

# Modeling Conditional Distributions For Functional Responses, With Application to Traffic Monitoring via GPS-enabled Mobile Phones\*

Kehui Chen

Department of Statistics, University of Pittsburgh,

and

Hans-Georg Müller

Department of Statistics, University of California, Davis

May 18, 2013

## Abstract

Motivated by problems involving a traffic monitoring system in which trajectory data are obtained from GPS-enabled mobile phones, we propose a novel approach to functional regression modeling, where instead of the usual mean regression the entire distribution of functional responses is modeled conditionally on predictors. An approach that sensibly balances flexibility and stability is obtained by assuming that the response functions are drawn from a Gaussian process, the mean and covariance function of which depend on predictors. The dependence of the mean function and covariance function of the response on the predictors is modeled additively. We demonstrate the proposed methods by constructing predicted curves and corresponding prediction regions for traffic velocity trajectories for a future time period, using current traffic velocity fields as predictor functions. The proposed functional regression and conditional distribution approach is of general interest for functional response settings, where in addition to predicting the conditional mean response function one is also interested in predicting the covariance surface of the random response functions, conditional on predictor curves.

*Keywords:* Additive Model, Conditional Covariance, Conditional Distribution in Function Space, Functional Data Analysis, Functional Regression, Prediction Region, Velocity Curve

technometrics tex template (do not remove)

---

\*This research was supported by NSF grants DMS-1104426 and DMS-1228369

# 1 Introduction

Increasingly, data requiring statistical analysis include processes that are continuously monitored over a continuum, such as time or spatial locations. While traditional univariate or multivariate analysis strategies can be useful to study such data, it is often more natural and efficient to view continuously observed processes as sample units. One can then apply specific techniques for data with functional components, such as functional data analysis (FDA) approaches (Ramsay and Silverman, 2005). Motivated by the need for data analysis for a newly developed traffic monitoring system, where a significant amount of trajectory data based on either spatial or temporal sampling strategies is collected, we propose here a novel way of applying FDA for the construction of conditional distributions when responses and predictors are functions. Applications include visualization, prediction and inference for traffic flows.

The motivating data are from the ‘Mobile Century’ experiment, where 165 UC Berkeley graduate students were hired to drive on a segment of the freeway I880 in the East Bay Area/Northern California, to demonstrate the feasibility of real-time highway traffic monitoring using GPS enabled mobile phones (Herrera et al., 2010). The follow-up project ‘Mobile Millennium’ was developed by the California Center for Innovative Transportation (CCIT), the Nokia Research Center (NRC), and the University of California (UC) at Berkeley (<http://traffic.berkeley.edu>). This pilot traffic monitoring system uses GPS enabled mobile phones to gather traffic information, aiming at real time estimation of traffic velocity on the monitored highway segments.

The individual trip data used in this paper focus on the highway segment between Decoto Rd. to the south (Postmile 21) and Winton Ave. to the north (Postmile 27.5), in the Northbound direction. For the  $i$ th trip,  $i = 1, \dots, n$ , the on-board mobile phone recorded GPS-measured locations  $s_{ij}$ , recorded in postmiles, and speeds  $V_{ij}$ , recorded in miles per hour, at a grid of times  $t_{ij}$ , where  $t_{ij}$  refers to clock time, with  $j = 1, \dots, N_i$ .

Combining all records and relabeling the data, they can be written in the form

$$(\{t_l, s_l\}, V_l)_{l=1, \dots, N}, \quad (1)$$

where  $N = \sum_i N_i$ . One can apply a two-dimensional smoothing procedure for these combined data to recover a smooth random velocity field  $V(t, s)$  along the highway as an ex-

