Journal of Statistical Planning and Inference 147 (2014) 53–65

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Predicting the continuation of a function with applications to call center data

Y. Goldberg [a,*], Y. Ritov [b], A. Mandelbaum [c]

[a] Department of Statistics, University of Haifa, Haifa 31705, Israel
[b] Department of Statistics and the Center for the Study of Rationality, The Hebrew University, Jerusalem 91905, Israel
[c] Industrial Engineering and Management, Technion—Israel Institute of Technology, Haifa 32000, Israel

ARTICLE INFO

Article history:
Received 24 January 2013
Received in revised form
17 November 2013
Accepted 18 November 2013
Available online 1 December 2013

Keywords:
Functional data analysis
Call center data
Workload process
B-spline
Knot insertion algorithm

ABSTRACT

We show how to construct the best linear unbiased predictor (BLUP) for the continuation of a curve, and apply the proposed estimator to real-world call center data. Using the BLUP, we demonstrate prediction of the workload process, both directly and based on prediction of the arrival counts. The Matlab code and all data sets in the presented examples are available in the supplementary material.

1. Introduction

Many data sets consist of a finite number of observations, where each of these observations is a sequence of points. It is often natural to assume that each sequence is a set of noisy measurements of points on a smooth curve. In such cases, it can be advantageous to address the observations as functional data rather than as a multiple series of data points. This approach was found useful, for example, in noise reduction, missing data handling, and in producing robust estimations (see the books of Ramsay and Silverman, 2002, 2005, for a comprehensive treatment of functional data analysis). In this work we consider the problem of forecasting the continuation of a curve using functional data techniques.

The problem we consider here is relevant to longitudinal data sets, in which each observation consists of a series of measurements over time that are sampled from an underlying curve, possibly with noise. Examples of such curves are growth curves of different individuals and arrival rates of calls to a call center or of patients to an emergency room during different days. We assume that such curves, or measurement series that approximate these curves, were collected previously. We would like to estimate the continuation of a new curve given its beginning, using the behavior of the previously collected curves.

The forecasting of curve continuation suggested here is based on finding the best linear unbiased predictor (BLUP) (Robinson, 1991). We assume that the curves are governed by a small number of underlying functional patterns, possibly with additional noise. These underlying functional patterns determine the main variation between the different curves.

* Corresponding author. Tel.: +972 48288507.
E-mail addresses: yair.goldy@gmail.com, ygoldberg@stat.haifa.ac.il (Y. Goldberg), yaacov.ritov@huji.ac.il (Y. Ritov), avim@tx.technion.ac.il (A. Mandelbaum).

0378-3758/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2013.11.006

The computation of the predictor is performed in two steps. First, the underlying functional patterns' coefficients are estimated from the beginning of the new curve, which is defined on the "past" segment. Second, the prediction is obtained by computing the representation of the curve patterns on the "future" segment. We prove that the resulting estimator is indeed the BLUP and that it is a smooth continuation of the beginning of the curve (at least in the absence of noise). From a computational point of view, we discuss the use of B-splines in the representation of the curves on the different segments. We explain why a B-spline representation ensures an efficient and stable way to compute the mean function and covariance operators on different partial segments.

We apply the proposed forecasting procedure to call center data. We forecast the continuation of two processes: the arrival process and the workload process (i.e., the amount of work in the system; see, for example, Aldor-Noiman et al., 2009). In call centers, the forecast of the arrival process plays an important role in determining staffing levels. Optimization of the latter is important since salaries account for about 60–70% of the cost of running a call center (Gans et al., 2003). Usually, call center managers utilize forecasts of the arrival process and knowledge of service times, along with some understanding of customer patience characteristics (Zeltyn, 2005), to estimate future workload and determine the staffing level (Aldor-Noiman et al., 2009). The disadvantage of this approach is that the forecast of the workload is not performed directly, and instead it is obtained using the forecast of the arrival process. Reich (in preparation) showed how the workload process can be estimated explicitly, thereby enabling direct forecast of the workload. In this work, we forecast the continuation of both the arrival and workload processes, given past days' information and the information up to a certain time of day. Since the actual processes are not smooth, we first approximate these processes with smooth curves. We compare direct and indirect forecasting results for the workload process. We also compare our results for the arrival process to those of other forecasting techniques, namely, to the techniques that were introduced by Weinberg et al. (2007) and Shen and Huang (2008).

This paper has two main contributions. First, we present a novel functional-data prediction method for continuation of a curve. We show that the proposed method is the best unbiased linear predictor for continuation of a curve. The proposed estimator is fast and easy to compute, and is a continuous process in time, thus enabling prediction for any given future time. Second, we demonstrate how to predict the workload process directly, and compare this direct method to the usual indirect ones that are based on prediction of the arrival process.

Forecasting of the continuation of a function was considered in previous works. Aguilera et al. (1997) proposed to predict the continuation of the curve by regression of the principal components of the second part of the interval on the principal components of the first part of the interval. Shen (2009), in the context of time series data, proposed to first forecast the new curve entirely, and then update this forecast based on the given curve beginning. Both of these methods do not discuss curve continuity at the point dividing the interval, or optimality of the estimator. In a different context, Yao et al. (2005) proposed a functional data method for sparse longitudinal data that enables prediction of curves, even if only a few measurements are available for each curve. Although this method can be used to forecast the continuation of a curve, it was not designed to optimize such prediction. This is also evident in the case study in Section 4, in which we compare the method of Yao et al. (2005) to the proposed BLUP.

The paper is organized as follows. The functional model and notation are presented in Section 2. In Section 3 we show how to construct the BLUP for the continuation of a curve. In Section 4 we apply the estimator to real-world data, comparing direct and indirect workload forecasting, and our results to other techniques. Concluding remarks appear in Section 5. Proofs are provided in the Appendix. A link for the code and data sets used for the case study appear in the Supplemental Materials.

2. The functional framework

In this section we present the functional model and notation that will be used for the construction of the BLUP.

2.1. The functional model

Assume that we observe random i.i.d. functions $Y^{(1)},\ldots,Y^{(M)}$ that are defined on the segment $S=[0,T]$. We assume that these functions have a basis expansion with respect to some N-dimensional continuous function space $\mathcal{S}$ (N can possibly be large). For now we do not impose any specific structure on the space $\mathcal{S}$, but in Sections 2.3 and 2.4 we will focus on spline functional spaces as an important example. Given a new function $Y^{(M+1)}$ which is observed only on the segment $S_1=[0,U]$, for some $0<U<T$, we would like to estimate the continuation of this function on the segment $S_2=[U,T]$.

We assume that, up to some noise, the functions $\{Y^{(m)}(t)\}_m$ are contained in some low-dimensional subspace of $\mathcal{S}$. More specifically, we assume that each function can be written as

$$Y^{(m)}(t)=\mu(t)+\sum_{i=1}^{p} h_i^{(m)}\phi_i(t)+\varepsilon^{(m)}(t)=\mu(t)+\boldsymbol{h}^{(m)'}\boldsymbol{\phi}(t)+\varepsilon^{(m)}(t), \tag{1}$$

where $\mu(t)\in\mathcal{S}$ is the mean function, $\boldsymbol{h}^{(m)}=(h_1^{(m)},\ldots,h_p^{(m)})'$ is a random vector with mean zero and covariance matrix L, $\boldsymbol{\phi}(t)=(\phi_1(t),\ldots,\phi_p(t))'$ is a vector of orthonormal functions in $\mathcal{S}$; $\varepsilon^{(m)}(t)$ is the noise which is defined to be the part of Y that is not in the span of the $\boldsymbol{\phi}(t)$. We assume that p, the dimension of the subspace, is much smaller than N, the dimension of $\mathcal{S}$. Such decomposition can arise, for example, when using principal component analysis for functional data (Ramsay and

Silverman, 2005, Chapter 8). Finally, note that by definition, the noise term $\varepsilon^{(m)}(t)$ is anything within $\mathcal{S}$ that cannot be explained by the model functions μ and $\boldsymbol{\phi}$.

