HHS Public Access

Author manuscript

J Bus Econ Stat. Author manuscript; available in PMC 2023 January 01.

Published in final edited form as:

J Bus Econ Stat. 2022 ; 40(4): 1523–1537. doi:10.1080/07350015.2021.1938085.

Unified Principal Component Analysis for Sparse and Dense Functional Data under Spatial Dependency

Haozhe Zhang,

Microsoft Corporation, Redmond, United States

Yehua Li[*]

Department of Statistics, University of California, Riverside

Abstract

We consider spatially dependent functional data collected under a geostatistics setting, where locations are sampled from a spatial point process. The functional response is the sum of a spatially dependent functional effect and a spatially independent functional nugget effect. Observations on each function are made on discrete time points and contaminated with measurement errors. Under the assumption of spatial stationarity and isotropy, we propose a tensor product spline estimator for the spatio-temporal covariance function. When a coregionalization covariance structure is further assumed, we propose a new functional principal component analysis method that borrows information from neighboring functions. The proposed method also generates nonparametric estimators for the spatial covariance functions, which can be used for functional kriging. Under a unified framework for sparse and dense functional data, infill and increasing domain asymptotic paradigms, we develop the asymptotic convergence rates for the proposed estimators. Advantages of the proposed approach are demonstrated through simulation studies and two real data applications representing sparse and dense functional data, respectively.

Keywords

covariance estimation; dimension deduction; infill asymptotics; nugget effect; spatio-temporal; tensor product splines

1 Introduction

1.1 Literature review

Modern technology and data collection methods produce massive data with repeated measurements over time and space, thus give rise to functional data (Ramsay and Silverman, 2005; Horváth and Kokoszka, 2012; Kokoszka and Reimherr, 2017). In many applications, functional data collected at different times or locations are naturally correlated. There have been a lot of recent theory and methodology developments for dependent functional data,

[*] yehuali@ucr.edu .

Supplemental Materials

The online Supplementary Material contains detailed proofs of the theoretical results, additional figures and tables for the simulation studies and real data analysis, and the codes implementing the proposed methods.

including multi-level functional data (Crainiceanu et al., 2009; Xu et al., 2018), functional time series (Hörmann and Kokoszka, 2010; Aue et al., 2015), and spatially dependent functional data (Staicu et al., 2010; Zhou et al., 2010; Gromenko et al., 2012; Zhang et al., 2016; Kuenzer et al., 2020; Liang et al., 2020). There has also been some work on modeling spatio-temporal point process data using a functional data approach (Li and Guan, 2014).

Functional data are commonly viewed as infinite dimensional random vectors in a Hilbert space, and dimension reduction is crucial for visualization, interpretation and inference on these data (Hsing and Eubank, 2015). There has been a lot of methodological and theoretical developments on dimension reduction for independent data using the functional principal component analysis (FPCA) (Yao et al., 2005; Hall et al., 2006; Li and Hsing, 2010). The functional principal component scores are also widely used as predictors in linear or nonlinear regression models to predict other variables of interest (Cai and Hall, 2006; Wong et al., 2019).

There has also been some work on FPCA on spatially dependent functional data. Hörmann and Kokoszka (2013) provide some theoretical justification on spatial FPCA, assuming the functions are fully observed. In practice, however, functional data are often observed on discrete time points and the measurements are contaminated with errors. Based on the number of observations on each curve, functional data are traditionally classified as sparse functional data (Yao et al., 2005) and dense functional data (Hall et al., 2006). For independent functional data, it is known that the convergence rates for various functional estimators (such as the mean, covariance and principal components) are different under different sampling schemes. Wang et al. (2018) show that nonparametric hypothesis tests have different properties under sparse and dense functional data, in terms of asymptotic null distribution and power. However, sparse and dense functional data are asymptotic concepts, which are not clearly defined in any practical contexts. A lot of recent research efforts were focused on developing unified estimation and inference strategies for all types of functional data (Li and Hsing, 2010; Zhang and Wang, 2016; Wang et al., 2018). No such results yet exist for spatially dependent functional data.

There is also a large volume of recent literature on spatial and spatio-temporal data analysis which is relevant to our work, see the textbooks of Cressie (1993); Banerjee et al. (2004); Schabenberger and Gotway (2017); Cressie and Wikle (2015) for comprehensive accounts of this area. Some recent work also include Lu et al. (2009); Lu and Tjøstheim (2014); Kuusela and Stein (2018); Al-Sulami et al. (2017); AL-SULAMI et al. (2019); Jiang et al. (2020), just to name a few.

1.2 Motivating data examples

Our work is motivated by two real data examples from business applications, representing sparse and dense spatially dependent functional data, respectively.

Example 1: sparse functional data on London house price. The data are public records of home sales from the UK government website (https://www.gov.uk/government). The dataset includes all houses with at least 5 transactions between Jan 1, 1995 and Dec 31, 2018 in the Greater London Area. Each transaction record contains information on the price, date,

and property address. Exact locations, including longitudes and latitudes, of the houses are obtained by searches of the property addresses on Google Map API. The house locations are shown in Panel (a) of Figure 1.

The value of a house changes continuously over time, the trajectory of which we model as functional data. However, the value is measured by the market only when a sale is made, and the number of sale transactions per house ranges between 5 and 12. The house price trajectories are shown in Panel (b) of Figure 1. As we can see, the transaction times are sparse, irregular and house-specific.

Example 2: dense functional data from Zillow Real Estate. Zillow (https://www.zillow.com/research) publishes real estate data for research purposes for all major cities in the US. Our variable of interest is the "home price-to-rent ratio", defined as the ratio of residential real estate price to the annual rent, which has attracted broad interests in economics and social sciences (Campbell et al., 2009; Kishor and Morley, 2015). It has strong relationships with market fundamentals, and has been widely used as an indicator for housing market bubbles. This variable is updated monthly for geographical units called "neighborhoods" defined by Zillow.

The dataset we analyze consists of monthly median price-to-rent ratios from 234 neighborhoods in the San Francisco Bay Area from October 2010 to August 2018, with 95 observations on each curve at a missing rate of 1.48%. Figure 2 illustrates the geographic locations of these neighborhoods and their price-to-rent ratio trajectories.

1.3 Our contributions

We propose a unified FPCA method that is applicable to both sparse and dense functional data collected under a geostatistics setting, where locations are sampled from a spatial point process. We assume that the trajectory of a random function is determined by two effects: a temporal process that is spatially correlated with neighboring functions and a location-specific random process independent from neighbors. The location-specific random process is also interpreted as the "nugget" effect following classic geostatistics literature(Cressie, 1993). Observations on each function are made on discrete time points and contaminated with measurement errors. Under the assumption of spatial stationarity and isotropy, we propose a tensor product spline estimator for the spatio-temporal covariance function. If a coregionalization covariance structure (Banerjee et al., 2004; Gelfand et al., 2004) is further assumed, we propose a new FPCA method that borrows information from neighboring functions. Byproducts of our approach also include nonparametric estimators for the spatial covariance functions of the principal component scores. Under a unified framework that combines both infill and increasing domain asymptotic paradigms, we develop unified asymptotic convergence rates for the proposed estimators which demonstrate a phase transition from sparse to dense functional data.

The rest of the paper is organized as follows. We introduce the model and framework in Section 2, propose our estimation procedure in Section 3, and investigate the theoretical properties of the proposed estimators in Section 4. We address some important implementation issues in Section 5 and further extend our method for functional kriging

in Section 6. Numerical performance of the proposed methods is illustrated by simulation studies in Section 7, where we also show existing methods ignoring the functional nugget effect can lead to biased results. We analyze the two motivating data examples in Section 8 and provide concluding remarks in Section 9. Technical proofs of the main theorems and additional figures from our numerical studies are collected in the online Supplementary Material.

2 Model and assumptions

2.1 Random field modeling for spatially dependent functional data

Suppose random functions of time defined on a time domain T are sampled from locations in a spatial domain $\mathscr{D}_n \subseteq \mathbb{R}^2$. Let $Y_{ij} = Y(\boldsymbol{s}_i, t_{ij})$ be the discrete observation at time t_{ij} on the random curve sampled at spatial location $\boldsymbol{s}_i$, $i = 1, \ldots, N$, $j = 1, \ldots, M_i$, and assume the following model

$$Y(\boldsymbol{s}_i, t_{ij}) = X(\boldsymbol{s}_i, t_{ij}) + U_i(t_{ij}) + \epsilon_{ij}, \tag{1}$$

where $X(\cdot,\cdot)$ is a spatio-temporal process on $\mathscr{D}_n \times T$ representing a spatially correlated functional effect, $\{U_i(\cdot)\}$ are zero-mean, independent temporal processes called the functional nugget effects, and $\{\epsilon_{ij}\}$ are the independent measurement errors with $E(\epsilon_{ij}) = 0$ and $\mathrm{var}(\epsilon_{ij}) = \sigma_\epsilon^2$. The functional nugget effects $U_i(\cdot)$ characterize local variations that are not correlated with neighboring functions, with the covariance function denoted by $\Lambda(t_1, t_2) = \mathrm{cov}\{U(t_1), U(t_2)\}$. The three model components $X(\cdot,\cdot)$, $U(\cdot)$ and ϵ are mutually independent.

Assuming that the spatial dependency is second-order stationary and isotropic, the general covariance function of $X(\boldsymbol{s},t)$ can be written as

$$R(\| \boldsymbol{s}_1 - \boldsymbol{s}_2 \|, t_1, t_2) = \mathrm{cov}\{X(\boldsymbol{s}_1, t_1), X(\boldsymbol{s}_2, t_2)\}, \tag{2}$$

for any $(\boldsymbol{s}_1, t_2)$, $(\boldsymbol{s}_2, t_2) \in \mathscr{D}_n \times T$. We consider $X_{\boldsymbol{s}}(t) = X(\boldsymbol{s},t)$ as spatial replicates of a temporal process with a standard Karhunen-Loève expansion

$$X_s(t) = \mu(t) + \sum_{j=1}^{\infty} \xi_j(\boldsymbol{s})\psi_j(t), \tag{3}$$

where $\mu(t) = E\{X_{\boldsymbol{s}}(t)\}$ is the mean function, $\psi_j(\cdot)$'s are orthonormal functions known as the principal components, and the principal component score $\xi_j(\boldsymbol{s}) = \int_T \{X(\boldsymbol{s}, t) - \mu(t)\}\psi_j(t)dt$ is the loading of $X(\boldsymbol{s}, t)$ on the jth principal component. We assume $(\xi_j(\boldsymbol{s}))$ are zero-mean, second-order stationary and isotropic random fields, that are uncorrelated across different j. Spatial dependence among the function data is induced by the dependence within each $\xi_j(\boldsymbol{s})$. Denote the spatial covariance function of $\xi_j(\boldsymbol{s})$ as $\mathscr{C}_j(\| \boldsymbol{s}_1 - \boldsymbol{s}_2 \|) = \mathrm{cov}\{\xi_j(\boldsymbol{s}_1), \xi_j(\boldsymbol{s}_2)\}$, for any $\boldsymbol{s}_1, \boldsymbol{s}_2 \in \mathscr{D}_n$, then the covariance function for $X(\boldsymbol{s}, t)$ can be written as

$$R(\| \boldsymbol{s}_1 - \boldsymbol{s}_2 \|, t_1, t_2) = \operatorname{cov}\left\{ \sum_{j=1}^{\infty} \xi_j(\boldsymbol{s}_1)\psi_j(t_1), \sum_{j=1}^{\infty} \xi_j(\boldsymbol{s}_2)\psi_j(t_2) \right\}$$
$$= \sum_{j=1}^{\infty} \mathscr{C}_j(\| \boldsymbol{s}_1 - \boldsymbol{s}_2 \|)\psi_j(t_1)\psi_j(t_2)\,. \tag{4}$$

Denote $\varpi_j = \mathscr{C}_j(0)$ as the marginal variance for $\xi_j(\boldsymbol{s})$, and assume the principal components are ordered according to their magnitudes such that $\varpi_1 \geq \varpi_2 \geq \cdots > 0$. It is easy to see that ϖ_j's and $\psi_j(\cdot)$'s are the eigenvalues and eigenfunctions of the covariance function $R(0,\cdot,\cdot)$, which reveals an important connection between our model and classic models for independent functional data. The functional nugget effect $U_i(\cdot)$, on the other hand, may have an entirely different covariance structure with different eigenvalues and eigenfunctions.

