{s}) + \Theta_t(\mathbf{s})\Delta_t^{-1}(\mathbf{Y}_t - \widehat{\mathbf{Y}}_{bft}), \tag{11a}$$

$$\Omega_{t+1}(\mathbf{s}, \mathbf{s}') = F\Omega_t(\mathbf{s}, \mathbf{s}')F^\top + C^{HV}(\mathbf{s}, \mathbf{s}') - \Theta_t^\top(\mathbf{s})\Delta_t^{-1}\Theta_t(\mathbf{s}'), \tag{11b}$$

where

$$\begin{aligned} \Delta_t &= \text{Var}(\mathbf{Y}_t - \widehat{\mathbf{Y}}_t), \\ &= \text{Var}(G(\mathbf{X}_t - \widehat{\mathbf{X}}_t)), \\ &= \begin{bmatrix} G\Omega_t(\mathbf{s}_1, \mathbf{s}_1)G^\top & \cdots & G\Omega_t(\mathbf{s}_1, \mathbf{s}_n)G^\top \\ \vdots & \ddots & \vdots \\ G\Omega_t(\mathbf{s}_1, \mathbf{s}_n)G^\top & \cdots & G\Omega_t(\mathbf{s}_n, \mathbf{s}_n)G^\top \end{bmatrix}, \\ \Theta_t^\top(\mathbf{s}) &= \text{cov}(X_{t+1}(\mathbf{s}), \mathbf{Y}_t - \widehat{\mathbf{Y}}_t), \\ &= \begin{bmatrix} \text{cov}(X_{t+1}(\mathbf{s}), Y_t(\mathbf{s}_1) - \widehat{Y}_t(\mathbf{s}_1)) \\ \vdots \\ \text{cov}(X_{t+1}(\mathbf{s}), Y_t(\mathbf{s}_n) - \widehat{Y}_t(\mathbf{s}_n)) \end{bmatrix}^\top = \begin{bmatrix} F\Omega_t(\mathbf{s}, \mathbf{s}_1)G^\top \\ \vdots \\ F\Omega_t(\mathbf{s}, \mathbf{s}_n)G^\top \end{bmatrix}^\top, \\ \widehat{Y}_t(\mathbf{s}) &= \mathbb{E}(Y_t(\mathbf{s})|\mathbf{Y}_1, \dots, \mathbf{Y}_t) = \mathbf{G}\widehat{\mathbf{X}}_t(\mathbf{s}), \\ C^{HV}(\mathbf{s}, \mathbf{s}') &= \text{cov}(HV_t(\mathbf{s}), HV_t(\mathbf{s}')) = HC^V(\xi)H^\top = \begin{pmatrix} C^\eta(\xi) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \end{aligned}$$

where $\Theta_t^\top(\mathbf{s}) \in \mathbb{M}_{\mathbb{R}((m+1) \times n)}$ and $C^{HV}(\mathbf{s}, \mathbf{s}') \in \mathbb{M}_{\mathbb{R}((m+1) \times (m+1))}$. Let θ be a parameter vector specifying model (10), then the log-likelihood function $\mathcal{L}(\cdot)$ (up to a constant), can be obtained from (11),

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

where $\epsilon_t(\theta) = (\mathbf{Y}_t - \widehat{\mathbf{Y}}_t)$ is the innovation vector, and $\Delta_t(\theta)$ is the innovation covariance matrix at time t obtained using the parameter value θ . Hence, the approximate maximum likelihood estimates (MLE) provided by the Kalman equations (11) is given by $\widehat{\theta} = \arg \max_{\theta \in \Theta} \mathcal{L}(\theta)$, where Θ is a parameter space. Note that the Kalman equations (11) can be applied directly to the general state-space representation (5), yielding in this case an exact MLE. In order to obtain predictions of unobserved values at a location \mathbf{s}_0 , we define the best linear predictor as

$$\widehat{Y}_{T+k}(\mathbf{s}_0) = \mathbb{E}(Y_{T+k}(\mathbf{s}_0)|\mathbf{Y}_1, \dots, \mathbf{Y}_T), \quad \mathbf{s}_0 \in D, \tag{12}$$

which is the k -step-out-sample predictor based on the finite past for $k = 1, \dots, K$. These forecasts and their mean squared prediction error are obtained from the Kalman recursive equations given by (11), as follows

$$\begin{aligned} \widehat{Y}_{T+k}(\mathbf{s}_0) &= \mathbb{E}(Y_{T+k}(\mathbf{s}_0)|\mathbf{Y}_1, \dots, \mathbf{Y}_T) = G \mathbb{E}(X_{T+k}(\mathbf{s}_0)|\mathbf{Y}_1, \dots, \mathbf{Y}_T) \\ &= GF \mathbb{E}(X_{T+k-1}(\mathbf{s}_0)|\mathbf{Y}_1, \dots, \mathbf{Y}_T) \\ &\vdots \\ &= GF^{k-1} \mathbb{E}(X_{T+1}(\mathbf{s}_0)|\mathbf{Y}_1, \dots, \mathbf{Y}_T) = GF^k \widehat{X}_T(\mathbf{s}_0). \end{aligned} \tag{13}$$

Additionally, its mean squared prediction error is calculated using the following recursive relations

$$\begin{aligned} \Omega_{T+1}(\mathbf{s}_0, \mathbf{s}_0) &= F\Omega_T(\mathbf{s}_0, \mathbf{s}_0)F^\top + C^{HV}(\mathbf{s}_0, \mathbf{s}_0) \\ \Omega_{T+2}(\mathbf{s}_0, \mathbf{s}_0) &= F^2\Omega_T(\mathbf{s}_0, \mathbf{s}_0)(F^\top)^2 + FC^{HV}(\mathbf{s}_0, \mathbf{s}_0)(F)^\top + C^{HV}(\mathbf{s}_0, \mathbf{s}_0) \\ &\vdots \\ \Omega_{T+k}(\mathbf{s}_0, \mathbf{s}_0) &= F^k\Omega_T(\mathbf{s}_0, \mathbf{s}_0)(F^\top)^k + \sum_{j=0}^{k-1} F^j C^{HV}(\mathbf{s}_0, \mathbf{s}_0)(F^\top)^j. \end{aligned}$$

Furthermore, the prediction error variance $\Delta_{T+k}(\mathbf{s}_0, \mathbf{s}_0)$ satisfies

$$\begin{aligned} \Delta_{T+k}(\mathbf{s}_0, \mathbf{s}_0) &= \text{Var}(Y_{T+k}(\mathbf{s}_0) - \widehat{Y}_{T+k}(\mathbf{s}_0) | \mathbf{Y}_1, \dots, \mathbf{Y}_T) \\ &= \text{Var}(GX_{T+k}(\mathbf{s}_0) + W_{T+k}(\mathbf{s}_0) - G\widehat{X}_{T+k} | \mathbf{Y}_1, \dots, \mathbf{Y}_T) \\ &= G\Omega_{T+k}(\mathbf{s}_0, \mathbf{s}_0)G^\top \\ &= GF^k\Omega_T(\mathbf{s}_0, \mathbf{s}_0)(F^\top)^kG^\top + G\sum_{j=0}^{k-1} F^j C^{HV}(\mathbf{s}_0, \mathbf{s}_0)(F^\top)^jG^\top. \end{aligned} \tag{14}$$

3.2 Missing observations

The analysis of missing observations in time series is an issue that has been treated by several authors, see [Harvey \(1989\)](#) and [Durbin and Koopman \(2012\)](#), among others. The SS method and its associated Kalman filter algorithm provides a simple methodology for handling missing values.