Denote the basis of $\mathcal{S}$ by $\boldsymbol{b}=(b_1,\ldots,b_N)'$. Write $\mu(t)=\boldsymbol{b}(t)'\boldsymbol{\mu}$ and $\boldsymbol{\phi}(t)=A'\boldsymbol{b}(t)$, for some $N\times 1$ vector $\boldsymbol{\mu}$ and $N\times p$ loading matrix A. Define

$$X^{(m)}(t)=\mu(t)+\boldsymbol{h}^{(m)'}\boldsymbol{\phi}(t)=\boldsymbol{b}(t)'(\boldsymbol{\mu}+A\boldsymbol{h}^{(m)}) \tag{2}$$

to be the noise-free part of $Y^{(m)}$ contained in the low-dimensional subspace spanned by $\boldsymbol{\phi}$. With respect to the basis $\boldsymbol{b}$ we can write $X^{(m)}(t)=\boldsymbol{b}(t)'\boldsymbol{x}^{(m)}$ and similarly $Y^{(m)}(t)=\boldsymbol{b}(t)'\boldsymbol{y}^{(m)}$ where $\boldsymbol{x}^{(m)}$ and $\boldsymbol{y}^{(m)}$ are the $N\times 1$ random vectors.

Let the random functions $Y_1^{(m)}$ and $Y_2^{(m)}$ be the restrictions of $Y^{(m)}$ to the segments $S_1=[0,U]$ and $S_2=[U,T]$, respectively, and similarly for $X_1^{(m)}$ and $X_2^{(m)}$. Let $\boldsymbol{b_i}(t)=(b_{i1}(t),\ldots,b_{iN_i}(t))$ be a basis of $\mathcal{S}_i$, the restriction of $\mathcal{S}$ to the segment S_i. Let $\boldsymbol{\mu_i}$, $\boldsymbol{x_i^{(m)}}$, $\boldsymbol{y_i^{(m)}}$ and A_i be the coefficient representations of μ, $X^{(m)}$, $Y^{(m)}$, and A, respectively, in the basis $\boldsymbol{b_i}$.

2.2. The covariance structure

Let $u(s,t)=\mathrm{Cov}(X(s),X(t))$ and $v(s,t)=\mathrm{Cov}(Y(s),Y(t))$, and let $\boldsymbol{b}(s)'g\boldsymbol{b}(t)$ and $\boldsymbol{b}(s)'G\boldsymbol{b}(t)$ be their respective matrix representations with respect to the basis $\boldsymbol{b}$. For $s\in S_i$ and $t\in S_j$, $i,j=1,2$, let $u_{ij}(s,t)=\boldsymbol{b_i}(s)'g_{ij}\boldsymbol{b_j}(t)$ and $v_{ij}(s,t)=\boldsymbol{b_i}(s)'G_{ij}\boldsymbol{b_j}(t)$ be the matrix representations of the restriction of the covariance functions to the partial segments. Finally, we define the operators γ_{ij} and Γ_{ij} from $\mathcal{S}_j$ to $\mathcal{S}_i$, for $i,j=1,2$, by

$$\begin{aligned}(\gamma_{ij}f)(t)&=\int_{S_j}u(s,t)f(s)\,\mathrm{d}s=\boldsymbol{b_i}(t)'g_{ij}W_j\boldsymbol{f}\\(\Gamma_{ij}f)(t)&=\int_{S_j}v(s,t)f(s)\,\mathrm{d}s=\boldsymbol{b_i}(t)'G_{ij}W_j\boldsymbol{f},\end{aligned} \tag{3}$$

where $W_j=\int_{S_j}\boldsymbol{b_j}(s)\boldsymbol{b_j}(s)'\,\mathrm{d}s$, and $\boldsymbol{f}$ is the expansion of the function f in $\boldsymbol{b_j}$. More details on the covariance structure can be found in the Appendix.

Note that even when given the basis $\boldsymbol{b}$ on the full segment, there is not necessarily an easy and efficient way to compute the bases $\boldsymbol{b_1}$ and $\boldsymbol{b_2}$. Similarly, there is not necessarily an easy and efficient way to compute the coefficient representations for the functions $X_i^{(m)}$ and $Y_i^{(m)}$, $m=1,\ldots,M$. Also finding the matrices g_{ij} and G_{ij}, even given the matrices g and G, can be challenging. This differs from the Euclidean case, where g_{ij} is a submatrix of g. In the following section we represent the functions $X^{(m)}(t)$ and $Y^{(m)}(t)$ using appropriate B-spline bases. We show that for B-spline function spaces there is an efficient way to compute the bases $\boldsymbol{b_1}$ and $\boldsymbol{b_2}$, the vectors of coefficients of $X_i^{(m)}(t)$ and $Y_i^{(m)}(t)$, and the matrices g_{ij} and G_{ij}.

2.3. B-spline spaces

Until now we assumed that the random functions $X^{(m)}(t)$ and their noisy version $Y^{(m)}(t)$ are contained in some N-dimensional continuous function space $\mathcal{S}$. Here, we suggest that the space $\mathcal{S}$ be chosen as a spline space, i.e., a space of piecewise polynomial functions. The use of splines is common in functional data analysis due to the simplicity of spline computation, and the ability of splines to approximate smooth functions (see, for example, Ramsay and Silverman, 2005). There are two more advantages of using finite-dimensional spline functional spaces in our case. First, the functional space restriction from the whole segment to a partial segment (the "past" segment or the "future" segment) has a natural B-spline basis that has a lower number of elements. This solves collinearity problems which can render any projection on the partial segment basis unstable. Second, the knot-insertion algorithm (see, de Boor, 2001, Chapter 11) ensures an efficient and stable way to compute the mean function and covariance operators on different partial segments. In the following we discuss shortly spline spaces and the knot-insertion algorithm. We then explain how to use this algorithm in order to restrict functions in $S_{k,\tau}$ to the subsegments S_1 and S_2. We refer the reader to de Boor (2001) for more details regarding splines, B-splines and the knot-insertion algorithm.

Let $S_{k,\tau}$ be a spline space, where k denotes the splines' order and where τ is a fixed knot sequence on $[0,T]$. Let $\boldsymbol{b}=(b_1,\ldots,b_N)'$ be the B-spline basis of $S_{k,\tau}$. Let τ_1 and τ_2 be knot sequences that agree with τ on the segments $[0,U)$ and $(U,T]$, respectively, and have knot multiplicity of k at U. Let S_{k,τ_i} for $i=1,2$ be the k-ordered spline space with knot sequence τ_i, and let $\boldsymbol{b_i}(t)=(b_{i1}(t),\ldots,b_{iN_i}(t))$ be its corresponding B-spline basis. We wish to represent $X_i^{(m)}$ and $Y_i^{(m)}$ $(i=1,2;\ m=1,\ldots,M)$ using the representations of $X^{(m)}$ and $Y^{(m)}$.

It is enough to represent the functions $X_i^{(m)}$ and $Y_i^{(m)}$ using the bases $\boldsymbol{b_i}$. Recall that $X^{(m)}(t)=\mu(t)+\boldsymbol{b}(t)'A\boldsymbol{h}^{(m)}$, and $\mu(t)=\boldsymbol{b}(t)'\boldsymbol{\mu}$ for some vector of coefficients $\boldsymbol{\mu}$. Using the knot-insertion algorithm we can obtain new vectors $\boldsymbol{\mu_i}$ such that (a) $\mu(t)=\boldsymbol{b_i}(t)'\boldsymbol{\mu_i}$ for all t on which $\boldsymbol{b_i}$ is defined, and (b) $\boldsymbol{\mu_i}$ is obtained from $\boldsymbol{\mu}$ by truncation and a change of at most k coefficients. Similarly, by truncation and a change of at most pk coefficients, we can obtain the loading matrices A_i such that $\boldsymbol{b}(t)'A=\boldsymbol{b_i}(t)'A_i$ for all t on which $\boldsymbol{b_i}$ is defined. Note that these truncations applied only to the shared components of the model, i.e., the mean vector, the loading matrices and covariance matrices; the unique coefficients that determine each

random function are not touched. Hence we obtain

$$X_i^{(m)}(s) = \boldsymbol{b_i}(s)'(\boldsymbol{\mu_i} + A_i \boldsymbol{h}^{(m)}) \tag{4}$$

for $i=1,2$ and for all $s \in S_i$. In a similar way, we can write $Y_i(s) = \boldsymbol{b_i}(s)'\boldsymbol{y}_i^{(m)}$.

