



Prediction of spatial functional random processes: comparing functional and spatio-temporal kriging approaches

Johan Strandberg¹ · Sara Sjöstedt de Luna¹ · Jorge Mateu²

Published online: 17 July 2019
© The Author(s) 2019

Abstract

We present and compare functional and spatio-temporal (Sp.T.) kriging approaches to predict spatial functional random processes (which can also be viewed as Sp.T. random processes). Comparisons with respect to computational time and prediction performance via functional cross-validation is evaluated, mainly through a simulation study but also on a real data set. We restrict comparisons to Sp.T. kriging versus ordinary kriging for functional data (OKFD), since the more flexible functional kriging approaches pointwise functional kriging (PWFK) and the functional kriging total model coincide with OKFD in several situations. Here we formulate conditions under which we show that OKFD and PWFK coincide. From the simulation study, it is concluded that the prediction performance of the two kriging approaches in general is rather equal for stationary Sp.T. processes. However, functional kriging tends to perform better for small sample sizes, while Sp.T. kriging works better for large sizes. For non-stationary Sp.T. processes, with a common deterministic time trend and/or time varying variances and dependence structure, OKFD performs better than Sp.T. kriging irrespective of the sample size. For all simulated cases, the computational time for OKFD was considerably lower compared to those for the Sp.T. kriging methods.

Keywords Functional kriging · Prediction · Spatial functional random processes · Spatio-temporal kriging

1 Introduction

In many fields, such as environmental, forestry, climatology, meteorology and medical sciences, it may be of interest to predict a curve at a new spatial location given that such curves have been observed at n other locations, using the information inherent in the spatial dependence between curves. Kriging predictors have a long history of being used to predict objects at new locations based on information observed at a set of other locations, especially for objects that are real- or vector-valued, see e.g. Chilès and Delfiner (2012), Cressie (2015), Cressie and Wikle (2015), and references therein. A kriging predictor is a weighted sum of the objects observed at the n spatial

locations, defined to be the best linear unbiased predictor (BLUP) minimizing the mean squared prediction error. Functional kriging predictors, used when the objects are random functions with infinite dimension, were initially discussed by Goulard and Voltz (1993), and further proposed by Giraldo et al. (2010, 2011) and Nerini et al. (2010). In these papers, the expected value of the random functions is assumed to be independent of the spatial location, the so called ordinary functional kriging. More recently, Caballero et al. (2013), Menafoglio et al. (2013), Ignaccolo et al. (2014), and Reyes et al. (2015) have investigated functional kriging methods where the expected value of the random functions may also depend on location.

Here, two kriging approaches to predict spatial functional random processes are compared. A functional random process is a process with stochastic functional objects (curves) $\chi_s = \chi_s(t), t \in T$ over the “time” domain $T \subset \mathbf{R}$ at each spatial location $s \in D \subset \mathbf{R}^d$. Given that the process has been observed at n different locations, a curve at a new location s_0 can be predicted by a *functional kriging approach*, i.e. as a

✉ Johan Strandberg
johan.strandberg@math.umu.se

¹ Department of Mathematics and Mathematical Statistics, Umeå University, Umeå, Sweden

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

linear combination of the n observed curves. A spatial functional process can also be viewed as a spatio-temporal (Sp.T.) random process $\{Z(s, t) = \chi_s(t), (s, t) \in D \times T\}$, and hence, a *Sp.T. kriging approach* could also be used, see e.g. Cressie and Wikle (2015) and Montero et al. (2015). The curve $\chi_{s_0}(t), t \in T$ would then be predicted at a dense grid of values over T , based on linear combinations of a time-grid of values over the observed curves. The question of which approach, functional or Sp.T. kriging, should be used to analyze a particular data set is an important one (with no optimal answer), as pointed out by Delicado et al. (2010). In this paper we compare the two approaches with respect to prediction performance and computational time, mainly by a simulation study but also using a real data set. Prediction performance is evaluated by functional cross-validation. Estimation of the kriging models is made without relying on distributional assumptions.

In Sect. 2 notation and definitions are given. Section 3 presents the functional and Sp.T. kriging approaches, including how to estimate the dependence structure. We also discuss how the functional kriging methods relate to each other, and under which circumstances they may coincide. In particular, we state conditions under which the two functional kriging methods ordinary kriging for functional data and pointwise functional kriging coincide, with proofs given in Appendix 1. A simulation study, comparing the two kriging approaches, is presented in Sect. 4, see also Appendix 2. In Sect. 5 both kriging approaches are applied to the Canadian temperature data, previously analyzed e.g. by Giraldo (2009), Giraldo et al. (2010) and Menafoglio et al. (2013). A discussion and concluding remarks are found in Sect. 6.

2 Preliminaries

A *spatial functional random process* $\{\chi_s : s \in D \subset \mathbf{R}^d\}$, is a process where, for each given $s \in D$, the observed random element is a functional random variable, χ_s , taking values in an infinite dimensional space, or function space (Giraldo et al. 2010; Delicado et al. 2010). We consider the case where χ_s for every fixed s is a real-valued function, $\chi_s(t), t \in T \subset \mathbf{R}$, from the compact set T to \mathbf{R} and with $s \in D \subset \mathbf{R}^2$. It is usually assumed that the realizations of the curves (functions) $\chi_s(t), t \in T, s \in D$ belong to a separable Hilbert space \mathbf{H} of square integrable functions defined on T . Our main focus is on *second-order isotropic and stationary* spatial functional random processes, that satisfy,

$$\begin{aligned} \text{(i)} \quad & E[\chi_s(t)] = m(t) \text{ and } \text{Var}[\chi_s(t)] = \sigma^2(t) \quad \forall s \in D \text{ and } \forall t \in T, \\ \text{(ii)} \quad & \text{Cov}[\chi_s(r), \chi_v(t)] = C(\|s - v\|, r, t) \quad \forall s, v \in D \text{ and } \forall r, t \in T, \end{aligned} \tag{1}$$

where $\|\cdot\|$ denotes the (Euclidean) distance measure. A spatial functional random process can also be viewed as a Sp.T. process $Z(s, t) = \chi_s(t)$, where $Z(s, t)$ takes values in \mathbf{R} , and is mapped from $(s, t) \in D \times T$, cf. Cressie and Wikle (2015). A Sp.T. process is said to be *second-order stationary and spatially isotropic* if

$$\begin{aligned} \text{(i)} \quad & E[Z(s, t)] = m \text{ and } \text{Var}[Z(s, t)] = \sigma_Z^2 \quad \forall s \in D \text{ and } \forall t \in T, \\ \text{(ii)} \quad & \text{Cov}[Z(s, r), Z(v, t)] = C_Z(\|s - v\|, |r - t|) \quad \forall s, v \in D \text{ and } \forall r, t \in T. \end{aligned} \tag{2}$$

Note that the class of stationary Sp.T processes is a subset of the class of stationary functional random processes. Section 4.3 gives examples of stationary functional random processes where (1) holds but not (2).

3 Kriging prediction

In this section two kriging approaches to predict spatial functional random processes are described. Section 3.1 presents different functional kriging methods, and under which circumstances they coincide. Section 3.2 describes the Sp.T. kriging approach. In the presentation below, estimation of the kriging models do not rely on distributional assumptions.

3.1 Functional kriging

Unless otherwise stated, we will assume that the spatial functional random process is second-order stationary and isotropic. Within the functional kriging framework, it is of interest to predict the complete random function $\chi_{s_0}(t), t \in T$, at a new location s_0 , given that a sample of random functions has been observed at n different locations, s_1, \dots, s_n . A functional kriging predictor, $\hat{\chi}_{s_0}(t), t \in T$, is the best linear unbiased predictor (BLUP) minimizing the mean integrated squared error (MISE)

$$\text{MISE}(s_0) = E \left[\int_T (\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))^2 dt \right]. \tag{3}$$

3.1.1 Ordinary kriging for functional data

Goulard and Voltz (1993) proposed one of the first functional kriging predictors,

$$\hat{\chi}_{s_0}(t) = \sum_{i=1}^n \lambda_i \chi_{s_i}(t), \quad t \in T, \tag{4}$$

which was further discussed by Giraldo et al. (2007, 2011) and there named *ordinary kriging for functional data* (OKFD). The optimal kriging weights, $\lambda = (\lambda_1, \dots, \lambda_n)^T \in \mathbf{R}^n$, that minimize (3) subject to the unbiasedness condition of the predictor, $\sum_{i=1}^n \lambda_i = 1$, satisfy

$$\begin{pmatrix} \Gamma_n & \mathbb{1}_n \\ \mathbb{1}_n^T & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ -\tau \end{pmatrix} = \begin{pmatrix} \mathbf{g}_n \\ 1 \end{pmatrix}, \tag{5}$$

where τ is the Lagrange multiplier. Here $\Gamma_n = \{\gamma(h_{ij})\}_{i,j=1}^n$, $\mathbf{g}_n = \{\gamma(h_{0j})\}_{j=1}^n$, and $\mathbb{1}_n = (1, \dots, 1)^T \in \mathbf{R}^n$, where

$$\gamma(h_{ij}) = \frac{1}{2} E \left[\int_T (\chi_{s_i}(t) - \chi_{s_j}(t))^2 dt \right], \quad \forall s_i, s_j \in D, \tag{6}$$

with $h_{ij} = \|s_i - s_j\|$, is called the *(isotropic) trace-semivariogram*. The trace-semivariogram often satisfies the properties of a classical semivariogram, being a conditional negative definite function (Menafoglio et al. 2013). The trace-semivariogram is in practice unknown and therefore needs to be estimated from the data. This is often done by first estimating the empirical trace-semivariogram for a set of h -values as

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{i,j \in N(h)} \int_T (\chi_{s_i}(t) - \chi_{s_j}(t))^2 dt, \tag{7}$$

where $N(h) = \{(s_i, s_j) : \|s_i - s_j\| \in (h - \epsilon, h + \epsilon)\}$, for some $\epsilon > 0$. A parametric variogram model $\gamma(h | \theta)$, is then fitted to a set of estimated values $\{\hat{\gamma}(h_l), h_l\}$, $l = 1, \dots, L$, by a least squares method, cf. Cressie (2015). Here, the ordinary least squares (OLS) method is used to estimate θ .

The random functions, $\chi_{s_i}(t)$, are typically observed only at a finite number of time points t_{i1}, \dots, t_{im_i} , $i = 1, \dots, n$. Goulard and Voltz (1993) suggested to fit a parametric model $\chi_{s_i}(\cdot | \alpha_i)$ to the observed values and replace $\chi_{s_i}(t)$ by $\chi_{s_i}(t | \hat{\alpha}_i)$ in (4) and (7). A non-parametric approach was suggested by Giraldo et al. (2011), where the observed random functions are represented by linear combinations of p known basis functions, $\mathbf{B}(t) = (B_1(t), \dots, B_p(t))^T$, as

$$\tilde{\chi}_{s_i}(t) = \sum_{k=1}^p a_{ik} B_k(t) = \mathbf{a}_i^T \mathbf{B}(t). \tag{8}$$

The basis functions could e.g. be B-splines, Fourier or Wavelets. The \mathbf{a}_i 's are typically determined by the least squares method, minimizing $\sum_{j=1}^{m_i} (\chi_{s_i}(t_{ij}) - \mathbf{a}_i^T \mathbf{B}(t_{ij}))^2$. In the final ordinary kriging predictor (4), the estimated trace-semivariogram values are plugged into the kriging weights (λ_i 's), with $\tilde{\chi}_{s_i}(t)$'s replacing the $\chi_{s_i}(t)$'s.

3.1.2 Pointwise functional kriging

Giraldo et al. (2008, 2010) suggested the *pointwise functional kriging predictor* (PWFK),

$$\hat{\chi}_{s_0}(t) = \sum_{i=1}^n \lambda_i(t) \chi_{s_i}(t), \quad t \in T,$$

which allows the λ_i 's to depend on t . In order to solve the infinite dimensional problem of finding the $\lambda_i(t)$ -functions that minimizes (3) subject to the unbiasedness constraint of the predictor, $\sum_{i=1}^n \lambda_i(t) = 1$, for all $t \in T$, Giraldo et al. (2008, 2010) represented the $\lambda_i(t)$ -functions by a linear combination of K known basis functions,

$$\lambda_i(t) = \sum_{k=1}^K b_{ik} B_{\lambda k}(t) = \mathbf{b}_i^T \mathbf{B}_\lambda(t), \quad i = 1, \dots, n, \tag{9}$$

and the $\chi_{s_i}(t)$'s as in (8). The optimization problem was thus reduced to a multivariate geostatistics problem. The system of $K(n + 1)$ equations to be solved in order to find the optimal \mathbf{b}_i 's is given by Giraldo et al. (2010) when $\mathbf{B}_\lambda(t) = \mathbf{B}(t)$, and for general $\mathbf{B}_\lambda(t)$ by (29) in Appendix 1, substituting $V[\chi_{s_i}(t)]$ and $Cov[\chi_{s_i}(t), \chi_{s_j}(t)]$ by $\mathbf{B}^T(t) Var[\mathbf{a}_i] \mathbf{B}(t)$ and $\mathbf{B}^T(t) Cov[\mathbf{a}_i, \mathbf{a}_j] \mathbf{B}(t)$, respectively. The optimal \mathbf{b}_i 's are functions of the covariances between the various \mathbf{a}_i 's, which in practice need to be estimated. Giraldo et al. (2010) suggest estimating these covariances via a linear model of coregionalization (Goulard and Voltz 1992). Note that the $\mathbf{B}_\lambda(t)$'s need to satisfy the unbiasedness condition

$$\sum_{i=1}^n \lambda_i(t) = \mathbf{c}^T \mathbf{B}_\lambda(t) = 1, \quad \text{for all } t \in T, \tag{10}$$

where $\mathbf{c} = \sum_{i=1}^n \mathbf{b}_i$. B-splines and Fourier basis functions are two admissible choices, that satisfy (10) when $\mathbf{c} = \mathbf{1}$ and $\mathbf{c} = (1, 0, \dots, 0)^T$, respectively. In fact any set of basis functions where one (the first say) basis function is a constant, $B_{\lambda 1}(t) = k$, satisfies (10) for $\mathbf{c} = (1/k, 0, \dots, 0)^T$.