In order to describe this procedure, we assume that for the set of missing observations $Y_t(\mathbf{s})$ for $t = n + 1, \dots, T - n$, the vector $\epsilon_t(\mathbf{s})$ and the matrix $\Theta_t(\mathbf{s})$ of the Kalman filter are set to zero, and the Kalman updates become

$$\begin{aligned} \widehat{X}_{t+1}(\mathbf{s}) &= FX_t(\mathbf{s}), \\ \Omega_{t+1}(\mathbf{s}, \mathbf{s}') &= F\Omega_t(\mathbf{s}, \mathbf{s}')F^\top + C^{HV}(\mathbf{s}, \mathbf{s}'), \quad \mathbf{s}, \mathbf{s}' \in D, \end{aligned}$$

for $t = n + 1, \dots, T - n$. This imputation procedure provides an alternative method to the forecasts of $Y_{T+k}(\mathbf{s})$ together with their forecast error, and this can be obtained merely by treating $Y_t(\mathbf{s})$ for $t = T + 1, \dots, T + k$ as missing observations, and continuing the Kalman filter beyond $t = T$ with $\epsilon_t(\mathbf{s}) = 0$ and $\Theta_t(\mathbf{s}) = 0$ for $t > T$. This procedure for forecasting is an elegant feature of SS methods for time series analysis.

4 Simulation studies

For the simulation studies, we used R free software (Core Team 2017) and C subroutines connected to R (Peng and de Leeuw 2002) through the interface called .C. The numerical optimization of the Gaussian log-likelihood function to obtain the QML estimates was carried out using the nlminb command of R. This method makes use of the subroutine “BFGS” corresponding to a quasi-Newton method (Broyden 1969; Fletcher 1970; Goldfarb 1970). We used nlminb because this optimizer can be implemented even when the sample size is small, obtaining convergence in the optimization process. In addition, it is less sensitive to initial values compared to other optimizers. We use Monte Carlo experiments to analyze the finite sample behavior of the Kalman filter estimator, for both short- and long-memory spatio-temporal processes, as detailed through Sect. 2. In particular, we consider two models, the first is an ARMA(1, 1), a short-memory case, and the second is an ARFIMA, a long-memory case.

4.1 Short-memory case

Consider an ARMA(1, 1) model for the errors defined by (1) with

$$\varepsilon_t(\mathbf{s}) = \phi\varepsilon_{t-1}(\mathbf{s}) + \theta\eta_{t-1}(\mathbf{s}) + \eta_t(\mathbf{s}), \quad \psi_j = (\phi - \theta)\phi^{j-1} \quad \text{for } j \geq 1, \quad (15)$$

where $\{\eta_t(\cdot)\}$ are independent over time, and follow a stationary, zero mean Gaussian spatial random process and covariance function given by Eq. (3). We consider the general class of Matérn covariance models defined by Eq. (4). In particular, we use two types of covariance functions generating two possible models:

- Model 1 $\nu = \frac{1}{2}$, corresponding to the exponential model

$$C^\eta(\xi; \boldsymbol{\theta}) = \sigma^2 \exp\{-\rho\xi\}, \quad \xi \geq 0, \quad \boldsymbol{\theta} = (\sigma^2, \rho, 1/2)^\top, \quad \text{and}$$

- Model 2 $\nu = \frac{3}{2}$, which leads to

$$C^\eta(\xi; \boldsymbol{\theta}) = \sigma^2 (1 + \rho\xi) \exp\{-\rho\xi\}, \quad \xi \geq 0, \quad \boldsymbol{\theta} = (\sigma^2, \rho, 3/2)^\top.$$

The choice of ν in C^η affects the mean square (m.s.) differentiability of the associate random field. For Model 1, the associated Gaussian random field will be a.s. continuous but no m.s. differentiable. For Model 2, it will be m.s. differentiable. We assume to observe a spatio-temporal process $\{\varepsilon_t(\mathbf{s}_i) : i = 1, \dots, n; t = 1, \dots, T\}$, on a regular, rectangular grid of $n \times n = N$ spatial locations in $[1, n]^2$, and at equidistant time points. For the data generation scheme, the process is generated recursively from (15) with initial values $\eta_1(\mathbf{s}) \sim N(0, C^\eta(0))$ and $\varepsilon_1(\mathbf{s}) = C^\eta(0)/(1 - \phi^2) + \eta_1(\mathbf{s})$. Each realization is of length $T = 100, 250$. In addition, the parameters of the covariance function are considered constant, with different values for σ^2 and ρ . Finally, we simulate each process 100 times, and for each simulation the Kalman filter estimates are evaluated by the relative bias (RelBias) and by the mean square error (MSE) defined as

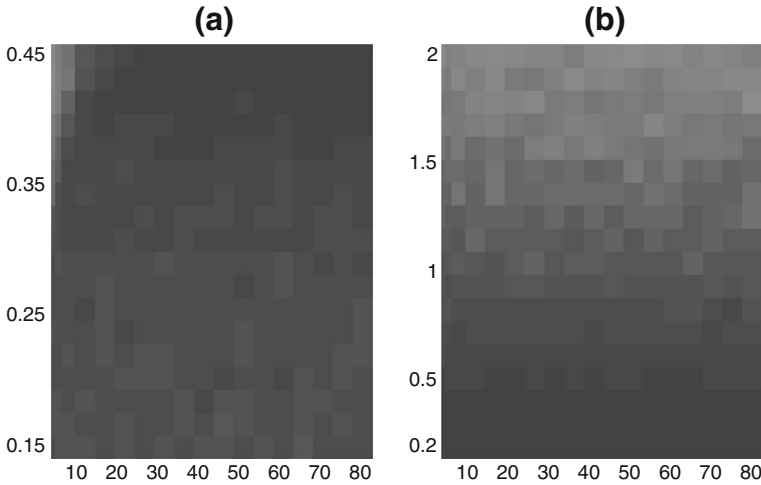


Fig. 1 MSE as a function of m and (ϕ, ρ) for the MA approximation of an AR(1) spatio-temporal process with covariance function following an exponential model. In **a** the empirical MSE of ϕ . In **b** the empirical MSE of the covariance parameter ρ