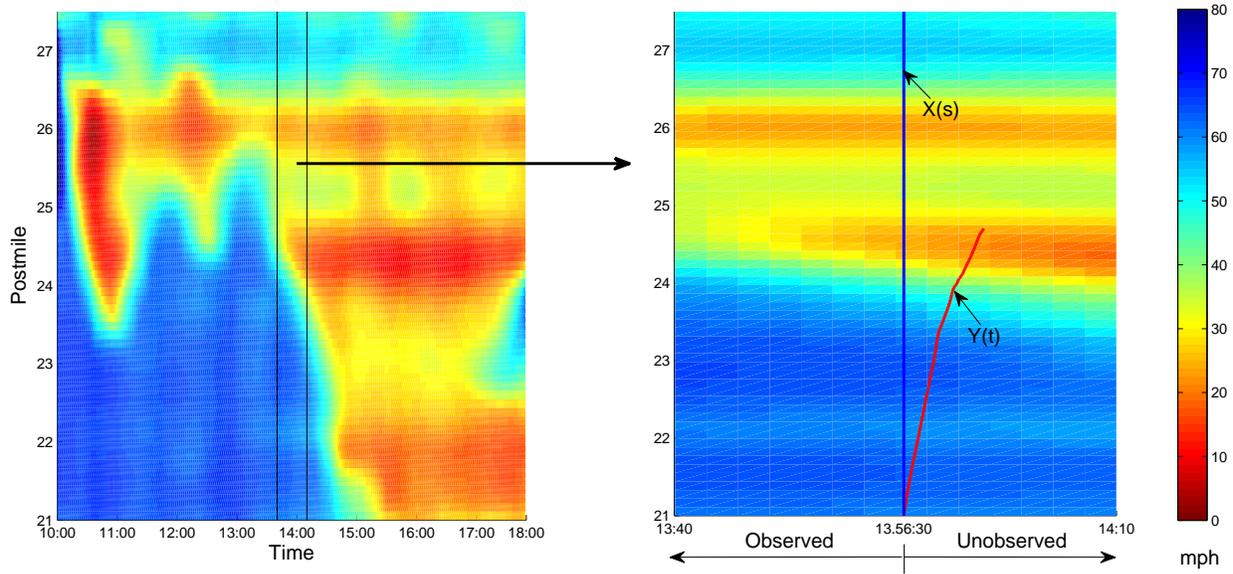


Figure 1: Left: Velocity field estimated from vehicle trajectories for postmiles 21 to 27.5 (on the vertical axis) between 10am and 6pm (horizontal axis). Color indicates velocity on a scale from 0 to 80 mph. Right: An enlarged section of the velocity field for a 30 minutes time interval to illustrate current speeds  $X(s)$  and actual speeds  $Y(t)$  when starting driving at time  $T$  at postmile 21. The latter is to be predicted from the former.

ploratory step. The velocity field reflects the speed of the traffic flow as a function of time  $t$  and postmile  $s$ ; see the left panel in Figure 1.

By continuous time monitoring of highway traffic conditions, for any current time  $T$  the local current speed  $X(s)$  on a section  $s \in [21, 27.5]$  along the highway can be recovered. It is then of interest to predict the actual speeds encountered by a vehicle that departs at time  $T$  from location  $s = 21$  and moves through the highway segment. The speed  $Y(t)$  over a future time period  $t \in [T, T + \Delta]$  for a given  $\Delta > 0$  involves future, unobserved velocity fields relative to current time  $T$ ; an illustration is given in the right panel of Figure 1, where  $T = 13:56:30$ . In this paper, we propose methods to construct a region in function space such that  $Y(t)$  will fall inside the region with high probability. We will base such functional prediction regions on the conditional distribution of response processes  $Y(t)$ , given the functional predictors

$X(s)$ .

Functional regression for models with functional responses so far has been discussed by various authors for mean regression relations, for which linear (Ramsay and Dalzell, 1991; Faraway, 1997; Shen and Faraway, 2004; Chiou et al., 2004) and additive (Müller and Yao, 2008) or nonparametric (Müller and Yao, 2006) approaches have been proposed, while the more general case of conditional distributions at given predictor levels has not yet been studied, to the best of our knowledge. For the simpler case of scalar responses paired with functional predictors, several approaches for quantile regression have been investigated (Cardot et al., 2005; Chen and Müller, 2011) and were shown to work well. The second of these approaches is based on inverting conditional distributions in order to obtain quantile functions for the scalar case, providing motivation to construct prediction regions via conditional distributions. Once conditional quantiles have been constructed, conditional prediction regions with coverage  $(1 - \alpha)$  are simply obtained as inter-quantile intervals defined by the  $\alpha/2$  and  $(1 - \alpha/2)$  quantiles.

Here we study a situation that to the best of our knowledge has remained unexplored: How to model and obtain conditional distributions and a functional equivalent of conditional prediction regions when responses are curves. As functional responses are increasingly encountered in studies across the sciences (Müller et al., 2008), it is natural to move beyond mean regression analysis and to target the conditional distributions of the responses in dependency on the predictor levels. Prediction regions may then be derived from conditional distributions. The situation involving functional responses is substantially more complex than the case of scalar responses, as the distribution of the functional responses is that of a stochastic process, and therefore one needs to model conditional distributions for stochastic processes in a way that is practically feasible. To make this task manageable, we make the assumption that, potentially after suitable transformations, the response processes for given predictors are Gaussian, with mean and covariance functions that depend on the predictors. We propose additive approaches for these tasks, including a novel additive conditional covariance model.

Traffic monitoring and forecasting is a challenging problem that has received much attention. Prior to the emergence of GPS enabled smart-phones, traffic monitoring data mainly came from dedicated equipment, such as loop detectors, cameras, and radars. These data of-

ten do not include direct speed measurements, so that speed needs to be calculated indirectly from other measurements. Many efforts went into developing models to estimate velocity fields and traffic flow (Wang and Nihan, 2000; Wang and Papageorgiou, 2005; Mihaylova et al., 2007), and also short time travel time prediction, see Nanthawichit et al. (2003), Rice (2004), Chu et al. (2005) and Bickel and Li (2007), among many others.

Utilizing GPS enabled smart phones, which potentially generate high quality position and speed data, provides new opportunities for highway traffic monitoring and forecasting (Herrera and Bayen, 2010a). We refer to Work et al. (2008), Wang et al. (2008), Herrera and Bayen (2010b) and Tao et al. (2012) for the estimation of velocity fields or traffic states from observed GPS data. Much less is known about short time traffic prediction from GPS data, and the existing approaches focus either on traffic condition prediction, such as congestion prediction (Herring et al., 2010) or on average traffic time (or average speed) prediction (Izadpanah et al., 2011; Mazare et al., 2012).

