

A Note on Bivariate Smoothing for Two-Dimensional Functional Data

Andrada E. Ivanescu¹

¹ Department of Biostatistics, East Carolina University, Greenville NC 27834, USA

Correspondence: Andrada E. Ivanescu, Department of Biostatistics, East Carolina University, Greenville NC 27834, USA. Tel: 1-252-744-6042. E-mail: ivanescua@ecu.edu

Received: April 1, 2013 Accepted: April 22, 2013 Online Published: April 25, 2013

doi:10.5539/ijsp.v2n2p102 URL: <http://dx.doi.org/10.5539/ijsp.v2n2p102>

Abstract

In this paper we study a bivariate smoothing approach for estimating multiple functional parameters for functional data with a two-dimensional domain. We present a penalized regression framework for smoothing with the purpose to: (a) facilitate the estimation of the smooth overall bivariate mean function of two-dimensional functional data, (b) enable the estimation of the functional effect of a scalar covariate, (c) accommodate completely or incompletely sampled data, (d) implement the fitting approach using available statistical software, and (e) construct pointwise approximate confidence intervals for multiple bivariate functional parameters. Implementation results from simulation studies show that these methods perform very well in practice. We illustrate the usefulness of the bivariate smoothing approach to several real datasets, including applications to electricity demand.

Keywords: bivariate functional parameters, bivariate splines, functional data analysis, mixed model, penalized regression, two-dimensional domain

1. Introduction

We study a statistical framework for smooth function estimation for functional data observed on a two-dimensional domain. Two-dimensional functional data are recorded in many disciplines and examples include human mortality rates (Gervini, 2010) and gel proteomic samples (Morris, Baladandayuthapani, Herrick, Sanna, & Gutstein, 2011). Analyzing two-dimensional functional data is a research area of dynamic interest. In this paper we examine a fitting approach for multiple smooth bivariate functional parameters.

We observe for the i^{th} sample the data values $\{[Y_i(t_{ij}, s_{ik})]_{jk}, W_i\}$, where $Y_i(t_{ij}, s_{ik})$ is the observed outcome at points $t_{ij} \in \mathcal{T}$ and $s_{ik} \in \mathcal{S}$, W_i is a scalar covariate, and $(t, s) \in \mathcal{T} \times \mathcal{S}$, where \mathcal{T} and \mathcal{S} are intervals on the real line. We consider the following model for the i^{th} observed outcome:

$$Y_i(t_{ij}, s_{ik}) = W_i \gamma(t_{ij}, s_{ik}) + f(t_{ij}, s_{ik}) + \epsilon_i(t_{ij}, s_{ik}), \quad (1)$$

for $1 \leq i \leq n$, $1 \leq j \leq m_i$ and $1 \leq k \leq q_i$. The bivariate mean function is modeled semi-parametrically (Ruppert, Wand, & Carroll, 2003; Wood, 2006) and comprises of two components: a linear function $W_i \gamma(\cdot, \cdot)$ to account for the scalar covariate W_i and an overall nonparametric function $f(\cdot, \cdot)$. The bivariate functional parameters f and γ are assumed smooth, and the errors ϵ_i are mean zero random processes. The model parameter $\gamma(t, s)$ is the bivariate functional coefficient of the scalar covariate and captures the effect of W_i varying smoothly on the two-dimensional domain (t, s) .

We wish to estimate f and γ in the setting of model (1). While the estimation of the overall mean function f has received some attention in the context of two-dimensional functional data, see Gervini (2010), the simultaneous estimation of both f and γ in the setting of model (1) has not been discussed. Important aspects that are still open to investigation in the setting of two-dimensional functional data include: estimating the bivariate smooth overall mean function for completely or incompletely sampled outcomes; incorporating the functional effect of a scalar covariate; constructing confidence intervals for multiple smooth bivariate functional parameters. Most smoothing methods for functional data are focused on functional data with a univariate domain (Ramsay & Silverman, 2005; Ramsay, Hooker, & Graves, 2009), and the use of smoothing methods in the setting of two-dimensional functional data is currently in the incipient stage. In this paper we undertake a mixed model framework with a fast-fitting approach to examine these aspects. This provides a statistically coherent analysis for important problems in the two-dimensional functional data literature.

The organization of the paper is as follows. Section 2 describes the methods for the bivariate smoothing approach. Section 3 outlines extensions to incompletely observed data scenarios. Section 4 presents numerical results for the empirical performance of the smoothing approach. Applications to several real datasets are illustrated in Section 5, and Section 6 is our conclusion.

2. Methods

In this section we describe the statistical modeling and fitting procedures. We begin with expressing model (1) as an additive model, followed by employing penalized regression and associated mixed model fitting framework. In this section it is considered that the i^{th} sampled outcome is observed at $(t_{ij}, s_{ik}) = (t_j, s_k)$ for all i , where $1 \leq j \leq m_i$ and $1 \leq k \leq q_i$ with $m_i = m$ and $q_i = q$ for all i . In section 3 we discuss the extension to other designs.

2.1 Statistical Modeling of Bivariate Functional Parameters Using Spline Bases

We express model parameters γ and f using a linear combination of tensor product spline basis functions. In particular, the parameter γ is represented as $\gamma(t, s) \approx \sum_{l_1=1}^{L_1} \sum_{l_2=1}^{L_2} \beta_{l_1, l_2} a_{l_1}(t) a_{l_2}(s)$, while f is expanded as $f(t, s) \approx \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} \mu_{k_1, k_2} \phi_{k_1}(t) \phi_{k_2}(s)$. The bases $a_{l_1}(\cdot)$ and $\phi_{k_1}(\cdot)$ are univariate bases in argument t , $a_{l_2}(\cdot)$ and $\phi_{k_2}(\cdot)$ are univariate bases in argument s , and β_{l_1, l_2} and μ_{k_1, k_2} are unknown coefficients corresponding to the tensor products $a_{l_1}(\cdot) a_{l_2}(\cdot)$ and $\phi_{k_1}(\cdot) \phi_{k_2}(\cdot)$, respectively. So, the term $W_i \gamma(t, s)$ can be represented as $\sum_{l_1=1}^{L_1} \sum_{l_2=1}^{L_2} \beta_{l_1, l_2} \psi_{l_1, i}(t) \psi_{l_2, i}(s)$, where $\psi_{l_1, i}(t) \psi_{l_2, i}(s) = a_{l_1}(t) a_{l_2}(s) W_i$.