$\text{RelBias}(\theta) = \frac{1}{100} \sum_{i=1}^{100} (\hat{\theta}_i / \theta_i - 1)$ and $\text{MSE}(\theta) = \frac{1}{100} \sum_{i=1}^{100} (\hat{\theta}_i - \theta_i)^2$, where $\hat{\theta}_i$ is the Kalman filter estimate of θ_i for the i th realization. A preliminary assessment of the SS approach is related to the choice of the truncation level, which has an influence on the parameter estimates. Figure 1 plots the estimated MSE for an AR(1) model for different values of ϕ and ρ as a function of the truncation level, m , for the $\text{MA}(\infty)$ decomposition. Figure 2 displays the MSE for different values of d and ρ as a function of the truncation level, m , based on a $\text{FN}(d)$ model. In both cases, we use the Model 1 as spatial covariance function with $\sigma^2 = 1$, $T = 250$ and $N = 100$ spatial locations. In these graphs, darker regions represent the minimal empirical MSE, while lighter regions indicate greater MSE values. Note that, for the long-memory case an improvement in terms of MSE is evidenced when the truncation parameter is $m \geq 10$. Similar results are obtained in Chan and Palma (1998) with $m > 6$ when considering only the temporal domain. The short-memory case requires a lower level of truncation ($m = 5$) on the state-space representation to guarantee the efficient performance of the truncated MLE estimates. In light of this evidence and due to space constraints, only a subset ($m = 5, 10$) of the results are presented; however, other results and codes are available from the authors upon request.

Table 1 shows the estimates of the parameters for two truncation levels, $m = 5, 10$ and $N = 100$ spatial locations. We have used a combination of parameter values that reproduce widely encountered situations in practical analysis. These scenarios are shown in Table 1. Note that the estimates are very close to their theoretical counterparts. Furthermore, it is noteworthy that goodness-of-fit criteria such as the standard deviation (SD), bias and $\sqrt{\text{MSE}}$ are very similar in relation to the level of truncation. In effect, if the truncation parameter is $m = 5$, the truncated Kalman filter works extremely well for both sample sizes.

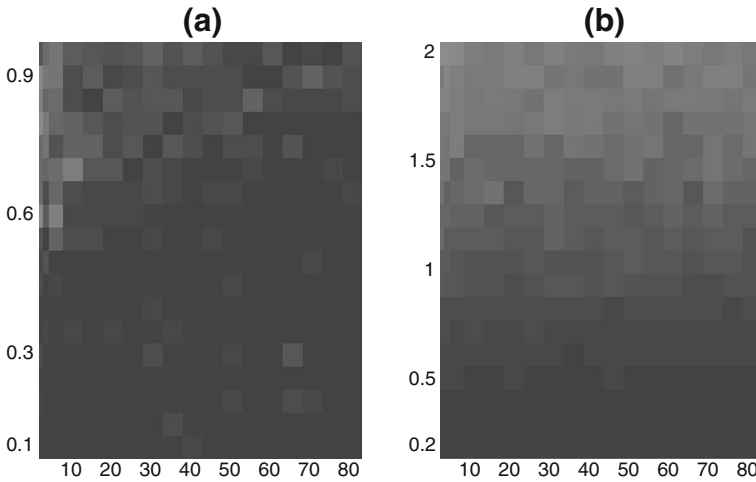


Fig. 2 MSE as a function of m and (d, ρ) for the MA approximation of a $FN(d)$ spatio-temporal process with covariance function following an exponential model . In **a** empirical MSE of the long-memory parameter d . In **b** empirical MSE of the covariance parameter ρ

4.2 Long-memory case

Consider now the following stationary ARFIMA(0, d , 1) model defined by

$$\varepsilon_t(\mathbf{s}) = (1 - \theta B)(1 - B)^{-d} \eta_t(\mathbf{s}), \quad \psi_j = \frac{\Gamma(j + d)}{\Gamma(j + 1)\Gamma(d)} + \theta \frac{\Gamma(j + d - 1)}{\Gamma(j)\Gamma(d)},$$

for $j \geq 1$, where $\Gamma(\cdot)$ is the Gamma function, d is the long-memory coefficient such that $0 < d < 1/2$ and θ is a moving average coefficient satisfying $|\theta| < 1$. Concerning the innovations $\{\eta_t(\mathbf{s})\}$, these were generated considering the same spatial structure as for the short-memory case. The samples from this ARFIMA process are generated using the innovation algorithm; see Brockwell and Davis (1991), page 172. In this implementation, the temporal covariance of the process $\{\varepsilon_t(\mathbf{s})\}$ is given by

$$\kappa_T(h) = \frac{\Gamma(1 - 2d)\Gamma(h + d)}{\Gamma(1 - d)\Gamma(d)\Gamma(h + 1 - d)} \times \left[1 + \theta^2 - \theta \frac{h - d}{h - 1 + d} - \theta \frac{h + d}{h + 1 - d} \right],$$

for $h > 0$. We offer some practical combination of parameter values commonly encountered in data set analysis with long-range dependence. Table 2 reports the results from the Monte Carlo simulations for several parameter values and two truncation levels, $m = 5, 10$. The simulations are based on sample sizes $T = 100, 250$ and 100 replications. As indicated in the previous case, the observed means for the estimates are close to their expected values. In contrast to the findings in the short-memory case, a long-memory process requires a higher level of truncation on the state-space representation to guarantee the efficient performance of the truncated MLE estimates.

Table 1 Results for the Kalman filter estimates for an ARMA(1, 1) model with observed locations on the square $[0, 10]^2$

Model	Parameter				Estimation				
	ϕ	θ	σ^2	ρ	$\hat{\phi}$	$\hat{\theta}$	$\hat{\sigma}^2$	$\hat{\rho}$	
<i>m</i> = 5					<i>T</i> = 100				
Model 1	0.45	0.3	0.5	0.35	Mean	0.4462	0.3009	0.5025	0.3465
					SD	0.0120	0.0136	0.0198	0.0168
					RelBias	-0.0085	0.0031	0.0051	-0.0101
					$\sqrt{\text{MSE}}$	0.0126	0.0135	0.0198	0.0171
	0.65	0.8	0.1	0.9	Mean	0.5722	0.7768	0.1020	0.8996
					SD	0.0047	0.0060	0.0019	0.0240
					RelBias	-0.1198	-0.0290	0.0197	-0.0004
					$\sqrt{\text{MSE}}$	0.0780	0.0240	0.0027	0.0239
Model 2	0.45	0.3	0.5	0.35	Mean	0.4459	0.3012	0.5028	0.3489
					SD	0.0118	0.0134	0.0291	0.0079
					RelBias	-0.0090	0.0039	0.0055	-0.0032
					$\sqrt{\text{MSE}}$	0.0125	0.0134	0.0297	0.0080
	0.65	0.8	0.1	0.9	Mean	0.5723	0.7762	0.1022	0.8985
					SD	0.0050	0.0061	0.0036	0.0148
					RelBias	-0.1195	-0.0298	0.0215	-0.0017
					$\sqrt{\text{MSE}}$	0.0778	0.0246	0.0042	0.0148
					<i>T</i> = 250				
Model 1	0.45	0.3	0.5	0.35	Mean	0.4476	0.3012	0.4976	0.3519
					SD	0.0165	0.0180	0.0186	0.0155
					RelBias	-0.0053	0.0041	-0.0049	0.0055
					$\sqrt{\text{MSE}}$	0.0166	0.0179	0.0187	0.0155
	0.65	0.8	0.1	0.9	Mean	0.5717	0.7739	0.1026	0.8984
					SD	0.0031	0.0035	0.0013	0.0171
					RelBias	-0.1204	-0.0327	0.0256	-0.0018
					$\sqrt{\text{MSE}}$	0.0783	0.0264	0.0029	0.0171
Model 2	0.45	0.3	0.5	0.35	Mean	0.4464	0.3033	0.5031	0.3492
					SD	0.0087	0.0095	0.0188	0.0047
					RelBias	-0.0080	0.0110	0.0061	-0.0024
					$\sqrt{\text{MSE}}$	0.0093	0.0100	0.0190	0.0048
	0.65	0.8	0.1	0.9	Mean	0.5713	0.7739	0.1025	0.8998
					SD	0.0038	0.0041	0.0023	0.0088
					RelBias	0.1210	-0.0327	0.0253	-0.0002
					$\sqrt{\text{MSE}}$	0.0787	0.0264	0.0034	0.0088