To our knowledge, the functional data perspective that we adopt in this paper is novel, specifically, viewing GPS recorded vehicle speeds that are sampled along the highway as discrete and noisy measurements, assumed to be generated by a smooth underlying random process. In an initial step, the velocity field is estimated by two dimensional smoothing. Current time velocity curves are then used as functional predictors to predict entire future velocity curves, rather than average travel times. The predicted velocity curves impact vehicles as they move along the highway. An important tool to recover the underlying current time velocity curves is functional principal component analysis.

The article is organized as follows. In Section 2, we briefly review the functional principal component analysis for predictor and response functions. In Section 3, we outline the proposed methods for modeling the conditional distribution of the response functions, describe estimating techniques for the model components and construct prediction regions for the response  $Y$  for Gaussian and non-Gaussian cases. Section 4 describes the application of these techniques to the traffic data. Section 5 contains a small simulation to demonstrate the coverage of the prediction regions under different settings. Additional discussion can be found in Section 6.

## 2 Review of Functional Principal Component Analysis

We begin with a brief outline of the population model for functional principal component analysis and then describe the practical implementation version. Functional predictors  $X(s)$  and responses  $Y(t)$  are assumed to be square integrable, with mean functions  $\mu_X(s) = EX(s)$  and  $\mu_Y(t) = EY(t)$  and covariance functions  $G_{XX}(s_1, s_2) = \text{cov}(X(s_1), X(s_2))$  and  $G_{YY}(t_1, t_2) = \text{cov}(Y(t_1), Y(t_2))$ . Then one has orthogonal expansions

$$G_{XX}(s_1, s_2) = \sum_k \lambda_k \phi_k(s_1) \phi_k(s_2), \quad G_{YY}(t_1, t_2) = \sum_j \gamma_j \psi_j(t_1) \psi_j(t_2) \quad (2)$$

with nonincreasing eigenvalues  $\lambda_k, \gamma_j$  and orthonormal eigenfunctions  $\phi_k, \psi_j$  of the linear Hilbert-Schmidt operators that are, respectively, defined by kernels  $G_{XX}$  and  $G_{YY}$ .

This then leads to the well-known Karhunen-Loève expansions for processes  $X(s)$  and  $Y(t)$ ,

$$X_i(s) = \mu_X(s) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(s), \quad Y_i(t) = \mu_Y(t) + \sum_{j=1}^{\infty} \zeta_{ij} \psi_j(t), \quad (3)$$

where  $\phi_k(s)$ ,  $k \geq 1$  and  $\psi_j(t)$ ,  $j \geq 1$  form an orthonormal basis in the respective function spaces and  $\xi_k$  and  $\zeta_j$  are sequences of functional principal components. These are respectively uncorrelated mean zero random variables with  $\text{var}(\xi_k) = \lambda_k$  and  $\text{var}(\zeta_j) = \gamma_j$ . The quantities  $\mu_X(s)$ ,  $\xi_{ik}$ ,  $\phi_k(s)$ , and  $\lambda_k$ , as well as the  $\mu_Y(t)$ ,  $\zeta_{ij}$ ,  $\psi_j(t)$  and  $\gamma_j$  can all be estimated from the data and we describe the details for  $X(s)$  in the following; the procedure for  $Y(t)$  is the same. We also remark that in some application areas, the practical implementation of the Karhunen-Loève expansion is known as the method of empirical orthogonal functions.

In practice, underlying smooth predictor and response trajectories  $X_i(s)$  and  $Y_i(t)$ ,  $i = 1, \dots, n$ , are usually recorded at a grid of points, where  $s$  and  $t$  can be different sampling units. The observations of the underlying random functions taken at these time points might be contaminated with additive measurement errors; this is the case in our traffic data example, due to aberrations of speed measurements for individual cars from smooth trajectories, which can be caused by small GPS location inaccuracies or small scale accelerations and decelerations by individual drivers.

To model functional data that are discretely measured and contaminated with measurement errors, denoting the observations of random trajectories  $X_i$  at times  $s_{ij}$  by  $V_{ij}$ , the

measurement errors by  $\varepsilon_{ij}$  and the number of observations made for the  $i$ th subject by  $N_i$ , a reasonable representation of the observed data is

$$V_{ij} = X_i(s_{ij}) + \varepsilon_{ij} \quad 1 \leq i \leq n, 1 \leq j \leq N_i, \quad (4)$$

where the errors  $\varepsilon_{ij}$  are assumed to be i.i.d. with  $E\varepsilon_{ij} = 0$  and  $E\varepsilon_{ij}^2 = \sigma^2$ . An analogous data model is also assumed for the observations of response processes  $Y_i$ . Response processes  $Y_i$  represent speed curves for individual cars and are available on a denser grid of observations and recorded with less error in comparison to the predictor processes.