3.1.3 Situations when OKFD and PWFK coincide

Here we present situations where PWFK and OKFD coincide. Consider spatial functional random processes that satisfy

- (i) $E[\chi_s(t)] = m(t) \forall s \in D$ and $\forall t \in T$,
 - (ii) $Cov[\chi_s(t), \chi_v(t)] = w^2(t) C(s, v) \forall s, v \in D$ and $\forall t \in T$,
- (11)

where $w(t)$ is a real-valued deterministic function. These type of processes include second-order (isotropic) stationary spatial functional random processes, but also e.g. when $\chi_s(t)$ can be expressed as $\chi_s(t) = m(t) + w(t)Z(s, t)$, where $Z(s, t)$ is a time stationary Sp.T. process with mean zero

and $Cov[Z(s, r), Z(v, t)] = C(s, v, |r - t|)$. Note that for Sp.T. stationary processes $m(t) = m$ and $w^2(t) = 1$. Let $\Sigma = \{C(s_i, s_j)\}_{i,j=1}^n \in \mathbf{R}^{n \times n}$, $\Sigma^{-1} = \{\alpha_{ij}\}_{i,j=1}^n$, and $\alpha_{\bullet\bullet} = \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij}$. The following proposition states conditions under which the two functional kriging methods OKFD and PWFK coincide.

Proposition 3.1 *Suppose that $\{\chi_s(t), t \in T, s \in D\}$ is a spatial functional random process satisfying (11). Further, assume that $\lambda_i(t) = \mathbf{b}_i^T \mathbf{B}_\lambda(t)$, that the $\mathbf{B}_\lambda(t)$'s satisfy (10) for some constant vector \mathbf{c} , that the inverse of the matrix \mathbf{W} exists, where*

$$\mathbf{W} = \int_T w^2(t) \mathbf{B}_\lambda(t) \mathbf{B}_\lambda^T(t) dt,$$

and that Σ^{-1} exists with $\alpha_{\bullet\bullet}$ being non-zero. Then the optimal kriging weights of PWFK that minimize (3) are unique and satisfy $\lambda_i(t) = \lambda_i$, with $\mathbf{b}_i = \lambda_i \mathbf{c}$, for all $i = 1, \dots, n$, and thus coincide with those of OKFD.

The existence of \mathbf{W}^{-1} and Σ^{-1} , with $\alpha_{\bullet\bullet} \neq 0$, ensure the existence of a unique solution, see further details in the proof of the proposition given in Appendix 1. In line with Giraldo (2009) we concluded that the computational time for PWFK (using R-code kindly provided by Giraldo et al. (2010)) was substantially larger than for OKFD [using the R-package *geofd*, see Giraldo et al. (2012)]. We also noted that the estimated PWFK kriging weights always became constant when $\mathbf{B}_\lambda(t) = \mathbf{B}(t)$ were cubic B-splines or Fourier basis functions (after correction of a bug in the R-code).

3.1.4 Functional kriging total model

Giraldo (2009, 2014), and independently Nerini et al. (2010), proposed the *functional kriging total model* (FKTM),

$$\hat{\chi}_{s_0}(t) = \sum_{i=1}^n \int_T \lambda_i(t, v) \chi_{s_i}(v) dv, \quad t \in T. \tag{12}$$

Assuming that the random functions $\chi_{s_i}(t)$ satisfy (8) and that the kriging weights satisfy

$$\begin{aligned} \lambda_i(t, v) &= \sum_{k=1}^p \sum_{l=1}^p c_{ik}^l B_k(t) B_l(v) \\ &= \mathbf{B}(t)^T \mathbf{C}_i \mathbf{B}(v), \quad i = 1, \dots, n, \end{aligned}$$

Giraldo (2014) proposed a way to determine the $\lambda_i(t, v)$'s (i.e. the \mathbf{C}_i 's) such that the predictor (12) is unbiased and minimizes (3). Also here, the \mathbf{C}_i 's are functions of the covariances between the \mathbf{a}_i 's, which in practice need to be estimated, see Giraldo (2014) for more details.

The FKTM method is computationally heavy compared to OKFD, just like the PWFK method (Giraldo 2009). Moreover, Menafoglio and Petris (2016) showed that if the realizations of $\chi_s(t)$ belong to the Hilbert space of square integrable functions on T (in fact also for general separable Hilbert spaces), and the functional second-order stationary random process is Gaussian, then the kriging weights of FKTM and OKFD agree a.s. for any orthonormal base $\mathbf{B}(t)$.

3.2 Spatio-temporal kriging

Since a spatial functional process also can be viewed as a Sp.T. process, $Z(s, t) = \chi_s(t)$, taking values in $(s, t) \in D \times T$, it could also be predicted by Sp.T. kriging methods. Given the observed values $\mathbf{Z} = (Z(s_1, t_{11}), \dots, Z(s_1, t_{1m_1}), \dots, Z(s_n, t_{n1}), \dots, Z(s_n, t_{nm_n}))^T \in \mathbf{R}^N$, $N = \sum_{i=1}^n m_i$, the Sp.T. kriging predictor at location s_0 and time point $t \in T$,

$$\hat{Z}(s_0, t) = \sum_{i=1}^n \sum_{j=1}^{m_i} \lambda_{ij}^t Z(s_i, t_{ij}), \tag{13}$$

is defined to be the BLUP minimizing the mean squared prediction error (MSPE)

$$MSPE(s_0, t) = E[(\hat{Z}(s_0, t) - Z(s_0, t))^2]. \tag{14}$$

For Sp.T. processes with constant mean value, the unbiasedness condition implies that $\sum_{i=1}^n \sum_{j=1}^{m_i} \lambda_{ij}^t = 1$. The optimal Sp.T. kriging weights $\boldsymbol{\lambda} = (\lambda_{11}^t, \dots, \lambda_{1m_1}^t, \dots, \lambda_{n1}^t, \dots, \lambda_{nm_n}^t)^T$ for processes with unknown constant mean, satisfy the system of $(N + 1)$ equations

$$\begin{pmatrix} C_N & \mathbb{1}_N \\ \mathbb{1}_N^T & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \tau \end{pmatrix} = \begin{pmatrix} \mathbf{k}_N \\ 1 \end{pmatrix}, \tag{15}$$

where τ is the Lagrange multiplier used to take into account the unbiasedness restriction, $C_N = Var[\mathbf{Z}] \in \mathbf{R}^{N \times N}$ is the variance-covariance matrix of \mathbf{Z} , and $\mathbf{k}_N = Cov[\mathbf{Z}, Z(s_0, t)] \in \mathbf{R}^N$. The predictor (13) is referred to as the *Sp.T. ordinary kriging predictor*, see e.g. Cressie and Wikle (2015). The dependence structure in practice needs to be estimated from the data and is then plugged into the kriging weights (λ_{ij}^t 's). For second-order stationary and spatially isotropic Sp.T. processes satisfying (2), we have that

$$\begin{aligned} Cov[Z(s, r), Z(v, t)] &= C_Z(\|s - v\|, |r - t|) \\ &= \sigma_Z^2 - \gamma_Z(\|s - v\|, |r - t|), \end{aligned} \tag{16}$$

where $\sigma_Z^2 = Var[Z(\cdot, \cdot)]$ and

$$\gamma_Z(\|s - v\|, |r - t|) = E[(Z(s, r) - Z(v, t))^2]/2, \\ s, v \in D \text{ and } r, t \in T,$$

is the (spatially isotropic) Sp.T. semivariogram. The dependence structure is typically estimated via the semivariogram as follows. First, an empirical (spatially isotropic) Sp.T. semivariogram is computed from lag classes as

$$\hat{\gamma}_Z(h, u) = \frac{1}{2|N(h, u)|} \sum_{(i,j,k,l) \in N(h,u)} (Z(s_i, t_{ik}) - Z(s_j, t_{jl}))^2,$$

where $N(h, u) = \{(s_i, t_{ik}), (s_j, t_{jl}) : \|s_i - s_j\| \in (h - \epsilon, h + \epsilon), \text{ and } |t_{ik} - t_{jl}| \in (u - \delta, u + \delta)\}$, for some $\epsilon, \delta > 0$, and $|N(h, u)|$ is the number of distinct elements in $N(h, u)$. A parametric semivariogram model, $\gamma(h, u|\theta)$, is then fitted to a set of $\{\hat{\gamma}_Z(h_l, u_l), (h_l, u_l)\}, l = 1, \dots, L$ by a least squares method.

In this paper we consider three commonly used types of stationary Sp.T. semivariogram (covariogram) models: *the separable model*

$$C_Z(h, u) = C_s(h)C_t(u), \tag{17}$$

modeling the Sp.T. covariance function by the product of a spatial and a temporal covariance function, *the product-sum model*,

$$C_Z(h, u) = kC_s(h)C_t(u) + C_s(h) + C_t(u),$$

with $k > 0$, and *the metric Sp.T. covariance model*

$$C_Z(h, u) = C_{joint}(\sqrt{h^2 + (\kappa u)^2}).$$

More generally, in Sp.T. kriging modeling, the process is often described as

$$Z(s, t) = \mu(s, t) + \epsilon(s, t),$$

where $\mu(s, t)$ is a deterministic trend, and $\epsilon(s, t)$ is a mean zero Sp.T. random field, usually assumed stationary. The trend is typically modeled by

$$\mu(s, t) = \beta^T \mathbf{x}(s, t), \tag{18}$$

where $\mathbf{x}(s, t) \in \mathbf{R}^M$ is a set of M known covariates, often chosen to be polynomials of s and t , and $\beta \in \mathbf{R}^M$ is an unknown parameter to be determined. When the Sp.T. process has a deterministic (unknown) non-constant trend of the form (18), then the BLUP (13) that minimizes (14) is called the *Sp.T. universal kriging predictor*, and the kriging weights are functions of both the dependence structure and the covariates evaluated at the observed and predicted locations, see e.g. Cressie and Wikle (2015) Section 4.1.2, page 148. An iterative weighted least squares method may be used to estimate β and the Sp.T. variogram parameter θ . Firstly, β can be estimated by the OLS method, minimizing

$$\sum_{i=1}^n \sum_{j=1}^{m_i} (Z(s_i, t_{ij}) - \beta^T \mathbf{x}(s_i, t_{ij}))^2.$$

Based on the resulting regression residuals, the Sp.T. semivariogram is then estimated by fitting a parametric Sp.T. semivariogram model to the corresponding empirical Sp.T. semivariogram by a least squares method. The parameter β is then re-estimated using a weighted least squares method, taking into account the estimated dependence structure of the residuals (Cressie 2015). The dependence structure (variogram) is again estimated based on the updated residuals, and the whole procedure iterated until convergence. Note that if the deterministic trend only depends on time, such that $\mu(s, t) = m(t)$, the functional kriging methods do not need to specify and estimate the trend, whereas the Sp.T. kriging methods need to.

4 A simulation study

We now present a simulation study that aims to bring light over the relative merits of Sp.T. and functional kriging, with particular focus on Gaussian second-order stationary functional processes in \mathbf{R}^2 . Since the functional kriging methods OKFD, PWFK, and FKTM often coincide for such processes (see Sects. 3.1.3 and 3.1.4) we restrict our comparisons to Sp.T. kriging versus OKFD. Inspired by the setups in Giraldo et al. (2012), Sun and Genton (2012) and Romano et al. (2015), we simulate data from Gaussian processes with three main types of covariance structures. The first two scenarios have stationary isotropic separable and non-separable covariance functions, respectively. The third scenario corresponds to second-order stationary functional (but non-stationary Sp.T.) processes with constant mean. For all three scenarios, several different cases are simulated, with varying strengths of spatial and temporal dependences, see Table 1. The different parameters in Table 1 control the Sp.T. correlation structure. Figure 1 illustrates examples of simulated data for six of the cases, all with constant means.

For each of the 24 cases, three different sample sizes were considered: *small* referring to $n = 6 \times 6$ spatial locations and $m = 12$ time points, *medium* referring to $n = 6 \times 6$ spatial locations and $m = 50$ time points, and *large* referring to $n = 15 \times 15$ spatial locations and $m = 50$ time points. The number of time points were equally distributed on $[0, 1]$ and the spatial locations were located on a regular grid in $[0, 1] \times [0, 1]$. Moreover, for cases 1–18, the presence of a deterministic time trend, $m(t) = 9 + 3 \sin(2\pi t)$, was also investigated.