We now compute the covariance matrices g_{ij} and G_{ij} from the covariance matrices g and G. Starting with g, recall that $\mathrm{Cov}(X(s),X(t)) = \boldsymbol{b}(s)'g\boldsymbol{b}(t) = \boldsymbol{b}(s)'ALA'\boldsymbol{b}(t)$ for some diagonal matrix L. Hence, since $\boldsymbol{b}(t)'A = \boldsymbol{b_i}(t)'A_i$, we obtain that

$$\mathrm{Cov}(X_i(s),X_j(t)) = \boldsymbol{b_i}(s)'A_iLA_j'\boldsymbol{b_j}(t),$$

i.e., $g_{ij} = A_iLA_j'$. The computation of G_{ij} is done similarly.

2.4. Estimation of the functional model

In the following we discuss the estimation of random functions $X^{(m)}$ and $Y^{(m)}$, and their restrictions $X_i^{(m)}$ and $Y_i^{(m)}$ $(i=1,2;\ m=1,\ldots,M)$ from M i.i.d. longitudinal data trajectories. We also discussed the vector of coefficients of the expectation function $\boldsymbol{\mu}$, the loading matrices A and V, the covariance matrices g and G, as well as their restrictions to subsegments. While there are many different estimation techniques, we restrict the discussion to smoothing splines. In this discussion we follow Ramsay and Silverman (2005, Chapter 5).

We assume that the data is given as a set of M longitudinal trajectories

$$\underline{y}^{(m)} = (Y^{(m)}(t_{1m}),\ldots,Y^{(m)}(t_{nm})), \quad m=1,\ldots,M$$

where $0 \le t_{1m} < \cdots < t_{nm} \le T$ are time points on which the evaluation of $Y^{(m)}$ is given. As before, let $S_{k,\tau}$ be a B-spline space where k denotes the splines' order and where τ is a fixed knot sequence on $[0,T]$. We estimate the coefficients of $Y^{(m)}$ with respect to the B-spline basis $\boldsymbol{b}$ using smoothing splines.

Using the coefficient vector representation of the functions $Y^{(1)},\ldots,Y^{(m)}$ with respect to the basis $\boldsymbol{b}$ we estimate the mean function $\mu(t)$ and the matrix G as follows. The mean function is estimated by $\boldsymbol{b}(t)'\hat{\boldsymbol{\mu}}$ where $\hat{\boldsymbol{\mu}}$ is the mean of the vectors $\hat{\boldsymbol{y}}^{(1)},\ldots,\hat{\boldsymbol{y}}^{(M)}$. Define the matrix C to be an M by N matrix where the m-th row of C is given by $(\hat{\boldsymbol{y}}^{(m)} - \hat{\boldsymbol{\mu}})'$. Then we define G as $M^{-1}C'C$, and consequently the covariance function is given by $v(s,t) = \boldsymbol{b}(s)'G\boldsymbol{b}(t)$.

We are now ready to estimate the loading matrix A. This is done by performing principal component analysis (PCA) for functional data (Ramsay and Silverman, 2005, Chapter 8). Let $\rho_1,\ldots,\rho_N$ and $\phi_1,\ldots,\phi_N$ be, respectively, the eigenvalues and eigenfunction of the operator

$$\Gamma f(s) = \int_0^T v(s,t)f(s)\,ds \equiv GW\boldsymbol{f},$$

where $W = \int_0^T \boldsymbol{b}(s)\boldsymbol{b}(s)'\,ds$, and $\boldsymbol{f}$ is the expansion of the function f in $\boldsymbol{b}$. Note that these eigenvalues and eigenfunctions can be obtained by solving the eigenvalue problem $W^{1/2}GW^{1/2}\boldsymbol{f} = \rho\boldsymbol{f}$ and computing $\phi(t) = \boldsymbol{b}(t)'W^{-1/2}\boldsymbol{f}$ for each eigenvector (see, Ramsay and Silverman, 2005, Chapter 8.4 for details).

Choosing p, the number of principal components that describe the model is a challenge. There are many techniques and rules of how to choose p, and we refer the reader to Jolliffe (2002, Chapter 6) for a survey of common techniques. Since our goal is to predict the continuation of a function, we choose the number of principal components as the number that best performs the prediction. In the numerical examples this was done using K-fold cross validation.

3. The construction of the BLUP

Given Y_1, the noisy version of the first part of the random function X, our goal is to find a *good* estimator for X_2, the continuation of X_1. For simplicity, we restrict the discussion to linear estimators.

Following Robinson (1991), we say that $\hat{X}_2$ is a *good* linear estimator of X_2 given Y_1 if the following criteria hold:

(C1) $\hat{X}_2$ is a linear function of Y_1.
(C2) $\hat{X}_2$ is unbiased, i.e., $E[\hat{X}_2(t)] = \mu(t)$.
(C3) $\hat{X}_2$ has minimum mean square error among the class of linear unbiased estimators.

Two additional demands regarding the estimator that seem desirable in our context are as follows:

(C4) The random function $\hat{X}_2$ lies in the space $\mathcal{S}_2$.
(C5) When no noise is introduced, i.e., when $Y_1 = X_1$, the concatenation of $\hat{X}_2$ to X_1 lies in $\mathcal{S}$. In other words, the combined function

$$\hat{X} = \begin{cases} X_1(t), & 0 \le t \le U \\ \hat{X}_2(t), & U < t \le T \end{cases}$$

is smooth enough.

An estimator that fulfills (C1)–(C5) will be referred to as a best linear unbiased predictor (BLUP). In this section we will show how to construct such a BLUP and prove that it is defined uniquely.

Remark. Note that the definition of an unbiased estimator in (C2) is not the usual definition. A more restrictive criterion is

(C2*) $\hat{X}_2$ is unbiased in the following sense: $E[\hat{X}_2(t)|Y_1] = E[X_2(t)|Y_1]$.

We will show that when Y is a Gaussian process, i.e., the random vectors $\boldsymbol{h}$ and $\boldsymbol{\epsilon}$ are multivariate normal, this criterion is fulfilled by the proposed BLUP as well.

Define the function

$$v_{11}^{+}(s,t) = \boldsymbol{b_1}(s)' W_1^{-1} G_{11}^{+} W_1^{-1} \boldsymbol{b_1}(t),$$

for every $s, t \in S_1$, where G_{11}^{+} is the Moore–Penrose pseudoinverse of G_{11}. Note that W_1 is invertible since it is a Gram matrix of basis functions (see Sansone, 1991, Theorem 1.5). Define the operator $\Gamma_{11}^{+} : \mathcal{S}_1 \to \mathcal{S}_1$ by

$$(\Gamma_{11}^{+} f)(t) = \int_{S_1} v_{11}^{+}(s,t) f(s)\, ds = \boldsymbol{b_1}(t)' W_1^{-1} G_{11}^{+} \boldsymbol{f},$$

where $\boldsymbol{f}$ is the expansion of the function f in the basis $\boldsymbol{b_1}$. We note that when the functional space $\mathcal{S}_1$ is infinite-dimensional, the operator Γ_{11} needs not to have a bounded pseudo-inverse. However, since we consider $\mathcal{S}_1$ to have a finite basis, this problem does not arise.

We are now ready to define the estimator for X_2 given Y_1, by

$$\hat{X}_2(t) = \mu(t) + \gamma_{21} \Gamma_{11}^{+}(Y_1 - \mu)(t) = \boldsymbol{b_2}(t)'(\boldsymbol{\mu_2} + g_{21} G_{11}^{+}(\boldsymbol{y_1} - \boldsymbol{\mu_1})), \tag{5}$$

for every $t \in S_2$. Then we have

Theorem 1. *Let X and its noisy version Y be random functions in an N-dimensional continuous function space $\mathcal{S}$, with basis expansions (2) and (1), respectively, and with covariance structure (3). Then, the estimator $\hat{X}_2$ meets criteria* (C1)–(C5) *and is unique up to equivalence. Moreover, if Y is a Gaussian process, then $\hat{X}_2$ meets criterion* (C2*) *as well.*

See proof in the Appendix.