As mean, covariance and eigenfunctions are assumed to be smooth, under the assumption that the time grid is reasonably dense for the pooled data, we may apply smoothing methods such as local linear smoothing to scatterplots  $\{(s_{ij}, V_{ij}), i = 1, \dots, n, j = 1, \dots, N_i\}$  to obtain the estimated mean functions  $\hat{\mu}_X(s)$ , and a two-dimensional weighted least squares smoother to scatterplots  $\{(s_{ij}, s_{il}), (V_{ij} - \hat{\mu}_X(s_{ij}))(V_{il} - \hat{\mu}_X(s_{il}))\}$ , omitting the diagonal elements, to obtain the estimated covariance surfaces  $\hat{G}_{XX}$  (for further details, we refer to Yao et al., 2005). Smoothing bandwidths may be chosen by cross-validation or generalized cross-validation, the latter being computationally faster.

Estimates of eigenfunctions  $\phi_k$  and of eigenvalues  $\lambda_k$  are then obtained as solutions of the eigen-equations,

$$\int G_{XX}(s_1, s_2)\phi_k(s_1)ds_2 = \lambda_k\phi_k(s_1),$$

where covariance function estimates  $\hat{G}_{XX}$  are substituted for  $G_{XX}$  and the eigenfunctions are subject to constraints  $\int \phi_k(s)^2 ds = 1$  and  $\int \phi_k(s)\phi_j(s)ds = 0$ , for  $k < j$ . The eigen-equations are invariably solved by discretization on a grid (for which we use a  $51 \times 51$  grid) and subsequent matrix spectral analysis. The functional principal components then are estimated based on the definition  $\xi_{ik} = \int (X_i(s) - \mu_X(s))\phi_k(s)dt$  and numerical integration. Alternatively, one can adopt a conditional expectation procedure that is described in Yao et al. (2005); one proceeds analogously for the eigencomponents  $\gamma_j, \psi_j$  and the corresponding functional principal components for response processes  $Y_i(t)$ . These procedures are implemented in the PACE package, accessible at <http://anson.ucdavis.edu/~mueller/data/pace.html>.

Once the first  $K$  functional principal components  $\xi_{ik}$ ,  $k = 1, \dots, K$ , for each predictor function  $X_i(s)$ , and the first  $P$  functional principal components  $\zeta_{ij}$ , eigenvalues  $\gamma_j$ , and eigenfunctions  $\psi_j(t)$ ,  $j = 1, \dots, P$ , for the response functions  $Y_i(t)$  have been obtained, this

leads to estimates

$$\begin{aligned}\hat{X}_i(s) &= \hat{\mu}_X(s) + \sum_{k=1}^K \hat{\xi}_{ik} \hat{\phi}_k(s) \\ \hat{Y}_i(t) &= \hat{\mu}_Y(t) + \sum_{j=1}^P \hat{\zeta}_{ij} \hat{\psi}_j(t)\end{aligned}\tag{5}$$

of predictor trajectories  $X_i(s)$ , resp., response trajectories  $Y_i(t)$ .

We note that applying this functional principal component approach means that no pre-smoothing step is needed to recover the trajectories of  $X(s)$  or  $Y(t)$ . The numbers  $K$  resp.  $P$  of included components for predictor resp. response expansions can be determined by controlling the fraction of variance explained (FVE), obtained for example for responses  $Y(t)$  by  $\sum_{j=1}^P \hat{\gamma}_j / \sum_{j=1}^M \hat{\gamma}_j$ , where  $M$  is chosen large, for modest sample sizes as for the traffic data we choose  $M = 20$ . Alternative methods for choosing  $K$  and  $P$  such as AIC and BIC type criteria have been discussed in Yao et al. (2005).

### 3 Modeling of Conditional Distributions for Response Functions

The study of conditional distributions of functional responses  $Y(t)$  given predictors  $X(s)$  poses specific difficulties, due to the complex nature of the response, which is a realization of a stochastic process. Therefore, suitable structural assumptions are needed that are manageable but do not compromise the desired flexibility. While we do not assume that  $(X, Y)$  are jointly Gaussian, a major simplification of the problem can be achieved by assuming that the conditional responses are Gaussian. This Gaussianity assumption, which we adopt from now on, reduces the problem of modeling conditional distributions to two components: The conditional mean function  $E(Y(t) | X)$  and the conditional covariance surface  $\text{cov}(Y(t_1), Y(t_2) | X)$ , as these jointly determine the distribution of the conditional Gaussian processes.

Even with this simplification, the modeling task remains complex and requires judicious choice of structural assumptions, due to the additional complexity that is due to the functional nature of the predictors. In the simpler case where functional responses are paired

with scalar predictors, nonparametric approaches have been proposed for conditional mean and covariance modeling (Cardot, 2007; Chiou and Müller, 2009; Jiang and Wang, 2010), but these methods do not extend to the case of high-dimensional or functional predictors. Fully nonparametric approaches are not promising when predictors are functional, due to the curse of dimensionality that results from very low values of the small ball probabilities in the function space  $L^2$  (Hall et al., 2009; Müller and Yao, 2010). Therefore, we propose here an additive modeling approach, which provides a good balance between the somewhat conflicting needs for flexibility and stability.