For each case, sample size (and trend type), 100 replicates were simulated, using the R-packages RandomFields

Table 1 The 24 different types (cases) of simulated Gaussian processes and their parameters: isotropic second-order stationary Sp.T. processes with separable (cases 1–9) and non-separable (cases 10–18) covariance functions, and second-order stationary functional (but non-stationary Sp.T.) processes (cases 19–24) with constant means

Generated data				Generated data				Generated data			
Case	Type	α	β	Case	Type	α	β	Case	Type	p	α
1	Separable	0.1	0.1	10	Non-separable	0.1	0.1	19	Non-stationary	7	0.1
2			1	11			1	20			0.5
3			10	12			10	21			2
4		0.5	0.1	13		0.5	0.1	22		15	0.1
5			1	14			1	23			0.5
6			10	15			10	24			2
7		2	0.1	16		2	0.1				
8			1	17			1				
9			10	18			10				

The larger the value of α and β the weaker the spatial and temporal correlation, respectively

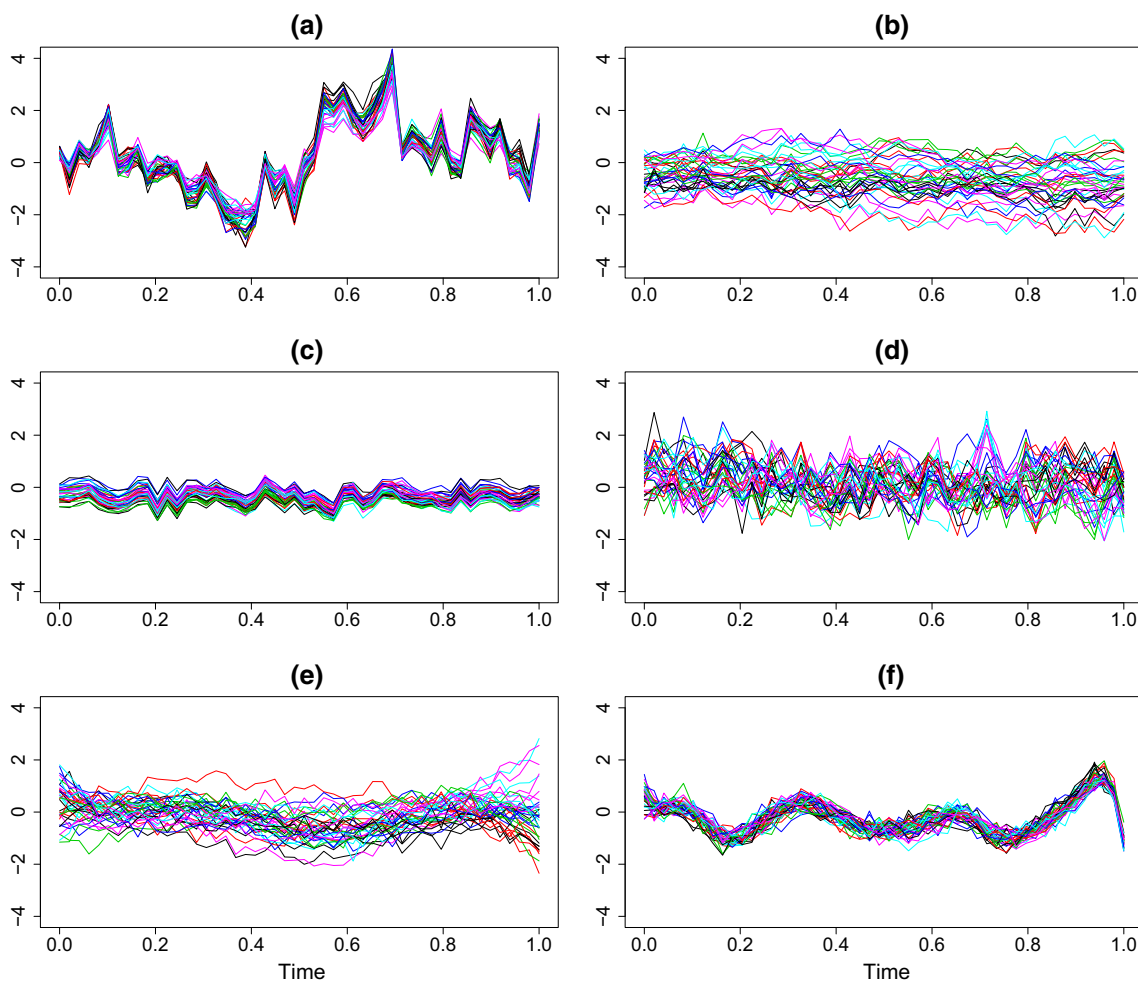


Fig. 1 Examples of simulated data considering medium sample sizes without a deterministic time trend for: **a** case 3 ($\alpha = 0.1, \beta = 10$), **b** case 7 ($\alpha = 2, \beta = 0.1$), **c** case 10 ($\alpha = 0.1, \beta = 0.1$), **d** case 18

($\alpha = 2, \beta = 10$), **e** case 21 ($\alpha = 2, p = 7$) and **f** case 22 ($\alpha = 0.1, p = 15$). The larger the value of α and β the weaker the spatial and temporal correlation, respectively

(Schlather et al. 2015) for cases 1–18 and fda (Ramsay et al. 2009) for cases 19–24. Estimation of the OKFD models was performed using the R-package geofd (Giraldo

et al. 2012). The Sp.T. kriging models were estimated using the R-packages gstat (Pebesma 2004) and spactime (Pebesma et al. 2012).

Functional cross-validation (FCV), suggested by Giraldo et al. (2010, 2011), was used to evaluate the prediction performance. In FCV, the data from each observed spatial location is removed, one at a time, and then predicted at all observed time points by the prediction method using the observed functional data at the remaining locations. The mean squared prediction error (MSPE) is computed as

$$MSPE = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} (Z(s_i, t_{ij}) - \hat{Z}^{-i}(s_i, t_{ij}))^2 / m_i, \tag{19}$$

where $\hat{Z}^{-i}(s_i, t_{ij})$ denotes the predicted value at location (s_i, t_{ij}) based on the functional data with the observations $Z(s_i, t_{ij}), j = 1, \dots, m_i$ excluded.

The three main scenarios are now presented in more detail, together with the simulated results.

4.1 Separable covariance function

Here we evaluate the prediction performance of OKFD and Sp.T kriging models, for the first nine cases in Table 1, which are Gaussian Sp.T processes with separable covariance functions. The spatial covariance function $C_s(h)$ in (17) was chosen to be the exponential covariance function with nugget effect,

$$C_s(h) = (1 - \nu) \exp(-\alpha h) + \nu I\{h = 0\}.$$

The nugget effect ν was set to 0.04. For parameter α , we considered the values 0.1, 0.5 and 2, corresponding to the effective ranges 30, 6 and 1.5 (very strong, medium and weak spatial correlation), respectively. The temporal covariance function $C_t(u)$ in (17) was chosen to be the stable covariance function

$$C_t(u) = \exp(-(\beta u)^\gamma). \tag{20}$$

Here, γ was fixed to 0.5, while the values for β were 0.1, 1, and 10, corresponding to the effective ranges 90, 9 and 0.9 (very strong, medium and weak temporal correlation), respectively.

Given the generated data $Z(s_i, t_j), i = 1, \dots, n, j = 1, \dots, m$, the OKFD model was estimated using Fourier or cubic B-splines with different numbers of basis functions, see Table 10 in Appendix 2 for a detailed specification. For each number and type of basis function the spherical, exponential and stable semivariogram models were fitted to the empirical trace-semivariogram. For each case (1–9) a total of 36, 42 and 42 OKFD models (two types of basis functions \times # different numbers of basis functions \times #trace-semivariograms) were estimated and fitted to the data for small, medium and large sample sizes, respectively. These models were then evaluated by FCV (functional cross-validation) in terms of the MSPE (19), and the minimum MSPE over the models registered. The

overall MSPE for each case and sample size was computed as the average minimum MSPEs over the 100 replicates.

Sp.T. ordinary kriging models were estimated for the data sets simulated without a deterministic time trend. First, the separable, product-sum and metric Sp.T. semi-variogram models were fitted to the empirical Sp.T. semivariograms. For these three models, all pairwise combinations of the exponential, spherical and stable variograms were considered for the spatial (isotropic), temporal and joint variogram models. It resulted in 9 separable, 9 product-sum and 3 metric Sp.T. semivariogram models. All the Sp.T. ordinary kriging models were evaluated by FCV, the minimum MSPE registered over the different models within each of the three subgroup Sp.T. semivariogram models was obtained, and the overall MSPE was also computed for each case (1–9) and sample size. The Sp.T. models with a product-sum and a metric covariance function were not evaluated for large size samples, due to the large computational time.

The overall MSPEs for the OKFD and the Sp.T. ordinary kriging models for cases 1–9 considering medium sample sizes are presented in Table 2. Corresponding results for small and large sample sizes are reported in Appendix 2, Tables 6 and 7. The last column in Table 2 reports p values from paired two-sided t tests comparing the overall MSPEs between the OKFD and the Sp.T. separable models. The Sp.T. separable kriging models in general had lower overall MSPEs compared to the Sp.T. product-sum and metric models. This was expected, since the simulated data were generated from Sp.T. models with separable covariance functions. Interestingly, the overall MSPE was often (significantly) lower for OKFD compared to the Sp.T. separable models, for small and medium sample sizes (Tables 2, 6). For large sample sizes, the estimated Sp.T. (separable) models often performed better than OKFD (Table 7).

Studying the overall MSPEs in more detail reveals that the weaker the spatial correlation and the stronger the temporal correlation, the better the OKFD performs in relation to the Sp.T. separable model, regardless of the sample size. Case 3 for example, with strong spatial and weak temporal correlation, has significantly lower overall MSPE for the Sp.T. separable model compared to the OKFD model for medium and large sample sizes (Tables 2, 7). On the other hand, for case 7, with weak spatial and strong temporal correlation, the result is reversed.

The numbers in parentheses in Tables 2, 6, and 7, report the average computational time (for estimation and FCV) in seconds over all estimated models and replications when run on a 3.5 GHz Intel Core i7 processor with 32 GB ram memory. It reveals that prediction by and estimation of an OKFD model is substantially faster than the Sp.T. kriging models, regardless of the sample size. The Sp.T. separable

Table 2 Prediction performance in terms of mean squared prediction errors (MSPEs) for the cases 1–9 (and 1*–9*) without (and with) a deterministic time trend considering medium sample sizes

Generated data				Overall MSPE				Comparison
Case	Type	α	β	OKFD	Sp.T. separable	Sp.T. product-sum	Sp.T. metric	p value
1	Separable, no trend	0.1	0.1	0.061 (0.2)	0.062 (26.7)	0.064 (88.0)	0.083 (90.3)	0.552
2			1	0.068 (0.2)	0.067 (26.1)	0.072 (89.0)	0.080 (89.9)	0.059
3			10	0.069 (0.2)	0.066 (24.7)	0.069 (92.2)	0.083 (90.3)	< 0.001
4		0.5	0.1	0.134 (0.2)	0.143 (23.6)	0.149 (85.4)	0.204 (85.7)	< 0.001
5			1	0.131 (0.2)	0.135 (29.8)	0.145 (102.9)	0.217 (98.5)	0.011
6			10	0.139 (0.2)	0.137 (27.6)	0.164 (103.3)	0.214 (95.7)	0.044
7		2	0.1	0.334 (0.2)	0.357 (29.0)	0.353 (100.6)	0.416 (97.6)	< 0.001
8			1	0.368 (0.2)	0.400 (29.3)	0.403 (106.1)	0.520 (99.2)	< 0.001
9			10	0.372 (0.2)	0.386 (28.7)	0.445 (104.3)	0.529 (97.9)	0.001
1*	Separable, trend	0.1	0.1	0.066 (0.2)	0.066 (27.5)	0.068 (89.4)	0.094 (91.5)	0.706
2*			1	0.063 (0.2)	0.063 (26.9)	0.066 (89.6)	0.084 (91.2)	0.882
3*			10	0.068 (0.2)	0.066 (25.8)	0.070 (93.8)	0.090 (92.0)	< 0.001
4*		0.5	0.1	0.135 (0.2)	0.141 (29.2)	0.143 (95.7)	0.220 (96.2)	0.002
5*			1	0.130 (0.2)	0.136 (30.4)	0.153 (100.3)	0.224 (96.0)	0.003
6*			10	0.135 (0.2)	0.133 (28.5)	0.157 (103.5)	0.182 (96.0)	0.007
7*		2	0.1	0.376 (0.2)	0.419 (29.1)	0.411 (100.8)	0.479 (96.2)	< 0.001
8*			1	0.377 (0.2)	0.394 (27.7)	0.408 (95.8)	0.524 (90.8)	0.001
9*			10	0.372 (0.2)	0.385 (29.7)	0.435 (103.4)	0.531 (96.5)	0.002

The smallest overall MSPE for each case is highlighted in bold. The numbers in parentheses represent the average computational time in seconds over the corresponding estimated models and replications. The last column shows p values from two-sided paired t tests comparing the overall MSPEs between the OKFD and the Sp.T. separable models. The larger the value of α and β the weaker the spatial and temporal correlation, respectively

models had lower computational time compared to the Sp.T. product-sum and metric models, due to simplifying (Kronecker product) structures of the variance covariance matrix.

