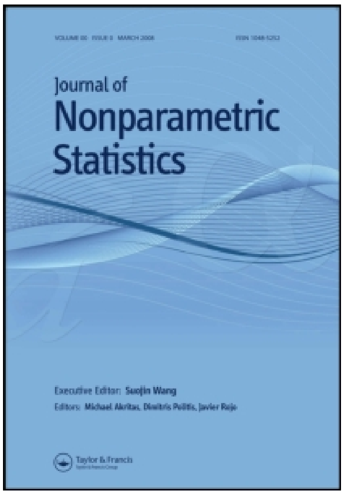

Journal of Nonparametric Statistics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713645758

Bivariate splines for spatial functional regression models

Serge Guillas [a]; Ming-Jun Lai [b]
[a] Department of Statistical Science & Aon Benfield UCL Hazard Research Centre, University College London, London, UK [b] Department of Mathematics, The University of Georgia, Athens, GA, USA

First published on: 01 December 2009

To cite this Article Guillas, Serge and Lai, Ming-Jun(2009) 'Bivariate splines for spatial functional regression models', Journal of Nonparametric Statistics,, First published on: 01 December 2009 (iFirst)
To link to this Article: DOI: 10.1080/10485250903323180
URL: http://dx.doi.org/10.1080/10485250903323180

Journal of Nonparametric Statistics
Vol. 00, No. 0, Month 2009, 1–21

Taylor & Francis
Taylor & Francis Group

Bivariate splines for spatial functional regression models

Serge Guillas[a]* and Ming-Jun Lai[b]

[a] *Department of Statistical Science & Aon Benfield UCL Hazard Research Centre, University College London, London WC1E 6BT, UK;* [b] *Department of Mathematics, The University of Georgia, Athens, GA 30602, USA*

(*Received 12 January 2009; final version received 8 September 2009*)

We consider the functional linear regression model where the explanatory variable is a random surface and the response is a real random variable, in various situations where both the explanatory variable and the noise can be unbounded and dependent. Bivariate splines over triangulations represent the random surfaces. We use this representation to construct least squares estimators of the regression function with a penalisation term. Under the assumptions that the regressors in the sample span a large enough space of functions, bivariate splines approximation properties yield the consistency of the estimators. Simulations demonstrate the quality of the asymptotic properties on a realistic domain. We also carry out an application to ozone concentration forecasting over the USA that illustrates the predictive skills of the method.

Keywords: functional data; regression; splines

AMS Subject Classification: 62G08; 65D07

1. Introduction

In various fields, such as environmental science, finance, geological science and biological science, large data sets are becoming readily available, e.g. by real time monitoring such as satellites circulating around the earth. Thus, the objects of statistical study are curves, surfaces and manifolds, in addition to the traditional points, numbers or vectors. Functional data analysis (FDA) can help represent and analyse infinite-dimensional random processes (Ramsay and Silverman 2005; Ferraty and Vieu 2006). FDA aggregates consecutive discrete recordings and views them as sampled values of a random curve or random surface, keeping track of order or smoothness. In this context, random curves have been the focus of many studies, but very few address the case of surfaces.

In regression, when the explanatory variable is a random function and the response is a real random variable, we can define the so-called functional linear model, see Chapter 15 in Ramsay and Silverman (2005) and references therein. In particular, Cardot, Ferraty, and Sarda (1999, 2003) introduced consistent estimates based on functional principal components, and decompositions in univariate splines spaces. The model can be generalised to the bivariate setting as follows.

*Corresponding author. Email: serge@stats.ucl.ac.uk

ISSN 1048-5252 print/ISSN 1029-0311 online

DOI: 10.1080/10485250903323180
http://www.informaworld.com

Let Y be a real-valued random variable. Let $\mathcal{D}$ be a polygonal domain in $\mathbf{R}^2$. The regression model is as follows:

$$Y = f(X) + \varepsilon = \langle g, X \rangle + \varepsilon = \int_{\mathcal{D}} g(s)X(s)\mathrm{d}s + \varepsilon, \tag{1}$$

where $g(s)$ is in a function space H (usually $= L^2(\mathcal{D})$), ε is a real random variable that satisfies $E\varepsilon = 0$ and $EX(s)\varepsilon = 0, \forall s \in \mathcal{D}$. One of the objectives in FDA is to determine or approximate g which is defined on a 2D spatial domain $\mathcal{D}$ from the observations on X obtained over a set of design points in $\mathcal{D}$ and Y.

This model in the univariate setting has been extensively studied using many different approaches. When the curves are supposed to be fully observed, it is possible to use the Karhunen–Loève expansion, or the principal components analysis for curves (Yao and Lee 2006; Cai and Hall 2006; Hall and Horowitz 2007). However, as pointed out by Hall, Muller, and Wang (2006), when the curves are not fully observed, which is obviously the case in practice, FDA would then proceed as though some smooth approximation of the observed curves were the collected ones. One typical approach is based on univariate splines (Cardot et al. 2003; Cardot, Crambes, and Sarda 2004; Cardot and Sarda 2005), whereas Cai and Hall (2006) and Hall and Horowitz (2007) use a local-linear smoother, which helps derive asymptotic results. Cardot et al. (2003) introduced the Penalised B-splines estimator and the smooth principal component regression estimator in one dimension. Finally, Crambes, Kneip, and Sarda (2009) considered the functional regression problem, using smoothing splines as well, but with a slightly modified penalty. They derived optimal rates of convergence for the error in the prediction based on random functions, as opposed to the case of a prediction error based on a fixed function covered in Cai and Hall (2006).

Motivated by the studies mentioned above, we investigate here the similar problem in the 2D setting. We consider a functional regression model where the explanatory variable is a random surface and the response is a real random variable. To express a random surface over a 2D irregular polygonal domain $\mathcal{D}$, we shall use bivariate splines which are smooth piecewise polynomial functions over a 2D triangulated polygonal domain $\mathcal{D}$. They are similar to univariate splines defined on piecewise subintervals. The theory of such bivariate spline functions has recently matured, see the monograph by Lai and Schumaker (2007). For example, we know the approximation properties of bivariate spline spaces and how to construct locally supported bases. Computational algorithms for scattered data interpolation and fitting are available in Awanou, Lai, and Wenston (2006). In particular, computing integrals with bivariate splines is easy, so it is now possible to use bivariate splines to build regression models for random surfaces. Certainly, it is possible to use the standard finite element (FE) method or thin-plate spline method for FDA, see Ramsay (2002) and Wood (2003) in a non-functional context. A FE analysis was carried out to smooth the data over complicated domains in Ramsay (2002) and thin-plate splines were used in regression in Wood (2003). Furthermore, it is also possible to use a tensor product of univariate splines or wavelets when the domain of interest is rectangular. Our bivariate splines are functions of piecewise polynomials which are more efficient than thin-plate splines. Also note that the basis functions for our spline spaces are Bernstein–Bézier polynomials over triangles which are locally supported and non-negative. The basis functions form a partition of unity, a stable basis and are suitable for computation. We find that our spline method is particularly easy to use, and hence will be used in our numerical experiments to be reported in the last section. We shall leave the investigation of using the FE method, thin-plate spline method and tensor product of univariate B-splines or wavelets for 2D FDA to the interested reader.

Our approach to FDA in the bivariate setting is a straightforward (called brute force) approach which is different from the approaches in Cardot et al. (1999, 2003, 2004) and Cardot and Sarda (2005). Mainly, we use the fact that the bivariate spline space can be dense in the standard $L_2(\mathcal{D})$ space and many other spaces as the size of triangulation decreases to zero. We can approximate

g and X in Equation (1) using spline functions and build a regression model. In our approach, we do not use the orthogonal expansion of covariance operator nor principal component analysis as in the standard functional regression approach. One significant difference of our spline approach for the functional linear model is that instead of using numerical quadrature, i.e. replacing $\int_0^1 \alpha(t)X(t)\mathrm{d}t$ by $\sum_{j=1}^{N} \alpha(t_j)X(t_j)s_j$ for some discrete points t_j with subinterval lengths $s_j = (t_j - t_{j-1})$, we approximate X by a spline fitting S_X based on the given data values $X(t_j)$ and data locations t_j (in our current research, these $t_j = (x_j, y_j)$ locate in a 2D domain) and approximate α by a spline function S_α (which may not be dependent on t_js) and then we compute $\int_\Omega S_{\alpha(x,y)} S_{X(x,y)} \mathrm{d}x\mathrm{d}y$ to approximate $\int_\Omega \alpha(x, y)X(x, y)\mathrm{d}x\mathrm{d}y$. Note that the inner products S_α and S_X can be computed easily based on our inner product formula for two polynomials over one triangle T (Lai and Schumaker 2007). In our approach, we may assume that the noise is bounded, or Gaussian, or unbounded under some moment assumptions, and we do not make explicit assumptions on the covariance structure of X. The only requirement in our approach is that all the random functions X span a large enough space so that g can be well estimated. It is a reasonable assumption. In this paper, we mainly derive rates of convergence in probability towards S_g, a spline approximation of g, of the empirical estimate when using bivariate splines to approximate X using a discrete least squares method and a penalised least squares method. We show that when the sample size n increases, empirical estimates converges to the spline estimator. In these theorems, the spline space dimension m is fixed. Indeed, the bivariate splines theory has already shown that as the size of triangulations goes to zero, and, thus, the dimension m of spline spaces becomes large, spline functions approximate any L_2 functions. We do know the convergence rate as m goes to infinity. However, in practice, we cannot make the size $|\triangle|$ as small as we wish due to the computing power and the limitation of the given data set. One has to fix a triangulation, degree d and smoothness r, and hence, the dimension m of spline space is fixed. The convergence of empirical estimates of S_g to g in the L_2 norm is currently under investigation by the authors with additional assumptions. We have implemented our approach using bivariate splines and performed numerical simulation, and forecasting with a set of real data. Comparison with univariate forecasting methods are given to show that our approach works very well. To our knowledge, our paper is the first piece of work on functional regression of a real random variable onto random surfaces.

The paper is organised as follows. After introducing bivariate splines in the preliminary section, we consider approximations of linear functionals with a penalty term in the next section. Then we address the case of discrete observations of random surfaces in Section 4. In order to illustrate the findings on an irregular region, in Section 5 we carry out simulations, and forecasting with real data, for which the domain is delimited by the US frontiers, and the sample points are the US Environmental Protection Agency (EPA) monitoring locations. Our numerical experiments demonstrate the efficiency and convenience of using bivariate splines to approximate linear functionals in functional data regression analysis.