Table 1 continued

Model	Parameter				Estimation				
	ϕ	θ	σ^2	ρ	$\hat{\phi}$	$\hat{\theta}$	$\hat{\sigma}^2$	$\hat{\rho}$	
$m = 10$					$T = 100$				
Model 1	0.45	0.3	0.5	0.35	Mean	0.4484	0.2987	0.5015	0.3476
					SD	0.0074	0.0148	0.0243	0.0217
					RelBias	-0.0055	-0.0051	0.0017	-0.0115
					$\sqrt{\text{MSE}}$	0.0123	0.0132	0.0152	0.0073
	0.65	0.8	0.1	0.9	Mean	0.6415	0.8014	0.0993	0.8991
					SD	0.0024	0.0041	0.0021	0.0213
					RelBias	-0.0146	0.0034	-0.011	-0.0010
					$\sqrt{\text{MSE}}$	0.0043	0.0131	0.0039	0.0065
Model 2	0.45	0.3	0.5	0.35	Mean	0.4501	0.3017	0.5031	0.3491
					SD	0.0133	0.0084	0.0123	0.0063
					RelBias	0.0008	0.0091	0.0062	-0.0045
					$\sqrt{\text{MSE}}$	0.0101	0.0204	0.0037	0.0103
	0.65	0.8	0.1	0.9	Mean	0.6465	0.8024	0.0993	0.9015
					SD	0.0102	0.0132	0.0055	0.0242
					RelBias	-0.0140	0.0016	-0.0068	0.0007
					$\sqrt{\text{MSE}}$	0.0156	0.0134	0.0053	0.0086
$T = 250$									
Model 1	0.45	0.3	0.5	0.35	Mean	0.4504	0.3075	0.4984	0.3495
					SD	0.0064	0.0232	0.0128	0.0121
					RelBias	0.0008	0.0159	-0.0068	-0.0045
					$\sqrt{\text{MSE}}$	0.0212	0.0149	0.0130	0.0084
	0.65	0.8	0.1	0.9	Mean	0.6477	0.8037	0.0998	0.9016
					SD	0.0027	0.0053	0.0018	0.0327
					RelBias	-0.0139	0.0031	-0.0015	0.0018
					$\sqrt{\text{MSE}}$	0.0046	0.0089	0.0026	0.0055
Model 2	0.45	0.3	0.5	0.35	Mean	0.4465	0.3032	0.5010	0.3498
					SD	0.0120	0.0181	0.0189	0.0080
					RelBias	-0.0009	0.0088	0.0020	-0.0005
					$\sqrt{\text{MSE}}$	0.0095	0.0173	0.0028	0.0080
	0.65	0.8	0.1	0.9	Mean	0.6415	0.8011	0.0997	0.9007
					SD	0.0085	0.0073	0.0038	0.0169
					RelBias	-0.0131	0.0013	-0.0035	0.0008
					$\sqrt{\text{MSE}}$	0.0120	0.0074	0.0038	0.0073

Table 2 Results for the Kalman filter estimates for an ARFIMA(0, d , 1) model with observed locations on the square $[0, 10]^2$

Model	Parameter				Estimation				
	d	θ	σ^2	ρ	\hat{d}	$\hat{\theta}$	$\hat{\sigma}^2$	$\hat{\rho}$	
$m = 5$					$T = 100$				
Model 1	0.15	0.45	1.5	0.55	Mean	0.1888	0.4840	1.5112	0.5483
					SD	0.0524	0.0535	0.0400	0.0191
					RelBias	0.2585	0.0756	0.0074	-0.0031
					$\sqrt{\text{MSE}}$	0.0648	0.0630	0.0413	0.0190
	0.35	0.2	0.8	0.75	Mean	0.3920	0.1870	0.8564	0.7494
					SD	0.0248	0.0342	0.0240	0.0288
					RelBias	0.1199	-0.0650	0.0705	-0.0008
					$\sqrt{\text{MSE}}$	0.0487	0.0364	0.0612	0.0287
Model 2	0.15	0.45	1.5	0.55	Mean	0.1828	0.4911	1.4476	0.5812
					SD	0.0862	0.0675	0.2414	0.1112
					RelBias	0.2190	0.0913	-0.0349	0.0568
					$\sqrt{\text{MSE}}$	0.0919	0.0787	0.2458	0.1150
	0.35	0.2	0.8	0.75	Mean	0.3897	0.1902	0.8598	0.7482
					SD	0.0177	0.0205	0.0367	0.0147
					RelBias	0.1134	-0.0489	0.0748	-0.0024
					$\sqrt{\text{MSE}}$	0.0434	0.0226	0.0701	0.0148
					$T = 250$				
Model 1	0.15	0.45	1.5	0.55	Mean	0.1810	0.4763	1.5072	0.5495
					SD	0.0269	0.0273	0.0270	0.0129
					RelBias	0.2069	0.0584	0.0048	-0.0009
					$\sqrt{\text{MSE}}$	0.0409	0.0378	0.0278	0.0129
	0.35	0.2	0.8	0.75	Mean	0.3863	0.1852	0.8555	0.7499
					SD	0.0126	0.0119	0.0144	0.0167
					RelBias	0.1037	-0.0742	0.0694	-0.0001
					$\sqrt{\text{MSE}}$	0.0384	0.0190	0.0573	0.0167
Model 2	0.15	0.45	1.5	0.55	Mean	0.2096	0.5065	1.5100	0.5499
					SD	0.0903	0.0924	0.0753	0.0113
					RelBias	0.3975	0.1255	0.0067	-0.0001
					$\sqrt{\text{MSE}}$	0.1079	0.1079	0.0756	0.0113
	0.35	0.2	0.8	0.75	Mean	0.3834	0.1862	0.8551	0.7499
					SD	0.0260	0.0237	0.0406	0.0164
					RelBias	0.0955	-0.0692	0.0689	-0.00013
					$\sqrt{\text{MSE}}$	0.0422	0.0273	0.0683	0.0163