Figure 2 presents how the type and number of basis functions used in the OKFD model affects prediction

performance (minimum MSPE over the three trace-semivariogram models, averaged over the 100 realizations) for cases 3 and 7 considering medium sample sizes. The number of basis functions turns out to be an important factor for prediction performance, in general with smaller prediction error the more basis functions used. On the other

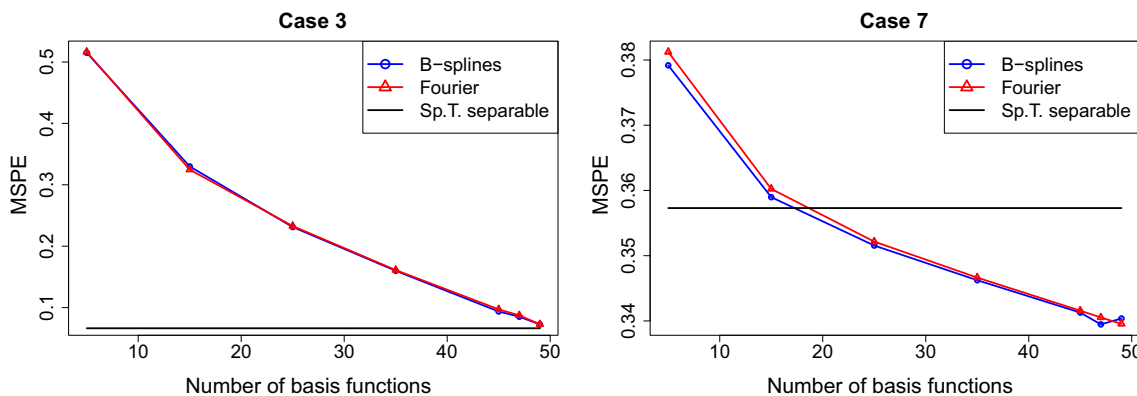


Fig. 2 Prediction performance (minimum MSPE over the three trace-semivariogram models, averaged over the 100 realizations) for cases 3 and 7 considering medium sample sizes without a deterministic time trend when the estimated OKFD model is based on different numbers

(p) of basis functions, being both Fourier and cubic B-spline. The solid black lines represent the corresponding overall MSPE of the Sp.T. separable model

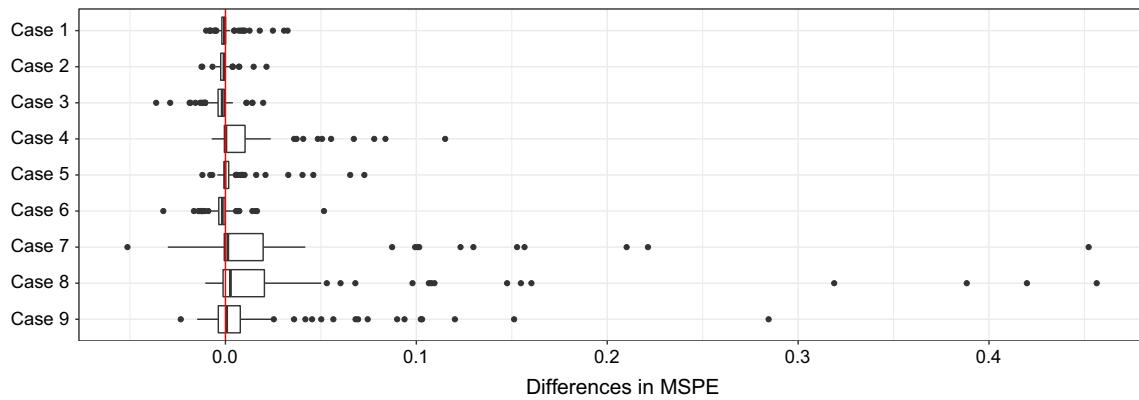


Fig. 3 Box-plots for cases 1–9 (considering medium sample sizes without a deterministic time trend) of the differences in (minimum) MSPE between the two kriging approaches (MSPE(Sp.T)-MSPE(OKFD)) for the 100 replicates

hand, the type of basis functions, Fourier or cubic B-splines, is of less importance. These findings are consistent with all cases (1–9) and for all considered sample sizes.

In Fig. 3 box-plots of the differences in (minimum) MSPE between the two kriging approaches (MSPE(Sp.T)-MSPE(OKFD)) for the 100 replicates considering medium sample sizes are presented. From this, it becomes clear that OKFD produces more robust predictions. The Sp.T. separable kriging models produced much higher MSPEs than OKFD for many realizations, especially for small and medium sample sizes.

For the cases 1–9 with a common deterministic (sinusoidal) time trend, the same OKFD models as specified above were used again since these are designed to handle situations where a common deterministic time trend is present. However, predictions by Sp.T. kriging were now performed by universal Sp.T. kriging instead of Sp.T. ordinary kriging, using the same Sp.T. semivariogram models as for the ordinary Sp.T. kriging models. The deterministic time trend in the universal Sp.T. kriging model was specified to be the same as the one simulated from.

Table 2 summarizes the prediction performance of the two kriging approaches for cases 1–9 with deterministic time trend considering medium sample sizes, presented as cases 1*–9*. Corresponding results for small and large sample sizes are reported in Appendix 2, Tables 6 and 7. From these tables, we see that the presence and estimation of a deterministic time trend did not have a large effect on the prediction performance, and more or less gave the same conclusions with respect to the relative performance of the two kriging approaches, regardless of the sample size.

4.2 Non-separable covariance function

Cases 10–18 in Table 1 correspond to Gaussian Sp.T. processes with non-separable covariance functions of the form

$$Cov_{NSEP}(h, u) = (1 - \nu)(2 - C_t(u))^{-\delta/2} \exp\left(-\frac{\alpha h}{\sqrt{2 - C_t(u)}}\right) + \nu I\{h = 0\},$$

with parameters set to $\delta = 2$, $\nu = 0.04$, and $\alpha = 0.1, 0.5$, and 2 . The covariance function $C_t(u)$ was chosen to be the stable covariance function (20) with $\gamma = 0.5$ and $\beta = 0.1, 1$, and 10 . The OKFD and the Sp.T. kriging models estimated in Sect. 4.1 were also fitted to the simulated data sets of cases 10–18. Prediction performance of the two kriging approaches was evaluated in the same way as described in Sect. 4.1 and is summarized in Table 3 for the non-separable cases 10–18 (and 10*–18*) for medium sample sizes without (and with) a deterministic time trend, respectively. Corresponding results for small and large sample sizes are reported in Appendix 2, Tables 8 and 9.

In general, we draw similar conclusions as in Sect. 4.1 for the separable cases 1–9 (and 1*–9*): the Sp.T. separable kriging models perform better than the Sp.T. product-sum and metric models; the weaker the spatial correlation and the stronger the temporal correlation, the better the OKFD performs in relation to the Sp.T. (separable) models; OKFD works better than the Sp.T. models for small to medium sample sizes whereas the Sp.T. separable kriging models perform better for large sample sizes; more basis functions in OKFD generally improve prediction performance; computational times are much lower for OKFD; the presence of a deterministic time trend does not change the conclusions.

A more detailed comparison of the overall MSPEs in Table 3, and Tables 8 and 9 reveals that prediction

Table 3 Prediction performance in terms of mean squared prediction errors (MSPEs) for the cases 10–18 (and 10*–18*) without (and with) a deterministic time trend considering medium sample sizes

Generated data				Overall MSPE				Comparison
Case	Type	α	β	OKFD	Sp.T. separable	Sp.T. product-sum	Sp.T. metric	p value
10	Non-separable, no trend	0.1	0.1	0.066 (0.2)	0.067 (26.2)	0.070 (87.6)	0.101 (89.9)	0.050
11			1	0.066 (0.2)	0.065 (25.7)	0.069 (87.5)	0.100 (89.6)	0.082
12			10	0.065 (0.2)	0.064 (24.4)	0.067 (90.2)	0.082 (89.8)	0.075
13		0.5	0.1	0.128 (0.2)	0.139 (25.8)	0.145 (92.8)	0.201 (92.9)	0.001
14			1	0.134 (0.2)	0.140 (24.2)	0.154 (90.2)	0.248 (87.0)	< 0.001
15			10	0.137 (0.2)	0.140 (27.4)	0.155 (99.5)	0.257 (93.3)	0.006
16		2	0.1	0.366 (0.2)	0.398 (26.8)	0.391 (95.9)	0.430 (92.4)	< 0.001
17			1	0.354 (0.2)	0.390 (24.8)	0.386 (91.8)	0.476 (86.5)	< 0.001
18			10	0.373 (0.2)	0.391 (27.4)	0.402 (95.2)	0.579 (90.4)	0.003
10*	Non-separable, trend	0.1	0.1	0.063 (0.2)	0.064 (26.4)	0.066 (88.5)	0.099 (90.5)	0.463
11*			1	0.065 (0.2)	0.065 (26.2)	0.070 (88.1)	0.100 (90.8)	0.976
12*			10	0.069 (0.2)	0.067 (25.8)	0.072 (86.1)	0.094 (86.6)	0.001
13*		0.5	0.1	0.137 (0.2)	0.147 (27.8)	0.152 (92.9)	0.212 (93.5)	0.001
14*			1	0.138 (0.2)	0.144 (30.3)	0.153 (97.4)	0.270 (95.3)	0.034
15*			10	0.136 (0.2)	0.140 (28.1)	0.154 (98.1)	0.230 (91.5)	0.002
16*		2	0.1	0.359 (0.2)	0.406 (26.1)	0.384 (96.0)	0.450 (91.7)	< 0.001
17*			1	0.367 (0.2)	0.414 (27.0)	0.406 (96.2)	0.516 (91.5)	< 0.001
18*			10	0.374 (0.2)	0.396 (27.5)	0.398 (96.3)	0.591 (91.9)	0.003

The smallest overall MSPE for each case is highlighted in bold. The numbers in parentheses represent the average computational time in seconds over the corresponding estimated models and replications. The last column shows p values from two-sided paired t tests comparing the overall MSPEs between the OKFD and the Sp.T. separable models. The larger the value of α and β the weaker the spatial and temporal correlation, respectively

performance of OKFD in general improves in comparison to the Sp.T. separable kriging models for the simulated data sets with non-separable covariance functions (cases 10–18) compared to those simulated from separable covariance functions (cases 1–9). This result was to be expected, since none of the fitted (Sp.T.) kriging models coincide with the models that generated the data for cases 10–18.

4.3 Non-stationary cases

Generation of simulated data sets of second-order isotropic stationary functional, but non-stationary Sp.T. Gaussian processes with constant mean (cases 19–24 in Table 1) were based on the model

$$\chi_{s_i}(t) = \mathbf{a}_i^T \mathbf{B}(t) + \epsilon_{s_i}(t), \quad i = 1, \dots, n. \tag{21}$$

The basis functions $\mathbf{B}(t) \in \mathbf{R}^p$ with $p = 7$, and 15 cubic B-splines, are defined on equally spaced knots on the interval $[0, 1]$. Moreover, $\mathbf{a}_i = (a_1(s_i), \dots, a_p(s_i))^T$, where $a_k(s), k = 1, \dots, p$, were chosen to be p independent identically distributed second-order stationary isotropic zero mean Gaussian processes in \mathbf{R}^2 with exponential covariance function $C(h) = \exp(-\alpha h)$ with $\alpha = 0.1, 0.5$ and 2 .

Hence, the vectors $(a_k(s_1), \dots, a_k(s_n))^T, k = 1, \dots, p$, are p independent realizations of a multivariate Gaussian random variable $N_n(\mathbf{0}, \Sigma)$, where the $n \times n$ covariance matrix equals $\Sigma = \{\exp(-\alpha \|s_i - s_j\|)\}$. The $\epsilon_{s_i}(t)$'s are white noise measurement errors, independent and identically normally distributed with mean 0 and variance 0.04, i.e. $\epsilon_{s_i}(t) \sim N(0, 0.04)$.

For each of the $2 \times 3 = 6$ cases (19–24) and for each such generated data set we fitted the same OKFD models as those fitted in Sect. 4.1. However, for medium and large sample sizes we extended the choices of number of basis functions (see Appendix 2 in Table 11 for a specification), yielding a total of 36, 90 and 90 different estimated OKFD models for small, medium and large sample sizes, respectively. For each case (19–24), sample size, and realization, predictions were made and evaluated by FCV for all models, and the minimum MSPE over the models registered. The overall MSPE for each case and sample size was finally computed as the average minimum MSPE over the 100 replicates. Furthermore, the same Sp.T. ordinary kriging models fitted to the data in Sect. 4.1, were also estimated for these data sets. Additionally, Sp.T. universal kriging models were fitted, with a deterministic time trend

specified by a linear combination of the same basis functions that were used to generate the data set. Hence, a total of 18 separable, 18 product-sum and 6 metric Sp.T. kriging models were fitted to the data; predictions evaluated by FCV, the minimum MSPE registered over the models within the three groups of dependence structures, and the overall MSPE computed for each case (19–24) and sample size. As in Sects. 4.1 and 4.2, the Sp.T. models with a product-sum and a metric covariance function were not evaluated for large sample sizes due to large computational times.