In many applications, including the two real data examples in Section 1, we are interested in the temporal processes defined on some spatially distributed entities, e.g. houses. These entities may not exist on all locations, and the random field framework is a tool of choice to describe the spatial dependence. Model (3) is also analogous to recent developments in factor models for high dimensional multivariate time series (Fan et al., 2018) in the sense that $\xi_j(\boldsymbol{s})$ can be considered as latent factors that govern the dynamics of the temporal process $X_{\boldsymbol{s}}(t)$ and provides reduced rank representations of these temporal processes. In some applications, the latent factors $\xi_j(\boldsymbol{s})$ are of interest and can used as predictors in a second stage regression analysis (Wong et al., 2019). Similar FPC expansion as (3) was also promoted by Horváth and Kokoszka (2012) for spatially dependent functional data, who argued that, even if stationarity in space is mildly violated, the mean and eigenfunctions still provide meaningful marginal summary statistics for the data. By allowing different orders of FPC score to have different spatial covariances, covariance function (4) is a "coregionalization" model (Banerjee et al., 2004; Gelfand et al., 2004), which is the sum of many separable spatio-temporal covariance functions, and it reduces to a separable structure if $\mathscr{C}_j(\cdot) = \varpi_j\rho(\cdot)$ for all j.

2.2 Sampling scheme for spatial locations and observation times

As illustrated by the two examples in Section 1, the spatial locations $\{\boldsymbol{s}_i\}$ are often irregular and random, and can be best described by a spatial point process $\mathcal{N}_s(\cdot)$. The simplest spatial point process is the inhomogeneous Poisson process, where given the total number the locations are independent and identically distributed. A point process can be used to describe more complicated location patterns, such as clustered or regular patterns (Cressie, 1993). The correlation between locations are described by the higher-order intensity functions.

For any location $\boldsymbol{s}$, let $d\boldsymbol{s}$ be a small neighborhood around $\boldsymbol{s}$, and denote $|d\boldsymbol{s}|$ as the area of $d\boldsymbol{s}$ and $\mathcal{N}_s(d\boldsymbol{s})$ as the number of locations sampled in $d\boldsymbol{s}$. The k-th order intensity function of $\mathcal{N}_s(\cdot)$ is defined as (Cressie, 1993)

$$\lambda_{s,k}(\boldsymbol{s}_1,\ldots,\boldsymbol{s}_k)=\lim_{\substack{|\,d\boldsymbol{s}_r\,|\to 0,\\ r=1,\ldots,k}}\frac{E\{\mathcal{N}_s(d\boldsymbol{s}_1)\ldots\mathcal{N}_s(d\boldsymbol{s}_k)\}}{|\,d\boldsymbol{s}_1\,|\ldots|\,d\boldsymbol{s}_k\,|}, \tag{5}$$

and we assume $\mathcal{N}_s$ has up to the 4th order intensity function well defined. The collection of observation time points on $Y(\boldsymbol{s},\cdot)$ is a realization of a temporal point process $\mathcal{N}_t(dt \mid \boldsymbol{s})$. Assume that temporal point processes at different locations are independent and identically distributed. Denote the first and second intensity functions of $\mathcal{N}_t(\cdot \mid \boldsymbol{s})$ as

$$\begin{aligned}\lambda_{t,1}(t)&=\lim_{|\,dt\,|\to 0}\frac{E\mathcal{N}_t(dt \mid \boldsymbol{s})}{|\,dt\,|},\ \lambda_{t,2}(t_1,t_2)\\&=\lim_{|\,dt_1\,|,\,|\,dt_2\,|\to 0}\frac{E\{\mathcal{N}_t(dt_1 \mid \boldsymbol{s})\mathcal{N}_t(dt_2 \mid \boldsymbol{s})\}}{|\,dt_1\,|\,|\,dt_2\,|},\end{aligned} \tag{6}$$

which are independent of $\mathcal{N}_s(d\boldsymbol{s})$. This setting also implies that the number of repeated measures on $Y(\boldsymbol{s}_i,\cdot)$ is a random variable $M_i := \int_T \mathcal{N}_t(dt \mid \boldsymbol{s}_i)dt$. We can also define the joint point process for sampling locations and times as $\mathcal{N}(d\boldsymbol{s},dt)=\mathcal{N}_s(d\boldsymbol{s})\mathcal{N}_t(dt \mid \boldsymbol{s})$.

As further discussed in Section 4, we do not require $\mathcal{N}_s(\cdot)$ or $\mathcal{N}_t(\cdot \mid \boldsymbol{s})$ to be stationary, but rather need the intensity functions of these point processes to be bounded from zero so that we have a positive chance to sample from any location and time. By allowing the intensity functions, $\lambda_{s,k}(t)$ and $\lambda_{t,k}(t)$, to diverge to infinity, we also allow the "infill" paradigm: the number of sampled locations in unit space and the number of measurements in unit time are allowed to diverge to infinity.

3 Estimation method

We now propose nonparametric estimators for various model components described in Section 2, where the core issue is estimating the spatio-temporal covariance function $R(\cdot,\cdot,\cdot)$ in (2). We then use the estimated covariance function to further derive estimators for the principal components $\psi_j(\cdot)$ and spatial covariance functions $\mathcal{C}_j(\cdot)$, which are of fundamental importance to dimension reduction and understanding the spatial dependence. We will also estimate the covariance function $\Lambda(\cdot,\cdot)$ for the functional nugget effect and the variance of the measurement error σ_ϵ^2, which will be further used in the functional kriging.

3.1 Estimation of the spatio-temporal covariance function

For ease of exposition, we assume $\mu(t)\equiv 0$ for Sections 3 and 4. In practice, one can estimate $\mu(t)$ using the smoothing method described in Section 5, center the response as $\widetilde{Y}(\boldsymbol{s}_i,t_{ij})=Y(\boldsymbol{s}_i,t_{ij})-\hat{\mu}(t_{ij})$, and then the rest of our methods and theory still apply.

We will only estimate $R(u,\cdot,\cdot)$ up to a pre-determined spatial distance $\Delta > 0$. As pointed out by many authors (Hall et al., 1994; Li et al., 2007), spatial dependency usually decays to zero beyond certain distance; the spatial covariance estimator at a large spatial lag tends to be highly variable, consisting of more noise than signal. To determine Δ, one needs

to get a rough estimate for the range of spatial dependency based on a pilot study, for example using the nonparametric method in Li et al. (2007) based on a more stringent separable spatio-temporal covariance structure. We consider $R(u, t_1, t_2)$ as a function over a 3-dimensional domain $H \coloneqq [0,\Delta]\times T\times T$, and propose to estimate it using 3-dimensional tensor product B-splines. For independent functional data, many nonparametric smoothing methods have been proposed to estimate the covariance function, including kernel methods (Yao et al., 2005; Li and Hsing, 2010) and penalized splines (Xiao et al., 2013). In this paper, we focus on tensor product regression spline methods for their computational merits (Huang and Yang, 2004), but our methods and theory can be naturally extended to other smoothers.

Without loss of generality, assume $T = [0,1]$. Let $\boldsymbol{B}_T(t) = \{B^{p_t}_{1,K_t}(t), B^{p_t}_{2,K_t}(t), \ldots, B^{p_t}_{K_t+p_t,K_t}(t)\}^{\mathrm{T}}$ be a vector of normalized B-spline functions (de Boor, 2001; Huang and Yang, 2004) of order p_t, defined on time domain T with equally spaced interior knots $\boldsymbol{\kappa}_j = j/(K_t+1)$, $j = 1,\ldots, K_t$, and denote the corresponding spline space as $\mathcal{S}^{p_t}_{K_t}[0,1]$. Similarly, let $\boldsymbol{B}_S(u) = \{B^{p_s}_{1,K_s}(u), B^{p_s}_{2,K_s}(u), \ldots, B^{p_s}_{K_s+p_s,K_s}(u)\}^{\mathrm{T}}$ be a vector of B-spline basis functions on $[0,\Delta]$ with equally spaced interior knots, where the order p_s and number of knots K_s can be different from p_t and K_t allowing different amount of smoothing in spatial and temporal directions. The assumption of knots being equally spaced is for ease of theoretical derivations, but can be relaxed in practice. Denote the spline space spanned by $\boldsymbol{B}_S(u)$ as $\mathcal{S}^{p_s}_{K_s}[0,\Delta]$. Then the 3-dimensional tensor product spline space is defined as $\mathcal{S}_{[3]} \equiv \mathcal{S}^{p_s}_{K_s}[0,\Delta]\otimes\mathcal{S}^{p_t}_{K_t}[0,1]\otimes\mathcal{S}^{p_t}_{K_t}[0,1]$, which is spanned by basis functions $B_{j_1j_2j_3}(u,t_1,t_2) = B^{p_s}_{j_1,K_s}(u)B^{p_t}_{j_2,K_t}(t_1)B^{p_t}_{j_3,K_t}(t_2)$. Pool the tensor product spline basis functions into a vector $\boldsymbol{B}_{[3]}(u, t_1, t_2) = \boldsymbol{B}_S(u)\otimes\boldsymbol{B}_T(t_1)\otimes\boldsymbol{B}_T(t_2)$, where $\otimes$ is the Kronecker product.

Define $\mathcal{N}_{s,2}(d\boldsymbol{s}_1, d\boldsymbol{s}_2) \coloneqq \mathcal{N}_s(d\boldsymbol{s}_1)\mathcal{N}_s(d\boldsymbol{s}_2)I(\boldsymbol{s}_1 \neq \boldsymbol{s}_2)$, and the tensor product spline estimator of the spatio-temporal covariance function is

$$\widehat{R}(\cdot,\cdot,\cdot) = \underset{g(\cdot,\cdot,\cdot)\in\mathcal{S}_{[3]}}{\operatorname{argmin}} \int_{\mathcal{D}_n}\int_{\mathcal{D}_n}\int_T\int_T \{Y(\boldsymbol{s}_1,t_1)Y(\boldsymbol{s}_2,t_2) - g(\|\boldsymbol{s}_1-\boldsymbol{s}_2\|,t_1,t_2)\}^2 \times I(\|\boldsymbol{s}_1-\boldsymbol{s}_2\|\le\Delta)\mathcal{N}_t(dt_1 \mid \boldsymbol{s}_1)\mathcal{N}_t(dt_2 \mid \boldsymbol{s}_2)\mathcal{N}_{s,2}(d\boldsymbol{s}_1, d\boldsymbol{s}_2), \tag{7}$$

where $I(\cdot)$ is the indicator function. The estimator above can be equivalently written as $\widehat{R}(u,t_1,t_2) = \boldsymbol{B}^{\mathrm{T}}_{[3]}(u,t_1,t_2)\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ minimizes

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^{N}\sum_{\substack{i'\neq i \\ \|\boldsymbol{s}_i-\boldsymbol{s}_{i'}\|\le\Delta}}\sum_{j=1}^{M_i}\sum_{j'=1}^{M_{i'}}\left\{Y_{ij}Y_{i'j'} - \boldsymbol{B}^{\mathrm{T}}_{[3]}(\|\boldsymbol{s}_i-\boldsymbol{s}_{i'}\|,t_{ij},t_{i'j'})\boldsymbol{\beta}\right\}^2. \tag{8}$$

The numbers of knots K_s and K_t decide the amount of smoothing and can be selected by data-driven methods described in Section 5.

3.2 Estimation of the functional principal components

When the coregionalization structure in (4) is assumed, define

$$\Omega(t_1, t_2) \coloneqq \int_0^{\Delta} R(u, t_1, t_2)\mathscr{W}(u)du = \sum_{j=1}^{\infty} \omega_j \psi_j(t_1)\psi_j(t_2), \tag{9}$$

where $\mathscr{W}(\cdot) \in L^2$ is a non-negative and bounded weight function and $\omega_j = \int_0^{\Delta} \mathscr{C}_j(u)\mathscr{W}(u)du$. For all numerical studies in this paper, we use a simple weight function $\mathscr{W}(u) \equiv 1$ for $u \in [0,\Delta]$ and 0 otherwise. It is easy to see that the FPCs $\psi_j(t)$ are eigenfunctions of $\Omega(\cdot,\cdot)$. An estimator of Ω is obtained as

$$\widehat{\Omega}(t_1, t_2) = \int_0^{\Delta} \widehat{R}(u, t_1, t_2)\mathscr{W}(u)du, \tag{10}$$

and the estimated eigenvalues and eigenfunctions of $\Omega(\cdot,\cdot)$, denoted as $\{\widehat{\omega}_j, \widehat{\psi}_j(t)\}$, are obtained by solving the eigen-decomposition problem

$$\int_T \widehat{\Omega}(t_1, t_2)\widehat{\psi}_j(t_1)dt_1 = \widehat{\omega}_j\widehat{\psi}_j(t_2), \quad j = 1, 2, \ldots, \tag{11}$$

subject to the orthonormal constraints $\int_T \widehat{\psi}_j(t)\widehat{\psi}_{j'}(t)dt = I(j = j')$.