Model (1) can thus be represented as an additive model:

$$Y_i(t, s) = \sum_{l_1=1}^{L_1} \sum_{l_2=1}^{L_2} \beta_{l_1, l_2} \psi_{l_1, i}(t) \psi_{l_2, i}(s) + \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} \mu_{k_1, k_2} \phi_{k_1}(t) \phi_{k_2}(s) + \epsilon_i(t, s), \quad (2)$$

where $\psi_{l_1, i}(t) \psi_{l_2, i}(s)$ and $\phi_{k_1}(t) \phi_{k_2}(s)$ are known quantities because W_i is observed and the bases $a_{l_1}(\cdot)$, $a_{l_2}(\cdot)$, $\phi_{k_1}(\cdot)$ and $\phi_{k_2}(\cdot)$ are chosen a priori. Functions f and γ are conveyed using rich tensor product bases to plausibly represent their complexity. The dimension of the bases is defined by choosing a large enough number of basis functions L_1 , L_2 , K_1 and K_2 .

2.2 Penalized Criteria for Fitting Smooth Bivariate Functional Parameters

In order to obtain smooth estimates for f and γ we impose quadratic penalties on their corresponding bases coefficients. Let $m_i((t, s); \beta, \mu)$ represent the mean of $Y_i(t, s)$, where β is the vector of all parameters β_{l_1, l_2} and μ is the vector of all parameters μ_{k_1, k_2} . Then the criterion to be penalized is:

$$\sum_{i,j,k} \|Y_i(t_j, s_k) - m_i((t_j, s_k); \beta, \mu)\|^2 + \text{Pen}(\beta) + \text{Pen}(\mu) \quad (3)$$

where $\text{Pen}(\beta)$ and $\text{Pen}(\mu)$ take the form of quadratic penalties for the bivariate smooths (Wood, 2006). For each parameter function we use smoothing parameters for both t and s dimensions. The penalty term for γ is $\text{Pen}(\beta) = \beta^T (\lambda_{\beta, t} \mathbf{P}_{\beta, t} + \lambda_{\beta, s} \mathbf{P}_{\beta, s}) \beta$. The penalty term for f is $\text{Pen}(\mu) = \mu^T (\lambda_{\mu, t} \mathbf{P}_{\mu, t} + \lambda_{\mu, s} \mathbf{P}_{\mu, s}) \mu$. Matrices $\mathbf{P}_{\beta, t}$, $\mathbf{P}_{\beta, s}$, $\mathbf{P}_{\mu, t}$ and $\mathbf{P}_{\mu, s}$ have known functional form, are fixed, and depend on the choice of spline bases.

The scalars $\lambda_{\beta, t}$, $\lambda_{\beta, s}$, $\lambda_{\mu, t}$ and $\lambda_{\mu, s}$ govern the smoothness of the fit and are unknown parameters. Fitting model (3) calls for estimation of these smoothing parameters, for which we opt for restricted maximum likelihood (REML) estimation (Reiss & Ogden, 2009) in a chosen mixed model and outline the approach next.

2.3 Mixed Model Representation for Bivariate Smoothing

The penalized criterion with the quadratic penalty choices discussed above is

$$\frac{1}{\sigma^2} \sum_{i,j,k} \|Y_i(t_j, s_k) - m_i((t_j, s_k); \beta, \mu)\|^2 + \frac{\lambda_{\beta, t}}{\sigma^2} \beta^T \mathbf{P}_{\beta, t} \beta + \frac{\lambda_{\beta, s}}{\sigma^2} \beta^T \mathbf{P}_{\beta, s} \beta + \frac{\lambda_{\mu, t}}{\sigma^2} \mu^T \mathbf{P}_{\mu, t} \mu + \frac{\lambda_{\mu, s}}{\sigma^2} \mu^T \mathbf{P}_{\mu, s} \mu,$$

where σ^2 is the variance of the errors in model (1). By letting $\sigma_{\beta, \cdot}^2 = \sigma^2 / \lambda_{\beta, \cdot}$ and $\sigma_{\mu, \cdot}^2 = \sigma^2 / \lambda_{\mu, \cdot}$ and employing the considerations presented in, for example, Ruppert et al. (2003, sec 4.9), one can establish that the solution to the penalized criterion (3) is the best linear unbiased predictor in the mixed model

$$\begin{aligned} Y_i(t, s) &\sim N\{m_i((t, s); \beta, \mu), \sigma^2\}; \\ \beta &\sim N(\mathbf{0}, (\sigma_{\beta, t}^2 \mathbf{P}_{\beta, t} + \sigma_{\beta, s}^2 \mathbf{P}_{\beta, s})^{-1}); \\ \mu &\sim N(\mathbf{0}, (\sigma_{\mu, t}^2 \mathbf{P}_{\mu, t} + \sigma_{\mu, s}^2 \mathbf{P}_{\mu, s})^{-1}), \end{aligned} \quad (4)$$

where $P_{\beta \cdot}$ and $P_{\mu \cdot}$ are known, and $\sigma_{\beta \cdot}^2$ and $\sigma_{\mu \cdot}^2$ control the smoothness of the functional parameters. In many instances the inverse is substituted by a generalized inverse or the coefficients are partitioned into penalized and unpenalized. This mixed model representation is a useful statistical framework that has several advantages. First, it provides a fast approach for bivariate smoothing that is carried out using available software. Second, elements of inference corresponding to mixed model analysis, such as pointwise confidence intervals (Nychka, 1998; Ruppert et al., 2003), can be obtained for the model parameters. We study pointwise confidence intervals for our bivariate functional parameters when we assume independent and identically distributed errors in model (1). Third, other design settings can be accommodated using this framework, such as incompletely observed two-dimensional functional data.

We display how to fit model (4) using statistical software based on REML estimation of the variance components. We employ the `gam()` function from the `mgcv` package (Wood, 2012) in R (R Development Team, 2012).