Table 4 summarizes the prediction performance of the two kriging approaches for cases 19–24 and all three sample sizes. Note that these simulated data sets have time varying variances and covariances, which the Sp.T. kriging approach is not designed to capture, whereas the OKFD model can handle such situations. We would therefore expect OKFD to perform better than the Sp.T. kriging approach, which is indeed the case. In fact, OKFD has significantly lower overall MSPE for all cases and sample sizes in Table 4 except for cases 22–23 considering small sample sizes. For these two cases the Sp.T. separable kriging model works better. This is coupled to the low number of observations (12) per location for small sample

sizes. When the functional representations of the data at each location is formed for the OKFD models, we can thus at most fit a linear combination of 12 basis functions, whereas, the data are generated by 15 B-splines. The functional representations may thus fail to capture the full temporal time dynamics. The Sp.T. universal kriging models on the other hand, do fit a common deterministic time trend using all 15 B-splines. From Table 4, it is also noted that Sp.T. kriging models with fitted metric variograms sometimes had better prediction performance than the Sp.T. separable kriging models, but still worse than the best OKFD models. Moreover, we again note that the computational time for OKFD is much lower than for the Sp.T. models.

Figure 4 illustrates how the type and number of basis functions used in the fitted OKFD models affect the prediction performance for cases 21 and 22 considering medium sample sizes. Case 21 corresponds to simulated data generated by 7 B-splines with weak spatial dependence, whereas case 22 corresponds to simulated data generated by 15 B-splines with strong spatial dependence. In contrast to the simulated stationary Sp.T. models (cases 1–18) where prediction performance typically increases with the number of basis functions used in the fitted OKFD

Table 4 Prediction performance in terms of mean squared prediction errors (MSPEs) for the cases 19–24 over the different sample sizes

Generated data					Overall MSPE				Comparison
Case	Type	Data size	p	α	OKFD	Sp.T. separable	Sp.T. product-sum	Sp.T. metric	p value
19	Non-stationary	Small	7	0.1	0.054 (0.2)	0.056 (8.0)	0.057 (11.5)	0.056 (6.3)	< 0.001
20				0.5	0.096 (0.2)	0.099 (8.4)	0.100 (14.1)	0.100 (6.1)	< 0.001
21				2	0.226 (0.2)	0.232 (9.0)	0.236 (15.7)	0.235 (6.4)	< 0.001
22			15	0.1	0.058 (0.2)	0.057 (7.4)	0.058 (13.2)	0.061 (6.1)	< 0.001
23				0.5	0.099 (0.2)	0.099 (7.8)	0.101 (15.2)	0.102 (6.0)	0.342
24				2	0.239 (0.2)	0.243 (8.7)	0.260 (16.4)	0.263 (6.5)	0.049
19	Medium	7	0.1	0.1	0.050 (0.2)	0.055 (24.8)	0.056 (83.8)	0.052 (84.1)	< 0.001
20				0.5	0.083 (0.2)	0.092 (24.4)	0.093 (79.5)	0.088 (79.6)	< 0.001
21				2	0.202 (0.2)	0.212 (28.7)	0.220 (91.2)	0.210 (88.0)	< 0.001
22			15	0.1	0.052 (0.2)	0.056 (26.1)	0.056 (81.9)	0.056 (84.2)	< 0.001
23				0.5	0.087 (0.2)	0.094 (28.0)	0.093 (90.2)	0.093 (87.3)	< 0.001
24				2	0.209 (0.2)	0.218 (28.2)	0.229 (92.6)	0.223 (87.7)	< 0.001
19	Large	7	0.1	0.1	0.044 (9.0)	0.047 (150.7)			< 0.001
20				0.5	0.055 (9.1)	0.061 (151.6)			< 0.001
21				2	0.097 (9.2)	0.105 (152.6)			< 0.001
22			15	0.1	0.045 (9.0)	0.047 (138.8)			< 0.001
23				0.5	0.057 (9.1)	0.061 (140.8)			< 0.001
24				2	0.100 (9.2)	0.106 (151.5)			< 0.001

The smallest overall MSPE for each case is highlighted in bold. The numbers in parentheses represent the average computational time in seconds over the corresponding estimated models and replications. The last column shows p values from two-sided paired t tests comparing the overall MSPEs between the OKFD and the best Sp.T. model

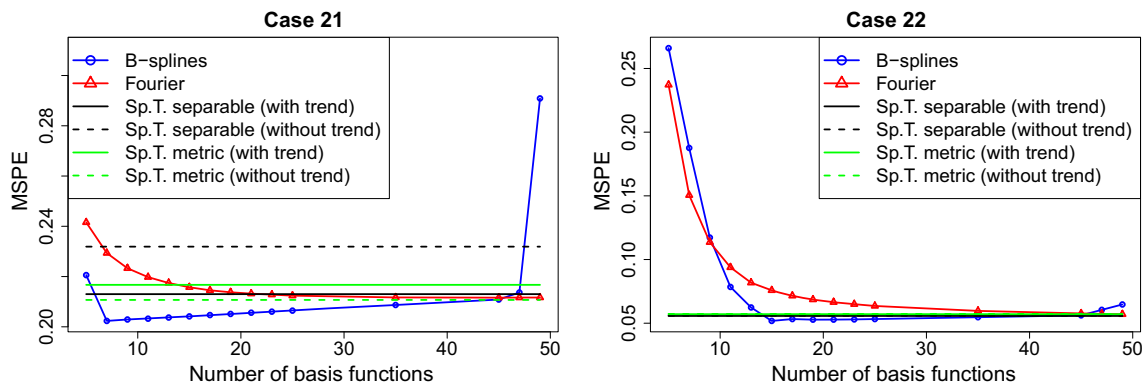


Fig. 4 Prediction performance (minimum MSPE over the three trace-semivariogram models, averaged over the 100 realizations) for cases 21 and 22 considering medium sample sizes when the estimated OKFD model is based on different numbers (p) of basis functions, being both Fourier and cubic B-spline. The solid and dashed black

lines represent the corresponding overall MSPE of the Sp.T. separable model with and without an estimated deterministic time trend, respectively. The solid and dashed green lines represent the corresponding overall MSPE of the Sp.T. metric model with and without an estimated deterministic time trend, respectively

models, here we observe this phenomena only when Fourier basis functions are used in the fitted OKFD models. For B-splines, the best prediction performance is (naturally) achieved using the same number of B-splines in the OKFD fitted models as used to generate the simulated data set (7 for case 21 and 15 for case 22). In fact, using too many B-splines may give substantially poorer predictions, especially when the spatial dependence is weak, as for case 21, cf. Fig. 4. It can also be noted that the best OKFD model using B-splines has significantly smaller MSPE than the best OKFD model using Fourier basis functions. If the simulated data sets would have been generated by a set of Fourier basis functions instead, we would most likely see the opposite behavior, i.e. that the same Fourier basis functions in the fitted OKFD model as in the data generation model probably would give the best prediction performance, and do better than the OKFD models using B-splines.

For the Sp.T. separable kriging models, it turned out that it was advantageous to use universal kriging, especially for the cases with weak spatial dependence, whereas the prediction performance was about the same for cases with strong spatial dependence (Fig. 4). For the Sp.T. metric model, we observed the opposite behavior, i.e., for cases with weak spatial dependence it was more advantageous to use ordinary kriging instead of universal kriging.

We have used FCV to evaluate the prediction performance, where predicted values are compared with the observed values of the process, which include measurement errors. In order to investigate whether comparisons with the true values would change the conclusions, we also computed the MSPEs with respect to the true realizations for cases 19–24 using small sample sizes, see Table 12 in Appendix 2. The MSPEs for the different cases in Table 12 are approximately 0.04 less than the corresponding MSPEs

based on the observed processes, presented in Table 4. The difference (0.04) corresponds to the variance of the white noise measurement errors. Furthermore, according to the p values in Tables 4 and 12, the results are consistent. Hence it illustrates that conclusions based on comparing predicted values with the observed values are consistent with those comparing with true values.

5 An application: temperature measurements in Canada

In this section we compare the prediction performance of the OKFD and the Sp.T. kriging models for a meteorological data set, available in the R-package `geofd` (Giraldo et al. 2012). The data consists of temperature measurements recorded at $n = 36$ weather stations at Canada's Atlantic coast in the Maritime Provinces (Fig. 5, top panel). At each station, the daily mean temperature averaged over the period 1960–1994 (February 29th combined with February 28th) has been recorded. The resulting functional data are displayed in Fig. 5 (bottom panel), connected by light grey lines. All the data except the measurements from Moncton are used to estimate the different models. This place is left as new location used to compare the predictions using the best OKFD and Sp.T. kriging models.

Using the R-package `geofd`, the OKFD model was first estimated using 51, 101, 151, 201, 251, 301 and 351 Fourier basis functions. For each number of Fourier basis functions three semivariogram models (exponential, spherical and stable) were fitted to the empirical trace-semivariogram by the OLS method. Thus, in total we estimated $7 \times 3 = 21$ OKFD models. Predictions were then made and evaluated by FCV in terms of their MSPEs (19).

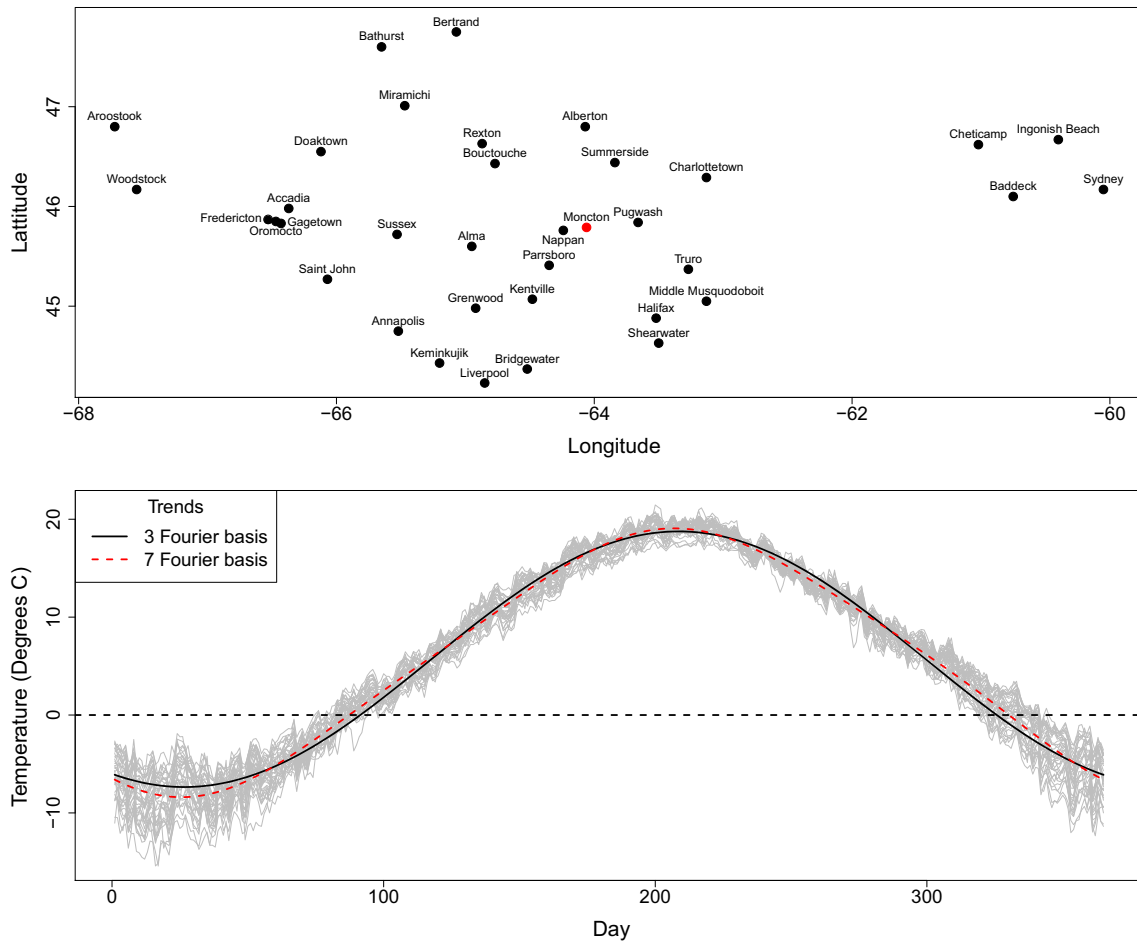


Fig. 5 Locations of the 36 weather stations in the Canadian Maritime provinces (top panel) where the average (over 30 years) daily temperature curves (bottom panel) were registered. The bottom panel

also presents the estimated common time trend specified as linear combinations of the first 3 and 7 Fourier basis functions, respectively

The best prediction performance was achieved using the stable trace-semivariogram (Fig. 6, left panel) for all considered numbers of Fourier basis. Figure 6 (right panel) clearly reveals that the prediction error (minimum MSPE over the three trace-semivariogram models) decreases with the number

of basis functions used in the fitted OKFD models. Thus, the best performance was attained with 351 Fourier basis functions and its MSPE was 0.5738. The average computational time for an estimated OKFD model based on 51 and 351 Fourier basis functions was less than one and three seconds, respectively.