From the right hand side of (10), it is easy to see that all B-splines in the spatial direction are integrated out, and $\widehat{\Omega}(\cdot, \cdot)$ is contained in a bivariate tensor product spline space $\mathcal{S}_{[2]}$ spanned by the basis $\boldsymbol{B}_{[2]}(t_1, t_2) \coloneqq \boldsymbol{B}_{\mathcal{T}}(t_1) \otimes \boldsymbol{B}_{\mathcal{T}}(t_2)$. Hence, the functional eigen-decomposition problem in (11) can be translated into a multivariate problem. Notice that our estimator $\widehat{\Omega}$ is inherently symmetric. We can arrange the coefficient vector into a symmetric matrix $\boldsymbol{S}$, so that $\widehat{\Omega}(t_1, t_2) = \boldsymbol{B}_T^{\mathrm{T}}(t_1)\boldsymbol{S}\boldsymbol{B}_T(t_2)$. Define an inner product matrix $\mathcal{J} = \int_T \boldsymbol{B}_T(t)\boldsymbol{B}_T^{\mathrm{T}}(t)dt$, then the eigen-decomposition problem in (11) is equivalent to the multivariate generalized eigenvalue decomposition

$$\boldsymbol{\phi}_j^{\mathrm{T}}\mathcal{J}\boldsymbol{S}\mathcal{J}\boldsymbol{\phi}_j = \widehat{\omega}_j, \quad \text{subject to} \quad \boldsymbol{\phi}_{j'}^{\mathrm{T}}\mathcal{J}\boldsymbol{\phi}_j = I(j = j'),$$

and $\widehat{\psi}_j(t) = \boldsymbol{B}_T^{\mathrm{T}}(t)\boldsymbol{\phi}_j$, $j = 1, 2, \ldots$

3.3 Estimation of the spatial covariance and correlation functions

By the orthogonality of $\psi_j(t)$'s and (4), $\mathscr{C}_j(u) = \int_T \int_T R(u, t_1, t_2)\psi_j(t_1)\psi_j(t_2)dt_1dt_2$, which motivates the following estimator of the spatial covariance function

$$\mathscr{C}_j(u) = \int_T \int_T \widehat{R}(u, t_1, t_2)\widehat{\psi}_j(t_1)\widehat{\psi}_j(t_2)dt_1dt_2\,. \tag{12}$$

We then estimate the variance of the jth FPC by $\widehat{\varpi}_j = \mathscr{C}_j(0)$ and estimate the spatial correlation function $\rho_j(u) = \mathscr{C}_j(u) \,/\, \mathscr{C}(0)$ by $\hat{\rho}_j(u) = \mathscr{C}_j(u) \,/\, \mathscr{C}_j(0)$.

3.4 Covariance estimation for the functional nugget effect

Define $\Gamma\{t_1, t_2) = R\{0, t_1, t_2) + \Lambda(t_1, t_2)$. By independence between $X\{\boldsymbol{s}_i, t)$ and the functional nugget effect $U_i(t)$, it is easy to see $\text{cov}\{Y(\boldsymbol{s}, t_1), Y(\boldsymbol{s}, t_2)\} = \Gamma(t_1, t_2)$ for $t_1 \neq t_2$, which motivates another spline estimator

$$\widehat{\Gamma}(\cdot, \cdot) = \underset{g(\cdot,\cdot) \in \mathcal{S}^{\Gamma}_{[2]}}{\text{argmin}} \int_{\mathscr{D}_n} \int_T \int_T \{Y(\boldsymbol{s}, t_1)Y(\boldsymbol{s}, t_2) - g(t_1, t_2)\}^2 I(t_1 \neq t_2) \mathscr{N}_t(dt_1 \mid \boldsymbol{s}) \mathscr{N}_t(dt_2 \mid \boldsymbol{s}) \mathscr{N}_s(d\boldsymbol{s}). \tag{13}$$

Here, $\mathcal{S}^{\Gamma}_{[2]}$ is a functional space of bivariate tensor product splines of order p_Γ defined on K_Γ interior knots. This spline space can be defined on a different set of temporal knots than those used to estimate $R(\cdot,\cdot,\cdot)$, thus allowing a different amount of smoothing. A natural covariance estimator for the functional nugget effect is

$$\widehat{\Lambda}(t_1, t_2) = \widehat{\Gamma}(t_1, t_2) - \widehat{R}(0, t_1, t_2), \tag{14}$$

where $\widehat{R}(0, t_1, t_2)$ is the estimator defined in (7) evaluated at $u = 0$.

3.5 Variance estimation for the measurement errors

The variance function of the response is $\sigma^2_Y(t) = \text{var}\{Y(\boldsymbol{s}, t)\} = R(0, t, t) + \Lambda(t, t) + \sigma^2_\epsilon = \Gamma(t, t) + \sigma^2_\epsilon$. We estimate $\sigma^2_Y(t)$ by the following spline estimator,

$$\widehat{\sigma}^2_Y(\cdot) = \underset{g(\cdot) \in \mathcal{S}^{\epsilon}_{[1]}}{\text{argmin}} \int_{\mathscr{D}_n} \int_T \left\{Y^2(\boldsymbol{s}, t) - g(t)\right\}^2 \mathscr{N}_t(dt \mid \boldsymbol{s}) \mathscr{N}_s(d\boldsymbol{s}), \tag{15}$$

where $\mathcal{S}^{\epsilon}_{[1]}$ is a univariate spline space of order p_ϵ defined on K_ϵ interior knots. The following variance estimator is similar in spirit with those proposed by Yao et al. (2005)

$$\widehat{\sigma}^2_\epsilon = \frac{1}{|\ T\ |} \int_T \{\widehat{\sigma}^2_Y(t) - \widehat{\Gamma}(t, t)\} dt. \tag{16}$$

Both $\widehat{\sigma}^2_\epsilon$ and $\widehat{\Lambda}$ are important quantities we will later use for functional kriging.

Remark. Our estimation procedure involves integration of (multivariate) spline functions, when calculating $\widehat{\Omega}(\cdot, \cdot)$, $\widehat{\psi}_j(\cdot)$, $\mathscr{C}_j(\cdot)$ and $\widehat{\sigma}_\epsilon$. In our R code that supplements this paper, we compute the exact values of these integrals, using close-form expressions for integrals and the Gram matrix of B-splines (de Boor, 2001).

4 Theoretical properties

One important theoretical challenge in our problem is that there is only one copy of the spatio-temporal random field and all data are correlated. Under such a setting, it is well-known that infill asymptotics may lead to inconsistent estimation of spatial covariance (Zhang and Zimmerman, 2005). We therefore adopt a theoretical framework that combines both the infill and increasing domain asymptotic paradigms. Lu and Tjøstheim (2014) proposed a different way to combine the increasing domain and infill paradigms, which does not rely on point process modeling of the sampling locations, but their message was in line with ours that we need to combine the two asymptotic paradigms for good statistical properties and flexible modeling of the data.

For any function $f(\cdot)$ (univariate or multivariate) defined on a compact support, denote $\| f \|_{L^2}$ and $\| f \|_\infty$ as its L^2 and L^∞ norms. For any positive sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$ if a_n / b_n is bounded above by a constant, and $a_n \asymp b_n$ if $C_1 \le a_n / b_n \le C_2$ for all n and some $C_1, C_2 > 0$. For any subset $E \subset \mathbb{R}^2$, let $\mathcal{F}_X(E)$ be the σ-algebra generated by $\{X(\boldsymbol{s}, t):(\boldsymbol{s}, t) \in E \times T\}$. Suppose the spatial dependence of the functional data can be described by the α-mixing coefficients (Rosenblatt, 1956):

$$\alpha_X(h) = \sup_{\substack{E_1, E_2 \subset \mathbb{R}^2 \\ dist(E_1, E_2) \ge h}} \sup_{\substack{A_1 \in \mathcal{F}_X(E_1), \\ A_2 \in \mathcal{F}_X(E_2)}} | P(A_1 \cap A_2) - P(A_1)P(A_2) | , \tag{17}$$

where $dist(E_1, E_2)$ denotes the minimal Euclidean distance between E_1 and E_2. We make the following assumptions for our theoretical investigation.

Assumption 1. While the time domain T is fixed, consider a sequence of spatial domains $\{\mathcal{D}_n\}$ with the same shape such that, as $n \to \infty$, $C_1 n \le | \mathcal{D}_n | \le C_2 n$, and $C_1 \sqrt{n} \le | \partial \mathcal{D}_n | \le C_2 \sqrt{n}$, for some $C_1, C_2 > 0$. Here, $| \mathcal{D}_n |$ and $| \partial \mathcal{D}_n |$ are the area and perimeter of $\mathcal{D}_n$.

Assumption 2. Assume $X(\boldsymbol{s}, t)$ is strictly stationary in $\boldsymbol{s}$ and, for some $\nu > 4$, $\sup_{t \in T} E | X(\boldsymbol{s}, t) |^\nu < \infty$ and $\sup_{t \in T} E | U(t) |^\nu < \infty$.

Assumption 3. The α-mixing coefficient (17) is well defined for $X(\boldsymbol{s}, t)$, and there exist constants $\delta_1 > 2\nu/(\nu-4)$ and $C > 0$ such that $\alpha_X(h) \le Ch^{-\delta_1}$ for all $h \ge 0$ (Guyon, 1995).

Assumption 4. Suppose $\mathcal{N}_s(d\boldsymbol{s})$ is also α-mixing with the coefficient, denoted as $\alpha_{\mathcal{N}}(h)$, similarly defined as (17), and assume $\alpha_{\mathcal{N}}(h) \le C \exp(-\delta_2 h)$ for some $C > 0$ and $\delta_2 > 0$. There exists a sequence of positive numbers $\{L_n\}$, that is either constant or monotonically increasing to infinity with n, and constants $C_2 > C_1 > 0$ such that $C_1 L_n^k \le \lambda_{s,k}(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_k) \le L_n^k C_2$ for $k = 1, \ldots, 4$ and all $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_4 \in \mathcal{D}_n$.

Assumption 5. Let M_n be a sequence of positive constants depending on n, such that there exist some $C_1, C_2 > 0$ such that $C_1 M_n^k \le \lambda_{t,k}(t_1, \ldots, t_k) \le C_2 M_n^k$ for all $t_1, t_2 \in T$ and $k = 1, 2$.

Assumption 6. As $n \to \infty$, both K_s and $K_t \to \infty$, and $K_s K_t^2 = o\left\{n \,/\, \log^2(n)\right\}$.

Assumption 7. Restricting $R(\cdot,\cdot,\cdot)$ on the compact 3-dimensional domain $H = [0, \Delta] \times T \times T$, for order $\boldsymbol{r} = (r_1, r_2, r_3)$ and $a > 0$, define the Hölder class of functions on H as

$$C_3^{r,a}(H) := \{f: \sup_{\boldsymbol{x}_1 \boldsymbol{x}_2 \in H} \mid f^{(\ell_1,\ell_2,\ell_3)}(\boldsymbol{x}_1) - f^{(\ell_1,\ell_2,\ell_3)}(\boldsymbol{x}_2) \mid \,/\, \| \boldsymbol{x}_1 - \boldsymbol{x}_2 \|^a < \infty, 0 \le \ell_i \le r_i, i. = 1,2,3\}$$

Assume that $R \in C_3^{p,a}$, where $\boldsymbol{p} = \{p_s, p_t, p_t\}$ is the order of the 3-dimensional tensor product spline function and $a > 0$.

Assumption 8. Define a class of bivariate Holder continuous functions on T^2 as

$$C_2^{\boldsymbol{r},a}(T^2) := \{f: \sup_{\boldsymbol{x}_1 \boldsymbol{x}_2 \in T^2} \mid f^{(\ell_1,\ell_2)}(\boldsymbol{x}_1) - f^{(\ell_1,\ell_2)}(\boldsymbol{x}_2) \mid \,/\, \| \boldsymbol{x}_1 - \boldsymbol{x}_2 \|^a < \infty, \boldsymbol{r} = (r_1, r_2), 0 \le \ell_1. \le r_1, 0 \le \ell_2 \le r_2\}$$

Assume that $\Gamma(\cdot,\cdot)$ and $\Lambda(\cdot\,,\cdot) \in C_2^{(p_t,p_t),a}\left(T^2\right)$, where $a > 0$.