2. Preliminary on bivariate splines

Let $\mathcal{D}$ be a polygonal domain in $\mathbf{R}^2$. Let $\triangle$ be a triangulation of $\mathcal{D}$ in the following sense: $\triangle$ is a collection of triangles $t \subset \mathcal{D}$ such that $\bigcup_{t\in\triangle} t = \mathcal{D}$ and the intersection of any two triangles $t_1, t_2 \in \triangle$ is either an empty set or their common edge of t_1, t_2 or their common vertex of t_1, t_2. For each $t \in \triangle$, let $|t|$ denote the longest length of the edges of t, and $|\triangle|$ the size of triangulation, which is the longest length of the edges of $\triangle$. Let $\theta_\triangle$ denote the smallest angle of $\triangle$. Next let $S_d^r(\triangle) = \{h \in C^r(\mathcal{D}), h|_t \in \mathbf{P}_d, t \in \triangle\}$ be the space of all piecewise polynomial functions h of degree d and smoothness r over $\triangle$, where $\mathbf{P}_d$ is the space of all polynomials of degree d. Such spline spaces have been studied in depth in the last 20 years and a basic theory and many important results are summarised in Lai and Schumaker (2007). Throughout the paper, $d \geq 3r + 2$. Then it is known (Lai and Schumaker 1998, 2007) that the spline space $S_d^r(\triangle)$ possesses an optimal approximation property: Let D_1 and D_2 denote the derivatives with respect to the first and second variables, $\|h\|_{L_p(\mathcal{D})}$ stand for the usual L_p norm of f over $\mathcal{D}$, $|h|_{m,p,\mathcal{D}}$ the L_p norm of the mth derivatives of h over $\mathcal{D}$ and $W_p^{m+1}(\mathcal{D})$ be the usual Sobolev space over $\mathcal{D}$.

THEOREM 2.1 *Suppose that $d \geq 3r + 2$ and $\triangle$ be a triangulation. Then there exists a quasi-interpolatory operator $Qh \in S_d^r(\triangle)$ mapping any $h \in L_1(\mathcal{D})$ into $S_d^r(\triangle)$ such that Qh achieves the optimal approximation order: if $h \in W_p^{m+1}(\mathcal{D})$,*

$$\|D_1^{\alpha} D_2^{\beta}(Qh - h)\|_{L_p(\mathcal{D})} \leq C|\triangle|^{m+1-\alpha-\beta}|h|_{m+1,p,\mathcal{D}} \tag{2}$$

for all $\alpha + \beta \leq m + 1$ with $0 \leq m \leq d$, where C is a constant which depends only on d and the smallest angle $\theta_{\triangle}$ and may be dependent on the Lipschitz condition of the boundary of $\mathcal{D}$.

Bivariate splines have been used for scattered data fitting and interpolation for many years. Typically, the minimal energy spline interpolation, discrete least squares splines for data fitting and penalised least squares splines for data smoothing as well as several other spline methods have been used. Their approximation properties have been studied and numerical algorithms for these data fitting methods have been implemented and tested. See Awanou et al. (2006), Lai (2007) and the references therein.

3. Approximation of linear functionals with penalty

In this section, we propose a new approach to study the functional f in model (1). We use a spline space $S_d^r(\triangle)$ with smoothness $r > 0$ and degree $d \geq 3r + 2$ over a triangulation $\triangle$ of a bounded domain $\mathcal{D} \subset \mathbf{R}^2$ with $|\triangle| < 1$ sufficiently small, i.e. enabling a good approximation (Awanou et al. 2006). The triangulation is fixed and, thus, the spline basis and its cardinality m as well. We study an approximation of the given functional f on the random functions X taking their values in H. Here H is a Hilbert space, for example, $H = W_2^{\nu}(\mathcal{D})$, the standard Sobolev space of all νth differentiable functions which are square integrable over $\mathcal{D}$ for an integer $\nu \geq r > 0$, where r is the smoothness of our spline space $S_d^r(\triangle)$.

We assume that X and Y follow the regression model (1). We seek a solution $\alpha \in H$ which solves the following minimisation problem:

$$\alpha = \arg\min_{\beta \in H} E[(Y - \langle \beta, X \rangle)^2] + \rho\|\beta\|_r^2, \tag{3}$$

where $\rho > 0$ is a parameter and $\|\beta\|_r^2$ denotes the semi-norm of β: $\|\beta\|_r^2 = \mathcal{E}_r(\beta, \beta)$, where $\mathcal{E}_r(\alpha, \beta) = \int_{\mathcal{D}} \sum_{k=0}^{r} \sum_{i+j=k} D_1^i D_2^j \alpha D_1^i D_2^j \beta$, and D_1 and D_2 stand for the partial derivatives with respect to the first and second variables. Unless the penalty is equal to zero, α is not necessarily equal to g. Since $S_d^r(\triangle)$ can be dense in H as $|\triangle| \to 0$, we consider a spline space $S_d^r(\triangle)$ for a smoothness $r \geq 0$ and degree $d > r$ over a triangulation $\triangle$ of $\mathcal{D}$ with $|\triangle|$ sufficiently small. Note that the triangulation is fixed and, thus, the spline basis and its cardinality m as well. We look for an approximation $S_{\alpha,\rho} \in S_d^r(\triangle)$ of α such that

$$S_{\alpha,\rho} = \arg\min_{\beta \in S_d^r(\triangle)} E[(Y - \langle \beta, X \rangle)^2] + \rho\mathcal{E}_r(\beta). \tag{4}$$

We now analyse how $S_{\alpha,\rho}$ approximates α in terms of the size $|\triangle|$ of triangulation and $\rho \to 0+$. Let $\{\phi_1, \ldots, \phi_m\}$ be a basis for $S_d^r(\triangle)$. We write $S_\alpha = \sum_{j=1}^m c_j\phi_j$. Then a direct calculation of the least squares solution of Equation (4) entails that the coefficient vector $\mathbf{c} = (c_1, \ldots, c_m)^{\mathrm{T}}$ satisfies a linear system $A\mathbf{c} = \mathbf{b}$ with A being a matrix of size $m \times m$ with entries $E(\langle \phi_i, X\rangle\langle \phi_j, X\rangle) + \rho\mathcal{E}_r(\phi_i, \phi_j)$ for $i, j = 1, \ldots, m$ and $\mathbf{b}$ being a vector of length m with entries $E(Y\langle \phi_j, X\rangle)$ for $j = 1, \ldots, m$.

Although we do not know how $X \in H$ is distributed, let us assume that only the zero polynomial is orthogonal to all functions in the collection $\mathcal{X} = X(\omega), \omega \in \Omega$ in the standard Hilbert space $L_2(\mathcal{D})$. This means that the random variables X are distributed in such a way that they generate a high-dimensional subspace of $L_2(\mathcal{D})$. In this case, A is invertible. Otherwise, we would have $\mathbf{c}^{\mathrm{T}} A\mathbf{c} = 0$, i.e.

$$E\left(\left\langle \sum_{i=1}^m c_i\phi_i, X\right\rangle\right)^2 + \rho\left\|\sum_{i=1}^m c_i\phi_i\right\|_r^2 = 0. \tag{5}$$

Since the second term in Equation (5) is equal to zero, $\sum_{i=1}^m c_i\phi_i$ is a polynomial of degree $< r$. As the first term in Equation (5) is also zero, this polynomial is orthogonal to X for all $X \in \mathcal{X}$. By the assumption, $\sum_{i=1}^m c_i\phi_i$ is a zero spline and hence, $c_i = 0$ for all i. Thus, we have obtained the following theorem.

THEOREM 3.1 *Suppose that only the zero polynomial is orthogonal to the collection $\mathcal{X}$ in $L_2(\mathcal{D})$. Then the minimisation problem* (4) *has a unique solution in $S_d^r(\triangle)$.*

To see that $S_{\alpha,\rho}$ is a good approximation of α, we let $\{\phi_j,\ j = m+1, m+2, \ldots\}$ be a basis of the orthogonal complement space of $S_d^r(\triangle)$ in $L_2(\mathcal{D})$. Then we can write $\alpha = \sum_{j=1}^\infty c_j\phi_j$. Note that the minimisation in Equation (3) yields $E(\langle \alpha, X\rangle\langle \phi_j, X\rangle) + \rho\mathcal{E}_r(\alpha, \phi_j) = E(f(X)\langle \phi_j, X\rangle)$ for all $j = 1, 2, \ldots$ while the minimisation in Equation (4) gives

$$E(\langle S_\alpha, X\rangle\langle \phi_j, X\rangle) + \rho\mathcal{E}_r(S_\alpha, \phi_j) = E(f(X)\langle \phi_j, X\rangle)$$

for all $j = 1, 2, \ldots, m$. It follows that

$$E(\langle \alpha - S_{\alpha,\rho}, X\rangle\langle \phi_j, X\rangle) + \rho\mathcal{E}_r(\alpha - S_{\alpha,\rho}, \phi_j) = 0 \tag{6}$$

for $j = 1, \ldots, m$. Let Q_α be the quasi-interpolatory spline in $S_d^r(\triangle)$ which achieves the optimal order of approximation of α from $S_d^r(\triangle)$ as in the preliminary section. Then Equation (6) implies that

$$\begin{aligned}
E((\langle \alpha - S_{\alpha,\rho}, X\rangle)^2) &= E(\langle \alpha - S_{\alpha,\rho}, X\rangle\langle \alpha - Q_\alpha, X\rangle) - \rho\mathcal{E}_r(\alpha - S_{\alpha,\rho}, Q_\alpha - S_{\alpha,\rho}) \\
&\le (E((\langle \alpha - S_{\alpha,\rho}, X\rangle)^2))^{1/2} E((\langle \alpha - Q_\alpha, X\rangle)^2)^{1/2} \\
&\quad - \rho\|\alpha - S_{\alpha,\rho}\|_r^2 + \rho\mathcal{E}_r(\alpha - S_{\alpha,\rho}, \alpha - Q_\alpha) \\
&\le \tfrac{1}{2}E((\langle \alpha - S_{\alpha,\rho}, X\rangle)^2) + \tfrac{1}{2}E((\langle \alpha - Q_\alpha, X\rangle)^2) \\
&\quad - \tfrac{1}{2}\rho\|\alpha - S_{\alpha,\rho}\|_r^2 + \tfrac{1}{2}\rho\|\alpha - Q_\alpha\|_r^2.
\end{aligned}$$

Hence $E((\langle \alpha - S_{\alpha,\rho}, X\rangle)^2) + \rho\|\alpha - S_{\alpha,\rho}\|_r^2 \le E((\langle \alpha - Q_\alpha, X\rangle)^2) + \rho\|\alpha - Q_\alpha\|_r^2$. The approximation of the quasi-interpolant Q_α of α (Lai and Schumaker 1998) gives:

THEOREM 3.2 *Suppose that $E(\|X\|^2) < \infty$ and suppose $\alpha \in C^\nu(\mathcal{D})$ for $\nu \ge r$. Then the solution $S_{\alpha,\rho}$ from the minimisation problem* (4) *approximates α: $E((\langle \alpha - S_{\alpha,\rho}, X\rangle)^2) \le C|\triangle|^{2\nu}E(\|X\|^2) + \rho C|\triangle|^{2(\nu-r)}$ where C is a positive constant independent of the size $|\triangle|$ of triangulation $\triangle$.*

Next we consider the empirical estimate of $S_{\alpha,\rho}$. Let X_i, $i = 1, \ldots, n$ be a sequence of functional random variables such that only the zero polynomial is perpendicular to the subspace spanned by $\{X_1, \ldots, X_n\}$ except on an event whose probability p_n goes to zero as $n \to +\infty$. The empirical estimate $\widehat{S_{\alpha,\rho,n}} \in S_d^r(\triangle)$ is the solution of

$$\widehat{S_{\alpha,\rho,n}} = \arg \min_{\beta \in S_d^r(\triangle)} \frac{1}{n} \sum_{i=1}^{n} (Y_i - \langle \beta, X_i \rangle)^2 + \rho \|\beta\|_r^2, \tag{7}$$

with $\rho > 0$ the smoothing parameter. The solution of the above minimisation is given by $\widehat{S_{\alpha,\rho,n}} = \sum_{i=1}^{m} c_{n,i} \phi_i$ with the coefficient vector $\mathbf{c}_n = (c_{n,i}, i = 1, \ldots, m)$ satisfying $\hat{A}_n \mathbf{c}_n = \hat{b}_n$, where

$$\hat{A}_n = \left[\frac{1}{n} \sum_{\ell=1}^{n} \langle \phi_i, X_\ell \rangle \langle \phi_j, X_\ell \rangle + \rho \mathcal{E}_r(\phi_i, \phi_j) \right]_{i,j=1,\ldots,m}$$

and

$$\hat{b}_n = \left[\frac{1}{n} \sum_{\ell=1}^{n} Y_\ell \langle \phi_j, X_\ell \rangle \right]_{j=1,\ldots,m} = \left[\frac{1}{n} \sum_{\ell=1}^{n} (f(X_\ell) + \epsilon_\ell) \langle \phi_j, X_\ell \rangle \right]_{j=1,\ldots,m}.$$

THEOREM 3.3 *Suppose that only the zero polynomial is perpendicular to the subspace spanned by $\{X_1, \ldots, X_n\}$ except on an event whose probability p_n goes to zero as $n \to +\infty$. Then there exists a unique $\widehat{S_{\alpha,\rho,n}} \in S_d^r(\triangle)$ minimising* (7) *with probability* $1 - p_n$.