### 3.1 Conditional Mean Modeling

To model the conditional mean function, we apply the functional additive model as described in Müller and Yao (2008). The population version of this model is

$$E(Y(t) | X) = \mu_Y(t) + \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} f_{jk}(\xi_k) \psi_j(t), \quad (6)$$

where  $\xi_k$  are the functional principal components of  $X(s)$  and  $\psi_j(t)$  are eigenfunctions of  $Y(t)$ , as described above. For identifiability, one requires

$$E f_{jk}(\xi_k) = 0, \quad k = 1, 2, \dots, j = 1, 2, \dots, \quad (7)$$

and in applications and implementations the sums in (6) are finitely truncated at suitable truncation levels.

While complex iterative procedures are required to fit a regular additive model (Mammen and Park, 2005), a straightforward estimation scheme is available to recover the component functions  $f_{jk}$  by a series of one-dimensional smoothing steps, under the assumption that the predictor functional principal components  $\xi_k$ ,  $k \geq 1$ , are independent (Müller and Yao, 2008). Since the functional principal components are always uncorrelated, this assumption is for example satisfied for the case where predictor processes are Gaussian. Then the basic functional additive model assumptions imply, for a suitably truncated version of model (6),

$$E(Y(t) | X) = \mu_Y(t) + \sum_{j=1}^P \sum_{k=1}^K E(\zeta_j | \xi_k) \psi_j(t), \quad (8)$$

i.e.,  $f_{kj}(\xi_k) = E(\zeta_j | \xi_k)$ , which leads to a simple implementation of the functional additive model. We adopt local linear fitting in this paper; other smoothing techniques could also be used.

Fitting a local linear regression to the estimated functional principal components of  $X$  resp.  $Y$ , given by  $\{(\hat{\xi}_{ik}, \hat{\zeta}_{ij}), i = 1, \dots, n\}$ , i.e., minimizing

$$\sum_{i=1}^n \mathcal{K}\left(\frac{\hat{\xi}_{ik} - x}{h_{jk}}\right) (\hat{\zeta}_{ij} - \beta_0 - \beta_1(x - \hat{\xi}_{ik}))^2$$

with respect to  $\beta_0$  and  $\beta_1$ , leads to  $\hat{f}_{jk}(x) = \hat{\beta}_0(x)$ , where  $h_{jk}$  is the bandwidth and  $\mathcal{K}$  is a kernel function. Then the fitted conditional mean function is

$$\hat{E}(Y(t) | X) = \hat{\mu}_Y(t) + \sum_{j=1}^P \sum_{k=1}^K \hat{f}_{jk}(\hat{\xi}_k) \hat{\psi}_j(t), \quad t \in \mathcal{T}, \quad (9)$$

where  $\hat{\xi}_k, k = 1, \dots, K$  are the first  $K$  estimated functional principal components of  $X$ .

### 3.2 Conditional Covariance Modeling

At the core of our approach is the modeling of conditional covariance functions for the responses, i.e., of  $\text{cov}(Y(t_1), Y(t_2) | X)$ , where we omit the argument  $s$  for the predictor function  $X$ . For this modeling task, we make the simplifying assumption that the influence of  $X$  on the conditional covariance function of  $Y$  consists in a rescaling of the conditional variance components, and the uncorrelatedness of response principal components  $\zeta_j, \zeta_k$  for  $j \neq k$  is preserved when conditioning on  $X$ , i.e.,  $\text{cov}(\zeta_j, \zeta_k | X) = 0$ , for  $j \neq k$ .

This leads to the conditional covariance function

$$\text{cov}(Y(t_1), Y(t_2) | X) = \sum_{j=1}^{\infty} \gamma_j(X) \psi_j(t_1) \psi_j(t_2), \quad (10)$$

where  $\gamma_j(X) = \text{var}(\zeta_j | X) = E(\zeta_j^2 | X) - (E(\zeta_j | X))^2$ . This model naturally connects to the common principal components approach (Flury, 1988; Benko et al., 2009), where the eigenfunctions (eigenvectors) of the conditional processes  $Y_X(t)$  are assumed to be the same across different values of predictor functions  $X$ . Again, here and in the following (10) refers to a population model and the sums are truncated at a finite number of components for practical implementations.

Noting that a completely nonparametric model for  $\gamma_j(X)$  is not feasible, due to the curse of dimensionality, as functional data are inherently infinite-dimensional, the functional additive approach is found to be a sensible compromise between flexibility and stability and is particularly useful for the proposed conditional covariance modeling. Observe that

$$E(\gamma_j(X)) = E(\text{var}(\zeta_j | X)) = \text{var}(\zeta_j) - \text{var}(E(\zeta_j | X)) = \gamma_j - E\left(\sum_{k=1}^{\infty} f_{jk}^2(\xi_k)\right),$$

where  $E(\zeta_j | X) = \sum_{k=1}^{\infty} f_{jk}(\xi_k)$  is implied by the assumed additive model for the conditional mean function, and  $f_{jk}(\xi_k) = E(\zeta_j | \xi_k)$  are the same smooth functions as in eq. (6).