Table 2 continued

Model	Parameter				Estimation				
	d	θ	σ^2	ρ	\hat{d}	$\hat{\theta}$	$\hat{\sigma}^2$	$\hat{\rho}$	
$m = 10$					$T = 100$				
Model 1	0.15	0.45	1.5	0.55	Mean	0.1594	0.4563	1.5096	0.5475
					SD	0.0252	0.0276	0.0446	0.0204
					RelBias	0.0624	0.0140	0.0064	-0.0045
					$\sqrt{\text{MSE}}$	0.0268	0.0281	0.0454	0.0205
	0.35	0.2	0.8	0.75	Mean	0.3829	0.1996	0.8327	0.7469
					SD	0.0318	0.0382	0.0323	0.0388
					RelBias	0.0941	-0.0021	0.0409	-0.0042
					$\sqrt{\text{MSE}}$	0.0457	0.0381	0.0459	0.0387
Model 2	0.15	0.45	1.5	0.55	Mean	0.1671	0.4599	1.5038	0.5501
					SD	0.0616	0.0705	0.1065	0.0175
					RelBias	0.1139	0.0222	0.0025	0.0002
					$\sqrt{\text{MSE}}$	0.0636	0.0708	0.1060	0.0174
	0.35	0.2	0.8	0.75	Mean	0.3896	0.2083	0.8333	0.7492
					SD	0.0176	0.0188	0.0349	0.0138
					RelBias	0.1131	0.0413	0.0416	-0.0010
					$\sqrt{\text{MSE}}$	0.0433	0.0205	0.0481	0.0138
					$T = 250$				
Model 1	0.15	0.45	1.5	0.55	Mean	0.1584	0.4574	1.5057	0.5500
					SD	0.0331	0.0371	0.0451	0.0206
					RelBias	0.0561	0.0165	0.0038	0.0000
					$\sqrt{\text{MSE}}$	0.0340	0.0377	0.0453	0.0205
	0.35	0.2	0.8	0.75	Mean	0.3837	0.2027	0.8280	0.7515
					SD	0.0226	0.0236	0.0255	0.0322
					RelBias	0.0964	0.0135	0.0350	0.0019
					$\sqrt{\text{MSE}}$	0.0405	0.0237	0.0378	0.0320
Model 2	0.15	0.45	1.5	0.55	Mean	0.1587	0.4544	1.5011	0.5505
					SD	0.0345	0.0377	0.0634	0.0097
					RelBias	0.0577	0.0098	0.0007	0.0010
					$\sqrt{\text{MSE}}$	0.0354	0.0378	0.0631	0.0097
	0.35	0.2	0.8	0.75	Mean	0.3852	0.2022	0.8287	0.7519
					SD	0.0232	0.0250	0.0366	0.0150
					RelBias	0.1005	0.0111	0.0359	0.0026
					$\sqrt{\text{MSE}}$	0.0421	0.0250	0.0464	0.0151

Another important aspect that should be assessed is the out-of-sample predictive ability of our methodology. To this end, we simulated a dataset under ARMA(1, 1) and ARFIMA(0, d , 1) models, and for both cases we used the Model 1 as spatial covariance structure. For each simulated dataset, we fitted three models, giving rise to three cases of comparison, namely AR(1), ARMA(1, 1) and ARFIMA(0, d , 1). We used cross-validation techniques to compare these three cases. Cross-validation is implemented on the observed data with K time points deleted for each location \mathbf{s}_i , i.e., $\{Y_t(\mathbf{s}_i) : t = 1, \dots, T - K; i = 1, \dots, n\}$, and then predicting $\{\widehat{Y}_{t_k}(\mathbf{s}_i)\}_{k=1}^K$ from the remaining data. We used the cross-validation statistics suggested by Carroll and Cressie (1997), given by

$$CR_1(\mathbf{s}_i) = \frac{\sum_{k=1}^K [Y_{t_k}(\mathbf{s}_i) - \widehat{Y}_{t_k}(\mathbf{s}_i)]}{\left[\sum_{k=1}^K \Delta_{t_k}(\mathbf{s}_i)\right]^{1/2}}, \quad CR_2(\mathbf{s}_i) = \left\{ \frac{\sum_{k=1}^K [Y_{t_k}(\mathbf{s}_i) - \widehat{Y}_{t_k}(\mathbf{s}_i)]^2}{\sum_{k=1}^K \Delta_{t_k}(\mathbf{s}_i)} \right\}^{1/2},$$

$$CR_3(\mathbf{s}_i) = \left\{ \sum_{k=1}^K \frac{[Y_{t_k}(\mathbf{s}_i) - \widehat{Y}_{t_k}(\mathbf{s}_i)]^2}{T} \right\}^{1/2},$$

where $\widehat{Y}_{t_k}(\mathbf{s}_i)$ is the prediction of the process at location \mathbf{s}_i and t_k ($k = 1, \dots, K$), and $\Delta_{t_k}(\mathbf{s}_i)$ is the corresponding prediction variance. These values are obtained from (13) and (14), respectively. CR_1 indicates the unbiasedness of the predictor and should be approximately equal to zero. CR_2 checks the accuracy of the standard deviation of the prediction error and should be approximately equal to one. CR_3 is a measure of goodness of prediction. One would like CR_3 to be small, which indicates that the predicted values are close to the true values. For the data generation scheme, ARMA(1, 1) is generated with parameters $(\phi, \theta, \sigma^2, \rho) = (0.45, 0.3, 0.5, 0.35)$ whereas for the ARFIMA(0, d , 1) we used $(d, \theta, \sigma^2, \rho) = (0.15, 0.45, 1.5, 0.55)$ with sample sizes of $T = 250$, $N = 25$ locations and truncation level, $m = 10$. Table 3 displays the average on the locations of the cross-validation statistics based on predictions for 5 days, i.e., $K = 5$. It can be noticed that all these measures favor our true simulated data.