Proof It is straightforward to see that the coefficient vector of $\widehat{S_{\alpha,\rho,n}}$ satisfies the above relations. To see that $\hat{A}_n \mathbf{c}_n = \hat{b}_n$ has a unique solution, we claim that if $\hat{A}_n c' = 0$, then $c' = 0$. It follows that $(c')^{\mathrm{T}} \hat{A}_n c' = 0$, i.e. $\sum_{\ell=1}^{n} (\langle \sum_{i=1}^{m} c_i' \phi_i, X_\ell \rangle)^2 = 0$. That is, $\sum_{i=1}^{m} c_i' \phi_i$ is orthogonal to X_ℓ, $\ell = 1, \ldots, n$. According to the assumption, $c' = 0$ except for an event whose probability p_n goes to zero when $n \to +\infty$. ■

We now prove that $\widehat{S_{\alpha,\rho,n}}$ approximates $S_{\alpha,\rho}$ in probability. For simplicity, we consider the case where the penalty is equal to zero as the entries of $A - \hat{A}_n$ and $\mathbf{b} - \hat{b}$ are exactly the same with or without penalty. To this end we need the following lemmas.

LEMMA 3.4 *Suppose that $\triangle$ is a β-quasi-uniform triangulation* (*cf. Lai and Schumaker 2007*). *There exist two positive constants C_1 and C_2 independent of $\triangle$ such that for any spline function $S \in S_d^r(\triangle)$ with coefficient vector $\mathbf{s} = (s_1, \ldots, s_m)^{\mathrm{T}}$ with $S = \sum_{i=1}^{m} s_j \phi_j$, $C_1 |\triangle|^2 \|\mathbf{s}\|^2 \leq \|S\|^2 \leq C_2 |\triangle|^2 \|\mathbf{s}\|^2$.*

A proof of this lemma can be found in Lai and Schumaker (1998, 2007). The following lemma is well known in numerical analysis (Golub and Van Loan 1989, p. 82).

LEMMA 3.5 *Let A be an invertible matrix and $\tilde{A}$ be a perturbation of A satisfying $\|A^{-1}\| \, \|A - \tilde{A}\| < 1$. Suppose that x and $\tilde{x}$ are the exact solutions of $Ax = b$ and $\tilde{A}\tilde{x} = \tilde{b}$, respectively. Then*

$$\frac{\|x - \tilde{x}\|}{\|x\|} \leq \frac{\kappa(A)}{1 - \kappa(A) \|A - \tilde{A}\| / \|A\|} \left[\frac{\|A - \tilde{A}\|}{\|A\|} + \frac{\|b - \tilde{b}\|}{\|b\|} \right].$$

Here, $\kappa(A) = \|A\| \|A^{-1}\|$ denotes the condition number of matrix A.

The next Lemma will be used to find the resulting upper bounds for the differences $\|S_\alpha - \widehat{S_{\alpha,\rho,n}}\|$.

LEMMA 3.6 *Let* $\beta = \|\boldsymbol{c} - \hat{c}_n\|/\|\boldsymbol{c}\|$, $\eta = \|A - \hat{A}_n\|/\|A\|$ *and* $\theta = \|\boldsymbol{b} - \hat{b}_n\|/\|\boldsymbol{b}\|$. *For all* $\delta \le 1$, *we have*

$$P\left(\frac{\|S_\alpha - \widehat{S_{\alpha,\rho,n}}\|}{\|S_\alpha\|} \ge \delta\right) \le 2P\left(\eta \ge \frac{\gamma\delta}{4\kappa(A)}\right) + P\left(\theta \ge \frac{\gamma\delta}{4\kappa(A)}\right)$$

where $\gamma = \sqrt{C_1/C_2}$ *from Lemma* 3.4.

Proof We first use Lemma 3.4 to get $P(\|S_\alpha - \widehat{S_{\alpha,\rho,n}}\|/\|S_\alpha\| \ge \delta) \le P(\|\mathbf{c} - \hat{c}_n\|/\|\mathbf{c}\| \ge \gamma\delta)$ where $\gamma = \sqrt{C_1/C_2}$. Then Lemma 3.5 implies that

$$\begin{aligned}
P(\beta \ge \gamma\delta) &\le P\left(\beta \ge \gamma\delta, \kappa(A)\eta \le \frac{1}{2}\right) + P\left(\beta \ge \gamma\delta, \kappa(A)\eta \ge \frac{1}{2}\right)\\
&\le P\left(\frac{\kappa(A)}{1-\kappa(A)\eta}(\eta+\theta) \ge \gamma\delta, \kappa(A)\eta \le \frac{1}{2}\right) + P\left(\kappa(A)\eta \ge \frac{1}{2}\right)\\
&\le P\left((\eta+\theta) \ge \frac{\gamma\delta}{2\kappa(A)}\right) + P\left(\kappa(A)\eta \ge \frac{1}{2}\right)\\
&\le P\left(\eta \ge \frac{\gamma\delta}{4\kappa(A)}\right) + P\left(\theta \ge \frac{\gamma\delta}{4\kappa(A)}\right) + P\left(\eta \ge \frac{\gamma\delta}{2\kappa(A)}\right)\\
&\le 2P\left(\eta \ge \frac{\gamma\delta}{4\kappa(A)}\right) + P\left(\theta \ge \frac{\gamma\delta}{4\kappa(A)}\right)
\end{aligned}$$

for all $\delta \le 1$. ■

Thus, we need to analyse the differences between the entries of A and $\hat{A}_n$ as well as the differences between $\mathbf{b}$ and $\hat{b}_n$. Let: $\xi^{(1)}_{i,j,l} = \langle\phi_i, X_l\rangle\langle\phi_j, X_l\rangle$, $\xi^{(2)}_{j,l} = f(X_l)\langle\phi_j, X_l\rangle$, and $\xi^{(3)}_{j,l} = \varepsilon_l\langle\phi_j, X_l\rangle$. We can find rates of convergence by applying exponential inequalities that will be valid uniformly over the entries of $\xi^{(p)}$ for $p = 1, 2, 3$.

To use Lemma 3.5, we employ for convenience (all norms are equivalent) the maximum norm for matrix $A - \hat{A}_n$ and vector $b - \hat{b}_n$. For simplicity, let us write

$$[a_{ij}]_{1\le i,j\le m} = A - \hat{A}_n = \left[\frac{1}{n}\sum_{\ell=1}^{n}(\xi^{(1)}_{i,j,l} - E(\xi^{(1)}_{i,j,l})\right]_{1\le i,j\le m}.$$

We have

LEMMA 3.7

$$P(\|[a_{ij}]_{1\le i,j\le m}\|_\infty \ge \delta) \le \sum_{i=1}^{m}\sum_{j=1}^{m} P\left(|a_{ij}| \ge \frac{\delta}{m})\right)$$

and, *if the probabilities* $P(|a_{ij}| \ge \delta/m))$ *are bounded for all* i, j *by the same quantity* $h(\delta, m)$,

$$P\left(\|[a_{ij}]_{1\le i,j\le m}\|_\infty \ge \delta\right) \le m^2 h(\delta, m).$$

Proof

$$P(\|[a_{ij}]_{1\le i,j\le m}\|_\infty \ge \delta) = P\left(\max_{1\le i\le m}\sum_{j=1}^m |a_{ij}| \ge \delta\right)$$
$$\le \sum_{i=1}^m P\left(\sum_{j=1}^m |a_{ij}| \ge \delta\right)$$
$$\le \sum_{i=1}^m \sum_{j=1}^m P\left(|a_{ij}| \ge \frac{\delta}{m}\right). \qquad \blacksquare$$

Similar to Lemma 3.7, we can estimate the entries of $\mathbf{b} - \hat{b}_n$. We denote its entries by $b_j = -(1/n)\sum_{\ell=1}^n f(X_\ell)\langle\phi_j, X_\ell\rangle - E(f(X)\langle\phi_j, X\rangle) + (1/n)\sum_{\ell=1}^n \epsilon_\ell\langle\phi_j, X_\ell\rangle$. Let us write $b_j = b_j^1 + b_j^2$ with b_j^1 and b_j^2 being the first and second terms, respectively. It is easy to see that $P(|b_j| \ge \delta) \le P(|b_j^1| \ge \delta/2) + P(|b_j^2| \ge \delta/2)$. Since the functional f is bounded, $|f(X_\ell)| \le F\|X_\ell\|$, with a finite constant F. We obtain immediately the following Lemma.

Lemma 3.8

$$P(\|\boldsymbol{b} - \hat{b}_n\|_\infty \ge \delta) \le \sum_{j=1}^m P\left(|b_j^1| \ge \frac{\delta}{2}\right) + P\left(|b_j^2| \ge \frac{\delta}{2}\right)$$

and, if the probabilities $P(|b_j^1| \ge \delta/2)$ and $P(|b_j^2| \ge \delta/2)$ are, respectively, bounded for all j by the same quantities $h^1(\delta)$ and $h^2(\delta)$,

$$P(\|\boldsymbol{b} - \hat{b}_n\|_\infty \ge \delta) \le m(h^1(\delta) + h^2(\delta)).$$

We consider the first case where the variables are bounded, for which we can apply the following Hoeffding's exponential inequality (Bosq 1998, p. 24).