The estimation of $\hat{X}_2$ has a simpler form when the noise is modeled in the following way. Assume that

$$Y_1(t) = \boldsymbol{b_1}(t)'(\boldsymbol{\mu_1} + A_1 \boldsymbol{h} + \boldsymbol{\epsilon_1}) \tag{6}$$

where $\boldsymbol{\epsilon_1}$ is an $N_1 \times 1$ mean zero random vector with $\sigma^2 I$ covariance matrix and I is the identity matrix. In this case

$$\hat{X}_2(t) = \boldsymbol{b_2}(t)'(\boldsymbol{\mu_2} + g_{21}(A_1 L A_1' + \sigma^2 I)^{-1}(\boldsymbol{x_1} - \boldsymbol{\mu_1})) \tag{7}$$

which has a similar structure to the ridge regression estimator (Hoerl and Kennard, 1970). However, it is important to note the difference between the role that the parameter σ^2 plays in the different models. In ridge regression, the parameter σ^2 determines the tradeoff between bias and variance, i.e., the smaller the σ^2, the smaller the bias (see, Gross, 2003, Chapter 3.4 for details). In our model the estimator $\hat{X}_2(t)$ is unbiased when σ^2 is known. When σ^2 is not known it can be estimated from the past data. Since the goal is to estimate $\hat{X}_2$, estimation of σ^2 can be done using cross validation, where the value of the parameter is chosen as the one that yields the best prediction for the past data (see also Section 4.3).

Note that the expression in (7) involves inverting an $N_1 \times N_1$ matrix $(A_1 L A_1' + \sigma^2 I)$; a simpler expression can be obtained using some matrix algebra (see, Robinson, 1991, Eq. (5.2)). We have

$$g_{21}(A_1 L A_1' + \sigma^2 I)^{-1} = A_2 L A_1'(A_1 L A_1' + \sigma^2 I)^{-1} = A_2(A_1' A_1 + \sigma^2 L^{-1})^{-1} A_1',$$

and hence

$$\hat{X}_2(t) = \boldsymbol{b_2}(t)'(\boldsymbol{\mu_2} + A_2(A_1' A_1 + \sigma^2 L^{-1})^{-1} A_1'(\boldsymbol{x_1} - \boldsymbol{\mu_1})), \tag{8}$$

which involves only the inverse of a $p \times p$ matrix. This final equation is the form used in the case study that appears in Section 4.

4. Case study

In this section we apply the estimator $\hat{X}_2$ to call center data. We are interested in forecasting the continuation of two processes: the arrival process and the workload process. The estimators of these two processes play an important role in determining the staffing level at call centers (see, for example, Aldor-Noiman et al., 2009; Shen and Huang, 2008; Reich, in preparation). Usually, staffing levels are determined in advance, at least one day ahead. Here we propose a method for updating the staffing level, given information obtained from the beginning of the day. As noted by Gans et al. (2003) and by Shen and Huang (2008), such updating is operationally beneficial and feasible. If the forecasting is performed appropriately, it could result in higher efficiency and service quality: based on the revised forecasts, a manager can adjust staffing levels by

calling in additional agents or dismissing agents as needed, and by transferring agents from or to other activities such as email inquiries and faxes.

This section is organized as follows. We first describe the arrival and workload processes (Section 4.1). We then describe the data (Section 4.2) and the forecast implementation (Section 4.3). The analysis appears in Sections 4.4–4.6. Finally, confidence bands are discussed in Section 4.7.

4.1. The arrival and workload processes

We define the arrival process of day j, $a_j(t)$, as the number of calls that arrive on day j during the time interval $[t-c,t]$, where t varies continuously over time and c is a fixed constant. Note that $a_j(t)$ itself is not a continuous function. When the call volume is large and this function does not change drastically over short time intervals, the function $a_j(t)$ can be well approximated by a smooth function. Moreover, it can be assumed that the function $a_j(t)$, for each day j, arises from some underlying deterministic smooth arrival rate function $\lambda(t)$ plus some noise (Weinberg et al., 2007). Using the notation of Section 2, we assume that a smooth version of $a_j(t)$ can be written as $\lambda(t)+\sum_{i=1}^{p} h_i^{(j)}\phi_i(t)+\varepsilon^{(j)}(t)$. We now describe the workload process $w_j(t)$ for each day j. The function $w_j(t)$ counts the number of calls that would have been handled by the call center on day j at time t, *assuming an unlimited number of agents and hence no abandonments*. From a management point of view, the advantage of looking at $w_j(t)$ over $a_j(t)$ is that $w_j(t)$ reflects the number of agents actually needed at each point in time. However, as opposed to the process $a_j(t)$, which is observable in real time, the computation of $w_j(t)$, for a specific time t, involves estimation of call durations for abandoned calls and can be performed only after all calls entered up to time t are actually served (see the discussion of Aldor-Noiman et al., 2009; Reich, in preparation).

4.2. The data

The data used for the forecasting examples were gathered at a call center of a large U.S. commercial bank. The bank has various types of operations such as retail banking, consumer lending and private banking. Since the call arrival pattern varies over different types of services, we restrict attention to retail services, which account for approximately 70% of the calls (see, Weinberg et al., 2007). The first two examples are of the arrival process and the workload process, for weekdays between March and October 2003. The data for the first example consists of the arrival counts at 5-min resolutions between 7:00 AM and 9:05 PM (i.e., $c=5$ in the definition of $a_j(t)$). The data for the second example consists of an average workload, also in 5-min resolutions, between 7:00 AM and 9:05 PM. There are 164 days in the data set after excluding some abnormal days such as holidays. Fig. 1 shows arrival count profiles for different days of the week.

The third example explores the arrival process during weekends between March and October 2003. There are 67 days in the data set (excluding one day with incomplete data). As can be seen from Fig. 1, the weekend behavior is different from

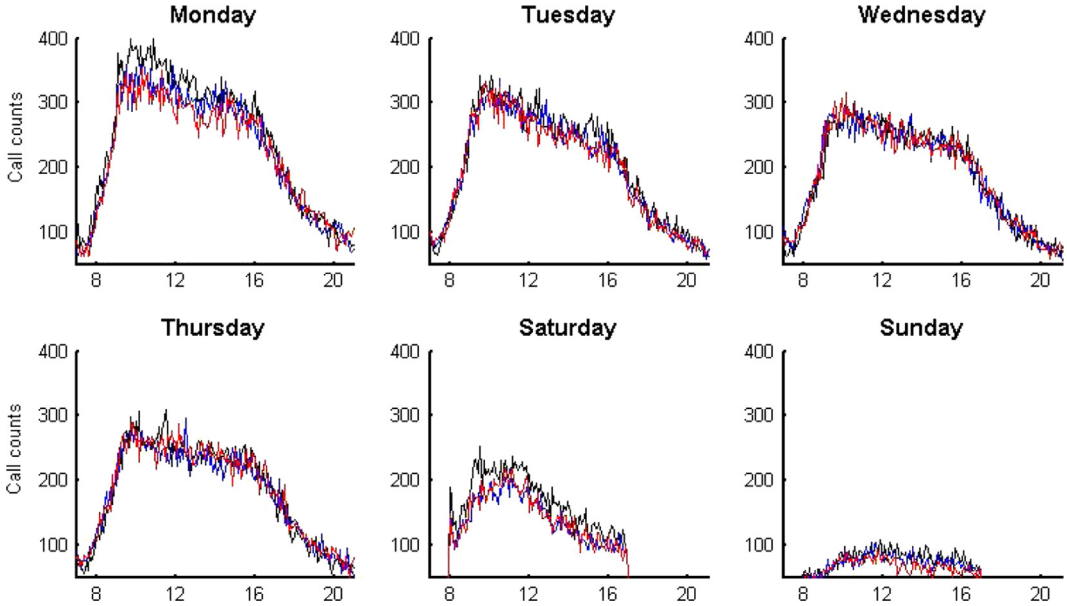

Fig. 1. Arrival count in 5-min resolutions for three successive weeks, grouped according to weekday (Friday was omitted due to space constraints). Each color represents a different week. There is a clear difference between workdays, Saturdays, and Sundays. For the working days, it seems that there is a common pattern. Between 7 AM and 10 AM the call count rises sharply to its peak. Then it decreases gradually until 4 PM. From 4 PM to 5 PM there is a rapid decrease followed by a more gradual decrease from 5 PM until 12 AM. The call counts are smaller for Saturday and much smaller for Sunday. Note also that the main activity hours for weekends are 8 AM to 5 PM, as expected.

that of the working days, and there is a Saturday pattern and a Sunday pattern. The data for this example consists of the arrival counts at 15-min resolutions between 8 AM and 5 PM. The change in interval length from the previous two examples is due to the decreased call-counts. The change in day length is due to the low activity in early morning and late afternoon hours on weekends (see Fig. 1).