2.4 Fitting Approach for Bivariate Smoothing via Mixed Models Software

In this section we provide implementation steps for the approach using the `mgcv` package (Wood, 2012) in R. As illustrated, model (3) can be viewed as a penalized additive model where the smooth bivariate model parameters are represented using the tensor product bases $a_{l_1}(\cdot)a_{l_2}(\cdot)$ and $\phi_{k_1}(\cdot)\phi_{k_2}(\cdot)$. These are spline bases and the model fit can be expressed via the following R code formula syntax:

```
fit <- gam(Y ~ te(t,s,by = W) + te(t,s), method = 'REML').
```

We now outline the components involved in the fitting procedure above while explaining the connection to our model and its notation. For two-dimensional smoothing, the data is organized into column vectors in order to implement the fitting approach. The observed two-dimensional functional data outcome samples $\{Y_i(t_j, s_k)\}_{ijk}$ are arranged as a column vector of size nmq , so that the i^{th} sample matrix with observed outcomes containing mq observations is $Y^{(i)} = (Y_i(t_1, s_1), Y_i(t_2, s_1), \dots, Y_i(t_m, s_1), Y_i(t_1, s_2), \dots, Y_i(t_m, s_q))^T$. All sample-particular matrices $\{Y^{(i)}\}_i$, $1 \leq i \leq n$ are row-stacked to form the column vector $Y = (Y^{(1)}, Y^{(2)}, \dots, Y^{(n)})^T$. The next step is to specify the grid of points in dimensions t and s that correspond precisely to the locations of the observed outcomes. The vector $\mathbf{t} = (t^{(1)}, t^{(2)}, \dots, t^{(n)})^T$ consists of row-stacking n copies of $t^{(i)} = (t_1, t_2, \dots, t_m, t_1, \dots, t_m)^T$. Similarly, we construct $\mathbf{s} = (s^{(1)}, s^{(2)}, \dots, s^{(n)})^T$, which consists of row-stacking n copies of $s^{(i)} = (s_1, s_1, \dots, s_1, s_2, \dots, s_q)^T$. The scalar covariate W_i is repeated mq times for each i to form the nmq dimensional column vector labeled \mathbf{W} . The expression `te(t,s,by = W)` thus fits a tensor product of spline basis functions at the grid of points specified in the vectors \mathbf{t} and \mathbf{s} while being weighted by the values of the scalar covariates W_i . Without the option `by = W` the expression `te(t,s)` constructs a tensor product of spline basis functions $a_{l_1}(t)a_{l_2}(s)$, whereas including `by = W` is taking into account the weights W_i , thus accounting for the effect of the scalar covariate via its functional coefficient varying over the t and s dimensions.

Fitting the part of the model that yields a smooth estimate for $f(t, s)$ involves the matrices \mathbf{t} and \mathbf{s} already constructed. Thus, the expression `te(t,s)` fits a bivariate tensor product spline $\phi_{k_1}(t)\phi_{k_2}(s)$ at the grid of points in the vectors \mathbf{t} and \mathbf{s} . The details of the `gam` fit can be adjusted as needed. For example, requiring the use of 10 basis functions for each dimension can be requested by including the term `k = 10` so that the expression now is `te(t,s,k = 10)`.

Using the approach described, the underlying `gam` fitting procedure equips the user to perform rapid bivariate smoothing and provides a useful toolbox for obtaining graphical and inferential methods readily accessible.

3. Extension to Incompletely Observed Two-Dimensional Functional Data

Incompletely observed two-dimensional functional data can be accommodated by the procedure described in section 2, provided some adjustments are made in the fitting step. In such settings the outcomes $\{Y_i(t_{ij}, s_{ik})\}_{ijk}$, $1 \leq i \leq n$, are sparsely observed at points $\{(t_{ij}, s_{ik})\}_{ijk}$ over the domain, for $1 \leq j \leq m_i$ and $1 \leq k \leq q_i$. Vectors \mathbf{Y} , \mathbf{t} , \mathbf{s} and \mathbf{W} all have dimension nm_iq_i and are constructed as follows. The vector of outcomes for the i^{th} sample has dimension m_iq_i and is constituted as $Y^{(i)} = (Y_i(t_{i1}, s_{i1}), Y_i(t_{i2}, s_{i1}), \dots, Y_i(t_{m_i}, s_{i1}), Y_i(t_{i1}, s_{i2}), \dots, Y_i(t_{m_i}, s_{i q_i}))^T$. The sample-specific matrices $\{Y^{(i)}\}_i$ are row-stacked to form $Y = (Y^{(1)}, Y^{(2)}, \dots, Y^{(n)})^T$. Next, the vector labeled \mathbf{t} has the form $\mathbf{t} = (t^{(1)}, t^{(2)}, \dots, t^{(n)})^T$ and consists of row-stacking the m_iq_i -dimensional vectors $t^{(i)} = (t_{i1}, t_{i2}, \dots, t_{im_i}, t_{i1}, \dots, t_{im_i})^T$ for all i . Further, $\mathbf{s} = (s^{(1)}, s^{(2)}, \dots, s^{(n)})^T$ is built by row-stacking all the sample-specific vectors of dimension m_iq_i defined as $s^{(i)} = (s_{i1}, s_{i1}, \dots, s_{i1}, s_{i2}, \dots, s_{i q_i})^T$. The vector of covariates \mathbf{W} is obtained by taking m_iq_i copies of W_i and row-stacking these vectors across all n samples. After these matrices are set up, the fitting approach proceeds as described in section 2. Implementation results obtained for this setting are provided in sections 4 and 5.

4. Simulation Study

We present numerical results from implementations to simulated data. Empirical performance for the fit of the estimators was evaluated for completely or incompletely observed two-dimensional data.