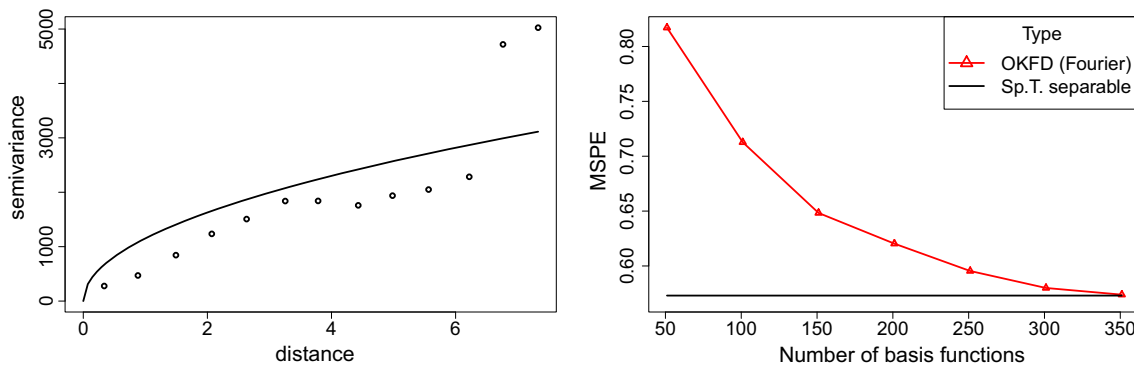


Fig. 6 Left panel: Empirical trace-semivariogram and the best fitted stable model for the Canadian temperature curves, represented by 351 Fourier basis functions. Right panel: Minimum MSPE over the three

trace-semivariogram models for OKFD, based on different numbers of Fourier basis functions. The solid black line represents the MSPE of the best Sp.T. model

The data was further predicted using Sp.T. kriging. Since the data show a clear time trend, universal Sp.T. kriging was first applied. The deterministic time trend was modeled by a linear combination of the 3 (and 7) first Fourier basis functions, and estimated by the OLS method. The dependence structure of the resulting residuals was then estimated by fitting Sp.T. second-order stationary and isotropic semivariogram models to the empirical Sp.T. semivariogram of the residuals. The Sp.T. semivariogram models (separable, product-sum and metric) described in Sect. 3.2 were estimated, letting their corresponding spatial, temporal and joint semivariogram models be altered between the exponential, spherical and stable semivariogram models. This resulted in 9 separable, 9 product-sum and 3 metric Sp.T. semivariogram models. As a comparison we also predicted the original data by Sp.T. ordinary kriging, using the same Sp.T. semivariogram models as for the universal Sp.T. kriging models. Thus, in total we investigated $(9 + 9 + 3) \times 3 = 63$ Sp.T. kriging models. All models were fitted to the data and predictions evaluated by FCV.

Table 5 presents the best (smallest MSPE) Sp.T. models, within each of the three groups of dependence structure (separable, product-sum and metric), with and without an estimated trend. The numbers in brackets report the corresponding average computational time in seconds over the estimated models. Many of the Sp.T. models have about the same prediction performance, with the exceptions of the Sp.T. metric models with estimated trend, which worked less well. The best Sp.T. models have approximately the same magnitude of MSPE as the best OKFD model (MSPE being 0.5738), but in terms of computational time, an OKFD model (taking 1–3 s to compute) was 100–10,000 times faster to compute compared to a Sp.T. kriging model.

Figure 7 presents the observed daily temperatures at locations Bertrand (the location with the largest prediction error) and Moncton (new location), together with the corresponding predicted values using the best OKFD and Sp.T. kriging models. It emphasizes that there are very small differences between the best OKFD and Sp.T. models in terms of prediction performance.

This data set has previously been analyzed by e.g. Giraldo (2009) with the objective to demonstrate and compare the functional kriging methods OKFD, PWFK and FKTM.

Giraldo (2009) concluded that the three methods have similar FCV prediction performance when the first 65 Fourier basis functions are used in (8) to represent the $\chi_{s_i}(t)$'s. In Menafoglio et al. (2013) this data set was used to investigate the effect of using universal kriging for functional data (UKFD) instead of OKFD, also by representing the functional data with the first 65 Fourier basis functions. Geodesic distance instead of the Euclidean distance was used to take into account the approximately spherical shape of the Earth. They concluded that UKFD performed better in terms of FCV compared to OKFD. The FCV performance was there computed with respect to the fitted data, thus differing from ours, where raw data has been used.

6 Concluding remarks

In this paper we have presented and compared functional and Sp.T. kriging approaches to predict spatial functional random processes. Comparisons with respect to prediction performance and computational time have been performed, mainly through a simulation study but also using a real data set. We restricted the comparison to Sp.T. kriging versus the functional kriging method OKFD, since the more flexible functional kriging approaches PWFK and FKTM coincide with OKFD in several situations (Sects. 3.1.3 and 3.1.4). Here we also contribute with new knowledge by proving that the kriging weights of OKFD and PWFK coincide under certain conditions, e.g. for all stationary spatial functional random processes, but also for more general processes, cf. Sect. 3.1.3.

The purpose of this study has been to bring light on the relative merits of functional and Sp.T. kriging methods for prediction of spatial functional random processes. While functional kriging predicts complete curves on a given (time) domain, given observations on the same domain, the Sp.T. kriging methods make (a raster of) pointwise predictions of the curves and are not restricted to a given (time) domain. For non-stationary Sp.T. (but stationary functional) processes, e.g. under the presence of a common deterministic time trend and/or time varying variances and dependence structure, functional kriging does not demand any extra modeling, whereas identification and modeling of

Table 5 Prediction performance of different Sp.T. kriging models for the Canadian weather data

Trend	MSPE		
	Sp.T. separable	Sp.T. product-sum	Sp.T. metric
No trend	0.5730 (1.8×10^2)	0.5861 (1.3×10^4)	0.5730 (1.3×10^4)
3 Fourier basis	0.5730 (1.8×10^2)	0.5731 (1.3×10^4)	1.1126 (1.4×10^4)
7 Fourier basis	0.5734 (1.6×10^2)	0.5731 (1.3×10^4)	1.0670 (1.4×10^4)

For each type of trend and Sp.T. variogram model, the (minimum) MSPE is reported. The numbers in parentheses represent the average computational time in seconds over the corresponding estimated models

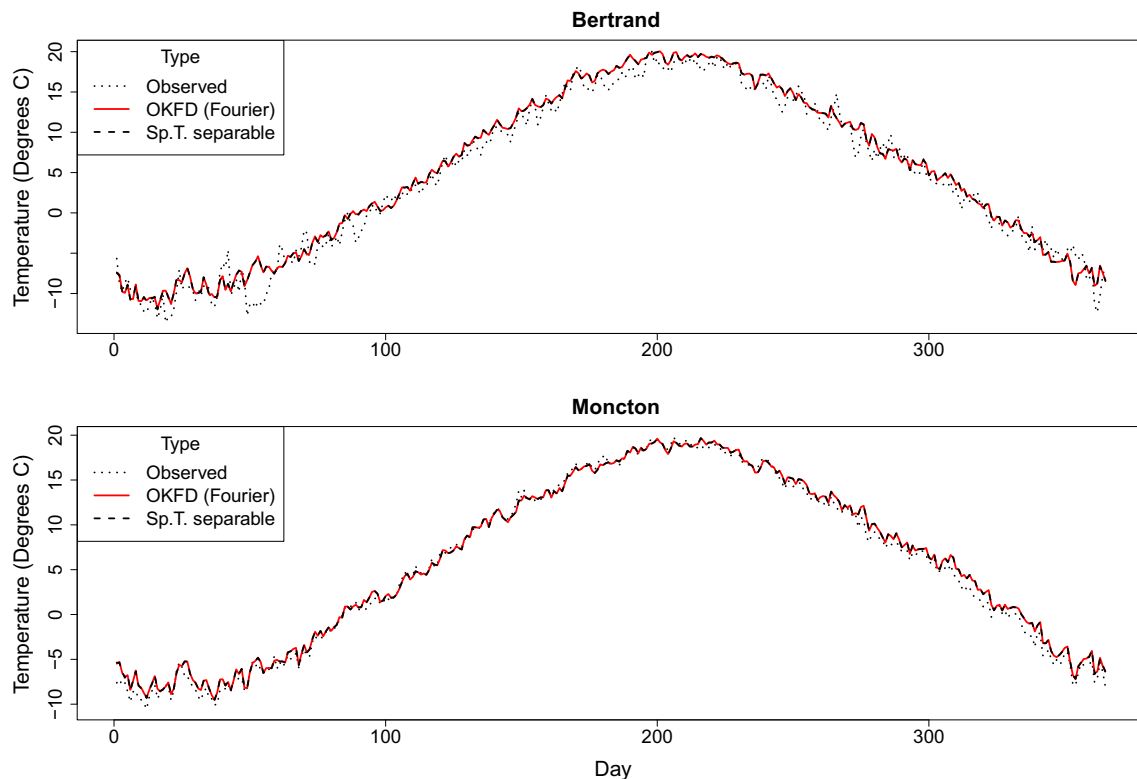


Fig. 7 Predicted temperatures at locations Bertrand (top) and Moncton (bottom) obtained by the best OKFD model (solid grey line) and the best Sp.T. model (dashed black line) together with the observed (dotted) values

trend and/or time varying dependence is necessary for Sp.T. kriging. From a modeler perspective, the Sp.T. kriging methods demand more work with a larger risk of choosing a suboptimal model. It should also be pointed out that the functional approach has the possibility to consider other embeddings for the data, i.e. other geometries than L2. In fact, in some cases, the functional approach may completely outperform Sp.T. kriging just because it can account for other data features, such as differential properties in the data or data constraints.

Based on the simulation study and the analysis of the data set, we observed that the prediction performance of OKFD normally was improved when the number of basis functions used to represent the functional data increased. Furthermore, for all considered cases, OKFD was computationally considerably faster than the Sp.T. kriging models. The large matrices that need to be inverted in order to perform Sp.T. kriging prediction at each location, is the major reason for this fact. One way to reduce the computational time for the Sp.T. kriging models could be to use only the local neighborhood (e.g. the k closest neighboring locations) when prediction is made. This can often be done without much loss in prediction performance. Computational time might also be saved if estimation methods based on distributional assumptions are used, both for the functional and Sp.T. kriging methods.

Experience from this study concludes that with respect to prediction performance, OKFD typically performed similarly or better than the Sp.T. kriging models for small and medium sample sizes. This is likely due to the more complex task of finding good estimates of the Sp.T. variogram compared to the trace-variograms used in OKFD, since trace-variograms have one dimension less. The large number of choices of Sp.T. variogram models and parameters to estimate makes the Sp.T. estimation process more vulnerable, especially for small data sets. For larger sample sizes, the Sp.T. kriging starts to perform better for the stationary Sp.T. processes, whereas OKFD continues to work best for the non-stationary Sp.T. (but stationary functional) processes. We also noted a clear tendency for OKFD to perform better relative to Sp.T. kriging, the stronger the temporal- and the weaker the spatial dependence considered.

An interesting extension of this work would be to develop and compare both parametric (relying on distributional assumptions) and non-parametric functional and Sp.T. kriging methods that deals with non-stationary functional processes. In this study we have focused on non-parametric functional and Sp.T. kriging approaches for prediction of stationary spatial functional random processes. The results highlight that OKFD is a good candidate in most situations for prediction of stationary spatial functional random processes, not only for its prediction performance but also for its speed and ease to use.

Acknowledgements Open access funding provided by Umea University. This work was supported by the Swedish Research Council (Project id 340-2013-5203) and J.Mateu has been partially funded by Grants MTM2016-78917-R from the Spanish Ministry of Science, and P1-1B2015-40 from University Jaume I.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix 1: Proofs

Below we present the proof of Proposition 3.1. However, as the proposition relies on Lemma 1.1 below, we start by stating and proving the lemma before presenting the proof of the proposition.