Therefore, the additive modeling for  $\gamma_j(X)$  is

$$\gamma_j(X) = \gamma_j + \sum_{k=1}^{\infty} \{g_{jk}(\xi_k) - f_{jk}^2(\xi_k)\}, \quad (11)$$

where  $g_{jk}$  are smooth functions satisfying  $E(g_{jk}(\xi_k)) = 0$ .

As predictor functional principal components  $\xi_k$  are independent, one finds that

$$g_{jk}(\xi_k) = E(\zeta_j^2 - \gamma_j | \xi_k).$$

By plugging  $\gamma_j(X)$  into eq. (10) we arrive at the following form of the conditional covariance function,

$$\text{cov}(Y(t_1), Y(t_2) | X) = G_{YY}(t_1, t_2) + \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \{g_{jk}(\xi_k) - f_{jk}^2(\xi_k)\} \psi_j(t_1) \psi_j(t_2), \quad (12)$$

where  $G_{YY}(t_1, t_2)$  is the continuous covariance function of  $Y(t)$  with eigenfunctions  $\psi_j(t)$ ,  $j \geq 1$ , and  $\xi_k$ ,  $k \geq 1$  are the functional principal components of the predictor processes  $X(s)$ .

Conditional response processes, at predictor level  $X$ , then are the Gaussian processes that are characterized by the mean function  $E(Y(t) | X)$  given by eq. (6) and the covariance function  $\text{cov}(Y(t_1), Y(t_2) | X)$  given by eq. (12). The characterizing components of these processes thus depend on eigenfunctions and eigenvalues of response processes  $Y(T)$  and the smooth regression functions  $f_{jk}(\cdot)$  and  $g_{jk}(\cdot)$  in eq. (11). The assumptions that are central for the proposed modeling approach are explicitly listed and discussed in Section 5.

Regarding the estimation of  $g_{jk}(\cdot)$ , we fit a local linear regression to the data  $\{(\hat{\xi}_{ik}, \hat{\zeta}_{ij}^2 - \hat{\gamma}_j), i = 1, \dots, n\}$ , by minimizing

$$\sum_{i=1}^n \mathcal{K}\left(\frac{\hat{\xi}_{ik} - x}{h_{jk}}\right) (\hat{\zeta}_{ij}^2 - \hat{\gamma}_j - \beta_0 - \beta_1(x - \hat{\xi}_{ik}))^2$$

with respect to  $\beta_0$  and  $\beta_1$ , leading to  $\hat{g}_{jk}(x) = \hat{\beta}_0(x)$ , where  $h_{jk}$  is the bandwidth. Invoking suitable truncation points, the fitted conditional covariance is

$$\begin{aligned} \widehat{\text{cov}}(Y(t_1), Y(t_2) | X) &= \hat{G}_{YY}(t_1, t_2) + \sum_{j=1}^P \sum_{k=1}^K \{\hat{g}_{jk}(\hat{\xi}_k) - \hat{f}_{jk}^2(\hat{\xi}_k)\} \hat{\psi}_j(t_1) \hat{\psi}_j(t_2) \\ &= \sum_{j=1}^P (\hat{\gamma}_j + \sum_{k=1}^K \{\hat{g}_{jk}(\hat{\xi}_k) - \hat{f}_{jk}^2(\hat{\xi}_k)\}) \hat{\psi}_j(t_1) \hat{\psi}_j(t_2), \end{aligned} \quad (13)$$

where  $\hat{G}_{YY}(t_1, t_2)$  is obtained from the functional principal component analysis step for processes  $Y(t)$ , and  $\hat{\xi}_k$  are the estimated functional principal components for predictor functions  $X(s)$ . We project these nonparametric surface estimates on the space of symmetric positive definite covariance surfaces (see Hall et al., 2008).

Asymptotic results for the conditional mean and conditional covariance estimates require studying the convergence properties of the estimated additive components  $\hat{f}_{jk}$  and  $\hat{g}_{jk}$ , which are based on the estimated functional principal components  $\hat{\xi}_{ik}$  and  $\hat{\zeta}_{ij}$ ,  $k = 1, \dots, K$ ,  $j = 1, \dots, P$ . Established convergence results for the estimated population components such as mean function, eigenfunction and eigenvalue estimates can be found in Yao et al. (2005) and Hall and Hosseini-Nasab (2006). Based on such results, Müller and Yao (2008) provided upper bounds for  $|\hat{\xi}_{ik} - \xi_{ik}|$  and  $|\hat{\zeta}_{ij} - \zeta_{ij}|$  under suitable regularity conditions, and analogous arguments as those presented in the proof of Theorem 1 in that paper can be used to show the convergence of  $\hat{f}_{jk}$  and  $\hat{g}_{jk}$  to their target values in probability as  $n \rightarrow \infty$ .