Lemma 3.9 *Let $\{\xi_l\}_{l=1}^n$ be n independent random variables. Suppose that there exists a positive number M such that for each l, $|\xi_l| \le M < \infty$ almost surely. Then $P(|1/n\sum_{\ell=1}^n(\xi_l - E(\xi_l))| \ge \delta) \le 2\exp(-n\delta^2/2M^2)$ for $\delta > 0$.*

Theorem 3.10 *Suppose that X_ℓ, $\ell = 1, \ldots, n$ are independent and identically distributed and X_1 is bounded almost surely. Suppose that the ϵ_ℓ are independent and bounded almost surely. Assume that $f(X)$ is a bounded linear functional. Then $\widehat{S_{\alpha,\rho,n}}$ converges to S_α in probability with convergence rate*

$$P\left(\frac{\|S_\alpha - \widehat{S_{\alpha,\rho,n}}\|}{\|S_\alpha\|} \ge \delta\right) \le 4m^2\exp\left(-\frac{n\gamma^2\delta^2}{32\kappa(A)^2m^2M^2}\right) + 2m\exp\left(-\frac{n\gamma^2\delta^2}{128\kappa(A)^2M_b^2}\right)$$
$$+ 2m\exp\left(-\frac{n\gamma^2\delta^2}{128\kappa(A)^2M_\epsilon^2}\right). \tag{8}$$

Proof The basis spline functions ϕ_j can be chosen to be bounded in $L_2(\mathcal{D})$ for all j independent of triangulation $\triangle$ (Lai and Schumaker 2007). The $\xi_{i,j,\ell}^{(1)}$ are bounded. Indeed, let $M = \max_{ij} \max_\ell |\langle\phi_i, X_\ell\rangle\langle\phi_j, X_\ell\rangle| \leq \max_{ij} \max_\ell \|\phi_i\|\|\phi_j\|\|X_\ell\|^2$. For each i, j, $|\xi^{(1)}_{i,j,\ell}| \leq M < \infty$ almost surely. We can apply Lemma 3.9 to get

$$P(\|[a_{ij}]_{1\leq i,j\leq m}\|_\infty \geq \delta) \leq 2m^2 \exp\left(-\frac{n\delta^2}{2m^2M^2}\right). \tag{9}$$

By Lemma 3.9, we also have $P(|b^1_j| \geq \delta/2) \leq 2\exp(-n\delta^2/8M_b^2)$, where $M_b = \max_j |f(X_\ell)\langle\phi_j, X_\ell\rangle| \leq F\|X_\ell\|\|\phi_j\|\,\|X_\ell\|$ which is a finite quantity since $\|X_\ell\|$ is bounded almost surely. Regarding the second term b^2_j, since the random noises ϵ_ℓ are bounded almost surely, we apply Lemma 3.9 to $\xi^3_{j,\ell}$ and it yields: $P(|b^2_j| \geq \delta/2) \leq 2\exp(-n\delta^2/8M_\epsilon^2)$ where $M_\epsilon = \max_j |\langle\phi_j, \epsilon_\ell X_\ell\rangle| \leq \max_j \|\phi_j\||\epsilon_\ell|\|X_\ell\|$ which is finite under the assumption that both $\|X_\ell\|$ and $|\epsilon_\ell|$ are bounded almost surely.

Thus, we have by Lemma 3.8

$$P(\|\mathbf{b} - \hat{b}_n\|_\infty \geq \delta) \leq 2m \exp\left(-\frac{n\delta^2}{8M_b^2}\right) + 2m \exp\left(-\frac{n\delta^2}{8M_\epsilon^2}\right). \tag{10}$$

We combine the estimates (9) and (10) to get Equation (8). ■

As an example, if we choose $m = n^{1/4}$, we get a convergence rate of $n^{1/2}\exp(-\sqrt{n}\gamma^2\delta^2/32\,\kappa(A)^2M^2)$ which is the slower of the terms.

We are now ready to consider the case where ϵ_ℓ is a Gaussian noise $N(0, \sigma_\ell^2)$ for $\ell = 1, \ldots, n$. Instead of Lemma 3.9, it is easy to prove.

Lemma 3.11 *Suppose that ϵ_ℓ is a Gaussian noise $N(0, \sigma_\ell^2)$ for $\ell = 1, \ldots, n$. Then*

$$P\left(\left|\frac{1}{n}\sum_{\ell=1}^n \epsilon_\ell\right| > \delta\right) \leq \exp\left(-\frac{n^2\delta^2}{2\sum_{\ell=1}^n \sigma_\ell^2}\right).$$

Theorem 3.12 *Suppose that X_ℓ, $\ell = 1, \ldots, n$ are independent and identically distributed random variables and X_1 is bounded almost surely. Suppose ϵ_ℓ are independent and identically distributed as a Gaussian noise $N(0, \sigma^2)$ and $f(X)$ is a bounded linear functional. Then $\widehat{S_{\alpha,\rho,n}}$ converges to S_α in probability with the convergence rate:*

$$P\left(\frac{\|S_\alpha - \widehat{S_{\alpha,\rho,n}}\|}{\|S_\alpha\|} \geq \delta\right) \leq 4m^2 \exp\left(-\frac{n\gamma^2\delta^2}{32\kappa(A)^2m^2M^2}\right) + 2m \exp\left(-\frac{n\gamma^2\delta^2}{128\kappa(A)^2M_b^2}\right)$$
$$+ 2m \exp\left(-\frac{n\gamma^2\delta^2}{128\kappa(A)^2\sigma^2C^2}\right). \tag{11}$$

Proof By Lemma 3.11, $P(|(1/n)\sum_{\ell=1}^n(\epsilon_\ell Z_\ell))| \geq \delta) \leq \exp(-n\delta^2/2\sigma^2C^2)$ for $\delta > 0$, under the assumption that Z_ℓ are independent random variables which are bounded by C, i.e. $\|Z_\ell\| \leq C$. Similar to the proof of Theorem 3.10, with $Z_\ell = \langle\phi_j, X_\ell\rangle$ in that case, we obtain the convergence rate in Equation (11). ■

We now extend the results to the case where both the explanatory variables X_n and noise ε_n are dependent and unbounded. We state two types of results based on dependence conditions of either association or mixing types that involve making moment assumptions on the variables of interest.

By definition, a sequence of real-valued variables $Y_1, Y_2, \ldots$ is positively associated (PA) (Esary, Proschan, and Walkup 1967) if, for every integer n, and every function f, g $\mathbf{R}^n \to \mathbf{R}$ is coordinatewise increasing, we have:

$$\mathrm{Cov}(f(Y_1, \ldots, Y_n), g(Y_1, \ldots, Y_n)) \geq 0.$$

The resulting rates are usually not exponential but geometric in the 'exponential' inequalities (Oliveira 2005, Theory 5.1). We note that Henriques and Oliveira (2005) relate assumptions of positive association for a transformation of a process. We could try to set up a new definition of positive association for Hilbert-valued random variables and see what it implies on the variables $\xi_l^{(p)}$ (we drop for convenience the other indices i, j in the sequel.) However, it would require a thorough study of these quantities $\xi_l^{(p)}$ for $p = 1, 2, 3$, and is beyond the scope of this paper. We have the following result.

THEOREM 3.13 *Suppose that for $p = 1, 2, 3$, the time series $\xi_l^{(p)}$, $\ell = 1, \ldots, n$, are strictly stationary and PA. Suppose that they all satisfy the following assumptions uniformly in i, j. Suppose that Oliveira* (2005), *Equation* (13) *is satisfied as follows:*

$$\frac{1}{p_n \log n} \exp\left(\left(\frac{\tau n \log n}{2p_n}\right)^{1/2}\right) \sum_{l=p_n+2}^{\infty} \mathrm{Cov}(\xi_1^{(p)}, \xi_l^{(p)}) \leq C_0 < \infty, \tag{12}$$

where $p_n = n\varepsilon^2/54\tau \log^3 n$, for $\varepsilon > 0$ small enough. Assume also that there exists $\lambda > \tau$ such that $\sup_{|t|\leq\lambda} E[\exp(t\xi_1^{(p)})] \leq M_\lambda < \infty$. Then $\widehat{S_{\alpha,\rho,n}}$ converges to S_α in probability, for $\delta > 0$ small enough and n large enough:

$$P\left(\frac{\|S_\alpha - \widehat{S_{\alpha,\rho,n}}\|}{\|S_\alpha\|} \geq \delta\right) \leq m^2\left(2\left(1 + \frac{4}{\tau}C_0\right) + \frac{192M_\lambda m^2 \kappa(A)^2}{\tau\|A\|^2\gamma^2\delta^2}\right)n^{1-\tau}$$
$$+2m\left(2\left(1 + \frac{4}{\tau}C_0\right) + \frac{768M_\lambda \kappa(A)^2}{\tau\|b\|^2\gamma^2\delta^2}\right)n^{1-\tau}. \tag{13}$$

As a result, $\widehat{S_{\alpha,\rho,n}}$ converges to S_α in probability with convergence rate in Equation (13).

Proof We use Lemmas 3.6–3.8 and then employ Oliveira (2005), Theory 5.1 for PA unbounded variables to get an exponential inequality in these cases. ■

An example of this situation is the autoregressive case, when $\mathrm{Cov}(\xi_1^{(p)}, \xi_l^{(p)}) = \rho_0\rho^n$, for some $\rho_0 > 0$ and $0 < \rho^n < 1$. See the discussion in Oliveira (2005, Section 5).

Another possibility is to consider mixing assumptions. However, these are more difficult to check than covariance-based conditions, see Doukhan and Louhichi (1999) for a discussion. Classical autoregressive moving average (ARMA) processes have mixing coefficients which decrease to zero at an exponential rate. For a strictly stationary time series (X_n), the strong mixing (or α-mixing) coefficient of order k is

$$\alpha(k) = \sup_{B\in\sigma(X_s, s\leq n),\quad C\in\sigma(X_s, s\geq n+k)} |P(B\cap C) - P(B)P(C)|$$

We apply Bosq (1998), Theory 1.4 in the strictly stationary case to derive the following result:

THEOREM 3.14 *Suppose that for* $p = 1, 2, 3$, *the time series* $\xi_l^{(p)}$, $\ell = 1, \ldots, n$, *are strictly stationary, and there exists* $c > 0$ *such that*

$$E[|\xi_1^{(p)}|] \leq c^{k-2} k! E[(\xi_1^{(p)})^2] < \infty,$$

for all $k \geq 3$. *Then for* $n \geq 2$, *each integer* $q \in [1, n/2]$, *for all* $\delta > 0$, $k \geq 3$,

$$P\left(\frac{\|S_\alpha - \widehat{S_{\alpha,\rho,n}}\|}{\|S_\alpha\|} \geq \delta\right) \leq 2m^2\left(a_1(\varepsilon_1)\exp\left(-\frac{q\varepsilon_1^2}{25m_2^2 + 5c\varepsilon_1}\right) + a_2(\varepsilon_1, k)\alpha\left(\left[\frac{n}{q+1}\right]\right)\right)$$
$$+ 2m\left(a_1(\varepsilon_2)\exp\left(-\frac{q\varepsilon_2^2}{25m_2^2 + 5c\varepsilon_2}\right) + a_2(\varepsilon_2, k)\alpha\left(\left[\frac{n}{q+1}\right]\right)\right), \tag{14}$$

where

$$a_1(\varepsilon) = 2\frac{n}{q} + 2\left(1 + \frac{\varepsilon^2}{25m_2^2 + 5c\varepsilon}\right)$$
$$a_2(\varepsilon, k) = 11n\left(1 + \frac{5m_k^{k/(2k+1)}}{\varepsilon}\right)$$

with $\varepsilon_1 = \|A\|\gamma\delta/4m\kappa(A)$, $\varepsilon_2 = \|b\|\gamma\delta/8\kappa(A)$, $m_2^2 = E[(\xi_1^{(p)})^2]$ *and* $m_k = (E[(\xi_1^{(p)})^k])^{1/k}$.

As a result, $\widehat{S_{\alpha,\rho,n}}$ *converges to* S_α *in probability with convergence rate in Equation* (14).