Assumption 1 describes a typical increasing domain asymptotic framework (Guan et al., 2004). A rectangular or circular spatial domain $\mathscr{D}_n$ with the same shape but increasing area would satisfy Assumption 1. Assumption 2 is a standard moment condition in functional data analysis (Li and Hsing, 2010). Assumption 3 allows the spatial dependency in $X(\boldsymbol{s}, t)$ to decay in a slow polynomial rate. In Assumption 4, we assume that the sampling spatial point process is also weakly dependent and there is a positive chance to sample any four points in $\mathscr{D}_n$. A homogenous Poisson process would satisfy Assumption 4. By allowing $L_n \to \infty$, our framework also accommodates the infill paradigm, meaning we allow $\lambda_{s,k}(\cdot)$ and hence the expected number of sampling points on any unit space to diverge to infinity. It is also worth pointing out that the expected number of repeated measures on $Y(\boldsymbol{s}_i, \cdot)$ is $\int_T \lambda_{t,1}(t)dt \asymp M_n$ under Assumption 5. When M_n are bounded by a constant, the data are spatially correlated sparse functional data; on the other hand, if $M_n \to \infty$ fast enough as a function of n, the data are dense functional data. In all of our theoretical results below, we allow M_n to be of any rate relative to n, thus admit all types of functional data in a unified framework. Assumption 6 is a standard assumption on the number of knots and sets a range for the tuning parameters. Assumptions 7 and 8 govern the smoothness of the functions that we estimate.

The following theorem provides the asymptotic convergence rate for the tensor-product spline estimator of the spatio-temporal covariance function.

Theorem 4.1. Under the model framework in Section 2 and Assumptions 1 – 7,

$$\| \hat{R} - R \|_{L^2} = O_p\left[\mid \mathscr{D}_n \mid^{-1/2} \{\sqrt{K_s} + \sqrt{K_s K_t \,/\, (M_n L_n)} + \sqrt{K_s K_t^2 \,/\, (M_n^2 L_n^2)}\} + K_s^{-p_s} + K_t^{-p_t}\right].$$

Remark (Effect of Infill). Theorem 4.1 implies that the most dominating factor in achieving consistent covariance estimation is the domain size $\mid \mathscr{D}_n \mid$. The infill factor L_n only plays a secondary role in the convergence rate: letting $L_n \to \infty$ but holding $\mid \mathscr{D}_n \mid$ fixed will

result in an inconsistence covariance estimator, which is in agreement with the results of Zhang and Zimmerman (2005) and Hörmann and Kokoszka (2013). Intuitively, increasing the sampling locations in a unit spatial domain will result in increasingly correlated data but not more information that is equivalent to independent samples. The factor $L_n M_n$ measures the number of spatio-temporal measurements in a unit spatial neighborhood. In an ideal case $L_n \to \infty$ in a fast enough rate so that we can choose $K_s \lesssim K_t \lesssim L_n M_n$, the dominant terms in $\| \hat{R} - R \|_L{}^2$ are of order $O_p(K_s^{1/2} \mid \mathcal{D}_n \mid^{-1/2} + K_s^{-p_s})$.

Remark (Phase Transition from Sparse to Dense Functional Data). For simplicity, the following discussion is restricted to a standard increasing domain framework where $\mid \mathcal{D}_n \mid \to \infty$ and L_n is a fixed constant. For sparse functional data where M_n is a bounded constant, assume $K_s = K_t \equiv K$ and $p_s = p_t \equiv p$ for simplicity, then the result in Theorem 4.1 can be simplified to $\| \hat{R} - R \|_L{}^2 = O_p(K^{3/2} \mid \mathcal{D}_n \mid^{-1/2} + K^{-p})$. Since $\mid \mathcal{D}_n \mid \asymp \mathrm{E}(N)$ is proportional to the sample size (i.e. the number of functions) under this setting, such a rate is the classic convergence rate for a 3-dimensional nonparametric regression using splines (Stone, 1994). For dense functional data with $M_n \gtrsim n^{1/(2\,p_t)}$ and choosing $K_t \asymp M_n$, we have $\| \hat{R} - R \|_L{}^2 = O_p(K_s^{1/2} \mid \mathcal{D}_n \mid^{-1/2} + K_s^{-p_s})$, which is the nonparametric convergence rate for estimating a stationary, isotropic spatial covariance function (Li et al., 2007). This result suggests $M_n \asymp n^{1/(2\,p_t)}$ is a transition point (Li and Hsing, 2010; Zhang and Wang, 2016; Wang et al., 2018), where estimating the 3-dim spatio-temporal covariance function is as efficient as estimating a 1-dim spatial covariance, and further increasing the number of repeated measures on each curve would not improve the convergence rate of $\hat{R}$.

The bivariate function $\Omega(\cdot,\cdot)$ in (10) is of fundamental importance to our FPCA methodology, where we borrow spatial information up to a distance $\Delta > 0$. The following theorem provides the convergence rate of $\widehat{\Omega}$.

Theorem 4.2. Under the assumptions in Theorem 4.1 and the coregionalization structure in (4), $\|\widehat{\Omega} - \Omega\|_L{}^2 = O_p\Big[\mid \mathcal{D}_n \mid^{-1/2} \{1 + \sqrt{K_t / (M_n L_n)}\} + K_s^{-p_s} + K_t^{-p_t} \Big]$.

Remark. By integrating over the spatial dimension of $\hat{R}$, we apply another step of smoothing and therefore obtain a faster convergence rate for $\widehat{\Omega}$ than $\hat{R}$. By undersmoothing in the spatial direction letting $K_s \gtrsim n^{1/(2\,p_s)}$, the $O_p(K_s^{-p_s})$ nuisance of estimating spatial covariance becomes negligible, then the rate in Theorem 4.2 is comparable to the classic covariance estimation convergence rate (Li and Hsing, 2010) for independent functional data using kernel smoothing. The convergence rate above becomes a typical bivariate spline smoothing rate $O_p(K_t / \mid \mathcal{D}_n \mid^{1/2} + K_t^{-p_t})$ when the data are sparse (the total number of measurements in a unit area $L_n M_n$ is bounded); and the root-n convergence rate, $\|\widehat{\Omega} - \Omega\|_L{}^2 = O_p(\mid \mathcal{D}_n \mid^{-1/2})$, is attainable, if the data are dense enough with $L_n M_n \gtrsim n^{1/(2\,p_t)}$ and if we choose $K_t \asymp L_n M_n$.

The convergence rate for $\widehat{\psi}_j(t)$ is a direct result from the perturbation theory in Hall and Hosseini-Nasab (2006) and is provided in the following theorem.

Theorem 4.3. Under the assumptions in Theorem 4.2 and suppose all eigenvalues of $\Omega(\cdot,\cdot)$ are distinct,

$$\| \widehat{\psi}_j - \psi_j \|_L^2 = O_p\Big[| \mathcal{D}_n |^{-1/2} \{1 + \sqrt{K_t / (M_n L_n)}\} + K_s^{-p_s} + K_t^{-p_t} \Big],$$

for $j = 1, 2, \ldots, J$, up to any fixed order J.

Remark. Results in Theorem 4.3 are comparable to those in Hall et al. (2006) and Li and Hsing (2010) for independent functional data. For sparse functional data where $L_n M_n$ is bounded by a constant, by adopting an undersmoothing strategy in the spatial direction (i.e. $K_s \gtrsim n^{1/(2p_s)}$, we get $\| \widehat{\psi}_j - \psi_j \|_L^2 = O_p\{(K_t / | \mathcal{D}_n |)^{1/2} + K_t^{-p_t}\}$. This is a 1-dim spline smoothing convergence rate, even though $\widehat{\psi}_j(t)$ is a byproduct of a 2-dim nonparametric estimator $\widehat{\Omega}(\cdot, \cdot)$ that converges in a slower 2-dim rate. For dense functional data ($L_n M_n \gtrsim n^{1/(2p_t)}$), by choosing $K_t \asymp L_n M_n$, we get $\| \widehat{\psi}_j - \psi_j \|_L^2 = O_p(| \mathcal{D}_n |^{1/2})$, which is a root-$n$ rate.

Restricting $\mathcal{C}_j(u)$ and $\mathcal{C}_j$ on $[0,\Delta]$, the following theorem provides convergence rates for the estimated spatial covariance functions.

Theorem 4.4. Under the assumptions of Theorem 4.3,

$$\| \mathcal{C}_j - \mathcal{C}_j \|_L^2 = O_p\Big[| \mathcal{D}_n |^{-1/2} \{\sqrt{K_s} + \sqrt{K_t / (M_n L_n)}\} + K_s^{-p_s} + K_t^{-p_t} \Big],$$

for $j = 1, 2, \ldots, J$ up to any fixed order J.

Remark. Suppose the covariance function R is smoother in the temporal directions than the spatial direction, i.e. $p_t \geq p_s$, by choosing $K_s^{p_s/p_t} \lesssim K_t \lesssim K_s$, the convergence rate in Theorem 4.4 becomes $O_p\Big\{(K_s / | \mathcal{D}_n |)^{1/2} + K_s^{-p_s}\Big\}$, which is comparable to the results in Li et al. (2007) developed for 1-dimensional spatial domain, multivariate response and under a rather stringent separable covariance assumption.

With the additional smoothness conditions in Assumption 8, we have the following results on the covariance estimator $\widehat{\Lambda}$ for the functional nugget effect and the variance estimator $\widehat{\sigma}_\epsilon^2$ for the measurement errors.

Theorem 4.5. Under Assumptions 1–8 and assume $K_\Gamma \asymp K_t$ and $p_\Gamma = p_t$,

$$\| \widehat{\Lambda} - \Lambda \|_L^2 = O_p\Big[| \mathcal{D}_n |^{-1/2} \{\sqrt{K_s} + \sqrt{K_s K_t / (M_n L_n)} + \sqrt{K_t^2 / (M_n^2 L_n)} + \sqrt{K_s K_t^2 / (M_n^2 L_n^2)}\} + K_s^{-p_s} + K_t^{-p_t} \Big].$$

Theorem 4.6. Under Assumptions 1 – 8 and further assume $K_\Gamma \asymp K_\epsilon \asymp K_t$ and

$$p_\Gamma = p_\epsilon = p_t,\ \hat{\sigma}_\epsilon^2 - \sigma_\epsilon^2 = O_p\Big[\ |\ \mathcal{D}_n\ |^{-1/2}\ \{1 + \sqrt{K_t\ /\ (M_n L_n)}\} + K_t^{-p_t}\Big].$$

Remark. The convergence rate of $\hat{\sigma}_\epsilon^2$ in Theorem 4.6 is comparable to Theorem 3.4 of Li and Hsing (2010) for independent functional data. Both (HTML translation failed) and $\hat{\sigma}_\epsilon^2$ are important quantities we will later use for functional kriging.

5 Implementation

5.1 Positive semi-definite adjustment for spatial covariance functions

The spatial covariance functions $\{\mathscr{C}_j(u): j = 1, \cdots, J\}$ are required by definition to be positive semi-definite in $\mathbb{R}^2$, meaning $\iint \mathscr{C}_j(\| s_1 - s_2 \|)a(s_1)a(s_2)ds_1 ds_2 \geq 0$, for any integrable functions $a(\cdot)$ defined on $\mathbb{R}^2$. The spline estimators $\hat{C}_j(u)$ defined in (12), even though consistent, are not guaranteed to be positive semidefinite. Nevertheless, this violation can be easily corrected using a correction procedure similar to that used in Hall et al. (1994).

By Bochner's theorem (Schabenberger and Gotway, 2017, p. 141), $\mathscr{C}_j(u)$ is positive semidefinite if $\mathscr{C}_j^+(\theta) \geq 0$ for all θ, where $\mathscr{C}_j^+(\theta) = \int_0^\infty \mathscr{C}_j(u)J_0(\theta u)udu$ is the Hankel transformation of $\mathscr{C}_j(\cdot)$ and $J_0(\cdot)$ is the Bessel function of the first kind with order 0. This motivates us to take a nonnegative truncation on the Hankel transformation of $\mathscr{C}_j(\cdot)$, i.e., $\mathscr{C}_j^+(\theta) = \max\{\int_0^\infty \mathscr{C}_j(u)J_0(\theta u)udu, 0\}$. In practice, $\mathscr{C}_j(u)$ decays to zero beyond the range of spatial dependence and $\mathscr{C}_j(u)$ is unstable for a large u. We therefore multiply $\mathscr{C}_j$ by a weight function $w(u) \leq 1$ when taking the Hankel transformation,

$$\mathscr{C}_j^+(\theta) = \max\left\{\int_0^\infty \mathscr{C}_j(u)J_0(\theta u)w(u)udu, 0\right\}. \tag{18}$$

Possible choices of $w(\cdot)$ suggested by Hall et al. (1994) are $w_1(u) = I(|u| \leq D)$ for a threshold $D > 0$; and $w_2(u) = 1$ if $|u| < D_1$, $(D_2 - |u|)\ /\ (D_2 - D_1)$ for $D_1 \leq |u| \leq D_2$ and 0 if $|u| > D_2$. Then the adjusted covariance estimators are the inverse Hankel transformations

$$\mathscr{C}_j(u) = \int_0^\infty \mathscr{C}_j^+(\theta)J_0(\theta u)\theta d\,\theta\,. \tag{19}$$

And the correlation functions are adjusted as $\tilde{\rho}_j(u) = \mathscr{C}_j(u)\ /\ \mathscr{C}_j(0)$ and an adjusted estimator for the spatio-temporal covariance function $R(\cdot,\cdot,\cdot)$ can be constructed as

$$\tilde{R}(u, t_1, t_2) = \sum_{j=1}^{J} \mathscr{C}_j(u)\hat{\psi}_j(t_1)\hat{\psi}_j(t_2), \tag{20}$$

where J is a large enough number such that the first J principal components capture most of the variation in the data. For the choice of the weight function in (18), we use $w_1(u) = I(|u| \leq D)$ and set $D = \Delta$ in all of our numerical studies, which leads to satisfactory results.