### 3.3 Prediction Regions For Response Functions

We consider a representation of conditional response processes truncated at  $P$  functional principal components by the expansion

$$\tilde{Y}_X(t) = \mu_{Y|X}(t) + \sum_{j=1}^P \zeta_j(X) \psi_j(t). \quad (14)$$

Due to the Gaussian assumptions for the conditional processes,  $\zeta_j(X)$ ,  $j = 1, \dots, P$ , are conditionally independent random components, with distributions  $\mathcal{N}(0, \gamma_j(X))$ . A natural ellipsoid type prediction region with  $1 - \alpha$  coverage is then obtained for  $\boldsymbol{\zeta}_X = (\zeta_1(X), \dots, \zeta_P(X))$  in a straightforward manner,

$$\Omega_{X,\alpha} = \{(\zeta_1(X), \dots, \zeta_P(X)) : \sum_{j=1}^P \frac{\zeta_j(X)^2}{\gamma_j(X)} \leq \mathcal{C}_{X,\alpha}^2\}, \quad (15)$$

where in the case that  $(\zeta_1(X), \dots, \zeta_P(X))$  are jointly Gaussian,  $\mathcal{C}_{X,\alpha} = \mathcal{C}_\alpha = \sqrt{\chi_{P,1-\alpha}^2}$ .

After constructing a prediction region for functional principal components  $\zeta_X$ , one needs to project the region onto the time domain, in order to obtain equivalent prediction regions for  $\tilde{Y}_X$  in the function space  $L^2$ . Eq. (14) motivates the construction of conditional prediction regions of the form  $\{(L(t), U(t)), t \in \mathcal{T}\}$ . The upper bound function  $U(t)$  is found by solving the maximization problems

$$\max_{\zeta_X \in \Omega_{X,\alpha}} \left\{ \mu_{Y|X}(t) + \sum_{j=1}^P \zeta_j(X) \psi_j(t) \right\}, \quad \text{for all } 0 < t < 1. \quad (16)$$

This is a quadratic constrained linear problem, which can be easily solved through the Lagrange dual problem. Utilizing estimates for  $\mu_{Y|X}(t)$  from eq. (9) and for  $\text{var}(\tilde{Y}_X(t))$  from eq. (13), one obtains

$$\hat{U}(t) = \hat{\mu}_{Y|X}(t) + \left\{ \mathcal{C}_\alpha^2 \sum_{j=1}^P \hat{\gamma}_j(X) \hat{\psi}_j^2(t) \right\}^{1/2} = \hat{\mu}_{Y|X}(t) + \mathcal{C}_\alpha \left\{ \hat{\text{var}}(\tilde{Y}_X(t)) \right\}^{1/2}. \quad (17)$$

Analogously,  $\hat{L}(t)$  is seen to be

$$\hat{L}(t) = \hat{\mu}_{Y|X}(t) - \mathcal{C}_\alpha \left\{ \hat{\text{var}}(\tilde{Y}_X(t)) \right\}^{1/2}. \quad (18)$$

These two bands then yield global prediction regions that are given by

$$P(\{\hat{L}(t) \leq \tilde{Y}_X(t) \leq \hat{U}(t)\}, t \in \mathcal{T}) \geq 1 - \alpha.$$

Here, Eq. (17) and (18) indicate that the sizes of prediction regions are determined by the estimated conditional mean and conditional covariance, up to a scale constant. These estimates can be shown to be consistent for the  $P$ -truncated version of the conditional processes, and the uncertainties induced by the measurement errors do not need to be additionally considered. The nominal coverage is attained asymptotically, as sample size increases.

While under the Gaussian assumption for the conditional response processes, theory indicates that  $\mathcal{C}_\alpha = \sqrt{\chi_{P,1-\alpha}^2}$ , in applications, the Gaussian assumption often is difficult to verify, in which case the joint distribution of the  $\zeta_j(X)$ ,  $j = 1, \dots, P$ , is unknown, and we may not know whether the conditional functional principal components are uncorrelated. Motivated by the Gaussian case, we still consider prediction regions for the vector of  $P$

functional principal components which are of ellipsoidal shape  $\Omega_{X,\alpha}$  as defined in Eq. (15), aiming at  $P(\zeta_X \in \Omega_{X,\alpha}) = 1 - \alpha$ , for each  $X$ .

A problem in this more general setting is the determination of an appropriate value for the constant  $\mathcal{C}_{X,\alpha}$ , which is hard to determine in the general case because it might vary with predictors  $X$ , in contrast to the Gaussian case. To address this problem, for the case where  $X$  might be assumed to be (intrinsically) very low dimensional, one could consider a local approximation of  $\mathcal{C}_{X,\alpha}$  by aggregating information from similar predictor levels  $X$ . However, in general scenarios where the  $X$  are infinite-dimensional predictor functions, such local constructions of  $\mathcal{C}_{X,\alpha}$  often will not be appropriate, due to the curse of dimensionality.