Proof We use Lemmas 3.6–3.8 and then employ Bosq (1998), Theory 1.4 for unbounded variables with mixing assumptions to get an exponential inequality in these cases. ■

By choosing say $q = \log n$ or $q = n/4$, we could achieve an explicit convergence rate if the strong mixing coefficients converge to zero. And depending on the case, one could find optimal values for q to achieve the best rates possible balancing the terms in the right-hand side. Note also that Bosq (2000), Theory 2.13 gives a similar result but for Hilbert-valued random variables directly. As a remark for future research, one could make assumptions on the Hilbert-valued processes themselves to retrieve similar rates. Furthermore, we could mix the various cases covered in this section, for instance by assuming that some of the processes $\xi_l^{(p)}$, $\ell = 1, \ldots, n$ satisfy the assumptions of Theorem 3.13 or 3.14, leading to various combined rates of convergence. Finally, one could derive almost sure convergence theorems as well via Borel–Cantelli's lemma in a straightforward way in these cases, assuming that the stronger conditions on the probabilities are met.

4. Approximation of linear functionals based on discrete observations

In practice, we do not know X completely over the domain $\mathcal{D}$. Instead, we have observations of X over some designed points s_k, $k = 1, \ldots, N$ over $\mathcal{D}$. Let S_X be the discrete least square fit spline approximation (Awanou et al. 2006) of X, assuming that s_k, $k = 1, \ldots, N$ are evenly distributed over $\triangle$ of $\mathcal{D}$ with respect to $S_d^r(\triangle)$. We consider α_S that solves the following minimisation problem:

$$\alpha_S = \arg\min_{\beta\in H} E[(Y - \langle\beta, S_X\rangle)^2] + \rho\|\beta\|_r^2. \tag{15}$$

Also we look for an approximation $S_{\alpha_S} \in S_d^r(\triangle)$ of α_S such that

$$S_{\alpha_S} = \arg\min_{\beta\in S_d^r(\triangle)} E[(Y - \langle\beta, S_X\rangle)^2] + \rho\|\beta\|_r^2. \tag{16}$$

We first analyse how α_S approximates α. It is easy to see that

$$F(\beta) = E[(Y - \langle\beta, X\rangle)^2]$$

is a strictly convex function and so is $F_S(\beta) = E\left[(Y - \langle\beta, S_X\rangle)^2\right] + \rho\|\beta\|_r^2$. Note that S_X approximates X very well as in Theorem 2.1 as $|\triangle| \to 0$. Thus, $F_S(\beta)$ approximates $F(\beta)$ for each β. Since the strictly convex function has a unique minimiser and both $F(\beta)$ and $F_S(\beta)$ are continuous, α_S approximates α. Indeed, if $\alpha_S \to \beta \neq \alpha$, then $F(\alpha) < F(\beta) = F_S(\beta) + \eta_1 = F_S(\alpha_S) + \eta_1 + \eta_2 \leq F_S(\alpha) + \eta_1 + \eta_2 = F(\alpha_S) + \eta_1 + \eta_2 + \eta_3$ for arbitrary small $\eta_1 + \eta_2 + \eta_3$. Thus, we would get the contradiction $F(\alpha) < F(\alpha)$.

We now begin to analyse how S_{α_S} approximates α_S in terms of the size $|\triangle|$ of triangulation. Recall that $\{\phi_1, \ldots, \phi_m\}$ forms a basis for $S_d^r(\triangle)$. We write $S_{\alpha_S} = \sum_{j=1}^m c_{S,j}\phi_j$. Then its coefficient vector $\mathbf{c}_S = (c_{S,1}, \ldots, c_{S,m})^{\mathrm{T}}$ satisfies $A_S\mathbf{c}_S = \mathbf{b}_S$ with A_S being a matrix of size $m \times m$ with entries $E(\langle\phi_i, S_X\rangle\langle\phi_j, S_X\rangle)$ for $i, j = 1, \ldots, m$ and $\mathbf{b}_S$ being a vector of length m with entries $E((Y)\langle\phi_j, S_X\rangle)$ for $j = 1, \ldots, m$. We can show that A_S converges to A as $|\triangle| \to 0$ because $E(\langle\phi_i, S_X\rangle\langle\phi_j, S_X\rangle) \to E(\langle\phi_i, X\rangle\langle\phi_j, X\rangle)$ as $S_X \to X$ by Theorem 2.1. That is, we have $\|S_X - X\|_{\infty,\mathcal{D}} \leq C|\triangle|^\nu|X|_{\nu,\infty,\mathcal{D}}$ for $X \in W_2^\nu(\mathcal{D})$ with $\nu \geq r > 0$.

To see that S_{α_S} is a good approximation of α_S, we let $\{\phi_j, j = m+1, m+2, \ldots\}$ be a basis of the orthogonal complement space of $S_d^r(\triangle)$ in H as before. Then we can write $\alpha_S = \sum_{j=1}^\infty c_{S,j}\phi_j$. Note that the minimisation in Equation (15) yields $E(\langle\alpha, S_X\rangle\langle\phi_j, S_X\rangle) = E((Y)\langle\phi_j, S_X\rangle)$ for all $j = 1, 2, \ldots$ while the minimisation in Equation (16) gives

$$E(\langle S_{\alpha_S}, S_X\rangle\langle\phi_j, S_X\rangle) = E(Y\langle\phi_j, S_X\rangle)$$

for all $j = 1, 2, \ldots, m$. It follows that

$$E(\langle\alpha_S - S_{\alpha_S}, S_X\rangle\langle\phi_j, S_X\rangle) = 0, \tag{17}$$

for all $j = 1, 2, \ldots, m$. Let Q_α be the quasi-interpolatory spline in $S_d^r(\triangle)$ which achieves the optimal order of approximation of α_S from $S_d^r(\triangle)$ as in the preliminary section. Then Equation (17) implies that

$$\begin{aligned} E((\langle\alpha_S - S_{\alpha_S}, S_X\rangle)^2) &= E(\langle S_\alpha - S_{\alpha_S}, S_X\rangle\langle\alpha_S - Q_{\alpha_S}, S_X\rangle) \\ &\leq (E((\langle\alpha_S - S_{\alpha_S}, S_X\rangle)^2))^{1/2} E((\langle\alpha_S - Q_{\alpha_S}, S_X\rangle)^2)^{1/2}. \end{aligned}$$

It yields $E((\langle\alpha_S - S_{\alpha_S}, S_X\rangle)^2) \leq E((\langle\alpha_S - Q_{\alpha_S}, S_X\rangle)^2) \leq \|\alpha_S - Q_{\alpha_S}\|_H^2 E(\|S_X\|^2)$. The convergence of S_X to X implies that $E(\|S_X\|^2)$ is bounded by a constant dependence on $E(\|X\|^2)$. The approximation of the quasi-interpolant Q_{α_S} of α_S (Theorem 2.1) gives:

THEOREM 4.1 *Suppose that $E(\|X\|^2) < \infty$ and suppose $\alpha \in C^r(\mathcal{D})$ for $r \geq 0$. Then the solution S_{α_S} from the minimisation problem* (16) *approximates α_S in the following sense: $E((\langle\alpha_S - S_{\alpha_S}, S_X\rangle)^2) \leq C|\triangle|^{2r}$ for a constant C dependent on $E(\|X\|^2)$, where $|\triangle|$ is the maximal length of the edges of $\triangle$.*

Next we consider the empirical estimate of S_α based on discrete observations of random surfaces X_i, $i = 1, \dots, n$. The empirical estimate $\widetilde{S_{\alpha,\rho,n}} \in S_d^r(\triangle)$ is the solution of

$$\widetilde{S_{\alpha,\rho,n}} = \arg\min_{\beta \in S_d^r(\triangle)} \frac{1}{n} \sum_{i=1}^{n} (Y_i - \langle \beta, S_{X_i} \rangle)^2 + \rho \|\beta\|_r^2.$$

In fact the solution of the above minimisation is given by $\widetilde{S_{\alpha,\rho,n}} = \sum_{i=1}^m \widetilde{c_{n,i}} \phi_i$ with coefficient vector $\widetilde{\mathbf{c}_n} = (\widetilde{c_{n,i}}, i = 1, \dots, m)$ satisfying $\widetilde{A_n} \widetilde{\mathbf{c}_n} = \widetilde{b_n}$, and

$$\widetilde{A_n} = \left[\frac{1}{n} \sum_{\ell=1}^{n} \langle \phi_i, S_{X_\ell} \rangle \langle \phi_j, S_{X_\ell} \rangle + \rho \mathcal{E}_r(\phi_i, \phi_j) \right]_{i,j=1,\dots,m},$$

where S_{X_ℓ} is the discrete least squares fit of X_ℓ and

$$\widetilde{b_n} = \left[\frac{1}{n} \sum_{\ell=1}^{n} Y_\ell \langle \phi_j, S_{X_\ell} \rangle \right]_{j=1,\dots,m}.$$

Recall the definition of $\hat{A}_n$ in Section 3. We have

$$\widetilde{A_n} - \hat{A}_n = \left[\frac{1}{n} \sum_{\ell=1}^{n} \langle \phi_i, S_{X_\ell} \rangle \langle \phi_j, S_{X_\ell} \rangle - \frac{1}{n} \sum_{\ell=1}^{n} \langle \phi_i, X_\ell \rangle \langle \phi_j, X_\ell \rangle \right]_{i,j=1,\dots,m}.$$

As S_{X_ℓ} converges X_ℓ as $|\triangle| \to 0$, i.e. $S_{X_\ell} - X_\ell = O(|\triangle|^\nu)$, we can show that $\|\widetilde{A_n} - \widehat{A_n}\|_\infty = O(|\triangle|^{\nu-2})$ and hence, $\|\widetilde{A_n} - \widehat{A_n}\|_\infty \to 0$ if $\nu > 2$. Likewise, $\widetilde{b_n} - \widehat{b_n}$ converges to 0. We consider here the case with no penalty for convenience. Lemma 3.5 implies that $\widetilde{S_{\alpha,\rho,n}}$ converges to $\hat{S}_{\alpha,\rho,n}$ as $|\triangle| \to 0$ under certain assumptions on X_ℓ, $\ell = 1, \dots, n$ with $n > m$ and $\nu > 4$. Indeed, let us assume that the surfaces X_ℓ, $\ell = 1, \dots, n$ are orthonormal and span a space which contains $S_d^r(\triangle)$ (or form a tight frame of a space which contains $S_d^r(\triangle)$.) Then we can show that the condition numbers $\kappa(\widehat{A_n})$ are bounded by n. Note that the condition number of $\kappa(\widehat{A_n})$ can be computed as the modulus of the ratio of the largest and smallest eigenvalues of the matrix. It is known that the largest eigenvalue $\lambda_{\max}$ and smallest eigenvalue $\lambda_{\min}$ of the matrix $\widehat{A_n}$ satisfy

$$\lambda_{\min} = \min_{\mathbf{c} \in \mathbf{R}^m} \frac{\mathbf{c}^{\mathrm{T}} \widehat{A_n} \mathbf{c}^{\mathrm{T}}}{\mathbf{c}^{\mathrm{T}} \mathbf{c}} \le \max_{\mathbf{c} \in \mathbf{R}^m} \frac{\mathbf{c}^{\mathrm{T}} \widehat{A_n} \mathbf{c}^{\mathrm{T}}}{\mathbf{c}^{\mathrm{T}} \mathbf{c}} = \lambda_{\max}.$$