5.2 Choosing the number of B-spline knots

The amount of smoothing in our spline covariance estimator $\widehat{R}$ is governed by the numbers of knots K_s and K_t. Following Huang and Yang (2004), we choose these tuning parameters by minimizing the following Bayesian Information Criterion (BIC)

$$\mathrm{BIC}(K_s, K_t) = \widetilde{N}\log\{\mathscr{L}(\boldsymbol{\beta})\} + df \times \log(\widetilde{N}), \tag{21}$$

where $\mathscr{L}(\cdot)$ is the square loss function defined in (8), the degree of freedom $df = (K_s + p_s)(K_t + p_t)^2$ is the total number of tensor product B-spline basis functions, and $\widetilde{N} = \int_{\mathscr{D}_n}\int_{\mathscr{D}_n}\int_T\int_T I(\|\boldsymbol{s}_1 - \boldsymbol{s}_2\| \leq \Delta)\mathcal{N}_t(dt_1 \mid \boldsymbol{s}_1)\mathcal{N}_t(dt_2 \mid \boldsymbol{s}_2)\mathcal{N}_{s,2}(d\boldsymbol{s}_1, d\boldsymbol{s}_2)$ is the total sample size for estimating $R(\cdot,\cdot,\cdot)$. Similar BIC criteria are used to choose the number of knots in $\widehat{\Gamma}(\cdot,\cdot)$ and $\widehat{\sigma}^2_Y(\cdot)$.

5.3 Estimation of the mean function

Up to this point, we assume $\mu(t) \equiv 0$. In practice, we first estimate $\mu(t)$ by

$$\widehat{\mu}(\cdot) = \underset{g(\cdot) \in \mathcal{S}^{p_m}_{K_m}[0,1]}{\operatorname{argmin}} \int_{\mathscr{D}_n}\int_T \{Y(\boldsymbol{s},t) - g(t)\}^2 \mathcal{N}_t(dt \mid \boldsymbol{s})\mathcal{N}_s(d\boldsymbol{s}), \tag{22}$$

where $\mathcal{S}^{p_m}_{K_m}[0,1]$ is a spline space with order p_m and K_m interior knots, and then proceed with the methods described in Section 3 using the centered response $\widetilde{Y}(\boldsymbol{s}_i, t_{ij}) = Y(\boldsymbol{s}_i, t_{ij}) - \widehat{\mu}(t_{ij})$. For fully observed functional data with simple parametric spatial covariance and no measurement error, Kokoszka and Reimherr (2017) proposed a method to improve estimation efficiency for the mean function taking into account the spatial dependence. However, it is not yet clear how to extend this method to the discretely observed functional data with non-separable covariance structures in our paper, especially with the complication of functional nugget effect and measurement error.

6 Kriging of spatially dependent functional data

Spatial prediction or kriging is a major interest in spatial statistics (Stein, 2012) and there has been some recent work on kriging for spatially dependent functional data. The FPCA-then-kriging two-step procedure (Nerini et al., 2010; Menafoglio et al., 2016) is to first perform the classic FPCA (Yao et al., 2005) ignoring any spatial dependence and then perform co-kriging on the estimated FPC scores by fitting parametric spatial covariance models such as those in the Matérn family. There are several issues with this procedure: first, it does not consider functional nugget effect and, as shown in our simulation studies, may suffer from large estimation biases; second, the estimated FPC scores are contaminated with estimation errors, which bring a lot of nuisance into spatial covariance estimation; third, the spatial covariance models are limited to a few parametric families which may be

mis-specified. The trace kriging method (Giraldo et al., 2011; Menafoglio et al., 2013) does not depend on dimension reduction (e.g. FPCA) and requires fully observed functional data without measurement error nor nugget effect.

We now propose a new functional kriging method under our model. Let $s_0 \in \mathcal{D}_n$ be a new location where no data are observed, and our goal is to predict the unobserved functional data $X(s_0, t)$ using information from neighboring locations. Under our framework, $X(s_0, t) = \mu(t) + \sum_{j=1}^{\infty} \xi_j(s_0)\psi_j(t)$. In practice, the infinite principal component expansion of $X(s_0, t)$ needs to be truncated at a finite order J, which can be determined by a simple "percentage of variation explained" method (Yao et al., 2005). We then predict $X(s_0, t)$ by $\widehat{X}(s_0, t) = \hat{\mu}(t) + \sum_{j=1}^{J} \hat{\xi}_j(s_0)\hat{\psi}_j(t)$, where $\hat{\xi}_j(s_0)$ is the Best Linear Unbiased Predictor (BLUP) of $\xi_j(s_0)$ using data collected from locations close to s_0.

Let $\mathcal{N}(s_0, \Delta)$ be the collection of sampled locations within a distance Δ from s_0, and $\boldsymbol{Y}_{s_0,\Delta} = \{Y(s_i, t_{ij}), s_i \in \mathcal{N}(s_0, \Delta)\}^{\mathrm{T}}$ be the vector of observed data from the neighboring locations. Similarly, let $\boldsymbol{X}_{s_0,\Delta} = \{X(s_i, t_{ij}), s_i \in \mathcal{N}(s_0, \Delta)\}^{\mathrm{T}}$ and $\boldsymbol{U}_{s_0,\Delta} = \{U_i(t_{ij}), s_i \in \mathcal{N}(s_0, \Delta)\}^{\mathrm{T}}$ be the latent random vectors in $\boldsymbol{Y}_{s_0,\Delta}$. Suppose $\boldsymbol{R}_{s_0,\Delta} = \mathrm{cov}(\boldsymbol{X}_{s_0,\Delta})$ is the covariance matrix interpolated from the spatio-temporal covariance function $R(\cdot,\cdot,\cdot)$, $\boldsymbol{\Lambda}_{s_0,\Delta} = \mathrm{cov}(\boldsymbol{U}_{s_0,\Delta})$ is a block diagonal matrix representing the covariance of the functional nugget effect, then $\boldsymbol{\Sigma}_{s_0,\Delta} = \mathrm{cov}(\boldsymbol{Y}_{s_0,\Delta}) = \boldsymbol{R}_{s_0,\Delta} + \boldsymbol{\Lambda}_{s_0,\Delta} + \sigma_\epsilon^2 \boldsymbol{I}$ is the covariance matrix of the observed data within the neighborhood $\mathcal{N}(s_0, \Delta)$. Define $\boldsymbol{\Upsilon}_{s_0,j} = \mathrm{cov}\{\xi_j(s_0), \boldsymbol{Y}_{s_0,\Delta}\} = \{\mathscr{C}_j(\| s_i - s_0 \|)\psi_j(t_{i\ell}), s_i \in \mathcal{N}(s_0, \Delta)\}^{\mathrm{T}}$, then the BLUP for $\xi_j(s_0)$ is

$$\hat{\xi}_j(s_0) = \boldsymbol{\Upsilon}_{s_0,j}^{\mathrm{T}} \boldsymbol{\Sigma}_{s_0,\Delta}^{-1} (\boldsymbol{Y}_{s_0,\Delta} - \boldsymbol{\mu}_{s_0,\Delta}), \tag{23}$$

where $\boldsymbol{\mu}_{s_0,\Delta} = \mathrm{E}(\boldsymbol{Y}_{s_0,\Delta})$ is the mean vector interpolated from the mean function $\mu(t)$. The BLUP in (23) depends on unknown functions such as $R(\cdot,\cdot,\cdot)$, $\Lambda(\cdot,\cdot)$, $\mathscr{C}_j(\cdot)$, $\psi_j(\cdot)$ and $\mu(\cdot)$, which we replace with the estimators proposed in Sections 3 and 5.

7 Simulation studies

We now illustrate the proposed methodology using simulation studies. Data are generated from model (1) in the spatial domain $\mathcal{D} = [0, 10]^2$ and time domain $T = [0,1]$, with

$$X(s, t) = \mu(t) + \sum_{j=1}^{3} \xi_j(s)\psi_j(t),\ \mu(t) = 2t\sin(2\pi t),\ \psi_1(t) = \sqrt{2}\cos(2\pi t),\ \psi_2(t) = \sqrt{2}\sin(2\pi t)$$

and $\psi_3(t) = \sqrt{2}\cos(4\pi t)$. The principal component scores, $\xi_j(s)$, $j = 1, 2, 3$, are Gaussian random fields generated using the *RandomFields* package in R. The variances of ξ_j's are $(\varpi_1, \varpi_2, \varpi_3) = (3, 2, 1)$. Their spatial covariance functions are members of the Matérn family,

$\mathscr{C}_j(u;v,\rho) = \varpi_j \frac{2^{1-v}}{\Gamma(v)}(\sqrt{2v}u / \rho)^v K_v(\sqrt{2v}u / \rho)$, where $K_\nu(\cdot)$ is the modified Bessel function of the second kind with degree ν. We set the shape parameter ν to be 5.5, 3.5 and 1.5 and range parameter ρ to be 1, 0.5 and 0.5 respectively for the three principal components. The spatial locations $\{\boldsymbol{s}_i\}$ are sampled from a homogeneous spatial Poisson process over $\mathscr{D}$, with the first-order intensity $\lambda_s \equiv 10$; time of repeated measures on each function are sampled from a Poisson process over T with $\lambda_t = 10$. The measurement errors ϵ_{ij} are generated as iid Normal$(0, \sigma_\epsilon^2)$, where $\sigma_\epsilon^2 = 0.25$. We consider two scenarios for the functional nugget effect.

- Scenario A: functional nugget effect $U_i(t) = \sum_{j=1}^{2} \xi_{\text{nug},j}(\boldsymbol{s}_i)\psi_{\text{nug},j}(t)$, where $\psi_{\text{nug},1}(t)$ and $\psi_{\text{nug},2}(t)$ are the first two basis functions in the normalized Fourier-Bessel Series, $\xi_{\text{nug},j} \sim$ Normal$(0, \omega_{\text{nug},j})$, $j = 1, 2$, and $(\omega_{\text{nug},1}, \omega_{\text{nug},2}) = (2,1)$.
- Scenario B: no functional nugget effect, i.e., $Y(\boldsymbol{s}_i, t_{ij}) = X(\boldsymbol{s}_i, t_{ij}) + \epsilon_{ij}$.

We simulate 200 datasets for each scenario and apply the proposed estimation procedure (denoted as *sFPCA*) to each simulated dataset. We use tensor product of cubic B-splines to estimate the spatial-temporal covariance function. The tuning parameters are selected using the BIC described in Section 5 on some pilot datasets, then held fixed for massive simulations. For comparison, we also apply the classic FPCA method (Yao et al., 2005) to the simulated datasets. To the best of our knowledge, Liu et al. (2017) is the only exiting work on FPCA for discretely-observed, spatially-dependent functional data, and their method is identical to the classic FPCA method when it comes to estimating the eigenvalues and eigenfunctions. The classic FPCA, denoted as *iFPCA*, is implemented using the R package *fdapace*, which has built-in tuning parameter selection. Compared with our methods, *iFPCA* only estimates a bivariate temporal covariance function using observations at the same location $\boldsymbol{s}$, does not distinguish the functional nugget effect and does not borrow spatial information like what we do through integration in (10). Since our focus is on covariance estimation, estimation results for $\mu(t)$ are relegated to Figure S.1 in the Supplementary Material.