Table 3 Cross-validation statistics

Model	Case	Criterion		
		Ave(CR_1)	Ave(CR_2)	Ave(CR_3)
ARMA(1, 1)	1	-0.4734	0.8353	0.2704
	2	-0.3472	0.8771	0.2432
	3	-0.6397	0.7317	0.3206
ARFIMA(0, d , 1)	1	-0.3243	0.7915	0.3645
	2	-0.2905	0.7874	0.3663
	3	-0.1706	0.9915	0.2292

Case 1: AR(1), case 2 : ARMA(1, 1), case 3: ARFIMA(0, d , 1)

Table 4 Results for the Kalman filter estimates for an AR(1) with observed locations on the square $[0, 10]^2$

% NA	$T = 15$			$T = 50$		
	0%	10%	20%	0%	10%	20%
$\widehat{\phi}$	0.6390	0.6315	0.5994	0.6414	0.6354	0.6057
$\widehat{\rho}$	0.5177	0.5181	0.5075	0.4996	0.4995	0.4989
$\widehat{\sigma}^2$	1.0089	0.8810	0.8120	1.0056	0.9122	0.8207
$\sqrt{\text{MSE}(\widehat{\phi})}$	0.0244	0.0543	0.0659	0.0183	0.0456	0.0849
$\sqrt{\text{MSE}(\widehat{\rho})}$	0.0994	0.0708	0.0999	0.0012	0.0022	0.0025
$\sqrt{\text{MSE}(\widehat{\sigma}^2)}$	0.1010	0.1492	0.2097	0.0349	0.0935	0.1825

Case of a small number of observations and missing data

4.3 Estimation with a small number of observations and missing data

The performance of the Kalman filter estimator when the size of the time series is small and has missing observations is assessed in this section. For this case, we use an AR(1) for the temporal structure, and for the spatial dependence we use a third model defined as

- Model 3: $\nu = \infty$, corresponding to the Gaussian model

$$C^\eta(\xi; \theta) = \sigma^2 \exp\{-\rho^2 \xi^2\}, \quad \xi \geq 0, \quad \theta = (\sigma^2, \rho, \infty)^\top.$$

We consider a regular grid on the square $[0, 10]^2$ and level of truncation $m = 5$. In addition, we incorporate 10% and 20% of missing values, which have been randomly selected for each simulation over 10 locations specified in the cartesian plane with coordinates $(i, 8)$ for $i = 1, \dots, 10$. For the AR(1) case, we consider $\phi = 0.65$, variance scale $\sigma^2 = 1$ and spatial correlation $\rho = 0.5$. Finally, we work with sample sizes $T = 15, 50$, and 100 replications.

Table 4 shows the estimates of the parameters and the mean square errors. Note that these estimates are very close to their theoretical counterparts. As expected, the precision of the estimates worsens as the percentage of missing data increases.

5 Real data applications

5.1 TOMS data

The Kalman filter algorithm presented in Sect. 3.1 is now applied to Level 3 Total Ozone Mapping Spectrometer (TOMS) data. TOMS Level-3 data have been analyzed in some recent papers, including Jun and Stein (2008) and Porcu et al. (2015). We refer to these papers for a detailed description of the data. It is worth pointing out that TOMS data are located on a spatially regular grid of 1° latitude by 1.25° longitude away from the poles, i.e., from a latitude interval $[-89.5, 89.5]$ to a longitude interval $[-180, 180]$. We focus our analysis on selected 140 spatial points with all temporal

Table 5 Parameter estimates for the TOMS data

Model	Parameter				
	ϕ_1	ϕ_2	θ	σ^2	ρ
Model 1					
AR(1)	0.6145			48.3142	0.3156
AR(2)	0.6257	0.0021		46.6578	0.2738
ARMA(1, 1)	0.6347		0.0012	45.6745	0.8934
Model 2					
AR(1)	0.6245			46.8182	0.4209
AR(2)	0.6198	0.0032		45.8765	0.6367
ARMA(1, 1)	0.6265		0.0011	44.7924	0.9356
Model 3					
AR(1)	0.6246			47.1875	0.6235
AR(2)	0.6285	-0.0011		46.1764	0.5678
ARMA(1, 1)	0.6269		0.0017	45.8182	0.6899

observations for a total of 2100 observations, both in space and time. In addition, we converted coordinates from longitude/latitude to universal transverse mercator (utm) coordinates. The projections are obtained using `spTransform` from the `rgdal` package (Bivand et al. 2015) which uses the PROJ.4 projection library to perform the calculations. The temporal covariance was analyzed by considering three structures, namely AR(1), AR(2) and ARMA(1,1) models. For the spatial covariance, we considered Model 1, 2 and 3 as described in the previous Section. Table 5 reports the parameter estimates using the Kalman filter with truncation level $m = 5$. From Table 5, we note that the difference between the estimation performance is not so apparent. On the other hand, parameters ϕ_2 and θ provide relatively little information compared to the other parameters, indicating that the temporal model has a potential correlation structure of AR(1) type. We computed these statistics for predicting the last day of measurement, i.e., $K = 1$. The averages of the cross-validation statistics are presented in Table 6. We note no significant difference between all these measures. Nevertheless, we can conclude that for Model 1 and AR(1) model, the CR_3 exhibits a smaller value than the other cases. Figure 3 displays the performance of this model in terms of the marginal spatial semivariograms and marginal sample autocorrelation function (ACF). As shown in this figure, the Kalman filter estimation offers a better fit for the modeling of TOMS data.

5.2 Irish wind data

Wind energy has grown significantly in developed countries and supplying electricity depends on methods of predicting wind speed in certain locations. The Irish wind speed data have been studied by several authors, in particular, Haslett and Raftery (1989), Gneiting (2002), Stein (2005) and Bevilacqua et al. (2012). Following Haslett

Table 6 Cross-validation statistics for the TOMS data

Model	Ave(CR ₁)	Ave(CR ₂)	Ave(CR ₃)
Model 1			
AR(1)	-0.0128	1.1104	3.6452
AR(2)	-0.0178	1.1297	3.7494
ARMA(1, 1)	-0.0181	1.1284	3.7523
Model 2			
AR(1)	-0.0187	1.1121	3.7142
AR(2)	-0.0177	1.1210	3.7525
ARMA(1, 1)	-0.0179	1.1212	3.7519
Model 3			
AR(1)	-0.0137	1.1107	3.6875
AR(2)	-0.0167	1.1210	3.7521
ARMA(1, 1)	-0.0177	1.1219	3.7623

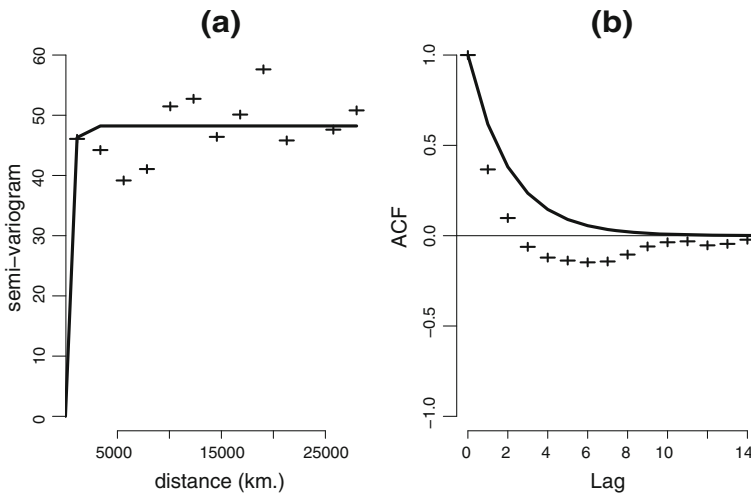


Fig. 3 TOMS data: **a** marginal spatial semivariogram versus estimated spatial semivariogram **b** marginal sample ACF versus estimated temporal AR(1) model

and Raftery (1989), we omitted the Rosslare station and then considered a square root transformation to have deseasonalized data. The seasonal component was estimated by calculating the average of the square roots of the daily means over all years and stations for each day of the year, and regressing the result on a set of annual harmonics. We refer to their paper for a detailed description of these data. Following Haslett and Raftery (1989), we use a long-memory process to model the temporal dependence, and an exponential model for the spatial covariance, defined as

$$C^\eta(\xi; \theta) = \begin{cases} \sigma^2 \exp\{-\rho\xi\} & \text{if } \xi \neq 0 \\ 1 & \text{if } \xi = 0, \end{cases}$$