Lemma 1.1 Consider the following square matrix

$$Q = \begin{pmatrix} \sigma_{11}W & \sigma_{12}W & \cdots & \sigma_{1n}W & \mathbf{I} \\ \sigma_{21}W & \sigma_{22}W & \cdots & \sigma_{2n}W & \mathbf{I} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{n1}W & \sigma_{n2}W & \cdots & \sigma_{nn}W & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \cdots & \mathbf{I} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \Sigma \otimes W & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{pmatrix}_{k(n+1) \times k(n+1)}, \tag{22}$$

where \otimes denotes kronecker product,

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

W is a square $k \times k$ matrix, \mathbf{I} is a $k \times k$ identity matrix, $\mathbf{0}$ is a $k \times k$ zero matrix, and the σ_{ij} 's are real values. If the inverses of Σ and W exist, and the sum of the n^2 elements of Σ^{-1} is non-zero, then the inverse of Q exists and satisfies

$$Q^{-1} = \begin{pmatrix} k_{11}W^{-1} & k_{12}W^{-1} & \cdots & k_{1n}W^{-1} & k_1\mathbf{I} \\ k_{21}W^{-1} & k_{22}W^{-1} & \cdots & k_{2n}W^{-1} & k_2\mathbf{I} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ k_{n1}W^{-1} & k_{n2}W^{-1} & \cdots & k_{nn}W^{-1} & k_n\mathbf{I} \\ l_1\mathbf{I} & l_2\mathbf{I} & \cdots & l_n\mathbf{I} & cW \end{pmatrix}_{k(n+1) \times k(n+1)}, \tag{23}$$

where the k_{ij} 's, k_i 's, l_i 's and c are constants, such that $\sum_{i=1}^n k_{ij}(= \sum_{j=1}^n k_{ij}) = 0$ for all j 's (and i 's), $\sum_{i=1}^n k_i = 1$ and $\sum_{i=1}^n l_i = 1$.

Note that the k_{ij} 's, k_i 's and the l_i 's also change with n . However, for notational simplicity we suppress the dependence of n in the k_{ij} 's, k_i 's and the l_i 's.

Proof of Lemma 1.1 The proof is based on the following block matrix inversion formula

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{pmatrix}, \tag{24}$$

where \mathbf{A} and \mathbf{D} are square matrices allowed to be of different size. Let

$$Q = \left(\begin{array}{c|c} \Sigma \otimes W & \mathbf{I} \\ \hline \mathbf{I} & \mathbf{0} \end{array} \right)_{k(n+1) \times k(n+1)} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix},$$

the elements of Σ^{-1} be denoted by

$$\Sigma^{-1} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nn} \end{pmatrix}_{n \times n},$$

and further let $\alpha_{\bullet j} = \sum_{i=1}^n \alpha_{ij}$, $\alpha_{i\bullet} = \sum_{j=1}^n \alpha_{ij}$ and $\alpha_{\bullet\bullet} = \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij}$. Then, by (24), together with the fact that $(\Sigma \otimes W)^{-1} = \Sigma^{-1} \otimes W^{-1}$, straightforward calculations confirm that

$$(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} = \frac{1}{\alpha_{\bullet\bullet}}W, \tag{25}$$

$$-\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} = \begin{pmatrix} \frac{\alpha_{1\bullet}}{\alpha_{\bullet\bullet}}\mathbf{I} \\ \vdots \\ \frac{\alpha_{n\bullet}}{\alpha_{\bullet\bullet}}\mathbf{I} \end{pmatrix}, \tag{26}$$

$$-(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} = \begin{pmatrix} \frac{\alpha_{\bullet 1}}{\alpha_{\bullet\bullet}}\mathbf{I} & \cdots & \frac{\alpha_{\bullet n}}{\alpha_{\bullet\bullet}}\mathbf{I} \end{pmatrix}, \tag{27}$$

and

$$\begin{aligned}
 & \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} \\
 &= \begin{pmatrix} \left(\alpha_{11} - \frac{\alpha_{\bullet 1}\alpha_{1\bullet}}{\alpha_{\bullet\bullet}}\right)\mathbf{W}^{-1} & \left(\alpha_{12} - \frac{\alpha_{\bullet 2}\alpha_{1\bullet}}{\alpha_{\bullet\bullet}}\right)\mathbf{W}^{-1} & \cdots & \left(\alpha_{1n} - \frac{\alpha_{\bullet n}\alpha_{1\bullet}}{\alpha_{\bullet\bullet}}\right)\mathbf{W}^{-1} \\ \left(\alpha_{21} - \frac{\alpha_{\bullet 1}\alpha_{2\bullet}}{\alpha_{\bullet\bullet}}\right)\mathbf{W}^{-1} & \left(\alpha_{22} - \frac{\alpha_{\bullet 2}\alpha_{2\bullet}}{\alpha_{\bullet\bullet}}\right)\mathbf{W}^{-1} & \cdots & \left(\alpha_{2n} - \frac{\alpha_{\bullet n}\alpha_{2\bullet}}{\alpha_{\bullet\bullet}}\right)\mathbf{W}^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ \left(\alpha_{n1} - \frac{\alpha_{\bullet 1}\alpha_{n\bullet}}{\alpha_{\bullet\bullet}}\right)\mathbf{W}^{-1} & \left(\alpha_{n2} - \frac{\alpha_{\bullet 2}\alpha_{n\bullet}}{\alpha_{\bullet\bullet}}\right)\mathbf{W}^{-1} & \cdots & \left(\alpha_{nn} - \frac{\alpha_{\bullet n}\alpha_{n\bullet}}{\alpha_{\bullet\bullet}}\right)\mathbf{W}^{-1} \end{pmatrix}. \tag{28}
 \end{aligned}$$

Combining (25)–(28), we have that \mathbf{Q}^{-1} is of the same form as in (23) with $k_{ij} = \alpha_{ij} - \frac{\alpha_{\bullet j}\alpha_{i\bullet}}{\alpha_{\bullet\bullet}}$, $k_i = \frac{\alpha_{i\bullet}}{\alpha_{\bullet\bullet}}$ and $l_i = \frac{\alpha_{\bullet i}}{\alpha_{\bullet\bullet}}$. In addition, it is easy to verify that

$$\begin{aligned}
 \sum_{i=1}^n k_i &= \sum_{i=1}^n \frac{\alpha_{i\bullet}}{\alpha_{\bullet\bullet}} = 1, \\
 \sum_{i=1}^n k_{ij} &= \sum_{i=1}^n \left(\alpha_{ij} - \frac{\alpha_{\bullet j}\alpha_{i\bullet}}{\alpha_{\bullet\bullet}}\right) = \alpha_{\bullet j} - \frac{\alpha_{\bullet j}\alpha_{\bullet\bullet}}{\alpha_{\bullet\bullet}} = 0,
 \end{aligned}$$

for all $j = 1, \dots, n$. Similarly it follows that $\sum_{i=1}^n l_i = 1$ and $\sum_{j=1}^n k_{ij} = 0$ for all $i = 1, \dots, n$, which completes the proof. \square

Proof of Proposition 3.1 Under the assumptions of Proposition 3.1 we here show that the coefficients of the functional kriging weights (9) of PWFK, which are obtained by minimizing (3) subject to the unbiasedness constraint of the predictor ($\sum_{i=1}^n \lambda_i(t) = 1$, for all $t \in T$), yield weights that are constant over time, i.e., $\lambda_i(t) = \mathbf{b}_i^T \mathbf{B}_\lambda(t) = \lambda_i$, $i = 1, \dots, n$. Giraldo et al. (2010) showed that the solution of the PWFK optimization problem is given by the solution to the system

$$\mathbf{Q}\boldsymbol{\beta} = \mathbf{J} \Rightarrow \boldsymbol{\beta} = \mathbf{Q}^{-1}\mathbf{J} \tag{29}$$

where

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_1 & \mathbf{Q}_{12} & \cdots & \mathbf{Q}_{1n} & \mathbf{I} \\ \mathbf{Q}_{21} & \mathbf{Q}_2 & \cdots & \mathbf{Q}_{2n} & \mathbf{I} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{Q}_{n1} & \mathbf{Q}_{n2} & \cdots & \mathbf{Q}_n & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \cdots & \mathbf{I} & \mathbf{0} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \\ \mathbf{m} \end{pmatrix}, \mathbf{J} = \begin{pmatrix} \mathbf{J}_1 \\ \mathbf{J}_2 \\ \vdots \\ \mathbf{J}_n \\ \mathbf{c} \end{pmatrix},$$

$$\mathbf{Q}_i = \int_T \mathbf{B}_\lambda(t) \text{Var}[\chi_{s_i}(t)] \mathbf{B}_\lambda^T(t) dt, \tag{30}$$

$$\mathbf{Q}_{ij} = \int_T \mathbf{B}_\lambda(t) \text{Cov}[\chi_{s_i}(t), \chi_{s_j}(t)] \mathbf{B}_\lambda^T(t) dt, \tag{31}$$

$$\mathbf{J}_i = \int_T \mathbf{B}_\lambda(t) \text{Cov}[\chi_{s_0}(t), \chi_{s_i}(t)] dt, \tag{32}$$

for $i, j = 1, \dots, n$, $\mathbf{m}^T = (m_1, \dots, m_K)$ are the K Lagrangian multipliers and \mathbf{c} is an unbiasedness constraint vector satisfying $\mathbf{c}^T \mathbf{B}_\lambda(t) = 1$.

From (11) we have that $\text{Var}[\chi_{s_i}(t)] = w^2(t)\sigma_{ii}$ and $\text{Cov}[\chi_{s_i}(t), \chi_{s_j}(t)] = w^2(t)\sigma_{ij}$ for all $i, j = 0, 1, \dots, n$. The expressions in (30) and (31) can thus be expressed as $\mathbf{Q}_i = \sigma_{ii}\mathbf{W}$ and $\mathbf{Q}_{ij} = \sigma_{ij}\mathbf{W}$ for $i, j = 1, \dots, n$, where $\mathbf{W} = \int_T w^2(t)\mathbf{B}_\lambda(t)\mathbf{B}_\lambda^T(t) dt$, and hence, the matrix \mathbf{Q} is of the form (22). By the assumptions in Proposition 3.1 and from Lemma 1.1 it follows that \mathbf{Q}^{-1} is of the form (23). Computing the solution $\boldsymbol{\beta} = \mathbf{Q}^{-1}\mathbf{J}$ yields that the \mathbf{b}_i 's are of the form

$$\mathbf{b}_i = \sum_{j=1}^n k_{ij}\mathbf{W}^{-1}\mathbf{J}_j + k_i\mathbf{c}, i = 1, \dots, n,$$

which can be rewritten as

$$\mathbf{W}(\mathbf{b}_i - k_i\mathbf{c}) = \sum_{j=1}^n k_{ij}\mathbf{J}_j, i = 1, \dots, n. \tag{33}$$

Thus, using (32) together with the expression for \mathbf{W} we may express (33) as

$$\int_T w^2(t)\mathbf{B}_\lambda(t)\mathbf{B}_\lambda^T(t)(\mathbf{b}_i - k_i\mathbf{c}) dt = \sum_{j=1}^n k_{ij}\sigma_{oj} \int_T w^2(t)\mathbf{B}_\lambda(t) dt$$

or equivalently,

$$\int_T w^2(t) \mathbf{B}_\lambda(t) \mathbf{B}_\lambda^T(t) \left(\mathbf{b}_i - \left(k_i + \sum_{j=1}^n k_{ij} \sigma_{0j} \right) \mathbf{c} \right) dt = \mathbf{0}, i = 1, \dots, n.$$

In the last expression we used the fact that $\mathbf{B}_\lambda^T(t) \mathbf{c}$ equals 1 for all values of $t \in T$ (the unbiasedness constraint). It can now be seen that the above equation holds if \mathbf{b}_i equals $(k_i + \sum_{j=1}^n k_{ij} \sigma_{0j}) \mathbf{c}$ and from the assumption of \mathbf{Q}^{-1} 's existence we also have that this solution is the only one. Thus, the weights satisfy

$$\lambda_i(t) = \mathbf{b}_i^T \mathbf{B}_\lambda(t) = \left(k_i + \sum_{j=1}^n k_{ij} \sigma_{0j} \right) \mathbf{c}^T \mathbf{B}_\lambda(t) = k_i + \sum_{j=1}^n k_{ij} \sigma_{0j} = \lambda_i,$$

for all values of $t \in T$ and $i = 1, \dots, n$, and consequently, the PWFK predictor coincides with the OKFD predictor. \square

Appendix 2: Tables of the simulation study

Tables 6, 7, 8 and 9 contain information about the prediction performance in terms of mean squared prediction errors (MSPEs) for the cases 1–18 for small and large sample sizes with and without a deterministic trend, respectively. The smallest overall MSPE for each case is highlighted in bold. The numbers in parentheses represent the average computational time in seconds over the corresponding estimated models and replications. The last column shows p values from two-sided paired t tests comparing the overall MSPEs between the OKFD and the Sp.T. separable models.

Tables 10 and 11 contain information about the number of basis functions used in the OKFD models concerning the stationary (corresponding to isotropic Sp.T. processes with a separable and non-separable covariance function) and non-stationary cases and the three different sample sizes. The (p) cubic B-splines were constructed based on $p - 4$ equally distributed interior knots on the interval $[0, 1]$.