4.1 Bivariate Smoothing for One Bivariate Functional Parameter

For each $1 \leq i \leq n$, we generated densely sampled two-dimensional outcome data according to the model

$$Y_i(t_j, s_k) = f(t_j, s_k) + \epsilon_i(t_j, s_k), \quad (5)$$

which were observed on an equally spaced grid with $t_j = (j-1)/m$ and $s_k = (k-1)/q$, $1 \leq j \leq m$, $1 \leq k \leq q$, where $m = 48$, $q = 50$, and $n = 60$. The values of m , q and n are chosen similar to the values in real data applications in sections 5.1 and 5.2. We considered three cases for the error terms: *Case R1* where $\epsilon_i(t_j, s_k)$ are independent and identically distributed $N(0, 2)$; and *Cases R2-R3* where $\epsilon_i(t_j, s_k) = Z_i(t_j, s_k) + e_{ijk}$, and where Z_i is a bivariate mean zero stochastic process with variance 1, and e_{ijk} are i.i.d. $N(0, 1)$. Independent samples $Z_i(\cdot, \cdot)$, $1 \leq i \leq n$, were simulated, using `GaussRF()` function in the `RandomFields` (Schlather, 2012) package, from a mean zero and variance one Gaussian Random Field with Matérn correlation function with scale parameter ρ and shape parameter $\nu = 0.5$. We considered two cases: *Case R2* $\rho = 0.05$, and *Case R3* $\rho = 0.15$.

The bivariate mean functional parameter was $f(t, s) = 1.6 \sin(\pi s) \cos(2\pi t)$, which has a similar shape to the periodic mean function in the application considered in section 5.2. Signal-to-noise ratio defined as $\text{SNR}_1 = \text{range}(f)/[\text{var}\{\epsilon(t, s)\}]^{1/2}$ was set to 2.2, where $\text{range}(f) = \max_{\mathcal{T} \times \mathcal{S}}\{f(t, s)\} - \min_{\mathcal{T} \times \mathcal{S}}\{f(t, s)\}$. Model (5) does not incorporate the effect of a scalar covariate and thus is a simpler model compared to model (1). Model (5) permits a comparison of several estimators for $f(\cdot, \cdot)$, which we discuss next. The first estimator, labeled E1, denotes the penalized splines estimator obtained by applying methods described in section 2. We facilitate accuracy of the fit comparisons with other estimators that can be used for estimation of f in the setting of model (5). Tensor products of cubic B-splines with 100 basis functions (for E1) and 25 basis functions (for E2 and E3) and second-order difference penalties in both directions were used to fit $f(\cdot, \cdot)$. The following estimators were computed:

E1. Pen Spl - Penalized splines estimator obtained by applying the methods presented in section 2.

E2. Smooth ensemble - Ensemble average was obtained $1/n \sum_{i=1}^n Y_i(t, s)$ followed by bivariate smoothing.

E3. Average of smooths - A bivariate smooth estimate for $Y_i(t, s)$ was fit for each i . Then, an ensemble average was computed.

E4. Assemble \hat{f}_t - For each fixed t , model (5) was fitted and a univariate penalized spline estimator was obtained. Then, these estimators were joined along t to form a bivariate estimator \hat{f} .

E5. Assemble \hat{f}_s - For each fixed s , model (5) was fitted and a univariate penalized spline estimator was obtained. Then, these estimators were joined along s to form a bivariate estimator \hat{f} .

We provide numerical results for the fit performance, as well as computational time aspects of the fitting procedures used. We computed the integrated mean square error (IMSE), integrated square bias (IBIAS²), and integrated variance (IVAR), where $\text{IMSE}\{\hat{f}(t, s)\} = \int_{\mathcal{T}} \int_{\mathcal{S}} E\{[\hat{f}(t, s) - f(t, s)]^2\} dt ds$, $\text{IBIAS}^2\{\hat{f}(t, s)\} = \int_{\mathcal{T}} \int_{\mathcal{S}} [E\{\hat{f}(t, s)\} - f(t, s)]^2 dt ds$, and $\text{IVAR}\{\hat{f}(t, s)\} = \int_{\mathcal{T}} \int_{\mathcal{S}} \text{Var}\{\hat{f}(t, s)\} dt ds$. Here $E\{\hat{f}(t, s)\}$ and $\text{Var}\{\hat{f}(t, s)\}$ were estimated by the empirical mean and variance of $\hat{f}(t, s)$ in $S = 250$ simulations.

Results for the fit of the estimation for each method are presented in Table 1 and can be summarized as follows. The IMSE for the penalized splines estimator E1 was smaller compared to the estimators E2-E5 for all R1-R3 choices. Figure 1 shows that the variability of the ISE of all estimators was comparable, and that the median ISE of E1 was smaller than for the other estimators. Compared to E4 and E5 estimators, which do not borrow strength in both t and s domains simultaneously, E1 performed better, as illustrated by smaller ISE; see Figure 1. IMSE was smaller for the case R1 of the errors. For the case of non-i.i.d. errors (cases R2 and R3) IMSE increased for all estimators, but E1 still outperformed E2-E5 estimators. Comparable IVAR among E1, E2 and E3 may be achieved for case R1 provided a similar number of basis functions is used for model fitting. For cases R2 and R3, the methods E1 and E3 display undersmoothing (of similar degree) compared to the uncorrelated errors case, and thus have larger IVAR compared to case R1. Bivariate smoothing is typically slow, but our approach was fast: about 11 seconds for the analysis of 60 two-dimensional data objects. Computational time considerations reveal that E1 is computationally fast compared to the other approaches; compare the average time over 250 model fits: E1 [11.8 sec], E2 [1.2 sec], E3 [101.4 sec], E4 [59.8 sec], E5 [57.4 sec].