Writing $\mathbf{c} = (c_1, \dots, c_m)^{\mathrm{T}}$, we let $S = \sum_{i=1}^m c_i \phi_i \in S_d^r(\triangle)$. Then by Lemma 3.4, $\lambda_{\max}$ and $\lambda_{\min}$ are equivalent to

$$\max_{S \in S_d^r(\triangle)} \frac{1}{n\|S\|_2^2} \sum_{\ell=1}^{n} |\langle S, X_\ell \rangle|^2 \le \frac{1}{n} \sum_{\ell=1}^{n} \|X_\ell\|_2^2 = 1$$

and

$$\min_{S \in S_d^r(\triangle)} \frac{1}{n\|S\|_2^2} \sum_{\ell=1}^{n} |\langle S, X_\ell \rangle|^2 = \frac{1}{n}.$$

Let us further assume that $n = Cm$ for some fixed constant $C > 1$. Next, we note that the dimension of $S_d^r(\triangle)$ is strictly less than $(d + 2/2)N$ with N being the number of triangles in $\triangle$ while N can be estimated as follows. Let $A_{\mathcal{D}}$ be the area of the underlying domain $\mathcal{D}$ and assume that the triangulation $\triangle$ is quasi-uniform (cf. Lai and Schumaker 2007). Then $N \leq C_1 A_{\mathcal{D}}/|\triangle|^2$ for a positive constant C_1. Thus, the condition number $\kappa(\widehat{A_n}) \leq Cm \leq CC_1 A_{\mathcal{D}}|\triangle|^{-2}$. That is, $\kappa(\widehat{A_n})\|\widetilde{A_n} - \widehat{A_n}\|_\infty/\|\widehat{A_n}\|_\infty = O(|\triangle|^{\nu-4})$. Therefore, Lemma 3.5 implies that the coefficients of $\widetilde{S_{\alpha,\rho,n}}$ converges to that of $\hat{S}_{\alpha,\rho,n}$ as $|\triangle| \to 0$ when $\nu > 4$. With Lemma 3.4, we conclude that $\widetilde{S_{\alpha,\rho,n}}$ converges to $\hat{S}_{\alpha,\rho,n}$.

A similar analysis can be carried out for the approximation with a penalised term. The details are omitted here. Instead, we shall present the convergence based on our numerical experimental results in the next section.

5. Numerical simulation and experiments

5.1. *Simulations*

In this subsection, we present a simulation example on a complicated domain, delimited by the US frontiers, which has been scaled into [0, 1] × [0, 1], see Figure 1. With bivariate spline functions, we can easily carry out all the experiments.

We illustrate the consistency of our estimators using the linear functional: $Y = \langle g, X\rangle$ with known function $g(x, y) = \sin(2\pi(x^2 + y^2))$ over the (scaled) US domain. The purpose of the simulation is to estimate g from the value Y based on random surfaces X. The bivariate spline space we employed is $S_5^1(\triangle)$, where $\triangle$ consists of 174 triangles (Figure 1).

We choose a sample size $n = 5$, 20, 100, 200, 500 and 1000. For each $i = 1, \ldots, n$, we first randomly choose a vector $\mathbf{c}_i$ of size m which is the dimension of $S_5^1(\triangle)$. This coefficient vector $\mathbf{c}_i$ defines a spline function S_i. We evaluate S_i over the (scaled) locations of 969 stations from the US EPA around the USA, and add a small noise with zero mean and standard deviation 0.4 at each location. We compute a least squares fit $\tilde{S}_i$ of the resulting 969 values by using the spline space $S_5^1(\triangle)$ and compute the inner product of g and $\tilde{S}_i$. We add a small noise of zero mean and standard deviation 0.0002 to get a noisy value Y_i of the functional.

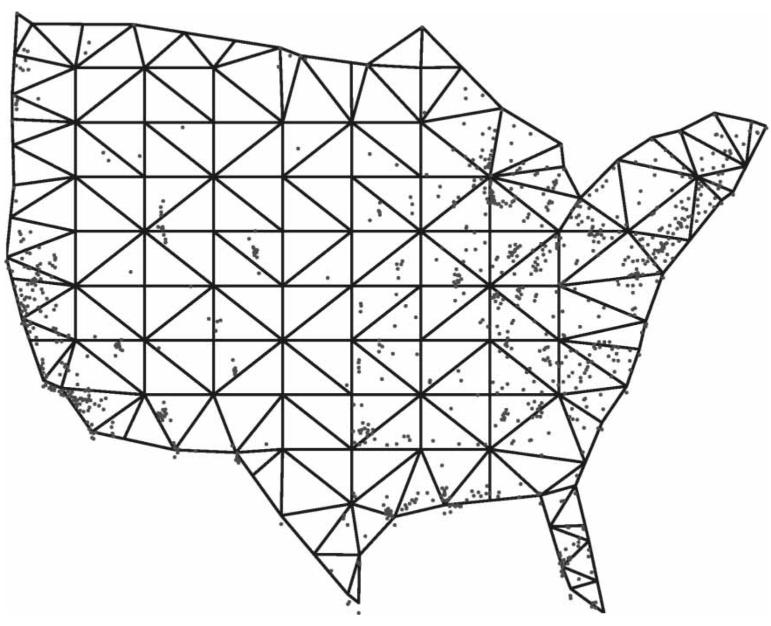

Figure 1. Locations of EPA stations and a triangulation.

Second, we build the associated matrix $\widetilde{A_n}$ as in Section 4 and the right-hand side vector $\widetilde{b_n}$, for which we use a penalty of $\rho = 10^{-9}$. Finally, we solve the linear equation to get the solution vector $\mathbf{c}$ and spline approximation $\widetilde{S_{g,\rho,n}}$ of g. We then evaluate g and $\widetilde{S_{g,\rho,n}}$ at locations which are the 101×101 equally spaced points over $[0, 1] \times [0, 1]$ that fall into the US domain, to compute their differences and find their maximum as well as L_2 norm. We carry out a Monte Carlo experiment over 20 different random seeds. The numerical results show that we approximate well the linear functional, see Table 1. An example of $S_{g,\rho,500}$ is shown in Figure 2. Note that in this study, the signal-to-noise ratio is around 10. We tried various large signal-to-noise ratios, with satisfying results not reported here. Further theoretical and applied studies of how the results of the estimation varies according to the signal-to-noise ratio are interesting. We leave them for future research.

Table 1. Errors for the differences $\widetilde{S_{\alpha,\rho,n}} - S_\alpha$ for the simulation and sample sizes n =5, 20, 100, 200, 500 and 1000 based on 20 Monte Carlo simulations and 174 triangles.

	L^2 error		
Sample size (n)	Min	Mean	Max
5	0.671	2.195	31.821
20	0.427	0.564	0.666
100	0.080	0.115	0.153
200	0.048	0.060	0.081
500	0.036	0.040	0.044
1000	0.029	0.032	0.035
	L^∞ error		
5	1.242	1.988	3.086
20	1.398	2.221	3.584
100	0.336	0.468	0.717
200	0.158	0.254	0.534
500	0.112	0.136	0.207
1000	0.092	0.102	0.123

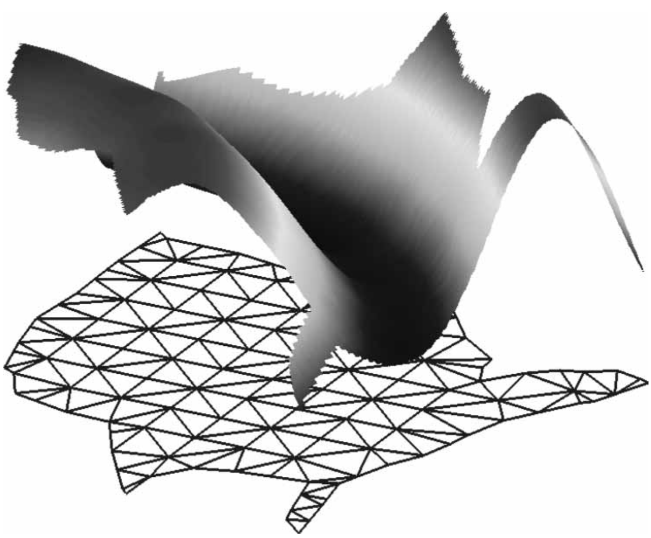

Figure 2. The surface of spline approximation $S_{g,n}$.

5.2. *Ozone concentration forecasting*

In this application, we forecast the ground-level ozone concentration at the center of Atlanta using the random surfaces over the entire US domain based on the measurements at various EPA stations from the previous days. Assume that the ozone concentration in Atlanta on one day at a particular time is a linear functional of the ozone concentration distribution over the US continent on the previous day. Also we may assume that the linear functional is continuous. These are reasonable assumptions as the concentration in Atlanta is proportional to the concentration distribution over the entire US continent and a small change in the concentration distribution over the US continent results in a small change of the concentration at Atlanta under a normal circumstance. Thus, we build one regression model of the type (1), where $f(X)$ is the ozone concentration value at the centre of Atlanta at one hour of one day, X is the ozone concentration distribution function over the entire US continent at the same hour but on the previous day, and g is estimated using the penalised least squares approximation with penalty ($=10^{-2}$) presented in the previous section. Let us outline our computational scheme as follows.

Step 1. Based on the observations X over 969 EPA station around the US at a given hour of a given day, we compute a penalised least squares fit spline S_X with penalised parameter $= 10^{-2}$, where S_X is a spline function of degree 5 and smoothness 1 over the triangulation given in the previous subsection. Let f_X be the ozone concentration at Atlanta at the given hour of the day after the given day.

Step 2. We find a spline function S_A of degree 5 and smoothness 1 over the same triangulation which solves the following minimisation problem

$$\min_{s \in S_5^1(\triangle)} \frac{1}{24N} \sum_{i=1}^{24N} (f_{X_i} - \langle s, S_{X_i} \rangle)^2$$

for N days. To predict the ozone value at Atlanta on September 8, we use all the observations over N days before and on September 6 as well as ozone values f_{X_i} at Atlanta on September 7.

Step 3. Based on the ozone values Z over the USA at a given hour on September 7, we compute a penalised least squares fit S_Z and then compute the inner product S_Z with S_A to predict the ozone value at the given hour on September 8. We compute the predictions based on N day learning period along these lines for various values of N. We use a penalised least squares fit S_X of X instead of the discrete least squares fit in the previous subsection to carry out the empirical estimate $\widetilde{S_{\alpha,\rho,n}}$ for S_g. See Awanou et al. (2006) for an explanation and discussion of bivariate splines for data fitting.

For computational efficiency, we actually used only one quarter of the triangulations of the whole US continent to generate the predictions. The triangulation of this region (southeastern region of the US) is shown in Figure 3. From Figures 4–10, it is easy to see that our spline predictions are very close to the true measurements. In particular, they are consistent for various learning periods. For more experimental results based on various size of triangulations and regions, see Ettinger (2009).