In the first and second examples, we used the first 100 weekdays as the training set and the last 64 weekdays as the test set. For each day from day 101 to day 164, we extracted the *same-weekday* information from the preceding 100 days. Thus, for each day of the week we have about 20 training days. For the third example, the test set consists of weekend days 41–67 while the training set for each day consists of its previous 40 weekend days. Thus, similarly, for each day we have about 20 training days. Additionally, we used the data from the start of the day, up to 10 AM and up to 12 PM. All forecasts were evaluated using the data after 12 PM, which enabled fair comparison between the results of the different cut points (10 AM and 12 PM). We also compare our results to the mean of the preceding days, from 12 PM on.

For a detailed description of the first example's data, the reader is referred to Weinberg et al. (2007), Section 2. For an explanation of how the second example's workload process was computed, the reader is referred to Reich (in preparation). The data for the third example was extracted using SEEStat, which is a software written at the Technion SEELab.[1] We refer the reader to Donin et al. (2006) for a detailed description of the U.S. commercial bank call-center data from which the data for all three examples was extracted. The U.S. bank call-center data is publicly accessible from the SEELab server (footnote 1).

4.3. Forecast implementation

The forecast was performed using a Matlab implementation of the BLUP algorithm from Section 3, where we follow the model discussed in (6). For the implementation we used the Matlab functional data analysis library written by Ramsay and Silverman.[2] We also used the Matlab library Spider for implementing cross-validation.[3] The Matlab code, as well as the data sets, are downloadable (see Supplemental Materials).

In all computations we used cubic-splines, i.e., fourth-order splines. We used knot sequences with a knot every hour. This means that in the first and the second examples below the dimension of the spaces S, S_1, and S_2 are 17, 6, and 14, respectively, for the 10 AM cut point, and 17, 8, and 12, respectively, for the 12 PM cut point. For the third example below, the dimension of the spaces S, S_1, and S_2 are 12, 5, and 10, respectively, for the 10 AM cut point, and 12, 7, and 8, respectively, for the 12 PM cut point.

We used a 5-fold-cross-validation to choose the dimension p of the subspace spanned by X, and the variance σ^2 (see Eq. (8)), among the set of pairs

$$(\sigma^2, p) = (10^{-2} \cdot 2^i, j), \quad i \in \{0,1,2\}, \;\; j \in \{1,2\}.$$

We quantified the results using both Root Mean Squared Error (RMSE) and Average Percent Error (APE), which are defined as follows. For each day j, let

$$RMSE_j = \left(\frac{1}{R}\sum_{r=1}^{R}(\mathcal{N}_{jr} - \hat{\mathcal{N}}_{jr})^2\right)^{1/2}; \quad APE_j = \frac{100}{R}\sum_{r=1}^{R}\frac{|\mathcal{N}_{jr} - \hat{\mathcal{N}}_{jr}|}{\mathcal{N}_{jr}},$$

where $\mathcal{N}_{jr}$ is the actual number of calls (mean workload) at the r-th time interval of day j in the arrival (workload) process application, $\hat{\mathcal{N}}_{jr}$ is the forecast of $\mathcal{N}_{jr}$, and R is the number of intervals. We computed local confidence bands with a 95% confidence level using cross-validation, as described in Section 4.7.

4.4. First example: arrival process for weekday data

In the first example we would like to predict the arrival process to a call center given the beginning of current day data and previous days' data. For the current day data we use two cut points, 10 AM and 12 PM, and also compare it to a forecast that is based only on the previous days' data.

Forecasting the arrival process for the first example data was studied by both Weinberg et al. (2007) and Shen and Huang (2008). Weinberg et al. assumed that the day patterns behave according to an autoregressive model. The algorithm they suggest starts by forecasting the current day based on previous days' data. The algorithm estimates the parameters in the autoregressive model using Bayesian techniques. An update for the continuation of the current day's forecast is obtained by conditioning on the data of the current day up to the cut point. We refer to this algorithm as Bayesian update (BU) for short. Similarly, the algorithm by Shen and Huang assumes an autoregressive model and gives a forecast for the current day. They then update this forecast using least-square penalization, assuming an underlying discrete process. We will refer to this algorithm as penalized least square (PLS). Yao et al. (2005) developed a functional data method for sparse longitudinal data

[1] SEELab: The Technion Laboratory for Service Enterprize Engineering. Webpage: http://ie.technion.ac.il/Labs/Serveng.
[2] The functional data analysis Matlab library can be downloaded from ftp://ego.psych.mcgill.ca/pub/ramsay/FDAfuns/Matlab/.
[3] The Matlab library Spider can be downloaded from http://people.kyb.tuebingen.mpg.de/spider/.

Table 1
Summary of statistics (minimum, lower quartile (Q1), median, mean, upper quartile (Q3), maximum) of RMSE for the forecast based on the mean of the previous days (mean), and BU, PLS, PACE, and BLUP, using data up to 10 AM and up to 12 PM for the call arrival data set. The results for BU and PLS were taken from the original papers. No maximum and minimum results were given for PLS. The table shows that the RMSE of BU, PLS, and BLUP, given the 10 AM data, is smaller than of previous days' mean, and it decreases further given 12 PM data. For the 12 PM data, the BLUP obtains the lowest RMSE in almost all categories. The prediction of the PACE algorithm does not improve as more information is gathered.

Example 1	Mean	10:00 AM				12:00 PM			
RMSE	BLUP	BU	PLS	PACE	BLUP	BU	PLS	PACE	BLUP
Minimum	12.37	**11.08**		11.44	11.67	**11.07**		11.42	12.05
Q1	14.07	14.00	**13.31**	14.57	13.48	13.56	13.33	14.62	**13.17**
Median	16.13	15.50	14.87	17.51	**14.40**	14.80	14.60	17.58	**14.17**
Mean	18.95	17.86	**16.48**	21.28	16.80	16.59	16.13	21.37	**15.99**
Q3	21.06	19.87	**17.26**	23.37	18.22	16.58	16.39	23.84	**16.23**
Maximum	68.60	57.72		78.18	**49.51**	53.66		78.55	**49.24**

that can handle continuation of a curve. This method, which we applied using the Matlab library PACE,[4] is referred to as PACE.

Comparison between the results of all four algorithms for the first data set appears in Table 1. Note that for BU, PLS, and BLUP, and all of the categories there is improvement in the 10 AM and 12 PM forecasts over the forecast based solely on past days. For these algorithms, the RMSE mean decreases by about 5–13% for the 10 AM forecast, and by 12–15% for the 12 PM forecast, depending on the algorithm. The fact that there is no improvement for PACE may be because this algorithm was not designed to optimize such prediction. It should be noted that the algorithms by Weinberg et al. and by Shen and Huang use information from all 100 previous days and the knowledge of the previous day call counts. For the methods that were designed for i.i.d. data, i.e., BLUP and PACE, we used only the same weekday information ($\sim$ 20 days instead of 100), and the previous day's information is not part of its training set.

The forecasting results for the week that follows Labor Day appear in Fig. 2. It can be seen that for the Tuesday that follows Labor Day (Monday) the call counts are much higher than usual. This is captured, to some degree, by the 10 AM forecast and much better by the 12 PM forecast. The same phenomenon occurs, with less strength, during the Wednesday and Thursday following Labor Day, until on Friday all the forecasts become roughly the same. It seems that the power of the continuation-of-curve forecasting is exactly for such situations, in which the call counts are substantially different than usual throughout the day, due to either predictable events, such as holidays, or unpredictable events.

4.5. Second example: workload process for weekday data

The second example consists of the workload process for weekday data for the same period as the first example. We forecast the workload process based on these sets of data: previous days' data, up to 10 AM data, and up to 12 PM data. We refer to this forecast as a *direct* workload forecast since we use past workload estimation as the basis for the forecast. An alternative (and simpler) workload forecasting method was proposed by Aldor-Noiman et al. (2009). Aldor-Noiman et al. (2009, Eq. (21)) suggest to forecast the workload by multiplying the forecasted arrival rate by the estimated average service time. We refer to this method as *indirect* workload forecasting.

For the indirect workload forecasting, one first needs to forecast the arrival rates. In this example, we estimated these quantities using the BLUP estimator. We note that the arrival rates could have been forecasted using other estimators. However, our goal is to compare between direct and indirect methods, and not between different estimators.