Table 1. Estimation results for the fit of the estimators for $f(\cdot, \cdot)$. Scenario: $m = 48$, $q = 50$, $SNR_1 = 2.2$, section 4.1 simulation setting. Results are based on 250 simulations

	$n = 60$ ($\times 100$)		
	\sqrt{IMSE}	\sqrt{IVAR}	$\sqrt{IBIAS^2}$
<i>Case R1</i>			
E1. Pen Spl	4.37	2.55	3.55
E2. Smooth ensemble	5.47	1.84	5.15
E3. Average of smooths	5.97	1.54	5.77
E4. Assemble \widehat{f}_t	5.64	5.58	0.83
E5. Assemble \widehat{f}_s	6.75	6.50	1.80
<i>Case R2</i>			
E1. Pen Spl	6.75	5.93	3.23
E2. Smooth ensemble	8.08	6.21	5.17
E3. Average of smooths	7.94	5.94	5.26
E4. Assemble \widehat{f}_t	9.21	9.16	0.92
E5. Assemble \widehat{f}_s	10.30	10.16	1.73
<i>Case R3</i>			
E1. Pen Spl	10.38	9.83	3.34
E2. Smooth ensemble	11.35	10.08	5.23
E3. Average of smooths	11.25	9.94	5.26
E4. Assemble \widehat{f}_t	11.55	11.49	1.17
E5. Assemble \widehat{f}_s	12.35	12.20	1.94

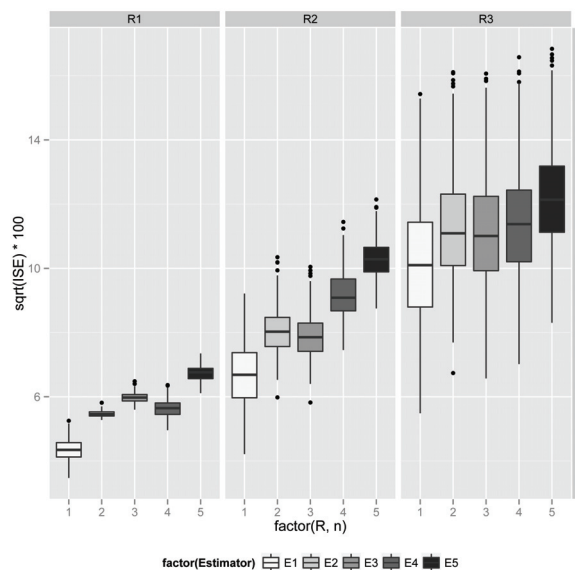


Figure 1. Displayed are boxplots of \sqrt{IMSE} corresponding to estimators E1 - E5 for f . Columns correspond to cases R1, R2 and R3. Scenario: $m = 48$, $q = 50$, $n = 60$, $SNR_1 = 2.2$, section 4.1 simulation setting. Results are based on 250 simulations

4.2 Bivariate Smoothing for Two Bivariate Functional Parameters

We generated two-dimensional functional outcomes according to model (1), where W_i was a binary variable such that $W_i = 1\{U_i \geq 0.75\}$, with $U_i \sim \text{Uniform}(0, 1)$, where $n = \{30, 60\}$. We considered the equally spaced grid design used in section 4.1. We generated $\epsilon_i(t, s)$ for cases R1 - R3 as in section 4.1.

For the bivariate functional parameters we set $\gamma(t, s) = \sin(\pi s) \cos(2\pi t)$ and $f(t, s) = 1.5 \exp^{-8s^2} + 0.6 \exp^{-8t^2}$, so that γ and f have comparable range. Our choice for $f(\cdot, \cdot)$ was motivated by the shape of the mean function for the application in section 5.3. Signal-to-noise ratio defined as $SNR_2 = \text{range}(.75\gamma + f) / [\text{var}\{\epsilon(t, s)\}]^{1/2}$ was set to 2.2. Because of the presence of the scalar covariate, only the estimators E1, E4 and E5 are employed for fitting model (1). Empirical results corresponding to the accuracy of the fit are presented in Table 2. Results show that the penalized splines estimator E1 outperformed the estimators E4 and E5 for both f and γ , for R1-R3 choices and for both n choices. The advantage of E1 was especially evident for case R1 of the errors. As n increased, the IMSE decreased, which indicates that the penalized splines estimator is consistent. E1 provides a smooth fit compared to the E4 and E5 estimators, which do not borrow strength in both t and s domains; see Figure 2 for the visual representation of the estimators E1, E4 and E5 for $f(\cdot, \cdot)$. Methods E4 and E5 comport like unpenalized (least squares) estimators in one of the two dimensions and thus sometimes yield slightly smaller IBIAS than E1, which is a penalized estimator in both t and s dimensions.

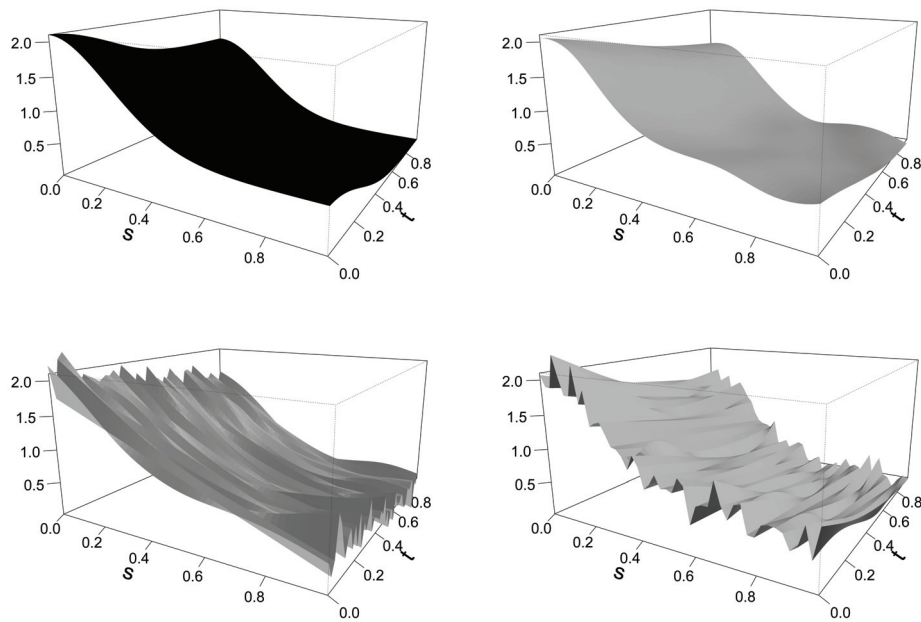


Figure 2. Estimates for the bivariate functional parameter f : true function (black: top left panel), E1 (gray: top right panel), E4 (gray: bottom left panel) and E5 (gray: bottom right panel). Scenario: $m = 48$, $q = 50$, $n = 60$, Case R1, section 4.2 simulation setting