This may be compared with the univariate functional autoregressive ozone concentration prediction method (Damon and Guillas 2002), but here with no exogenous variables. The idea is to consider a time series of functions which correspond to the ozone concentrations at the location of interest over 24 hours, and then build an autoregressive Hilbertian model for this time series. The estimation of the autocorrelation operator in a reduced subspace enables predictions. We selected only five functional principal components in the dimension reduction process to keep

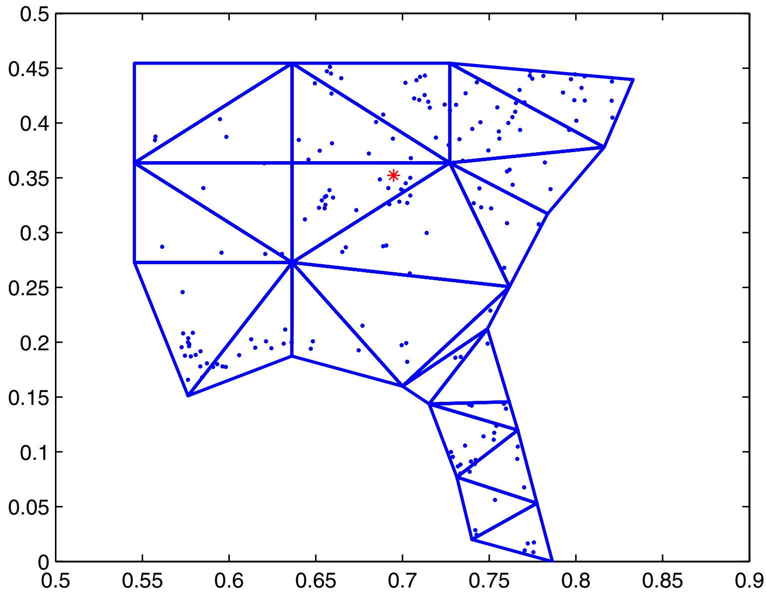

Figure 3. Locations of EPA stations and a triangulation of the Southeastern US. The star is the location of the Atlanta observation station used for predictions.

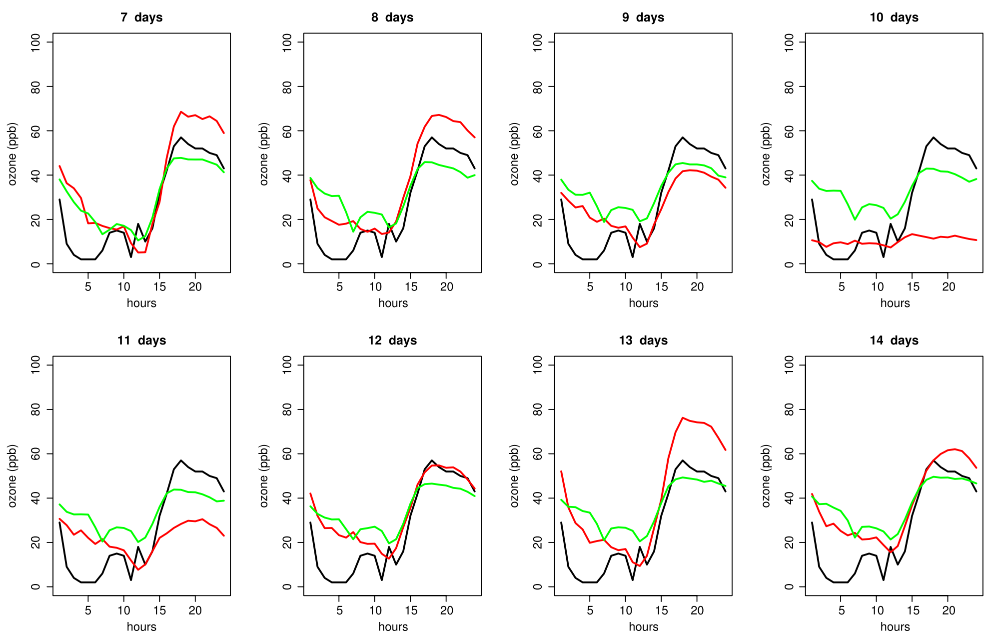

Figure 4. Ozone concentrations in Atlanta on 8 September 2005. Observations (black), forecast 1D (red), forecast 2D (green).

parsimony in our model, due to sample sizes (i.e number of days considered) of 7–14. As we see on Figures 5–8, the forecasts provided by the 2D spline strategy outperforms the univariate functional autoregressive method based on the same sizes of samples. This may be explained by the fact that the 2D approach uses more information to construct its forecasts. The comparisons show

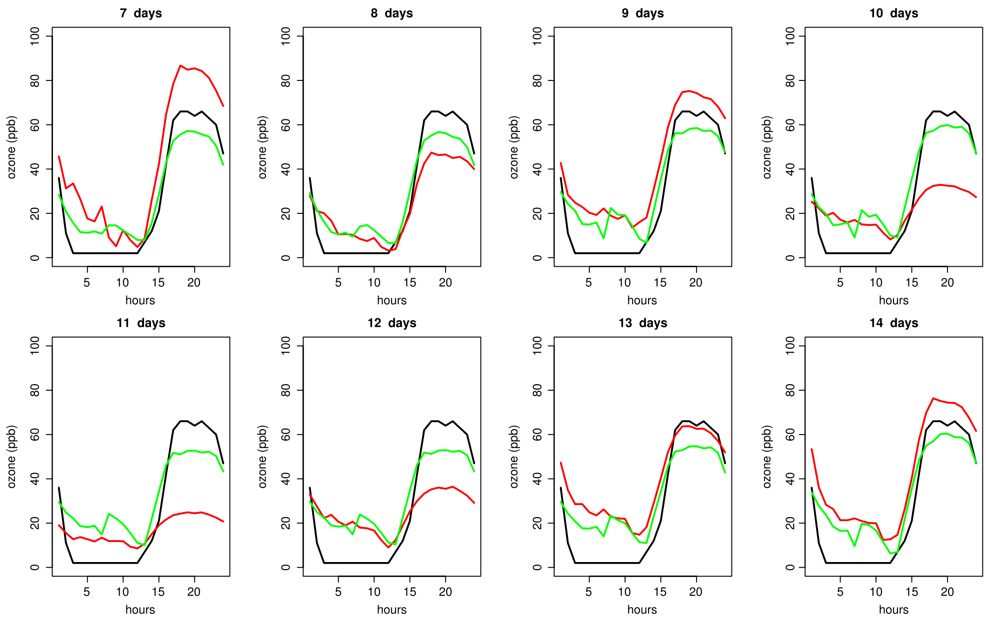

Figure 5. Ozone concentrations in Atlanta on 9 September 2005. Observations (black), forecast 1D (red), forecast 2D (green).

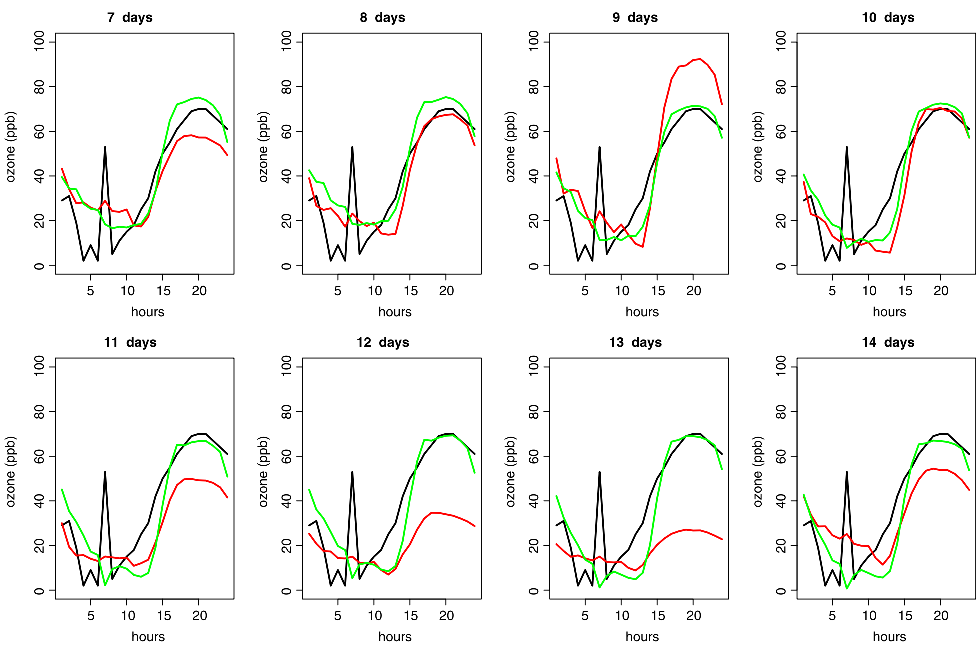

Figure 6. Ozone concentrations in Atlanta on 11 September 2005. Observations (black), forecast 1D (red), forecast 2D (green).

that our bivariate spline technique almost consistently predicts the ozone concentration values which are closer to the observed values for these 5 days for various learning periods, especially near the peaks. The 1D method presented in this paper, which is considered to be among the best of many 1D forecasting methods (Damon and Guillas 2002), is not consistent for various learning

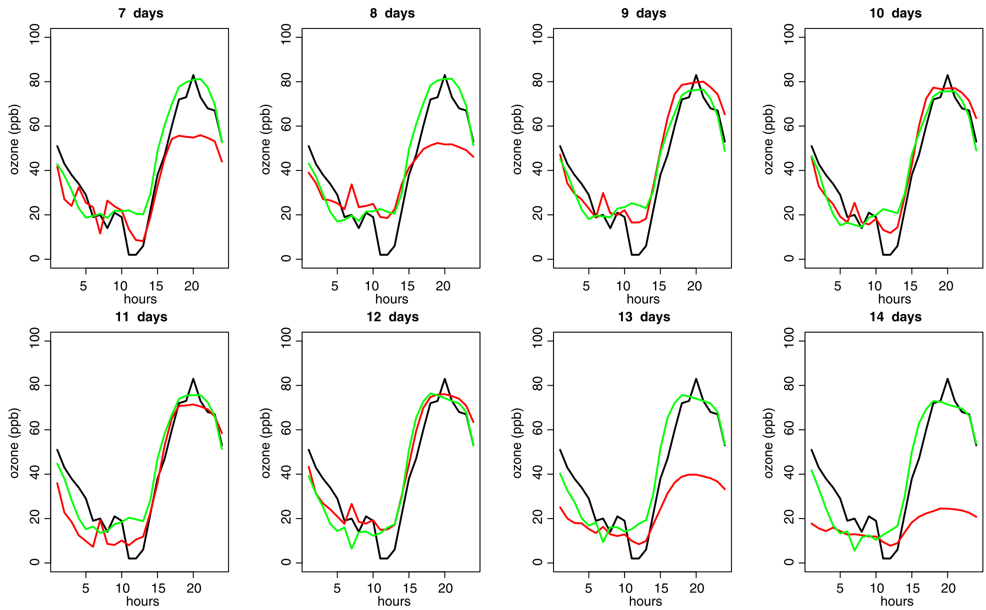

Figure 7. Ozone concentrations in Atlanta on 12 September 2005. Observations (black), forecast 1D (red), forecast 2D (green).