In Panels (a) - (f) of Figure 3, we summarize the estimation results of *sFPCA* under Scenario A for $\psi_j(\cdot)$ and $\mathscr{C}_j(\cdot,\cdot)$, $j = 1, 2, 3$. In each plot, we compare the mean of our estimator with the true function and provide confidence bands formed by pointwise 5% and 95% percentiles of the estimator. By taking a spectral decomposition of $\widehat{\Lambda}$ in (14), we also get estimators of $\psi_{\text{nug},j}(\cdot)$ and $\omega_{\text{nug},j}$. Graphical summaries of $\widehat{\psi}_{\text{nug},j}(t)$, $j = 1, 2$, are provided in Panels (g) and (h) of Figure 3; boxplots of scalar estimators $\widehat{\varpi}_j$ and $\widehat{\omega}_{\text{nug},j}$ are provided in Panel (i). As we can see, the *sFPCA* estimators behave reasonably well: all functional estimators exhibit very little bias and the confidence bands are tight around the true functions. The only functional estimator shows considerable variation is $\widehat{\psi}_{\text{nug},2}$, which is partially due to the fact that the convergence rate of $\widehat{\Gamma}$ in Theorem 4.5 is much slower compared with that of $\widehat{\Omega}$ in Theorem 4.2.

The *iFPCA* method does not produce estimates for the spatial covariance functions nor the eigenfunctions of the functional nugget effect, we therefore only provide graphical summaries of $\widehat{\psi}_j(t)$ for *iFPCA* under Scenario A in Figure 4. As we can see, these

functional estimators suffer from significant biases and large variation. The large biases can be explained by fact that *iFPCA* does not distinguish the functional nugget effect from the spatially dependent functional effect; the large variations, on the other hand, are due to strong spatial dependence and the fact that *iFPCA* does not borrow spatial information like we do through integration in (10). Under Scenario B, which is a simpler setting by removing the functional nugget effect $U_i(t)$ from Scenario A, both the classic *iFPCA* and our *sFPCA* methods provide consistent estimators for the eigenfunctions, and the differences between these methods are not as striking as in Scenario A. We therefore relegate graphical summaries under Scenario B to Figures S.2 and S.3 in the Supplementary Material. In theory, spectral decomposition of $\hat{R}(0, \cdot, \cdot)$ also provides consistent estimators of the eigenfunctions, however such a method suffers from the slow convergence rate of 3-dim spline smoothing and is not recommended. In simulation results not shown here, directly decomposing $\hat{R}(0, \cdot, \cdot)$ performs poorer than the proposed method under Scenario A and poorer than both methods presented under Scenario B.

We also summarize, in Table 1, the mean and standard deviation of integrated square error (ISE) for the functional estimators of *sFPCA* and *iFPCA*. These numerical summaries confirm our observations from the graphs that the *sFPCA* estimators behave overwhelmingly better than those of *iFPCA* under Scenario A, due to the existence of functional nugget effect. All estimators behave better under Scenario B due to smaller noises. However, even under Scenario B without functional nugget effects, *sFPCA* estimators of the eigenfunctions are still better than *iFPCA* because we borrow spatial information by including pairs of data in neighboring locations.

To illustrate the proposed *sFPCA* kriging method in Section 6, we randomly sample new functions from 100 new locations in each simulated dataset, and use the training data and the estimated covariance structure to predict $X(\boldsymbol{s}, t)$ at the new locations. The integrated square error (ISE), $\int \{\hat{X}(\boldsymbol{s}, t) - X(\boldsymbol{s}, t)\}^2 dt$, is averaged over all new locations and then repeated for each dataset. For comparison, we apply the *iFPCA+CoKriging* two-step procedure (Nerini et al., 2010) and the trace kriging method (Giraldo et al., 2011) to the simulated data. Both methods are implemented in R package *fdagstat*. For the *iFPCA+CoKriging* method, the number of principal components for *iFPCA* is selected to explain 99% of the variation and the spatial covariance functions are estimated using the Matérn models based on the estimated *iFPCA* scores. The trace kriging method requires fully observed functional data, we therefore treat the observed data as step functions with jumps at observed time points. The kriging results are summarized in Table 2, where we provide the mean and standard deviation of ISE for all competing methods. As we can see, our kriging method yields much smaller prediction errors than the two competing methods under both scenarios.

8 Data analysis

We now analyze the two motivating datasets described in Section 1.

8.1 Analysis of the London house price data

This dataset consists of 10, 980 transaction records of 2013 houses in the Greater London Area from Jan 1, 1995 to Dec 31, 2018. Figure S.4 in the Supplemental Material shows the empirical distributions for the number of transactions per house and the transaction dates. The estimated mean function, shown in Figure 1, demonstrates an overall increasing trend. Remarkably, the two dips on the mean curve reflect the impacts of the 2008 financial crisis and the 2016 Brexit.

A pilot study indicates that the range of spatial dependency is about 5.5 kilometers, which is also confirmed by the final estimators of the spatial correlations in Figure 5. We therefore estimate the spatio-temporal covariance function $R(\cdot,\cdot,\cdot)$ up to a spatial lag of $\Delta = 5.5$ km, using tensor product of cubic B-splines. The numbers of knots chosen by BIC are $K_s = 6$ and $K_t = 6$ in spatial and temporal directions, respectively.

Next, we perform FPCA to the data by a spectral decomposition of $\widehat{\Omega}$. The first two eigenvalues, $\widehat{\omega}_1 = 285.80$ and $\widehat{\omega}_2 = 21.52$, in total explain 99.42% of variation in $\widehat{\Omega}$. A contour plot of $\widehat{\Omega}(\cdot, \cdot)$ and the first two estimated eigenfunctions are shown in Figure 5 (a) and (c). The estimated spatial correlation functions and their positive semi-definite adjustments are shown in Figure 5 (e) and (f). As we can see, $\hat{\rho}_2(u)$ decays to 0 more sharply than $\hat{\rho}_1(u)$, indicating that the two principal components have different ranges of spatial dependence and the spatio-temporal covariance may not be separable. We also estimate the covariance function $\Lambda(\cdot,\cdot)$ of the functional nugget effect and the nugget principal components, the results of which are shown in Figure 5 (b) and (d). The noise-to-signal ratio of the functional nugget effect is $\|\widehat{\Lambda}(\cdot, \cdot)\|_L^2 \,/\, \|\widehat{R}(0, \cdot, \cdot)\|_L^2 = 1.11$. The first three eigenvalues, $\widehat{\omega}_{\text{nug},1} = 144.89$, $\widehat{\omega}_{\text{nug},2} = 80.50$, and $\widehat{\omega}_{\text{nug},3} = 29.23$, explain 98.77% of the total variation in the functional nugget effect. These results show that, for the London housing market, the house-specific effect is more important than the spatial dependent effect. These house-specific effects might be explained by factors such as size, year built, number of bedrooms, number of bathrooms, etc. These variables are not available in public records, hence not included in our analysis. It would be interesting to include these covariates in our future analysis, should an external data source becomes available.

8.2 Analysis of the Zillow real estate data

The spatial locations in this dataset are sampled from six regions in the Bay Area: *Fremont, Oakland, Palo Alto, San Francisco, San Jose*, and *San Mateo*. The estimated region-specific mean functions are presented in Figure S.5 of the Supplementary Material. To get rid of the regional effects, we center the trajectories in Figure 2 by subtracting their region-specific mean functions, and the residual trajectories are presented in Figure S.6. Our methodology is based on the spatially stationary assumption, but can be easily extended to piecewise-stationary settings, we therefore apply the proposed methodology to the residual trajectories.

Our pilot analysis on the Zillow data indicates that the spatial correlation diminishes at a distance of about 3 km. We therefore estimate the spatio-temporal covariance function $R(\cdot,\cdot,\cdot)$ up to a spatial lag of $\Delta = 3.5$ km, using tensor-product cubic B-splines. The number

of knots chosen by BIC are $K_s = 5$ and $K_t = 6$. Spectral analysis of $\widehat{\Omega}$ yields that the first two eigenvalues, $\widehat{\omega}_1 = 974.22$ and $\widehat{\omega}_2 = 18.59$, explain 97.97% of variation in $\widehat{\Omega}$. A contour plot of $\widehat{\Omega}(\cdot, \cdot)$ and the first two eigenfunctions are shown in Figure 6. Notice that $\widehat{\psi}_1(t)$, given by the solid curve in Figure 6 (c), is almost constant over time, which implies that the first FPC is a spatial random intercept – locations with high scores $\xi_1(\mathbf{s})$ on the first FPC has higher than average price-to-rent ratio. On the other hand, $\widehat{\psi}_2(t)$ represents a decreasing trend in time. Since the overall trend of price-to-rent ratio is increasing in Figure 2 (b), locations with high values of $\xi_2(\mathbf{s})$ has slower than average increase of price-to-rent ratio. The estimated spatial correlation functions and their positive semi-definite adjustments are shown in the lower panels of Figure 6. We also estimate the covariance function $\Lambda(\cdot,\cdot)$ of the functional nugget effect and the nugget principal components, the results of which are shown in Figure 6. The first three eigenvalues, $\widehat{\omega}_{\text{nug},1} = 92.72$, $\widehat{\omega}_{\text{nug},2} = 20.75$, and $\widehat{\omega}_{\text{nug},3} = 10.43$, explain 91.12% of the total variation in the functional nugget effect. The estimated variance of measurement errors is $\widehat{\sigma}^2_{\epsilon} = 0.246$.

We illustrate the performance of the proposed *sFPCA* kriging method by a leave-one-curve-out kriging experiment: leave one curve out as test data, use the rest of the data and the fitted model to predict the curve on the left out location, calculate the integrated squared error (ISE) for the prediction, and repeat this experiment for all locations. For comparison, we also perform the same kriging experiment for *iFPCA+Co-kriging* and *Trace Kriging*, described in Sections 6 and 7. After scaling the time domain to [0,1], the median prediction ISE is 1.85 for *sFPCA* kriging, 2.91 for *Trace Kriging*, and 3.61 for *iFPCA+Co-kriging*, which confirms that our proposed kriging method has much smaller prediction error than existing functional kriging methods.

8.3 Sensitivity Analysis

In Figures S.7 and S.8, we show contour plots of $\widehat{R}(u, \cdot, \cdot)$ at different values of u for the two data examples, respectively. To make different slices of this 3-dim function comparable, we standardize the contour plots by $\| \widehat{R}(u, \cdot, \cdot) \|_1 = \int | \widehat{R}(u, t_1, t_2) | \, dt_1 dt_2 / | T |^2$. For both datasets, the differences in the standardized contour plots show some evidence that the covariance structures are non-separable.

In Section S.6 in the Supplementary Material, we perform sensitivity analyses on both datasets to verify the assumption of spatial stationarity. We compare the FPCA estimates obtained from the whole spatial domain with those obtained from sub-domains. For the London data, we consider two sub-domains – regions to the north and south of River Thames; for the Zillow data, we divide the domain into two sub-domains: areas on the peninsula (*San Francisco, San Mateo* and *Palo Alto*) and those outside (*Fremont, Oakland* and *San Jose*). The fact that the FPCA estimates from the whole domain agree well with those from subdomains suggests that there is no serious violation of the stationarity assumption.

9 Discussion

As discussed in Section 2, spatial functional data analysis is deeply connected with spatio-temporal models, yet substantially different. In the two real data examples presented in this paper, our focus is to perform dimension reduction for temporal processes defined on real entities, which happen to be spatially correlated. We demonstrate how our model can be used for spatial prediction, but more importantly it extracts latent factors in the data, which can be used in further analysis, including a second stage regression.

We propose a three dimensional tensor product spline approach to estimate the spatio-temporal covariance function. Based on a coregionalization structural assumption, which is more flexible than the commonly used separable structure assumed in the literature, our three dimensional spline covariance estimator yields important byproducts, including nonparametric estimators of the principal components and the spatial covariance functions for the FPC scores. We also stress the importance of modeling the functional nugget effects, which model the local characteristics that are not dependent to the neighbors. We show in our simulation studies, ignoring the functional nugget effects can potentially cause large biases in the FPCA estimators. Our asymptotic study for the proposed methodology is quite comprehensive, where we combine both infill and increasing domain paradigms and accommodate both sparse and dense functional data. We found that, compared with the domain size, the effect of infilling locations in a unit spatial domain only has a secondary effect on the asymptotic convergence rate of the proposed estimators. We also establish phase transition in the convergence rates from sparse to dense functional data, which was not previously available for spatially dependent functional data.

Our method is based on three dimensional spline smoothing on the product of all data pairs within a prescribed distance, and hence computationally more intense than some of the existing method such as the *iFPCA* method implemented in the 'fdaPACE' package. In the Scenario A of our simulation study reported in Section 7, the average running time of *iFPCA* on a computer of 2.60GHz processor and 128 GB memory is 45.2 seconds, while the average running time for our method is 384.0 seconds. The extra computational cost is justifiable by the additional information we offer on the spatio-temporal covariance structure and being able to distinguish the functional nugget effect from the spatial functional effect. In our supplementary material, we also provide additional simulation results on the sensitivity of our method to the choice of Δ. We recommend to use a Δ approximately equal to the range of spatial dependency, where the spatial correlation decays to 0. In reality such a range is unknown and our results in Table S.1 suggest that our estimation results for the functional principal components are not sensitive to the choice of Δ. On the other hand, Table S.1 also summarizes the running time of *sFPCA* under different choices of Δ, and a larger Δ results in a longer running time. This is understandable because more data pairs are included into the three dimensional smoothing when a larger Δ is used.