Table 2. Estimation results for $\gamma(\cdot, \cdot)$ and $f(\cdot, \cdot)$ bivariate functional parameters. Scenario: $m = 48$, $q = 50$, $SNR_2 = 2.2$, section 4.2 simulation setting. Results are based on 250 simulations

	$n = 30$						$n = 60$					
	\sqrt{IMSE}	\sqrt{IVAR}	$\sqrt{IBIAS^2}$	\sqrt{IMSE}	\sqrt{IVAR}	$\sqrt{IBIAS^2}$ ($\times 100$)	\sqrt{IMSE}	\sqrt{IVAR}	$\sqrt{IBIAS^2}$	\sqrt{IMSE}	\sqrt{IVAR}	$\sqrt{IBIAS^2}$
Completely observed data												
Case R1												
E1. Pen Spl	5.69	5.01	2.69	4.49	3.94	2.16	5.38	4.74	2.55	3.76	3.34	1.72
E4. Assemble $\hat{\gamma}_t, \hat{f}_t$	17.34	15.98	6.74	15.62	14.14	6.63	12.20	11.46	4.17	11.04	10.14	4.37
E5. Assemble $\hat{\gamma}_s, \hat{f}_s$	16.15	15.93	2.64	13.22	13.03	2.23	11.80	11.61	2.09	9.67	9.45	2.06
Case R2												
E1. Pen Spl	23.60	23.47	2.44	20.12	20.04	1.84	15.13	14.88	2.74	11.51	11.34	1.96
E4. Assemble $\hat{\gamma}_t, \hat{f}_t$	26.07	25.63	4.76	23.04	22.59	4.52	18.88	18.61	3.14	16.77	16.46	3.22
E5. Assemble $\hat{\gamma}_s, \hat{f}_s$	26.56	26.42	2.73	22.36	22.30	1.60	19.15	19.04	2.04	15.91	15.85	1.42
Case R3												
E1. Pen Spl	34.37	34.27	2.60	29.93	29.85	2.10	23.51	23.33	2.86	19.69	19.57	2.25
E4. Assemble $\hat{\gamma}_t, \hat{f}_t$	34.58	34.28	4.56	30.53	30.24	4.26	24.95	24.76	3.08	22.04	21.81	3.19
E5. Assemble $\hat{\gamma}_s, \hat{f}_s$	35.08	34.95	3.03	30.13	30.07	1.89	25.17	25.06	2.35	21.43	21.36	1.73

4.3 Incompletely Observed Two-Dimensional Data

The bivariate smoothing method studied in this paper can be applied to incompletely two-dimensional sampled data. In this section we present a brief evaluation of the accuracy of the fit in such scenarios. Data was generated as

described in section 4.2 for Case R1. We randomly selected and kept sixty percent of the generated outcome data and this set of data was used as input values for model fitting. The approach described in section 3 was applied. Results for the fit of the estimators show that accuracy in estimation is affected by the sparsity in the data, but not by much; compare IMSE, IVAR and IBIAS columns in Table 3 to results in Table 2 for Case R1. As the number of samples n increased, the estimation accuracy improved, as expected; see IMSE columns in Table 3.

Table 3. Estimation results for estimator E1 for $\gamma(\cdot, \cdot)$ and $f(\cdot, \cdot)$ for incompletely observed two-dimensional data. Scenario: $m = 48$, $q = 50$, $SNR_2 = 2.2$, section 4.3 simulation setting. Results are based on 250 simulations

	$n = 30$						$n = 60$					
	\sqrt{IMSE}	$\frac{\gamma(t,s)}{\sqrt{IVAR}}$	$\sqrt{IBIAS^2}$	\sqrt{IMSE}	$\frac{f(t,s)}{\sqrt{IVAR}}$	$\sqrt{IBIAS^2}$ ($\times 100$)	\sqrt{IMSE}	$\frac{\gamma(t,s)}{\sqrt{IVAR}}$	$\sqrt{IBIAS^2}$	\sqrt{IMSE}	$\frac{f(t,s)}{\sqrt{IVAR}}$	$\sqrt{IBIAS^2}$
Incompletely observed data												
Case R1	6.85	6.22	2.87	5.55	4.92	2.57	6.40	5.79	2.72	4.56	4.04	2.11

4.4 Confidence Intervals for Smooth Bivariate Functional Parameters

One advantage of the fitting approach discussed in section 2 is obtaining pointwise confidence intervals for the model parameters as a consequence of the mixed model fitting procedure. We considered the same data generating process as in section 4.2 and Case R1. To characterize the properties of the pointwise confidence intervals we computed the integrated actual pointwise coverage (IAC) and integrated actual width (IAW). For the case of $f(\cdot, \cdot)$, $IAC = E[\int_T \int_S 1\{\widehat{f}(t, s) \in CI_p(t, s)\} dtds]$. An approximate 95% pointwise confidence interval $CI_p(t_0, s_0)$ for $f(t_0, s_0)$ can be constructed as $\widehat{f}(t_0, s_0) \pm 1.96 \widehat{sd}\{\widehat{f}(t_0, s_0)\}$. Since $\widehat{f}(t_0, s_0) = \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} \widehat{\mu}_{k_1, k_2} \phi_{k_1}(t_0) \phi_{k_2}(s_0)$, for every (t_0, s_0) , we obtain $\widehat{sd}\{\widehat{f}(t_0, s_0)\} = \sqrt{\phi_{k_1}(t_0) \phi_{k_2}(s_0) \widehat{\Sigma} [\phi_{k_1}(t_0) \phi_{k_2}(s_0)]^T}$, where $\widehat{\Sigma}$ is the estimated covariance matrix of $\widehat{\mu}$. We used the Bayesian posterior covariance matrix (Ruppert et al. 2003). The length of this confidence interval was $3.92 \widehat{sd}\{\widehat{f}(t_0, s_0)\}$ and $IAW = E[3.92 \int_T \int_S \widehat{sd}\{\widehat{f}(t, s)\} dtds]$. Empirical coverage results are presented in Table 4. The nominal level was set to 0.95.