Table 7 Cross-validation statistics for the Irish wind data

Model	Ave(CR ₁)	Ave(CR ₂)	Ave(CR ₃)
AR(1)	0.3567	0.5678	0.9235
ARMA(1, 1)	0.4572	0.4367	1.0456
FN(<i>d</i>)	0.2769	0.8269	0.6908
ARFIMA(1, <i>d</i> , 0)	0.2631	0.8179	0.6798
ARFIMA(2, <i>d</i> , 0)	0.2853	0.8254	0.6917

where $\theta = (\sigma^2, \rho)^\top$, with $\sigma^2 \in (0, 1]$ and $\rho > 0$. In this direction, we have proposed three different models for the temporal dependence, namely FN(*d*), ARFIMA(1, *d*, 0) and ARFIMA(2, *d*, 0). In order to evaluate a possible structure of short-memory on the temporal covariance, ARMA(1, 1) and AR(1) models are also proposed. In addition, we considered all data except the last week, which will be used as a validation set. In order to obtain the Kalman filter estimation, we used a truncation level $m = 10$. The parameters estimated for these cases are given as follows:

- For the AR(1) model, $\hat{\phi} = 0.8564$, $\hat{\sigma}^2 = 0.7654$ and $\hat{\rho} = 0.00438$, whereas for the ARMA(1, 1), $\hat{\phi} = 0.6523$, $\hat{\theta} = 0.3812$, $\hat{\sigma}^2 = 0.5841$ and $\hat{\rho} = 0.00817$.
- For the FN(*d*) model, the estimates are: $\hat{d} = 0.3373$, $\hat{\sigma}^2 = 0.98799$ and $\hat{\rho} = 0.00137$.
- For the ARFIMA(1, *d*, 0) model, we have that $\hat{d} = 0.3137$, $\hat{\phi} = 0.04386$, $\hat{\sigma}^2 = 0.98878$ and $\hat{\rho} = 0.00147$.
- For the ARFIMA(2, *d*, 0) model, $\hat{d} = 0.3251$, $\hat{\phi}_1 = 0.0101$, $\hat{\phi}_2 = -0.0599$, $\hat{\sigma}^2 = 0.99786$ and $\hat{\rho} = 0.00164$.

To choose the best model, we considered the previous defined cross-validation statistics. In this case, we perform predictions for 7 days, i.e., $K = 7$. From Table 7, we can see that for the ARFIMA(1, *d*, 0) model, the CR₃ exhibits a smaller value than the other cases. From these results, we focus on the goodness-of-fit analysis of the long-memory models. Figure 4 exhibits two panels exploring the correlations of the stations and the marginal sample ACF. Note that from panel (b) all the marginal sample ACF decay slowly, confirming a long-memory behavior. The dashed line represents the FN(*d*) model, the continuous line corresponds to the ARFIMA(1, *d*, 0) model, while the dotted line represents the behavior of the ARFIMA(2, *d*, 0) case. It seems that the ARFIMA(1, *d*, 0) model offers a better fit to the temporal sample ACF, whereas the behavior of the spatial correlations are very similar.

6 Discussion

In this article, we have proposed a state-space methodology to model spatio-temporal processes. In particular, we have proposed to model the temporal dependence structure both short- and/or long-memory through the infinite moving average representation MA(∞). In this context, we have incorporated the ARFIMA models to quantify the temporal correlation and Matérn covariance models to characterize the spatial correlation in the spatio-temporal processes. In terms of the estimation procedure, we have

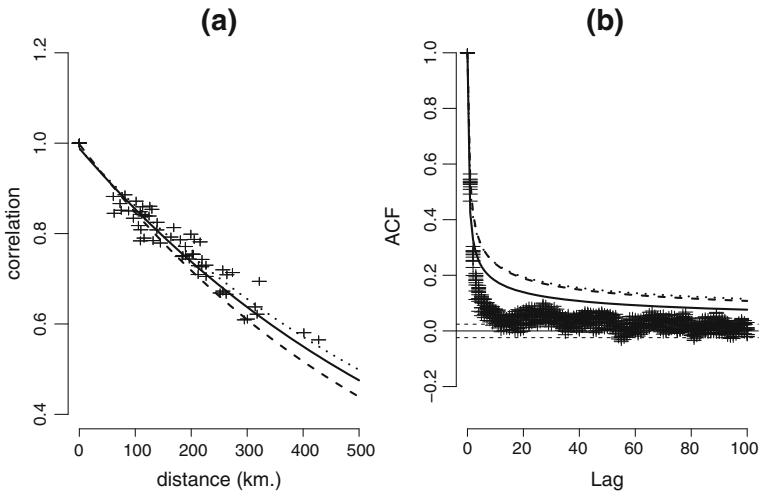


Fig. 4 Irish wind data: **a** distance correlation plot and **b** sample ACF. FN(d) model (*dashed line*), ARFIMA($1, d, 0$) (*continuous line*), ARFIMA($2, d, 0$) (*dotted line*)

proposed an approximation to the likelihood functions via truncation which provides an efficient means to calculate the MLE. Simulation studies evidenced that the proposed 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 addition, we used the Kalman filter algorithm to obtain the k -step ahead prediction and handle missing values without any additional assumption or additional procedure of imputation to fill in the missing values. These features provide clear advantages over alternatives procedures that deal with spatio-temporal models.

An interesting direction for future research is to use state-space models in the $MA(\infty)$ expansion to incorporate non-stationarity, by introducing time-varying models and/or location-dependence processes on the observation operator $G_t(\mathbf{s})$. Although this would imply a significant increase of parameters to be estimated, it would require only minor changes in the algorithms presented here. Rao (2008) proposed a local least squares method to estimate the parameters of a spatio-temporal model with location-dependent parameters which are used to describe spatial non-stationarity, and we have recently begun to work combining these two ideas.

Acknowledgements The first author would like to express his thanks for the support from DIUC 215.014.024-1.0, established by the Universidad de Concepción and Postdoctoral scholarship from Conicyt, Chile, 2014 (Folio 74150023). Jorge Mateu's research was supported by Grant MTM2013-43917-P from the Spanish Ministry of Science and Education, and Grant P1-1B2015-40 and Emilio Porcu's research was supported by Fondecyt Regular Project from Ministry of Science and Education, Chile.