Comparison between the two methods appears in Table 2. Following Aldor-Noiman et al. (2009), we estimated the average service time over a 30-min period for indirect workload computations. Note that the direct workload forecast results are slightly better than the indirect workload forecast in most of the categories. Also note that in almost all categories, there is an improvement in the 10 AM and 12 PM forecasts over the forecast based solely on past days. The RMSE mean decreases by about 16% (8%) for the 10 AM forecast, and by 16% (13%) for the 12 PM forecast for the direct (indirect) forecast. Fig. 3 presents a visual comparison between the direct and the indirect forecast methods on a specific day. The two forecasts look roughly the same, which is also true for all other days in this data set.

While in this example there is no significant difference between the direct and indirect workload forecasts, we expect these methods to obtain different forecasts when the arrival rate changes during an average service time. This is true, for example, for arrival and service of patients in emergency rooms. The arrival rates of patients to emergency rooms can change within an hour while the time that a patient spends in emergency room (the "service time") is typically on the order of hours. As pointed out by Rozenshmidt (2008, Section 6), in such cases, forecasting the workload by the arrival count multiplied by the average service time may not be accurate. This is because the number of customers in the system is

[4] The Matlab library PACE can be downloaded from http://www.stat.ucdavis.edu/PACE/.

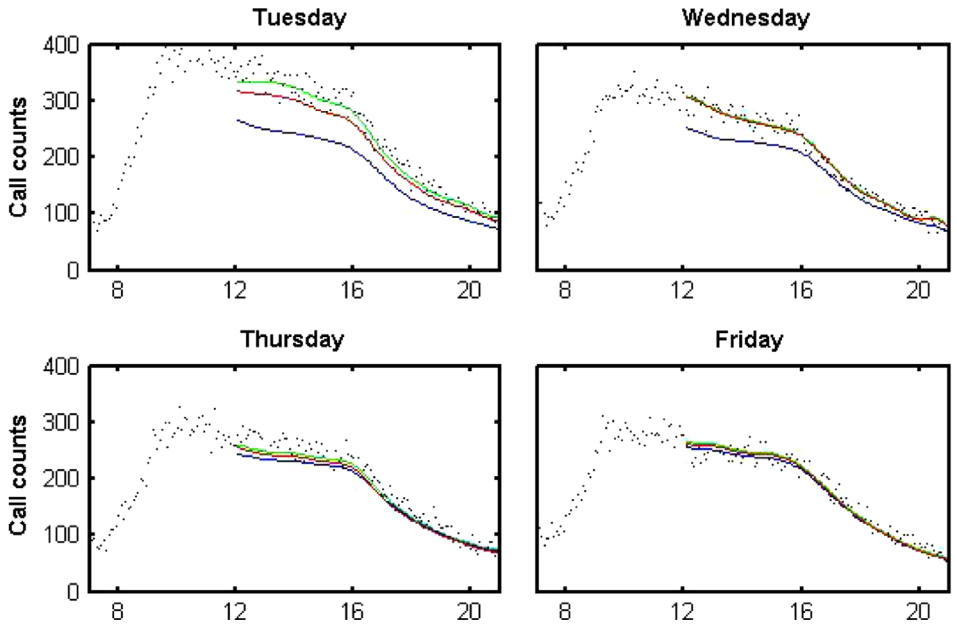

Fig. 2. Forecasting results for the week following Labor Day (September 2–5, 2003) for the call arrival process of the first example. Labor Day itself (Monday) does not appear since holiday data is not included in the data set. The black dots represent the true call counts in 5-min resolutions. The forecasts based on previous days, 10 AM data, and 12 PM data are represented by the blue, red, and green lines, respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

Table 2
Summary of statistics (minimum, lower quartile (Q1), median, mean, upper quartile (Q3), maximum) of RMSE for the forecast based on the mean of the previous days' data, up to 10 AM data and up to 12 PM data, for the workload data set, for both the indirect and the direct forecast methods using the BLUP.

Example 2	Day ahead		10:00 AM		12:00 PM	
RMSE	Workload (by arrivals)	Workload (explicitly)	Workload (by arrivals)	Workload (explicitly)	Workload (by arrivals)	Workload (explicitly)
Minimum	8.41	8.39	8.00	7.80	7.94	8.31
Q1	10.59	10.80	10.14	10.03	10.24	10.00
Median	11.99	12.29	11.60	10.97	11.39	11.27
Mean	15.79	15.96	14.60	13.36	13.95	13.24
Q3	15.08	15.34	14.33	14.16	13.65	13.20
Maximum	96.06	94.69	96.30	55.97	93.75	56.01

cumulative, while the arrival count counts only those who arrive in the current time interval. Thus, if the arrival count is lower than it was in the previous time interval and the average service time is long, the workload is underestimated. Similarly, if the arrival count is larger than the previous one, the workload is overestimated. Analyzing emergency room data would be interesting, but it is beyond the scope of this paper.

4.6. Third example: arrival process for weekend data

The third example is that of the weekend arrivals. The main difference between the first two examples and this one is that the data in this example cannot be considered as data from successive days, due to the six day difference between any Sunday and its successive Saturday. Note that the models considered by Weinberg et al. (2007) and Shen and Huang (2008) have an autoregressive structure, and hence are not directly applicable. Nevertheless, it would have been interesting to compare the performance of the BU and PLS prediction methods on this data. Since we do not have access to the code, the comparison to these methods was done only for the first example.

We forecasted arrival rate for the weekend data using both the BLUP and the PACE estimators. Note that even when the autoregressive structure does not hold, the results appearing in Table 3 reveal that forecasting for this data set is still beneficial. Indeed, for the BLUP method, the RMSE mean decreases by about 15% for the 10 AM forecast, and by 24% for the 12 PM forecast. These results are impressive since the curves in this example begin an hour later than the curves in the previous two examples.

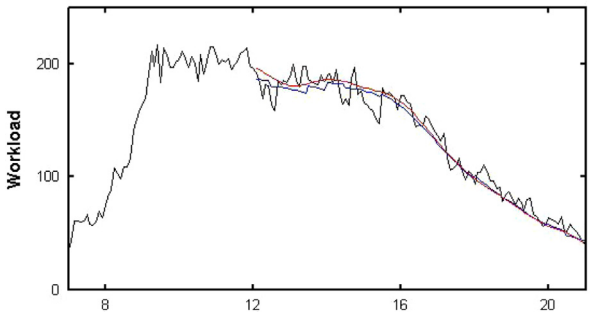

Fig. 3. Workload forecasting for Friday, September 5, 2003, using both the direct and the indirect methods. The black curve represents the workload process estimated after observing the data gathered throughout the day. The blue and red curves represent the workload forecasts for the indirect and direct forecasts, respectively, given data up to 12 PM. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

Table 3
Summary of statistics (minimum, lower quartile (Q1), median, mean, upper quartile (Q3), maximum) of RMSE for the forecast based on the mean of the previous days (mean), and PACE and BLUP using data up to 10 AM and up to 12 PM for the call arrival data set. The table shows that for BLUP there is an improvement in the prediction's RMSE given the 10 AM data, and further improvement given 12 PM data. The prediction of the PACE algorithm does not improve as more information is gathered.

Example 3	**Mean**	**10:00 AM**		**12:00 PM**	
RMSE	BLUP	PACE	BLUP	PACE	BLUP
Minimum	4.30	3.72	3.46	3.72	3.47
Q1	5.70	5.78	4.77	5.81	4.97
Median	9.01	7.03	6.27	7.07	6.13
Mean	8.87	8.35	7.53	8.43	6.71
Q3	10.56	10.45	9.13	10.7	7.71
Maximum	16.74	21.92	18.42	22.09	17.26

4.7. Confidence bands

Confidence bands are important for two main reasons. First, they enable one to assess the accuracy of the prediction. Second, which is more specific to our discussion, they enable one to choose the earliest cut point that provides a satisfying prediction. Clearly, the earlier the cut point time, the less accurate the prediction and hence the wider the confidence band. Using the width of the confidence band as a measure, one can determine the earliest cut point which provides the needed accuracy level.

We computed (local) confidence bands in the following way. We first estimated $\hat{D}(t) \equiv \mathrm{Var}(X_2(t)|Y_1)^{1/2}$. We then define the confidence band as the pair of functions

$$(\hat{X}_2(t) - C\hat{D}(t), \hat{X}_2(t) + C\hat{D}(t)), \tag{9}$$

where C is chosen using cross-validation. We omit the details.