Our approach is based on moderate model assumptions, such as spatial stationarity. As we demonstrate in our real data analysis, the stationarity assumption can be easily relaxed to piecewise stationarity. The second order stationarity assumption on the principal component scores can also be relaxed: suppose $\mathscr{C}_j(\boldsymbol{s}_1, \boldsymbol{s}_2) = \mathrm{cov}\{\xi_j(\boldsymbol{s}_1), \xi_j(\boldsymbol{s}_2)\}$

is non-stationary, but the averages of these covariance functions at distance u, $\mathscr{C}_j^*(u) = \lim_{n\to\infty} \frac{1}{2\pi u \mid \mathscr{D}_n \mid} \int_{\mathscr{D}_n} \int_{\|\boldsymbol{v}\|=1} \mathscr{C}_j(\boldsymbol{s}, \boldsymbol{s}+u\boldsymbol{v}) d\boldsymbol{v} d\boldsymbol{s}$, exist and are uniformly bounded, then under some weak dependence assumptions the proposed tensor spline covariance estimator consistently estimates $\mathscr{R}^*(u, t_1, t_2) = \sum_j \mathscr{C}_j^*(u)\psi_j(t_1)\psi_j(t_2)$. We still get legitimate principal component estimates, but spatial covariance function estimates become less interpretable. Our work based on the stationary assumption also paves the way for extensions to more sophisticated models, such as the locally stationary models (Kuusela and Stein, 2018), which can be applied to data collected from a large spatial region. Our methods also open up many new research questions, related to model selection and statistical inference for the proposed model. For instance, one important research question is how to select the number of principal components in the model. Aikaike information criterion such as that studied in Li et al. (2013) depends on evaluating the likelihood, which is difficult for spatially dependent functional data. It might also be possible to relax the isotropic assumption in our approach to a more flexible geometric anisotropy setting. All these questions and possible extensions call for future research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

Li's research was partially supported by National Institute on Aging, grant 5R21AG058198. We thank the two anonymous referees for their constructive comments and helpful suggestions, which lead to significant improvement of our paper.

References

Al-Sulami D, Jiang Z, Lu Z, and Zhu J (2017). Estimation for semiparametric nonlinear regression of irregularly located spatial time series data. Econometrics and Statistics, 2:22 – 35.

AL-SULAMI D, JIANG Z, LU Z, and ZHU J (2019). On a semiparametric data-driven nonlinear model with penalized spatio-temporal lag interactions. Journal of Time Series Analysis, 40:327 – 342.

Aue A, Norinho DD, and Hörmann S (2015). On the prediction of stationary functional time series. Journal of the American Statistical Association, 110(509):378–392.

Banerjee S, Carlin BP, and Gelfand AE (2004). Hierarchical Modeling and Analysis for Spatial Data. Chapman and Hall/CRC, New York.

Cai TT and Hall P (2006). Prediction in functional linear regression. The Annals of Statistics, 34(5):2159–2179.

Campbell SD, Davis MA, Gallin J, and Martin RF (2009). What moves housing markets: A variance decomposition of the rent–price ratio. Journal of Urban Economics, 66(2):90–102.

Crainiceanu CM, Staicu A-M, and Di C-Z (2009). Generalized multilevel functional regression. Journal of the American Statistical Association, 104(488):1550–1561. [PubMed: 20625442]

Cressie N and Wikle CK (2015). Statistics for Spatio-Temporal Data. John Wiley & Sons, Hoboken, NJ.

Cressie NAC (1993). Statistics for Spatial Data. Wiley, New York.

de Boor C (2001). A Practical Guide to Splines. Springer-Verlag, New York.

Fan J, Liu H, and Wang W (2018). Large covariance estimation through elliptical factor models. The Annals of Statistics, 46:1383–1414. [PubMed: 30214095]

Gelfand AE, Schmidt AM, Banerjee S, and Sirmans C (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. Test, 13(2):263–312.

Giraldo R, Delicado P, and Mateu J (2011). Ordinary kriging for function-valued spatial data. Environmental and Ecological Statistics, 18(3):411–426.

Gromenko O, Kokoszka P, Zhu L, and Sojka J (2012). Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. The Annals of Applied Statistics, 6(2):669–696.

Guan Y, Sherman M, and Calvin JA (2004). A nonparametric test for spatial isotropy using subsampling. Journal of the American Statistical Association, 99(467):810–821.

Guyon X (1995). Random Fields on a Network: Modeling, Statistics, and Applications. Springer-Verlag, New York.

Hall P, Fisher NI, and Hoffmann B (1994). On the nonparametric estimation of covariance functions. The Annals of Statistics, 22(4):2115–2134.

Hall P and Hosseini-Nasab M (2006). On properties of functional principal components analysis. Journal of the Royal Statistical Society: Series B, 68(1):109–126.

Hall P, Müller H-G, and Wang J-L (2006). Properties of principal component methods for functional and longitudinal data analysis. The Annals of Statistics, 34(3):1493–1517.

Hörmann S and Kokoszka P (2010). Weakly dependent functional data. The Annals of Statistics, 38(3):1845–1884.

Hörmann S and Kokoszka P (2013). Consistency of the mean and the principal components of spatially distributed functional data. Bernoulli, 19(5A):1535–1558.

Horváth L and Kokoszka P (2012). Inference for Functional Data with Applications. Springer, New York.

Hsing T and Eubank R (2015). Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators. Wiley.

Huang JZ and Yang L (2004). Identification of non-linear additive autoregressive models. Journal of the Royal Statistical Society: Series B, 66(2):463–477.

Jiang Z, Ling N, Lu Z, Tjøstheim D, and Zhang Q (2020). On bandwidth choice for spatial data density estimation. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82(3):817–840.

Kishor NK and Morley J (2015). What factors drive the price–rent ratio for the housing market? A modified present-value analysis. Journal of Economic Dynamics and Control, 58:235–249.

Kokoszka P and Reimherr M (2017). Introduction to Functional Data Analysis. CRC Press, New York.

Kuenzer T, Hörmann S, and Kokoszka P (2020). Principal component analysis of spatially indexed functions. Journal of the American Statistical Association, to appear.

Kuusela M and Stein ML (2018). Locally stationary spatio-temporal interpolation of argo profiling float data. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 474:20180400.

Li Y and Guan Y (2014). Functional principal component analysis of spatiotemporal point processes with applications in disease surveillance. Journal of the American Statistical Association, 109(507):1205–1215. [PubMed: 25368436]

Li Y and Hsing T (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. The Annals of Statistics, 38(6):3321–3351.

Li Y, Wang N, and Carroll RJ (2013). Selecting the number of principal components in functional data. Journal of the American Statistical Association, 108(504):1284–1294.

Li Y, Wang N, Hong M, Turner ND, Lupton JR, and Carroll RJ (2007). Nonparametric estimation of correlation functions in longitudinal and spatial data, with application to colon carcinogenesis experiments. The Annals of Statistics, 35(4):1608–1643.

Liang D, Zhang H, Chang X, and Huang H (2020). Modeling and regionalization of China's $PM_{2.5}$ using spatial-functional mixture models. Journal of the American Statistical Association, to appear.

Liu C, Ray S, and Hooker G (2017). Functional principal component analysis of spatially correlated data. Statistics and Computing, 27(6):1639–1654.

Lu Z, Steinskog D, Tjøstheim D, and Yao Q (2009). Adaptively varying-coefficient spatiotemporal models. Journal of Royal Statistical Society, Series B, 71:859–880.

Lu Z and Tjøstheim D (2014). Nonparametric estimation of probability density functions for irregularly observed spatial data. Journal of the American Statistical Association, 109(508):1546–1564.

Menafoglio A, Grujic O, and Caers J (2016). Universal kriging of functional data: Trace-variography vs cross-variography? Application to gas forecasting in unconventional shales. Spatial Statistics, 15:39–55.

Menafoglio A, Secchi P, and Dalla Rosa M (2013). A universal kriging predictor for spatially dependent functional data of a Hilbert space. Electronic Journal of Statistics, 7:2209–2240.

Nerini D, Monestiez P, and Manté C (2010). Cokriging for spatial functional data. Journal of Multivariate Analysis, 101(2):409–418.

Ramsay JO and Silverman BW (2005). Functional Data Analysis. Springer, New York.

Rosenblatt M (1956). Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics, 27(3):832–837.

Schabenberger O and Gotway CA (2017). Statistical Methods for Spatial Data Analysis. Chapman and Hall/CRC, Boca Raton.

Staicu A-M, Crainiceanu CM, and Carroll RJ (2010). Fast methods for spatially correlated multilevel functional data. Biostatistics, 11 (2):177–194. [PubMed: 20089508]

Stein ML (2012). Interpolation of Spatial Data: Some Theory for Kriging. Springer-Verlag, New York.

Stone CJ (1994). The use of polynomial splines and their tensor products in multivariate function estimation. The Annals of Statistics, 22(1):118–171.

Wang H, Zhong P-S, Cui Y, and Li Y (2018). Unified empirical likelihood ratio tests for functional concurrent linear models and the phase transition from sparse to dense functional data. Journal of the Royal Statistical Society: Series B, 80(2):343–364.

Wong R, Li Y, and Zhu Z (2019). Partially linear functional additive models for multivariate functional data. Journal of the American Statistical Association, 114:406–418.

Xiao L, Li Y, and Ruppert D (2013). Fast bivariate psplines: the sandwich smoother. Journal of the Royal Statistical Society: Series B, 75(3):577–599.

Xu Y, Li Y, and Nettleton D (2018). Nested hierarchical functional data modeling and inference for the analysis of functional plant phenotypes. Journal of the American Statistical Association, 113(522):593–606.

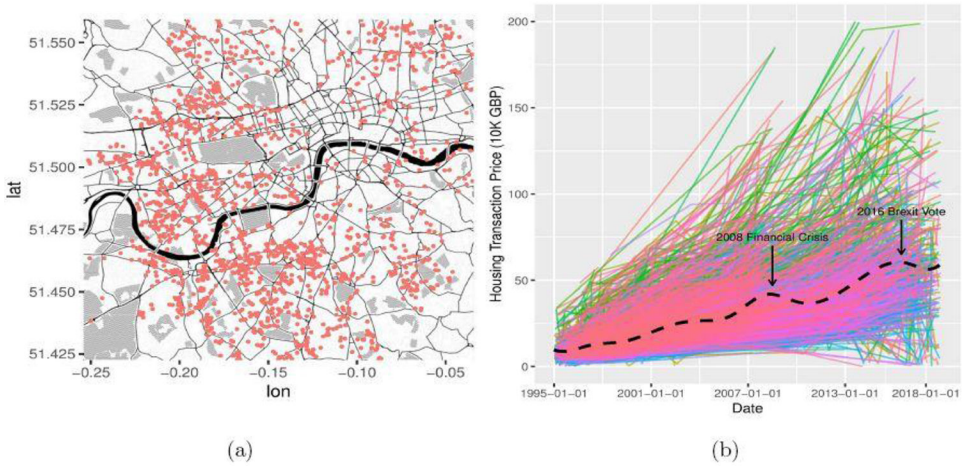
Yao F, Müller H-G, and Wang J-L (2005). Functional data analysis for sparse longitudinal data. Journal of the American Statistical Association, 100(470):577–590.

Zhang H and Zimmerman DL (2005). Towards reconciling two asymptotic frameworks in spatial statistics. Biometrika, 92(4):921–936.

Zhang L, Baladandayuthapani V, Zhu H, Baggerly KA, Majewski T, Czerniak BA, and Morris JS (2016). Functional car models for large spatially correlated functional datasets. Journal of the American Statistical Association, 111(514):772–786. [PubMed: 28018013]

Zhang X and Wang JL (2016). From sparse to dense functional data and beyond. The Annals of Statistics, 44(5):2281–2321.