Table 6 Simulated cases 1–9 (and 1*–9*) without (and with) a deterministic time trend considering small sample sizes

Generated data				Overall MSPE				Comparison
Case	Type	α	β	OKFD	Sp.T. separable	Sp.T. product-sum	Sp.T. metric	p value
1	Separable, no trend	0.1	0.1	0.068 (0.2)	0.068 (7.1)	0.070 (9.8)	0.106 (6.1)	0.176
2			1	0.065 (0.2)	0.065 (8.4)	0.069 (10.8)	0.080 (6.3)	0.561
3			10	0.063 (0.2)	0.064 (6.2)	0.065 (14.6)	0.070 (6.3)	0.409
4		0.5	0.1	0.135 (0.2)	0.145 (7.0)	0.150 (12.7)	0.204 (6.4)	< 0.001
5			1	0.135 (0.2)	0.139 (8.6)	0.147 (14.6)	0.196 (6.1)	< 0.001
6			10	0.134 (0.2)	0.139 (7.3)	0.154 (17.3)	0.154 (6.3)	0.204
7		2	0.1	0.377 (0.2)	0.400 (6.2)	0.395 (15.2)	0.452 (6.2)	< 0.001
8			1	0.356 (0.2)	0.386 (8.0)	0.399 (16.0)	0.476 (6.2)	< 0.001
9			10	0.365 (0.2)	0.378 (7.8)	0.436 (17.6)	0.421 (6.2)	< 0.001
1*	Separable, trend	0.1	0.1	0.066 (0.2)	0.066 (9.3)	0.068 (12.6)	0.072 (6.2)	0.107
2*			1	0.063 (0.2)	0.063 (8.8)	0.066 (11.9)	0.070 (6.1)	0.004
3*			10	0.066 (0.2)	0.066 (6.7)	0.068 (15.5)	0.066 (6.3)	0.441
4*		0.5	0.1	0.128 (0.2)	0.134 (9.7)	0.138 (14.6)	0.149 (6.4)	< 0.001
5*			1	0.135 (0.2)	0.140 (9.9)	0.148 (14.5)	0.146 (5.9)	0.029
6*			10	0.137 (0.2)	0.138 (7.8)	0.159 (17.3)	0.146 (6.4)	0.022
7*		2	0.1	0.380 (0.2)	0.398 (8.2)	0.413 (15.5)	0.457 (6.2)	< 0.001
8*			1	0.377 (0.2)	0.400 (9.4)	0.422 (15.6)	0.430 (5.9)	< 0.001
9*			10	0.387 (0.2)	0.405 (8.1)	0.442 (17.6)	0.415 (6.2)	< 0.001

Table 7 Simulated cases 1–9 (and 1*–9*) without (and with) a deterministic time trend considering large sample sizes

Generated data				Overall MSPE		Comparison
Case	Type	α	β	OKFD	Sp.T. separable	<i>p</i> value
1	Separable, no trend	0.1	0.1	0.051 (8.7)	0.050 (147.9)	< 0.001
2			1	0.055 (8.6)	0.053 (139.0)	< 0.001
3			10	0.054 (8.6)	0.051 (125.8)	< 0.001
4		0.5	0.1	0.078 (9.9)	0.080 (153.0)	0.029
5			1	0.079 (9.9)	0.078 (154.5)	0.064
6			10	0.081 (10.0)	0.079 (139.9)	< 0.001
7		2	0.1	0.164 (9.6)	0.168 (146.4)	0.207
8			1	0.160 (9.7)	0.161 (147.9)	0.147
9			10	0.167 (9.7)	0.167 (141.3)	0.846
1*	Separable, trend	0.1	0.1	0.054 (8.7)	0.053 (151.7)	< 0.001
2*			1	0.053 (8.7)	0.051 (143.9)	< 0.001
3*			10	0.054 (8.7)	0.051 (131.9)	< 0.001
4*		0.5	0.1	0.074 (9.9)	0.076 (158.6)	0.016
5*			1	0.077 (9.9)	0.077 (159.5)	0.465
6*			10	0.080 (9.9)	0.077 (146.3)	< 0.001
7*		2	0.1	0.156 (9.6)	0.164 (150.8)	0.151
8*			1	0.163 (9.7)	0.169 (150.5)	0.104
9*			10	0.165 (10.0)	0.166 (153.1)	0.479

Table 8 Simulated cases 10–18 (and 10*–18*) without (and with) a deterministic time trend considering small sample sizes

Generated data				Overall MSPE				Comparison
Case	Type	α	β	OKFD	Sp.T. separable	Sp.T. product-sum	Sp.T. metric	<i>p</i> value
10	Non separable, no trend	0.1	0.1	0.063 (0.2)	0.063 (6.6)	0.067 (9.0)	0.105 (6.1)	0.766
11			1	0.063 (0.2)	0.063 (8.5)	0.066 (9.6)	0.100 (6.3)	0.755
12			10	0.063 (0.2)	0.065 (6.2)	0.068 (10.4)	0.090 (6.2)	0.013
13		0.5	0.1	0.132 (0.2)	0.144 (6.3)	0.153 (13.0)	0.199 (6.3)	< 0.001
14			1	0.132 (0.2)	0.140 (8.2)	0.150 (14.1)	0.216 (5.9)	< 0.001
15			10	0.139 (0.2)	0.144 (9.0)	0.157 (16.6)	0.195 (6.4)	< 0.001
16		2	0.1	0.363 (0.2)	0.398 (5.9)	0.379 (14.9)	0.430 (6.1)	< 0.001
17			1	0.365 (0.2)	0.422 (7.1)	0.396 (16.0)	0.481 (6.2)	< 0.001
18			10	0.362 (0.2)	0.387 (8.2)	0.385 (16.8)	0.532 (6.5)	< 0.001
10*	Non separable, trend	0.1	0.1	0.063 (0.2)	0.063 (9.6)	0.064 (12.7)	0.066 (6.3)	0.598
11*			1	0.065 (0.2)	0.064 (9.0)	0.066 (11.1)	0.066 (6.2)	0.009
12*			10	0.066 (0.2)	0.065 (6.7)	0.071 (13.2)	0.069 (6.3)	0.049
13*		0.5	0.1	0.135 (0.2)	0.139 (9.7)	0.148 (14.5)	0.153 (6.5)	< 0.001
14*			1	0.133 (0.2)	0.136 (10.0)	0.146 (14.2)	0.153 (5.9)	< 0.001
15*			10	0.134 (0.2)	0.136 (8.9)	0.148 (16.6)	0.142 (6.2)	< 0.001
16*		2	0.1	0.362 (0.2)	0.375 (8.3)	0.384 (15.4)	0.447 (6.4)	< 0.001
17*			1	0.360 (0.2)	0.377 (9.4)	0.384 (15.9)	0.431 (6.1)	< 0.001
18*			10	0.377 (0.2)	0.402 (8.9)	0.412 (16.8)	0.423 (6.1)	< 0.001

Table 9 Simulated cases 10–18 (and 10*–18*) without (and with) a deterministic time trend considering large sample sizes

Generated data				Overall MSPE		Comparison
Case	Type	α	β	OKFD	Sp.T. separable	<i>p</i> value
10	Non separable, no trend	0.1	0.1	0.053 (8.7)	0.052 (150.1)	< 0.001
11			1	0.053 (8.7)	0.051 (146.5)	< 0.001
12			10	0.054 (8.6)	0.050 (134.6)	< 0.001
13		0.5	0.1	0.076 (9.3)	0.077 (156.4)	0.138
14			1	0.078 (9.6)	0.077 (163.3)	0.431
15			10	0.078 (10.3)	0.075 (177.5)	< 0.001
16		2	0.1	0.160 (9.7)	0.175 (145.6)	0.038
17			1	0.161 (9.7)	0.162 (147.7)	0.516
18			10	0.165 (9.7)	0.164 (148.3)	0.723
10*	Non separable, trend	0.1	0.1	0.053 (8.7)	0.051 (151.6)	< 0.001
11*			1	0.053 (8.7)	0.050 (150.0)	< 0.001
12*			10	0.055 (8.6)	0.051 (134.1)	< 0.001
13*		0.5	0.1	0.077 (10.3)	0.078 (190.2)	0.051
14*			1	0.077 (9.9)	0.076 (159.8)	0.288
15*			10	0.079 (10.0)	0.077 (155.8)	< 0.001
16*		2	0.1	0.163 (9.7)	0.166 (150.7)	0.029
17*			1	0.163 (9.7)	0.167 (152.6)	0.186
18*			10	0.163 (9.7)	0.162 (152.5)	0.213

Table 10 Stationary

	Sample sizes		
	Small	Medium	Large
Fourier	5, 7, 9, 11	5, 15, 25, 35, 45, 47, 49	5, 15, 25, 35, 45, 47, 49
B-splines	5, 6, 7, 8, 9, 10, 11, 12	5, 15, 25, 35, 45, 47, 49	5, 15, 25, 35, 45, 47, 49
Total	12	14	14

Table 11 Non-stationary

	Sample sizes		
	Small	Medium	Large
Fourier	5, 7, 9, 11	5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 35, 45, 47, 49	5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 35, 45, 47, 49
B-splines	5, 6, 7, 8, 9, 10, 11, 12	5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 35, 45, 47, 49	5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 35, 45, 47, 49
Total	12	30	30

Table 12 Prediction performance in terms of mean squared prediction errors (MSPEs) for the cases 19–24 using small sample sizes

Generated data					Overall MSPE		Comparison
Case	Type	Data size	p	α	OKFD	Sp.T. separable	p value
19	Non-stationary	Small	7	0.1	0.014	0.017	< 0.001
20				0.5	0.055	0.059	< 0.001
21				2	0.187	0.193	< 0.001
22			15	0.1	0.018	0.017	< 0.001
23				0.5	0.059	0.059	0.156
24				2	0.199	0.203	0.062

The comparison in terms of MSPE have been made with respect to the true realizations. The smallest overall MSPE for each case is highlighted in bold. The last column shows p values from two-sided paired t tests comparing the overall MSPEs between the OKFD and the Sp.T. separable models

References

- Caballero W, Giraldo R, Mateu J (2013) A universal kriging approach for spatial functional data. *Stoch Environ Res Risk Assess* 27(7):1553–1563
- Chilès J-P, Delfiner P (2012) *Geostatistics: modeling spatial uncertainty*, 2nd edn. Wiley, Hoboken
- Cressie N (2015) *Statistics for spatial data*. Wiley, Hoboken
- Cressie N, Wikle CK (2015) *Statistics for spatio-temporal data*. Wiley, Hoboken
- Delicado P, Giraldo R, Comas C, Mateu J (2010) Statistics for spatial functional data: some recent contributions. *Environmetrics* 21(3–4):224–239
- Giraldo R (2009) *Geostatistical analysis of functional data*. Ph.D. thesis, Universitat Politècnica da Catalunya, Barcelonaa
- Giraldo R (2014) Cokriging based on curves: prediction and estimation of the prediction variance. *InterStat* 2:1–30
- Giraldo R, Delicado P, Mateu J (2007) *Geostatistics for functional data: an ordinary kriging approach*. Technical report, Universitat Politècnica da Catalunya. <http://hdl.handle.net/2117/1099>
- Giraldo R, Delicado P, Mateu J (2008) Continuous time-varying kriging for spatial prediction of functional data: an environmental application. Technical report, Universitat Politècnica da Catalunya. <http://hdl.handle.net/2117/2167>
- Giraldo R, Delicado P, Mateu J (2010) Continuous time-varying kriging for spatial prediction of functional data: an environmental application. *J Agric Biol Environ Stat* 15(1):66–82
- Giraldo R, Delicado P, Mateu J (2011) Ordinary kriging for function-valued spatial data. *Environ Ecol Stat* 18(3):411–426
- Giraldo R, Mateu J, Delicado P (2012) geofd: an R package for function-valued geostatistical prediction. *Revista Colombiana de Estadística* 35(3):385–407
- Goulard M, Voltz M (1992) Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. *Math Geol* 24(3):269–286
- Goulard M, Voltz M (1993) Geostatistical interpolation of curves: a case study in soil science. In: *Geostatistics Tróia'92*. Springer, pp 805–816
- Ignaccolo R, Mateu J, Giraldo R (2014) Kriging with external drift for functional data for air quality monitoring. *Stoch Environ Res Risk Assess* 28(5):1171–1186
- Menafoglio A, Petris G (2016) Kriging for Hilbert-space valued random fields: the operatorial point of view. *J Multivar Anal* 146:84–94
- Menafoglio A, Secchi P, Dalla Rosa M (2013) A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space. *Electron J Stat* 7:2209–2240
- Montero JM, Fernández-Avilés G, Mateu J (2015) *Spatial and spatio-temporal geostatistical modeling and kriging*, vol 998. Wiley, Hoboken
- Nerini D, Monestiez P, Manté C (2010) Cokriging for spatial functional data. *J Multivar Anal* 101(2):409–418
- Pebesma EJ (2004) Multivariable geostatistics in S: the gstat package. *Comput Geosci* 30(7):683–691
- Pebesma E et al (2012) spactime: spatio-temporal data in R. *J Stat Softw* 51(7):1–30
- Ramsay JO, Hooker G, Graves S (2009) *Functional data analysis with R and MATLAB*. Springer, Berlin
- Reyes A, Giraldo R, Mateu J (2015) Residual kriging for functional spatial prediction of salinity curves. *Commun Stat Theory Methods* 44(4):798–809
- Romano E, Mateu J, Giraldo R (2015) On the performance of two clustering methods for spatial functional data. *ASTA Adv Stat Anal* 99(4):467–492
- Schlather M, Malinowski A, Menck PJ, Oesting M, Strokorb K (2015) Analysis, simulation and prediction of multivariate random fields with package randomfields. *J Stat Softw* 63(8):1–25
- Sun Y, Genton MG (2012) Adjusted functional boxplots for spatio-temporal data visualization and outlier detection. *Environmetrics* 23(1):54–64

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.