Following Weinberg et al. (2007), we introduce the 95% confidence band coverage (COVER) and the average 95% confidence band width (WIDTH). Specifically, for each day j, let

$$COVER_j = \frac{1}{R}\sum_{r=1}^{R} I\left(F_{L,jr} < \mathcal{N}_{jr} < F_{U,jr}\right)$$

$$WIDTH_j = \frac{1}{R}\sum_{r=1}^{R} \left(F_{U,jr} - F_{L,jr}\right),$$

where $(F_{L,jr}, F_{U,jr})$ is the confidence band of day j, evaluated at the beginning of the r-th interval. The mean coverage and mean width, for all three examples, are presented in Table 4. First, note that for all three examples, the width of the confidence band narrows down as more information is revealed. In other words, the width of the confidence band for the 12 PM forecast is narrower than the width for the 10 AM forecast which, in turn, is narrower than the width for the previous days' mean. We also see that the mean coverage becomes more accurate as more information is revealed. Fig. 4 depicts the confidence bands for the arrival process on a particular Sunday. Note that the 10 AM forecast confidence band is narrower than the confidence band for the mean of the previous days' forecast. Moreover, at 12 PM, when more information on this particular day becomes available, the confidence band narrows down even more and captures the underlying behavior.

Table 4
The mean confidence band coverage and the mean width for the forecasts based on the previous days' mean, the 10 AM cut and the 12 PM cut for the arrival process on the working days data set (Example 1), the workload process on the working days data set (Example 2) and the arrival process on the weekend data set (Example 3).

Data	Confidence band coverage			Confidence band width		
	Example 1 (%)	Example 2 (%)	Example 3 (%)	Example 1 (%)	Example 2 (%)	Example 3 (%)
Mean	93.8	91.4	98.9	82.75	63.93	51.02
10:00 AM	94.3	92.9	95.7	76.38	53.98	39.34
12:00 PM	95.0	93.2	94.4	74.76	54.60	30.49

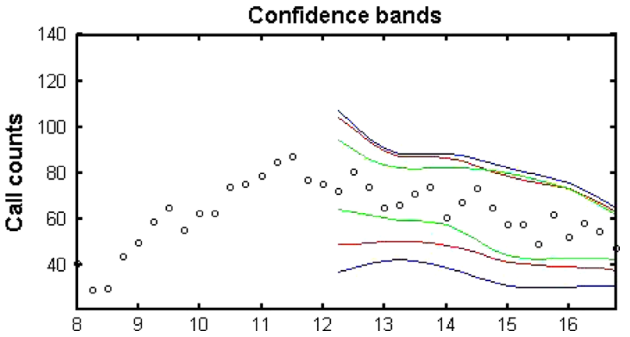

Fig. 4. Confidence bands for Sunday, August 17, 2003. The black dots represent the true call counts in 15-min resolutions. The confidence bands based on the mean of previous days, 10 AM data, and 12 PM data are represented by the blue, red, and green lines, respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

Summarizing, using call center data, we demonstrated that forecasting of curve continuation can be achieved successfully by the proposed BLUP. We also showed that confidence bands for such forecasts can be obtained using cross-validation.

5. Concluding remarks

In this work we forecasted the continuation of both the arrival and workload process using the best linear unbiased predictor (BLUP) for the continuation of a curve. We had shown that the proposed BLUP is a fast and simple alternative to existing methods for predicting the continuation of both the arrival and the workload functions.

As discussed in Feldman et al. (2008) and Reich (in preparation), the workload process is a more appropriate candidate than the arrival process, as a basis for determining staffing levels in call centers. This work, along with Aldor-Noiman et al. (2009) and Reich (in preparation), is the first step in exploring direct forecasting of the workload process, but more remains to be done (see, for example, Whitt, 1999; Zeltyn et al., 2011).

Acknowledgments

The authors are grateful to anonymous associate editor and reviewers for the helpful suggestions and comments. The first author would like to thank the Technion's SEELab team for the hospitality and support during many visits to the lab. The authors thank Michael Reich for helpful discussions and for providing us with the data for the workload example.

Appendix A. Proofs

Before we start the proofs, we need some additional notation. Recall that $Y(t)=X(t)+\varepsilon(t)$ where we assumed that ε is the part of Y that is not contained in the subspace spanned by the functions $\boldsymbol{\phi}$ (see Section 2). Write $\varepsilon(t)=\boldsymbol{b}(t)'B\boldsymbol{\epsilon}$ for an $N\times q$ loading matrix B and a $q\times 1$ random vector $\boldsymbol{\epsilon}$. Let Σ be the covariance matrix of $\boldsymbol{\epsilon}$ and note that G, the covariance matrix of Y, can be written as $ALA'+B\Sigma B'$.

Let A_i and B_i, $i=1,2$, be matrices such that for all $t\in S_i$, $X(t)=\boldsymbol{b_i}(t)'(\boldsymbol{\mu_i}+A_i\boldsymbol{h})$ and $Y(t)=\boldsymbol{b_i}(t)'(\boldsymbol{\mu_i}+A_ih+B_i\boldsymbol{\epsilon})$. A_i and B_i are the loading matrices adapted to the basis $\boldsymbol{b_i}$. Using the above notation we have $g_{ij}=A_iLA_j'$ and $G_{ij}=(A_iLA_j'+B_i\Sigma B_j')$.

We need the following two lemmas.

Lemma 2. *Let T be an $n\times p$ matrix of rank s and let L be a $p\times p$ positive definite diagonal matrix. Then the following assertions are true*

1. $T'T(T'T)^+T'=T'$;
2. $T'LT(T'LT)^+T'=T'$.

The proof is technical and thus omitted.

The following lemma justifies the notation of Γ_{11}^+ as a pseudoinverse operator.

Lemma 3. *With probability one,* $\Gamma_{11}\Gamma_{11}^+(Y_1-\mu)=\Gamma_{11}^+\Gamma_{11}(Y_1-\mu)=Y_1-\mu$.

Proof. Write $Y_1(t)-\mu(t)=\boldsymbol{b_1}(t)'(A_1\boldsymbol{h}+B_1\boldsymbol{\epsilon})$. Hence,

$$(\Gamma_{11}\Gamma_{11}^+(A_1\boldsymbol{h}+B_1\boldsymbol{\epsilon}))(t)=\boldsymbol{b_1}(t)'G_{11}W_1W_1^{-1}G_{11}^+(A_1\boldsymbol{h}+B_1\boldsymbol{\epsilon})=\boldsymbol{b_1}(t)'G_{11}G_{11}^+(A_1\boldsymbol{h}+B_1\boldsymbol{\epsilon})$$
$$=\boldsymbol{b_1}(t)'[A_1,B_1]\begin{bmatrix}L&0\\0&\Sigma\end{bmatrix}\begin{bmatrix}A_1'\\B_1'\end{bmatrix}\left([A_1,B_1]\begin{bmatrix}L&0\\0&\Sigma\end{bmatrix}\begin{bmatrix}A_1'\\B_1'\end{bmatrix}\right)^+\begin{bmatrix}A_1\boldsymbol{h}\\B_1\boldsymbol{\epsilon}\end{bmatrix}$$

and the result follows from Lemma 2. □

We are now ready to prove Theorem 1.

Proof of Theorem 1. We show that (C1)–(C5) hold, one by one.

(C1) holds because $\hat{X}_2$ is indeed a linear transformation of Y_1 as can be seen from (2).

(C2) holds since

$$E[\hat{X}_2(t)]=\boldsymbol{b_2}(t)'(\boldsymbol{\mu_2}+g_{21}G_{11}^+(E[\boldsymbol{y_1}-\boldsymbol{\mu_1}]))=\boldsymbol{b_2}(t)'\boldsymbol{\mu_2}=\mu(t).$$

(C3) states that $\hat{X}_2$ should minimize the mean square error among all the unbiased linear estimators. Let $\tilde{X}_2$ be another linear unbiased estimator. Then we can write $\tilde{X}_2=(\tilde{X}_2-\hat{X}_2)+\hat{X}_2$. Since both $\tilde{X}_2$ and $\hat{X}_2$ are unbiased, $\tilde{X}_2-\hat{X}_2$ is an unbiased linear estimator of zero, hence it is of the form $\boldsymbol{b_2}(t)'D(\boldsymbol{y_1}-\boldsymbol{\mu_1})$ for an $N_2\times N_1$ matrix D. Moreover, it can be shown that $\mathrm{Cov}(X_2-\hat{X}_2,\tilde{X}_2-\hat{X}_2)=0$. Indeed,

$$\mathrm{Cov}((X_2-\hat{X}_2)(s),(\tilde{X}_2-\hat{X}_2)(t))=E[(X_2-\hat{X}_2)(s)(\tilde{X}_2-\hat{X}_2)(t)]=\boldsymbol{b_2}(s)'E[(\boldsymbol{x_2}-\boldsymbol{\mu_2})(\boldsymbol{y_1}-\boldsymbol{\mu_1})']D'\boldsymbol{b_2}(t)$$
$$-\boldsymbol{b_2}(s)'E[\boldsymbol{\mu_2}+g_{21}G_{11}^+(\boldsymbol{y_1}-\boldsymbol{\mu_1})(\boldsymbol{y_1}-\boldsymbol{\mu_1})']D'\boldsymbol{b_2}(t)\boldsymbol{b_2}(s)'(g_{21}D'+g_{21}G_{11}^+G_{11}D'))\boldsymbol{b_2}(t)=0,$$

where the last equality follows from Lemma 3.