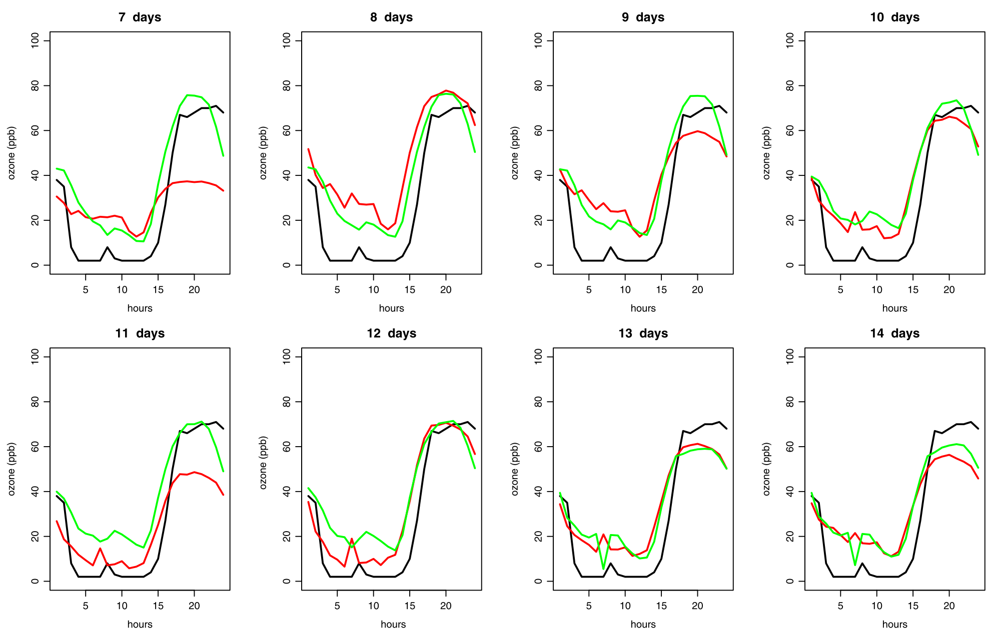

Figure 8. Ozone concentrations in Atlanta on 13 September 2005. Observations (black), forecast 1D (red), forecast 2D (green).

periods and the patterns based on the 1D method are not as close to the exact measurements as those based on the bivariate spline method most of the time. This could be explained by the very small sample size. The 2D method naturally borrows strength across space and does not suffer as much from the lack of data.

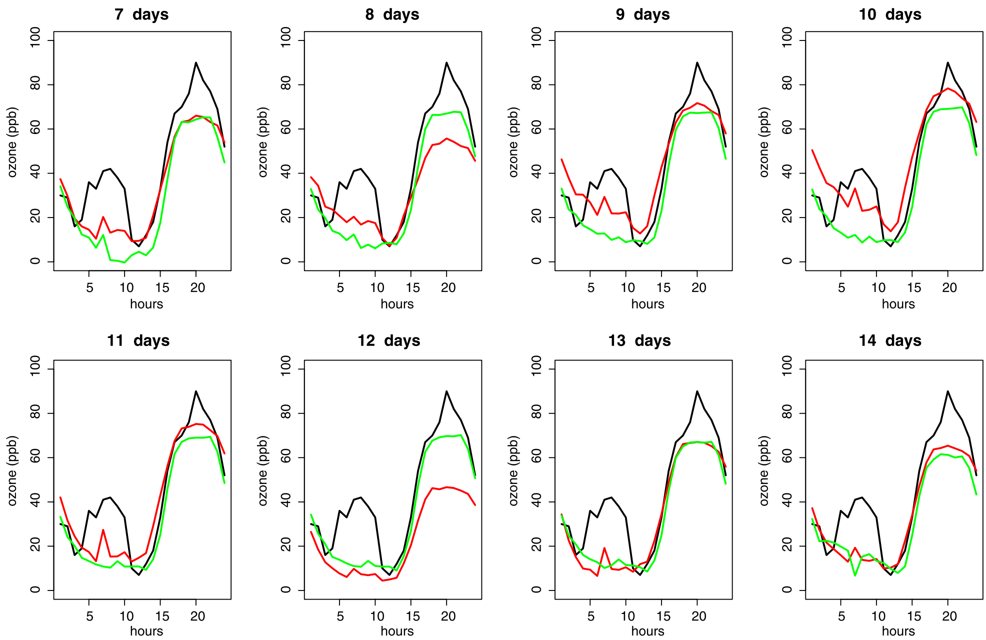

Figure 9. Ozone concentrations in Atlanta on 14 September 2005. Observations (black), forecast 1D (red), forecast 2D (green).

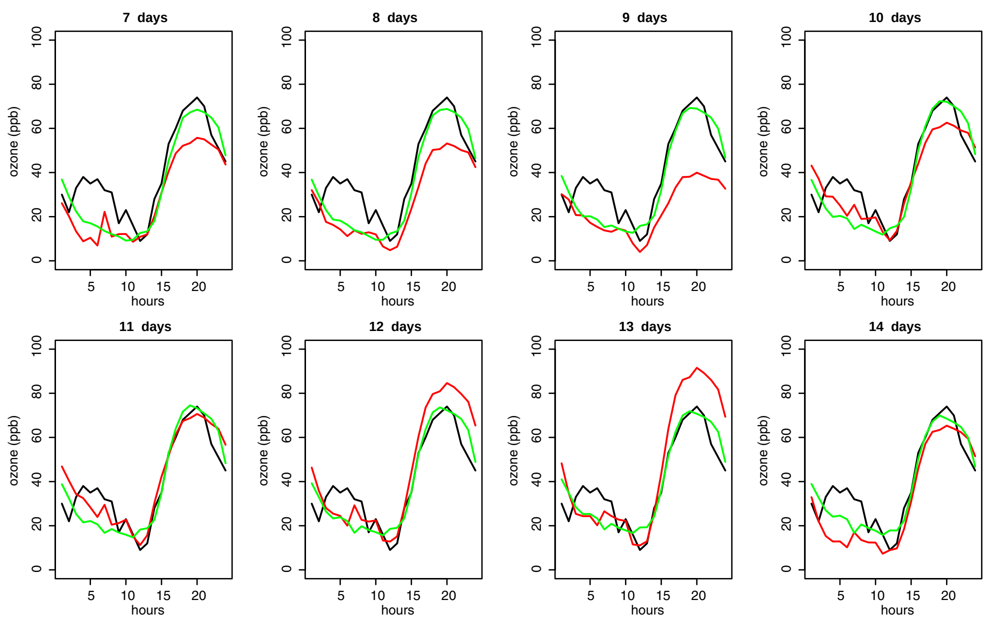

Figure 10. Ozone concentrations in Atlanta on 15 September 2005. Observations (black), forecast 1D (red), forecast 2D (green).

Finally, we remark that we are currently studying the autoregressive approach using orthonormal expansion in a bivariate spline space for the ozone concentration prediction (cf. Ettinger 2009) and numerical results as well as comparison of both approaches will be available soon. Our study shows that to determine how many eigenvalues and eigenfunctions should be used for the best prediction is not easy.

Acknowledgements

The authors would like to thank Ms Bree Ettinger for the help in performing the numerical experiments using 2D splines for ozone concentration predictions reported in this paper. We thank the referees for their very constructive comments that enabled us to greatly improve the results and the presentation of the paper.

References

Awanou, G., Lai, M.-J., and Wenston, P. (2006), 'The Multivariate Spline Method for Numerical Solution of Partial Differential Equations and Scattered Data Interpolation', in *Wavelets and Splines: Athens 2005*, eds. G. Chen and M.-J. Lai, Nashville, TN: Nashboro Press, pp. 24–74.

Bosq, D. (1998), *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, Lecture Notes in Statistics (Vol. 110), NewYork: Springer-Verlag.

Bosq, D. (2000), *Linear Processes in Function Spaces: Theory and Applications*, Lecture Notes in Statistics (Vol. 149), NewYork: Springer-Verlag.

Cai, T.T., and Hall, P. (2006), 'Prediction in Functional Linear Regression', *The Annals of Statistics*, 34(5), 2159–2179.

Cardot, H., and Sarda, P. (2005), 'Estimation in Generalized Linear Models for Functional Data via Penalized Likelihood', *Journal of Multivariate Analysis*, 92, 24–41.

Cardot, H., Ferraty, F., and Sarda, P. (1999), 'Functional Linear Model', *Statistics & Probability Letters*, 45, 11–22.

Cardot, H., Ferraty, F., and Sarda, P. (2003), 'Spline Estimators for the Functional Linear Model', *Statistica Sinica*, 13, 571–591.

Cardot, H., Crambes, C., and Sarda, P. (2004), 'Spline Estimation of Conditional Quantiles for Functional Covariates', *Comptes Rendus Mathematique*, 339, 141–144.

Crambes, C., Kneip, A., and Sarda, P. (2009), 'Smoothing Splines Estimators for Functional Linear Regression', *The Annals of Statistics*, 37(1), 35–72.

Damon, J., and Guillas, S. (2002), 'The Inclusion of Exogenous Variables in Functional Autoregressive Ozone Forecasting', *Environmetrics*, 13, 759–774.

Doukhan, P., and Louhichi, S. (1999), 'A New Weak Dependence Condition and Applications to Moment Inequalities', *Stochastic Processes and their Applications*, 84(2), 313–342.

Esary, J.D., Proschan, F., and Walkup, D.W. (1967), 'Association of Random Variables, With Applications', *The Annals of Mathematical Statistics*, 38(5), 1466–1474.

Ettinger, B. (2009), 'Bivariate Splines for Ozone Density Predictions', Ph.D. thesis, University of Georgia (in preparation).

Ferraty, F., and Vieu, P. (2006), *Nonparametric Functional Data Analysis: Theory and Practice*, Springer Series in Statistics, London: Springer-Verlag.

Golub, G.H., and Van Loan, C.F. (1989), *Matrix Computations*, Johns Hopkins Series in the Mathematical Sciences (Vol. 3), Baltimore, MD: Johns Hopkins University Press.

Hall, P., and Horowitz, J.L. (2007), 'Methodology and Convergence Rates for Functional Linear Regression', *The Annals of Statistics*, 35(1), 70–91.

Hall, P., Muller, H.G., and Wang, J.L. (2006), 'Properties of Principal Component Methods for Functional and Longitudinal Data Analysis', *The Annals of Statistics*, 34, 1493–1517.

Henriques, C., and Oliveira, P.E. (2005), 'Exponential Rates for Kernel Density Estimation Under Association', *Statistica Neerlandica*, 4, 448–466.

Lai, M.-J. (2007), 'Multivariate Splines for Data Fitting and Approximation', in *Approximation Theory XII: San Antonio 2007*, eds. M. Neamtu and L.L. Schumaker, Nashboro Press, pp. 210–228.

Lai, M.-J., and Schumaker, L.L. (1998), 'On the Approximation Power of Bivariate Splines', *Advances in Computational Mathematics*, 9(3–4), 251–279.

Lai, M.-J., and Schumaker, L.L. (2007), *Spline Functions Over Triangulations*, Cambridge, UK: Cambridge University Press.

Oliveira, P.E. (2005), 'An Exponential Inequality for Associated Variables', *Statistics & Probability Letters*, 73(2), 189–197.

Ramsay, T. (2002), 'Spline Smoothing Over Difficult Regions', *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 64(2), 307–319.

Ramsay, J., and Silverman, B.W. (2005), *Functional Data Analysis*, New York: Springer-Verlag.

Wood, S.N. (2003), 'Thin Plate Regression Splines', *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 65(1), 95–114.

Yao, F., and Lee, T.C.M. (2006), 'Penalized Spline Models for Functional Principal Component Analysis', *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 68, 3–25.