Table 4. Results for pointwise approximate confidence intervals. Displayed are IAC(IAW) results for 95% nominal confidence level. Scenario: $m = 48$, $q = 50$, $SNR_2 = 2.2$, Case R1, section 4.4 simulation setting. Results are based on 250 simulations

	$\gamma(t, s)$	$f(t, s)$
$n = 30$	0.97 (0.47)	0.97 (0.41)
$n = 60$	0.97 (0.44)	0.97 (0.38)

Table 4 results can be summarized as follows. The pointwise confidence intervals had a coverage probability exceeding the nominal level. As the number of samples n increased, the width of the intervals decreased for both parameter functions, while maintaining IAC above the nominal level. In conclusion, for Case R1, the pointwise approximate confidence intervals for the bivariate functional parameters had narrow width and achieved or exceeded the nominal coverage level.

5. Applications

We applied the bivariate smoothing methods discussed in section 2 to several real datasets. Implementations to datasets on electric energy demand, temperature and human mortality are presented.

5.1 Application to Electric Energy Demand in Adelaide, Australia

Electricity is an energy support for human progress. Electricity demand fluctuates as it is influenced by consumption patterns according to business hours, outside temperature and other weather conditions. Understanding electricity demand is useful for planning optimal electric energy supply depending on various factors, such as, time of day and week of year. Our goal is to estimate the electric energy demand patterns in Adelaide, Australia, as function of time of day t and week s . We use the Electricitydemand data, which is part of the functional datasets `fds` package (Shang & Hyndman, 2012) in R, and comprises of electricity demand recordings, in megawatts, that span the calendar years 1998 - 2006. Following the notations in section 2, $Y_i(t, s)$ corresponds to the electricity demand for i^{th} day at half-hourly time of day t and week of year s . We estimate the mean function $f(\cdot, \cdot)$ and the functional coefficient $\gamma(\cdot, \cdot)$ of weekend, where $W_i = 1\{\text{weekend day}\}$, for $1 \leq i \leq n$, from data consisting of $n = 63$ (corresponding to data for the seven days of the week, for nine years) two-dimensional data samples. The

two-dimensional domain is $\mathcal{T} \times \mathcal{S}$, where $t \in \mathcal{T} = \{1, 2, \dots, 48\}$ corresponds to 48 half-hourly time points during the course of a day, starting with the first half-hour after midnight, and where $s \in \mathcal{S} = \{1, 2, \dots, 52\}$ corresponds to the week of year, starting with the first week in January.

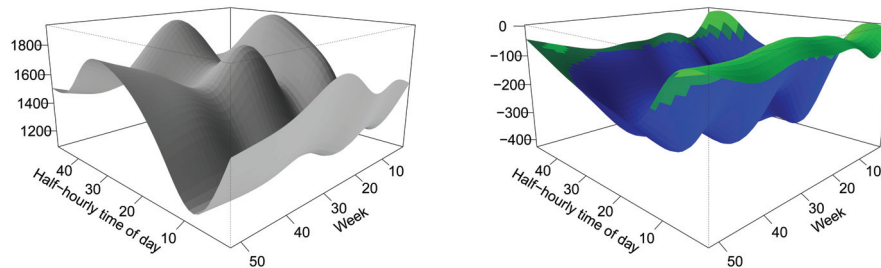


Figure 3. Electric energy demand data. Estimated overall mean electricity demand (left panel) and estimated functional coefficient corresponding to weekend days (right panel)

We applied the methods presented in section 2 to obtain smooth \hat{f} and $\hat{\gamma}$. Tensor products of cubic B-splines with 64 basis functions and second-order difference penalties in both directions were used to fit $f(\cdot, \cdot)$ and $\gamma(\cdot, \cdot)$. Results show that over the course of the year, the mean electricity demand was high for the beginning and end-year weeks (Australian summer) and in the mid-year weeks (Australian winter); see Figure 3. Over the course of the day, the electricity demand was higher during mid-day, at around half-hour 30. During the early morning and late evening hours of the day, more electricity was demanded during mid-year (Australian winter); this may be due to the use of heating for homes during the night hours. During mid-day more electricity was demanded during first and last weeks (Australian summer). The effect of weekend on the electricity demand resulted in a decrease by as much as 300 megawatts; see Figure 3. This decrease is more pronounced for the mid-day hours of the weekend, around half-hours 20 - 30, throughout the year. The mean width of the 95% pointwise confidence intervals was 17.7 for f and 32.9 for γ . Figure 3 displays information about the pointwise confidence intervals for γ : bright green corresponds to effects that were approximately 2 standard errors above -100, and bright blue corresponds to effects that were approximately 2 standard errors below -100.

5.2 Application to Temperature Patterns in Adelaide, Australia

Adelaide, Australia has an overall mild climate year-round. In this region, winter corresponds to the mid-year months, while the summer months occur during the weeks at the beginning and end of the year. The goal was to estimate the mean temperature pattern as a function of half-hourly time of day t and week s . Data `tempkent`, which is available in the functional datasets `fds` package (Shang & Hyndman, 2012) in R, was used for this purpose. The two-dimensional domain $\mathcal{T} \times \mathcal{S}$ in this application was the same as in the application to electricity demand. Here $Y_i(t, s)$ denotes the temperature in degrees Celsius recorded for i^{th} day at time of day t and week s . We applied the methodology in section 2 to estimate $f(\cdot, \cdot)$ using $n = 63$ two-dimensional data samples (corresponding to 7 days of week for years 1998 - 2006). Scalar covariates were not considered in this application. Tensor products of cubic B-splines with 100 basis functions and second-order difference penalties in both directions were used to fit $f(\cdot, \cdot)$. The estimator is illustrated in Figure 4. The features can be summarized as follows. Warm average temperature of about 27 degrees Celsius was observed for the beginning and end weeks of the year (Australian summer), especially during mid-day at half-hours 20 - 30. The mid-year weeks registered a decreased average temperature with a minimum average temperature of about 10 degrees Celsius during the beginning and evening hours of the day. The mean width of the 95% pointwise confidence intervals for f was 0.21.