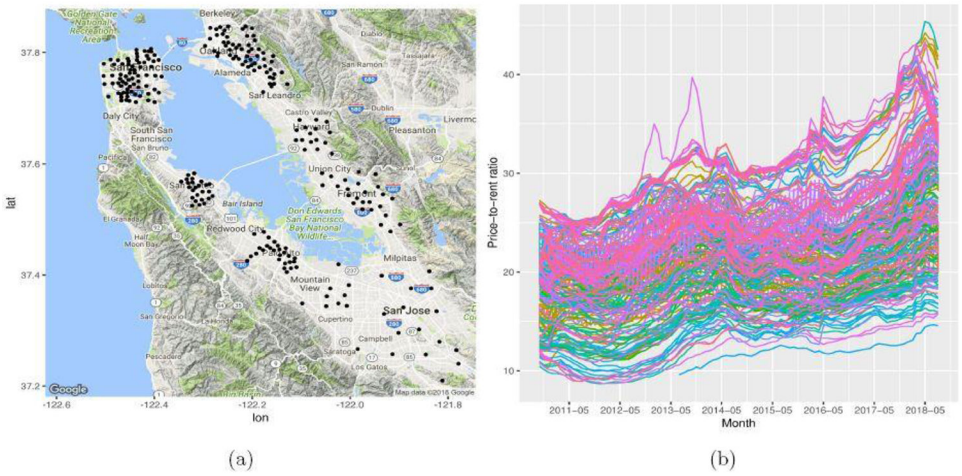
Zhou L, Huang JZ, Martinez JG, Maity A, Baladandayuthapani V, and Carroll RJ (2010). Reduced rank mixed effects models for spatially correlated hierarchical functional data. Journal of the American Statistical Association, 105(489):390–400. [PubMed: 20396628]



(a) (b)

Fig. 1.

London house price data. (a) Locations of houses in the Greater London Area; (b) trajectories of the house prices and the estimated mean function (dashed line).



(a) (b)

Fig. 2.

(a) The locations of the 234 neighborhoods in the San Francisco Bay Area; (b) trajectories of the home price-to-rent ratios, observed monthly from October 2010 to August 2018 in the 234 neighborhoods.

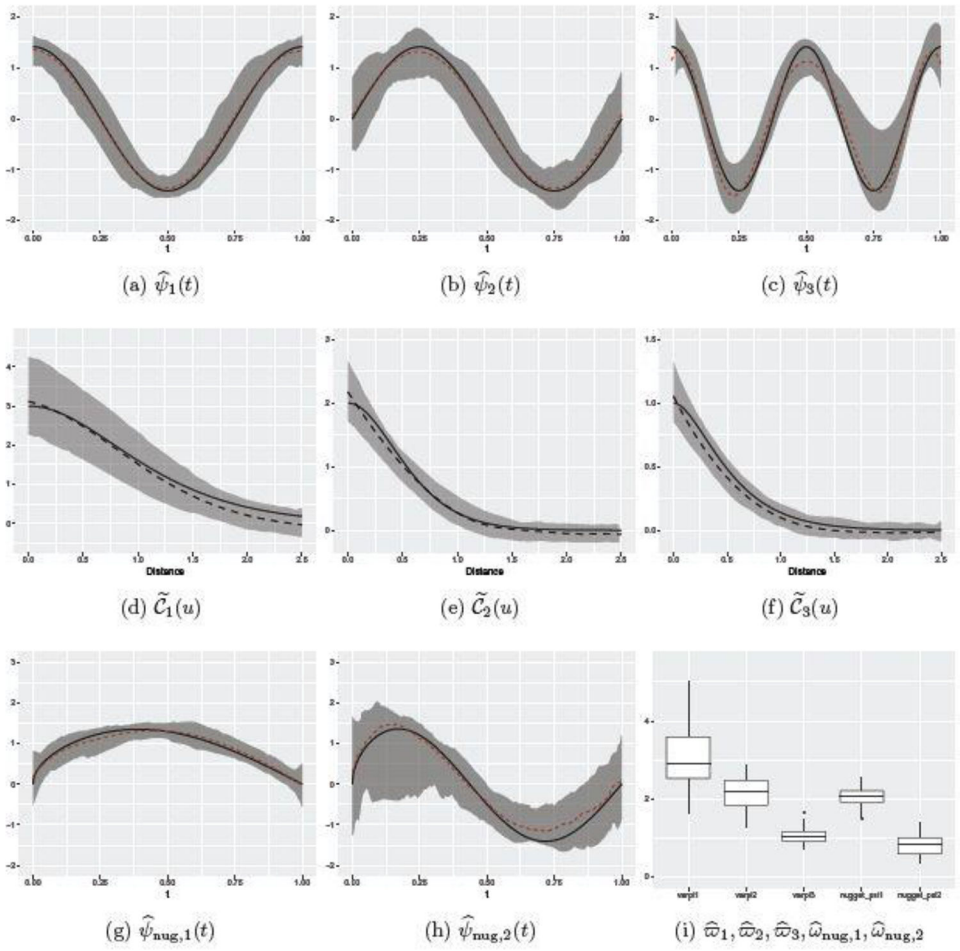


Fig. 3.

Estimation results of *sFPCA* under Scenario A. Panels (a) - (h) contain summaries of the functional estimators, as described in the labels. In each panel, the solid line is the true function; the dashed line is the mean of the functional estimator; and the shaded area illustrates the bands of pointwise 5% and 95% percentiles. Panel (i) contains the boxplots of $\widehat{\varpi}_1$, $\widehat{\varpi}_2$, $\widehat{\varpi}_3$, $\widehat{\omega}_{\mathrm{nug},1}$, and $\widehat{\omega}_{\mathrm{nug},2}$.

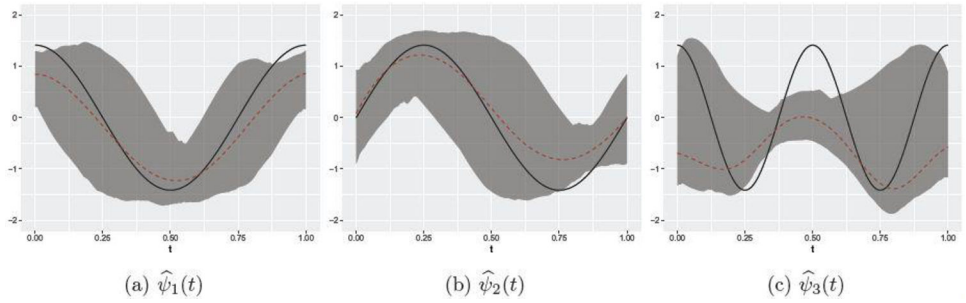


Fig. 4.

Estimation results of *iFPCA* under Scenario A. In each panel, the solid line is the true function; the dashed line is the mean of the functional estimator; and the shaded area illustrates the bands of pointwise 5% and 95% percentiles.

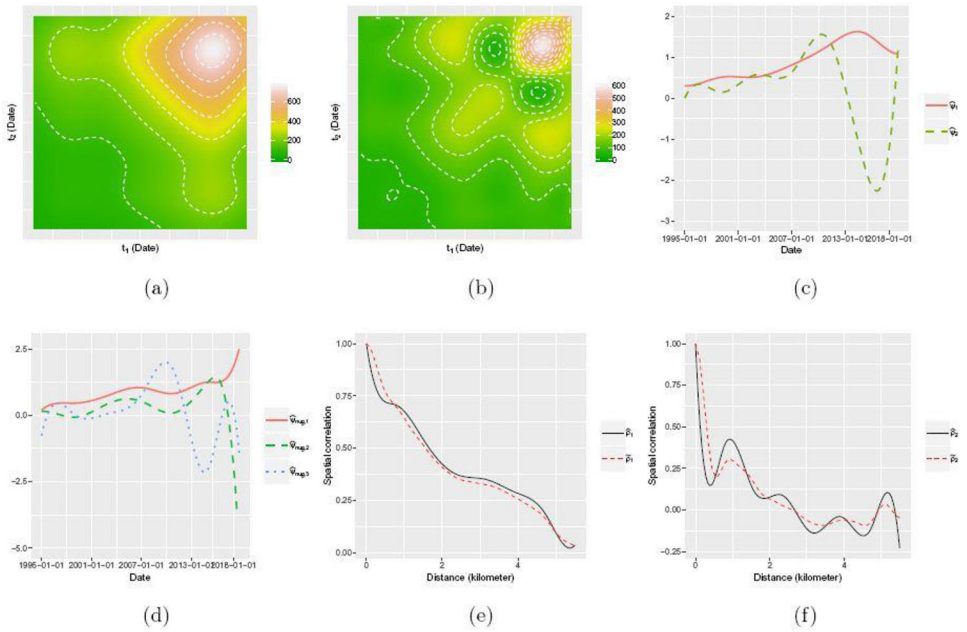


Fig. 5.

Results on the London housing price data: (a) contour plot of $\widehat{\Omega}(t_1, t_2)$; (b) contour plot of $\widehat{\Lambda}(t_1, t_2)$, covariance of the functional nugget effect; (c) the first two eigenfunctions of $\widehat{\Omega}(\cdot, \cdot)$; (d) the first three eigenfunctions of $\widehat{\Lambda}(\cdot, \cdot)$; (e) spatial correlation function $\hat{\rho}_1(\cdot)$ and its positive semi-definite adjustment $\tilde{\rho}_1(\cdot)$; (f) $\hat{\rho}_2(\cdot)$ and $\tilde{\rho}_2(\cdot)$.

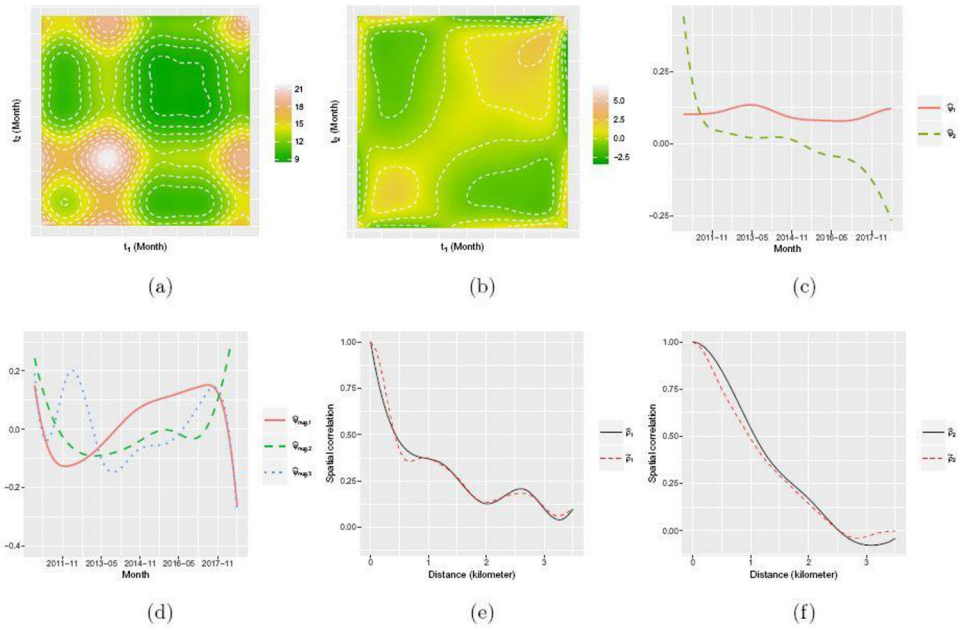


Fig. 6.

Results on the Zillow real estate data: (a) contour plot of $\widehat{\Omega}(t_1, t_2)$; (b) contour plot of $\widehat{\Lambda}(t_1, t_2)$, covariance function of the functional nugget effect; (c) the first two eigenfunctions (d) the first three eigenfunctions of $\widehat{\Lambda}(\cdot, \cdot)$ (e) the estimated spatial correlation function $\hat{\rho}_1(\cdot)$ and its positive semi-definite adjustment $\tilde{\rho}_1(\cdot)$; (f) $\hat{\rho}_2(\cdot)$ and $\tilde{\rho}_2(\cdot)$.

Table 1

Simulation results on the mean and standard deviation of integrated square errors for functional principal components estimated by *sFPCA* and *iFPCA*.

Simulation Scenario	**FPC**	***sFPCA***	***iFPCA***
Scenario A	ψ_1	0.076(0.104)	0.411(0.376)
	ψ_2	0.104(0.119)	0.367(0.369)
	ψ_3	0.077(0.071)	1.494(0.311)
	$\psi_{\text{nug},1}$	0.035(0.031)	–
	$\psi_{\text{nug},2}$	0.368(0.515)	–
Scenario B	ψ_1	0.073(0.114)	0.134(0.232)
	ψ_2	0.092(0.113)	0.123(0.232)
	ψ_3	0.061(0.043)	0.059(0.025)

Table 2

Kriging results in the simulation study: mean and standard deviation of the integrated squared errors for *sFPCA, iFPCA+CoKriging* and *Trace Kriging*.

Simulation Scenario	***sFPCA***	***iFPCA+CoKriging***	***Trace Kriging***
Scenario A	2.123(0.589)	5.147(0.989)	5.224(4.941)
Scenario B	1.563(0.704)	4.602(1.335)	5.073(4.846)