References

Bevilacqua M, Gaetan C, Mateu J, Porcu E (2012) Estimating space and space-time covariance functions for large data sets: a weighted composite likelihood approach. *J Am Stat Assoc* 107(497):268–280

- Bilonick RA (1985) The space–time distribution of sulfate deposition in the northeastern united states. *Atmos Environ* (1967) 19(11):1829–1845
- Bivand R, Keitt T, Rowlingson B (2015) rgdal: bindings for the geospatial data abstraction library. R Foundation for Statistical Computing, Vienna
- Bocquet M, Elbern H, Eskes H, Hirtl M, Zabkar R, Carmichael G, Flemming J, Inness A, Pagowski M, Pérez Camañó J et al (2015) Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models. *Atmos Chem Phys* 15(10):5325–5358
- Brockwell PJ, Davis RA (1991) Time series: theory and methods, 2nd edn. Springer, New York
- Brown PJ, Le ND, Zidek JV (1994) Multivariate spatial interpolation and exposure to air pollutants. *Can J Stat* 22(4):489–509
- Broyden CG (1969) A new double-rank minimization algorithm. *Not Am Math Soc* 16:670
- Cameletti M, Lindgren F, Simpson D, Rue H (2013) Spatio-temporal modeling of particulate matter concentration through the spde approach. *ASTA Adv Stat Anal* 97(2):109–131
- Carroll SS, Cressie N (1997) Spatial modeling of snow water equivalent using covariances estimated from spatial and geomorphic attributes. *J Hydrol (Amst)* 190(1):42–59
- Chan NH, Palma W (1998) State space modeling of long-memory processes. *Ann Stat* 26(2):719–740
- Cressie N, Wikle CK (2011) Statistics for spatio-temporal data. Wiley, New York
- Daley D, Porcu E (2014) Dimension walks and schoenberg spectral measures. *P Am Math Soc* 142(5):1813–1824
- Durbin J, Koopman SJ (2012) Time series analysis by state space methods. Number 38 in Oxford statistical science series. Oxford University Press, Oxford
- Eynon B, Switzer P (1983) The variability of rainfall acidity. *Can J Stat* 11:11–24
- Fasso A, Cameletti M, Nicolis O (2007) Air quality monitoring using heterogeneous networks. *Environmetrics* 18(3):245–264
- Ferreira G, Rodríguez A, Lagos B (2013) Kalman filter estimation for a regression model with locally stationary errors. *Comput Stat Data Anal* 62:52–69
- Fletcher R (1970) A new approach to variable metric methods. *Comput J* 13:317–322
- Gneiting T (2002) Nonseparable, stationary covariance functions for space–time data. *J Am Stat Assoc* 97(458):590–600
- Goldfarb D (1970) A family of variable metric methods derived by variational means. *Math Comput* 24:23–26
- Guyon X (1995) Random fields on a network: modeling, statistics, and applications. Springer, New York
- Handcock MS, Wallis JR (1994) An approach to statistical spatial-temporal modeling of meteorological fields. *J Am Stat Assoc* 89(426):368–378
- Hannan EJ, Deistler M (1988) The statistical theory of linear systems. Wiley, New York
- Harvey AC (1989) Forecasting structural time series and the Kalman filter. Cambridge University Press, Cambridge
- Haslett J, Raftery AE (1989) Space–time modelling with long-memory dependence: assessing Ireland’s wind power resource. *J R Stat Soc C Appl* 38(1):1–50
- Huang H-C, Cressie N (1996) Spatio-temporal prediction of snow water equivalent using the Kalman filter. *Comput Stat Data Anal* 22(2):159–175
- Hughes JP, Guttorp P, Charles SP (1999) A non-homogeneous hidden markov model for precipitation occurrence. *J R Stat Soc C Appl* 48(1):15–30
- Ippoliti L (2001) On-line spatio-temporal prediction by a state space representation of the generalized space time autoregressive model. *Metron Int J Stat* LIX:157–169
- Jun M, Stein ML (2008) Nonstationary covariance models for global data. *Ann Appl Stat* 2(4):1271–1289
- Kokoszka PS, Taqqu MS (1995) Fractional arima with stable innovations. *Stoch Proc Appl* 60(1):19–47
- Li B, Genton MG, Sherman M (2008) On the asymptotic joint distribution of sample space–time covariance estimators. *Bernoulli* 14(1):228–248
- Mardia KV, Goodall C, Redfern EJ, Alonso FJ (1998) The kriged Kalman filter. *Test* 7(2):217–282
- Matérn B (1986) Spatial variation, volume 36 of lecture notes in statistics. Springer, Berlin
- Mikosch T, Gadrich T, Kluppelberg C, Adler RJ (1995) Parameter estimation for arma models with infinite variance innovations. *Ann Stat* 23(1):305–326
- Militino A, Ugarte M, Goicoa T, Genton M (2015) Interpolation of daily rainfall using spatiotemporal models and clustering. *Int J Climatol* 35(7):1453–1464
- Oehlert GW (1993) Regional trends in sulfate wet deposition. *J Am Stat Assoc* 88(422):390–399

- Palma W (2007) Long-memory time series: theory and methods. Wiley series in probability and statistics. Wiley, Hoboken
- Palma W, Olea R, Ferreira G (2013) Estimation and forecasting of locally stationary processes. *J Forecast* 32(1):86–96
- Peng RD, de Leeuw J (2002) An introduction to the .C interface to R. UCLA, Academic Technology Services, Statistical Consulting Group
- Porcu E, Bevilacqua M, Genton MG (2015) Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. *J Am Stat Assoc*. doi:[10.1080/01621459.2015.1072541](https://doi.org/10.1080/01621459.2015.1072541)
- R Core Team (2017) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Rao SS (2008) Statistical analysis of a spatio-temporal model with location-dependent parameters and a test for spatial stationarity. *J Time Ser Anal* 29(4):673–694
- Stein ML (2005) Space–time covariance functions. *J Am Stat Assoc* 100(469):310–321
- Stroud JR, Stein ML, Lesht BM, Schwab DJ, Beletsky D (2010) An ensemble Kalman filter and smoother for satellite data assimilation. *J Am Stat Assoc* 105(491):978–990
- Waller LA, Carlin BP, Xia H, Gelfand AE (1997) Hierarchical spatio-temporal mapping of disease rates. *J Am Stat Assoc* 92(438):607–617
- Wikle CK (2003) Hierarchical models in environmental science. *Int Stat Rev* 71(2):181–199
- Wikle CK, Cressie N (1999) A dimension-reduced approach to space-time Kalman filtering. *Biometrika* 86(4):815–829
- Wikle CK, Berliner LM, Cressie N (1998) Hierarchical Bayesian space–time models. *Environ Ecol Stat* 5(2):117–154
- Xu K, Wikle CK (2007) Estimation of parameterized spatio-temporal dynamic models. *J Stat Plan Inference* 137(2):567–588
- Zes D (2014) Facile spatio-temporal modeling, forecasting with adaptive least squares and the Kalman filter. *J Environ Stat* 6(1). <http://jes.stat.ucla.edu/v06/i01>