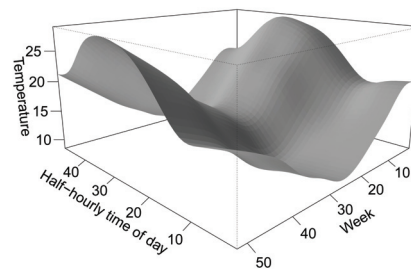


Figure 4. Temperature data. Smooth estimate for the overall bivariate functional mean temperature

5.3 Application to Australian Human Mortality Data

Data on human mortality for seven regions in Australia was used in this implementation. The seven regions are New South Wales, Northern Territory, Queensland, South Australia, Tasmania, Victoria, and Western Australia. Australian human mortality data is available for both female and male gender in the functional data sets `fds` package (Shang & Hyndman, 2012) in R. In this application $Y(t, s)$ is the log mortality for individuals of age s during year t . Lower values of log mortality correspond to less observed mortality. The purpose of this implementation was to obtain a smooth estimate for the overall mean log mortality $f(t, s)$, as a function of year t and age s , and a smooth estimate for the coefficient $\gamma(t, s)$ corresponding to the covariate gender, where $W_i = 1\{\text{female}\}$, $1 \leq i \leq 14$. The two-dimensional domain in this application was $\mathcal{T} \times \mathcal{S}$, where $\mathcal{T} = \{1901, 1921, \dots, 2003\}$ corresponds to the calendar year, and $\mathcal{S} = \{0, 1, \dots, 100\}$ corresponds to age. The methods presented in section 3 was employed to obtain smooth fitted model parameters. Tensor products of cubic B-splines with 100 basis functions and second-order difference penalties in both directions were used to fit $f(\cdot, \cdot)$ and $\gamma(\cdot, \cdot)$.

The estimated overall mean function displayed a decrease in infant mortality in recent years; see Figure 5. This observed effect was probably due to recent improvements in medical care for newborns and infants. After age 10, as age increased mortality also increased for most years. The estimated functional coefficient of gender shows that, overall, female mortality was lower than male mortality; see Figure 5. Female mortality for early youth had lower magnitude compared to male mortality and this decrease was larger in recent years. This effect was due, possibly, to the recent medical advancements that benefited the female population. The mean width of the 95% pointwise confidence intervals was 0.04 for f and 0.06 for γ . Figure 5 displays information about the pointwise confidence intervals for γ : bright green corresponds to effects that were approximately 2 standard errors above -0.5 and bright blue corresponds to effects that were approximately 2 standard errors below -0.5.

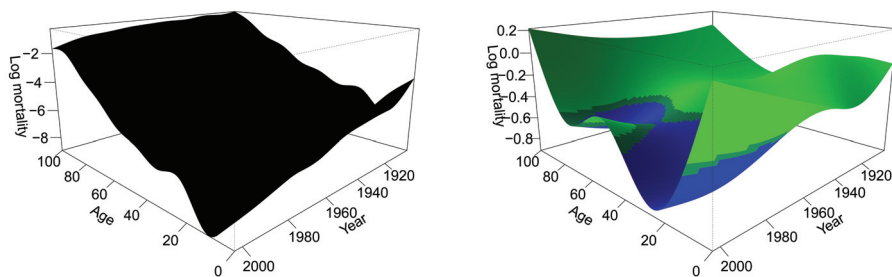


Figure 5. Human mortality data. Smooth overall mean log mortality estimate (left panel) and smooth functional coefficient estimate corresponding to gender (right panel)

6. Discussion

The approach we study in this paper is a useful statistical framework for fitting smooth bivariate functional parameters in the setting of two-dimensional functional data. Our bivariate smoothing research is a useful contribution

that: (i) yields smooth bivariate estimator for the overall mean functional parameter, (ii) produces an estimate for the bivariate smooth varying coefficient to account for the effect of a scalar covariate, (iii) can be applied to completely or incompletely observed two-dimensional functional data, (iv) is carried out using available software, (v) is computationally fast, and (vi) facilitates the construction of pointwise confidence intervals for smooth bivariate functional parameters. We applied the approach to various simulated data scenarios and the numerical results and visual considerations showed good performance in implementations. Results from our applications to several real datasets and from our simulation study illustrate the usefulness of the bivariate smoothing approach in practice.

References

- Gervini, D. (2010). The functional singular value decomposition for bivariate stochastic processes. *Computational Statistics and Data Analysis*, 54, 163-172. Retrieved from <https://pantherfile.uwm.edu/gervini/www/papers/FSVD.pdf>
- Morris, J. S., Baladandayuthapani, V., Herrick, R. C., Sanna, P., & Gutstein, H. B. (2011). Automated analysis of quantitative image data using isomorphic functional mixed models with application to proteomics data. *Annals of Applied Statistics*, 5(2A), 894-923. <http://dx.doi.org/10.1214/10-AOAS407>
- Nychka, D. (1988). Confidence Intervals for Smoothing Splines. *Journal of the American Statistical Association*, 83, 1134-1143. <http://dx.doi.org/10.1080/01621459.1988.10478711>
- R Development Core Team. (2012). *R: a Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from www.R-project.org
- Ramsay, J., & Silverman, B. (2005). *Functional data analysis* (2nd ed.). New York: Springer.
- Ramsay, J. O., Hooker, G., & Graves, S. (2009). *Functional Data Analysis with R and Matlab*. New York: Springer. <http://dx.doi.org/10.1007/978-0-387-98185-7>
- Reiss, P., & Ogden, T. (2009). Smoothing parameter selection for a class of semiparametric linear models. *Journal of the Royal Statistical Society, Series B*, 71(2), 505-523. <http://dx.doi.org/10.1111/j.1467-9868.2008.00695.x>
- Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge, UK: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511755453>
- Schlather, M. (2012). *RandomFields: Simulation and Analysis of Random Fields*. R package.
- Shang, H. L., & Hyndman, R. J. (2012). *fds: Functional data sets*. R package.
- Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. New York: Chapman & Hall/CRC.
- Wood, S. N. (2012). *mgcv: GAMs with GCV/AIC/REML Smoothness Estimation and GAMMs by PQL*. R package.