To see that $\hat{X}_2$ minimizes the mean square error, note that

$$E[(X_2-\tilde{X}_2)^2(t)]=E[(X_2-\hat{X}_2)^2(t)]+E[(\tilde{X}_2-\hat{X}_2)^2(t)]+2E[(X_2-\hat{X}_2)(t)(\hat{X}_2-\tilde{X}_2)(t)]$$
$$=E[(X_2-\hat{X}_2)^2(t)]+E[(\tilde{X}_2-\hat{X}_2)^2(t)]\geq E[(X_2-\hat{X}_2)^2(t)],$$

which proves that $\hat{X}_2$ minimizes the mean square error and is unique up to equivalence.

(C4) holds by construction.

(C5) states that when no noise is introduced, $\hat{X}_2$ is a smooth continuation of X_1 in the sense that the combined function is in the space $\mathcal{S}$. First, note that by Lemma 3

$$X_1(t)=\boldsymbol{b_1}(t)'(\boldsymbol{\mu_1}+G_{11}G_{11}^+(\boldsymbol{x_1}-\boldsymbol{\mu_1}))=\boldsymbol{b_1}(t)'(\boldsymbol{\mu_1}+A_1(LA_1'G_{11}^+)(\boldsymbol{x_1}-\boldsymbol{\mu_1})).$$

By definition we also have

$$\hat{X}_2(t)=\boldsymbol{b_2}(t)'(\boldsymbol{\mu_2}+g_{21}G_{11}^+(\boldsymbol{x_1}-\boldsymbol{\mu_1}))=\boldsymbol{b_2}(t)'(\boldsymbol{\mu_2}+A_2(LA_1'G_{11}^+)(\boldsymbol{x_1}-\boldsymbol{\mu_1})).$$

Define $\hat{X}(t)=\boldsymbol{b}(t)'(\mu(t)+A(LA_1'G_{11}^+)(\boldsymbol{x_1}-\boldsymbol{\mu_1}))$. It follows from the definitions of $\boldsymbol{\mu_i},A_i$ and $\boldsymbol{b_i}$ that $\hat{X}(t)$ agrees with X_1 on S_1 and with $\hat{X}_2$ on S_2. Since $\hat{X}\in\mathcal{S}$, the result follows.

Finally, if Y is a Gaussian process, then $\boldsymbol{y_1}$ and $\boldsymbol{x_2}$ are normally distributed such that $\mathrm{Var}(\boldsymbol{y_1})=G_{11}$ and $\mathrm{Cov}(\boldsymbol{x_2},\boldsymbol{y_1})=g_{21}$. Following Marsaglia (1964) we obtain

$$E[X_2(t)|Y_1]=\boldsymbol{b}(t)'E[\boldsymbol{x_2}|\boldsymbol{y_1}]=\boldsymbol{b}(t)'(\boldsymbol{\mu_2}+g_{21}G_{11}^+(\boldsymbol{y_1}-\boldsymbol{\mu_1}))=\hat{X}_2(t)=E[\hat{X}_2(t)|Y_1] \tag{A.1}$$

and criterion (C2*) is met. □

Appendix B. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org.10.1016/j.jspi.2013.11.006.

References

Aguilera, A.M., Ocaña, F.A., Valderrama, M.J., 1997. An approximated principal component prediction model for continuous-time stochastic processes. Appl. Stochastic Models Data Anal. 13 (2), 61–72.

Aldor-Noiman, S., Feigin, P.D., Mandelbaum, A., 2009. Workload forecasting for a call center: methodology and a case study. Ann. Appl. Stat. 3 (4), 1403–1447.
de Boor, C., 2001. A Practical Guide to Splines, Applied Mathematical Sciences, revised ed. Springer-Verlag, New York.
Donin, O., Feigin, P.D., Mandelbaum, A., Zeltyn, S., Trofimov, V., Ishay, E., Khudiakov, P., Nadjharov, E., 2006. The Call Center of "US Bank". Available at ⟨http://ie.technion.ac.il/Labs/Serveng/files/The_Call_Center_of_US_Bank.pdf⟩.
Feldman, Z., Mandelbaum, A., Massey, W.A., Whitt, W., 2008. Staffing of time-varying queues to achieve time-stable performance. Management Sci. 54 (2), 324–338.
Gans, N., Koole, G., Mandelbaum, A., 2003. Telephone call centers: tutorial, review, and research prospects. Manufacturing Ser. Oper. Management 5 (2), 79–141.
Gross, J., 2003. Linear Regression. Springer.
Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12 (1), 55–67.
Jolliffe, I.T., 2002. Principal Component Analysis, 2nd ed. Springer.
Marsaglia, G., 1964. Conditional means and covariances of normal variables with singular covariance matrix. J. Amer. Statist. Assoc. 59 (308), 1203–1204.
Ramsay, J., Silverman, B.W., 2002. Applied Functional Data Analysis: Methods and Case Studies. Springer Series in Statistics, 2nd ed. Springer-Verlag, New York.
Ramsay, J., Silverman, B.W., 2005. Functional Data Analysis. Springer Series in Statistics. Springer-Verlag, New York.
Reich, M., 2012. The Workload Process: Modelling, Inference and Applications M.Sc. thesis. Technion. ⟨http://ie.technion.ac.il/serveng/References/Michael_Reich_Thesis_withlinks.pdf⟩.
Robinson, G.K., 1991. That BLUP is a good thing: the estimation of random effects. Statist. Sci. 6 (1), 15–32.
Rozenshmidt, L., 2008. On Priority Queues with Impatient Customers: Stationary and Time-Varying Analysis (Master's thesis). Technion—Israel Institute of Technology. Available at ⟨http://iew3.technion.ac.il/serveng/References/thesis_Luba_Eng.pdf⟩.
Sansone, G., 1991. Orthogonal Functions, revised ed. Dover Publications, New York.
Shen, H., 2009. On modeling and forecasting time series of smooth curves. Technometrics 51 (3), 227–238.
Shen, H., Huang, J.Z., 2008. Interday forecasting and intraday updating of call center arrivals. Manufacturing Ser. Oper. Management 10 (3), 391–410.
Weinberg, J., Brown, L.D., Stroud, J.R., 2007. Bayesian forecasting of an inhomogeneous Poisson process with applications to call center data. J. Amer. Statist. Assoc. 102.
Whitt, W., 1999. Dynamic staffing in a telephone call center aiming to immediately answer all calls. Oper. Res. Lett. 24 (5), 205–212.
Yao, F., Müller, H.G., Wang, J.L., 2005. Functional data analysis for sparse longitudinal data. J. Amer. Statist. Assoc. 100, 577–590.
Zeltyn, S., 2005. Call Centers with Impatient Customers: Exact Analysis and Many-Server Asymptotics of the M/M/n+G Queue (Ph.D. thesis). Technion—Israel Institute of Technology. Available at ⟨http://ie.technion.ac.il/serveng/References/references.html⟩.
Zeltyn, S., Marmor, Y.N., Mandelbaum, A., Carmeli, B., Greenshpan, O., Mesika, Y., Wasserkrug, S., Vortman, P., Shtub, A., Lauterman, T., Schwartz, D., Moskovitch, K., Tzafrir, S., Basis, F., 2011. Simulation-based models of emergency departments: operational tactical and strategic staffing. ACM Trans. Model. Comput. Simul. 21 (4).