Instead, we opt to aim at the correct average coverage for the prediction regions, i.e., seek to find a constant  $\mathcal{C}_\alpha$  and regions  $\Omega_{X,\alpha}$ , so that

$$E\{P(\zeta_X \in \Omega_{X,\alpha} \mid X)\} = 1 - \alpha.$$

In practice, one can determine the constant  $\mathcal{C}_\alpha$  by aiming to keep the empirical leave-one-out average coverage at the desired level. We adopt this strategy in the simulations and the application to traffic analysis. After determining the ellipsoid for  $\zeta_X$ , we use the same projection method as described above for the Gaussian case to obtain  $(1 - \alpha)$  prediction regions  $(L(t), U(t))$  for the conditional processes.

## 4 Application to Traffic Data Obtained via GPS-enabled Mobile Phones

The Mobile Century data were collected on February 8, 2008 between 10:00am and 18:00pm (PST) on Interstate 880 in the East Bay Area in Northern California as part of a joint UC Berkeley - Nokia project, aimed at studying the use of GPS enabled mobile phones for traffic monitoring. An extensive description of the experiment and data is available in Herrera et al. (2010). As described in the introduction, the records can be written in the form

$$(\{t_l, s_l\}, V_l)_{l=1,\dots,N}, \tag{19}$$

where  $N = \sum_i N_i$ . One can use these combined data to recover a smooth random velocity field  $V(t, s)$  along the highway by two-dimensional local linear smoothing (Fan and Gijbels,

1996; Yao et al., 2005) on the scatterplots  $(\{t_l, s_l\}, V_l)_{l=1, \dots, N}$ . This leads to the velocity field estimate in Figure 1, providing a concise visualization of the traffic speed conditions in the given time-space domain and demonstrating the usefulness of nonparametric smoothing methods for traffic modeling.

For random times  $T$  within the time domain of interest, define processes  $X(s) = V(T, s)$  to be the velocity of traffic flow at time  $T$  as a function of location on the relevant highway segment  $s \in [21, 27.5]$ , as prescribed by the random field  $V$  at time  $T$ . Using predictor processes  $X(s)$ , which are observable at time  $T$ , we wish to predict the anticipated speed for a vehicle that enters the selected highway segment at this time  $T$ . The velocity curve  $Y(t)$  of this vehicle as a function of time will depend on velocities  $V(t, s)$  for which  $t > T$ , i.e., which are in the future of time  $T$ , as the vehicle travels along the freeway on a time interval that begins at  $T$ .

For a vehicle that enters the selected highway segment at time  $T$ , in addition to providing an estimate for the response curve  $Y(t)$ ,  $t \in [T, T + \Delta]$  for a  $\Delta > 0$ , the proposed model also leads to a prediction region for  $Y(t)$  over the future time period  $t \in [T, T + \Delta]$ . We choose  $\Delta = 300$  seconds for our application and assume that the sample of observed times  $T_i$  at which test drivers enter the selected segment of highway have the same distribution as  $T$ ,  $T_i \sim T$ . These considerations lead to pairs of predictor curves  $X_i(s) = V(T_i, s)$  and associated response curves  $Y_i(t)$ ,  $t \in [T_i, T_i + \Delta]$ , which are to be predicted.

A complication is that for an entry time  $T$ , we do not directly observe the entire function  $X(s)$ ,  $s \in [21, 27.5]$ , but only observe a subset of the data points in eq. (19),  $\{t_l, s_l, V_l\}$ , with  $t_l = T$ . In practice, this can be easily overcome by including all  $t_l \in [T - h, T]$ , which lie in a small window towards the past, relative to the random entry time  $T$ , where  $h$  is a very small bandwidth. In this way, one can assemble a reasonable number of records to recover  $X(s)$ ; we chose  $h = 10$  seconds for this application. Assuming there are a total of  $L$  records that satisfy  $t_l \in [T - h, T]$ , we then re-index these records as  $(s_j, V_j)_{j=1, \dots, L}$ .

Due to deviations of the velocities of specific vehicles, including the test vehicles, from the prevailing velocity field, the velocities  $V_j$  in practice are approximations of the actual velocity field of the freeway at location  $s_j$  at time  $T$ . For a sample of test cars with entry times  $T_i$ ,  $i = 1, \dots, n$ , data model (4) therefore applies to the data  $(\{s_{ij}, V_{ij}\}, j = 1, \dots, L)_{i=1, \dots, n}$ , and one can perform functional principal component analysis to obtain estimates for the

mean function  $\mu_X$ , eigenfunctions  $\phi_k$ , and functional principal components  $\xi_{ik}$ . Then one may recover the velocity trajectories  $X_i$  through the estimates  $\hat{X}_i$  that are defined in eq. (5).

We standardize  $X_i(s_0) = 0$ , so that speeds are valued relative to the speed recorded at the location corresponding to postmile 21. The number of included components  $K$  was chosen to be four, applying the criterion of fraction of variance explained (FVE) with threshold 85%. The estimated mean function and the first four eigenfunctions are illustrated in Figure 2. The mean function is seen to conform quite well with the velocity heat plot in Figure 1. Various eigenfunctions have peaks located before the velocity minimum, which points towards increased variability in certain eigenfunction directions in advance of the mean velocity minimum. This likely reflects various modes in which traffic slows down towards the area with minimum velocity.

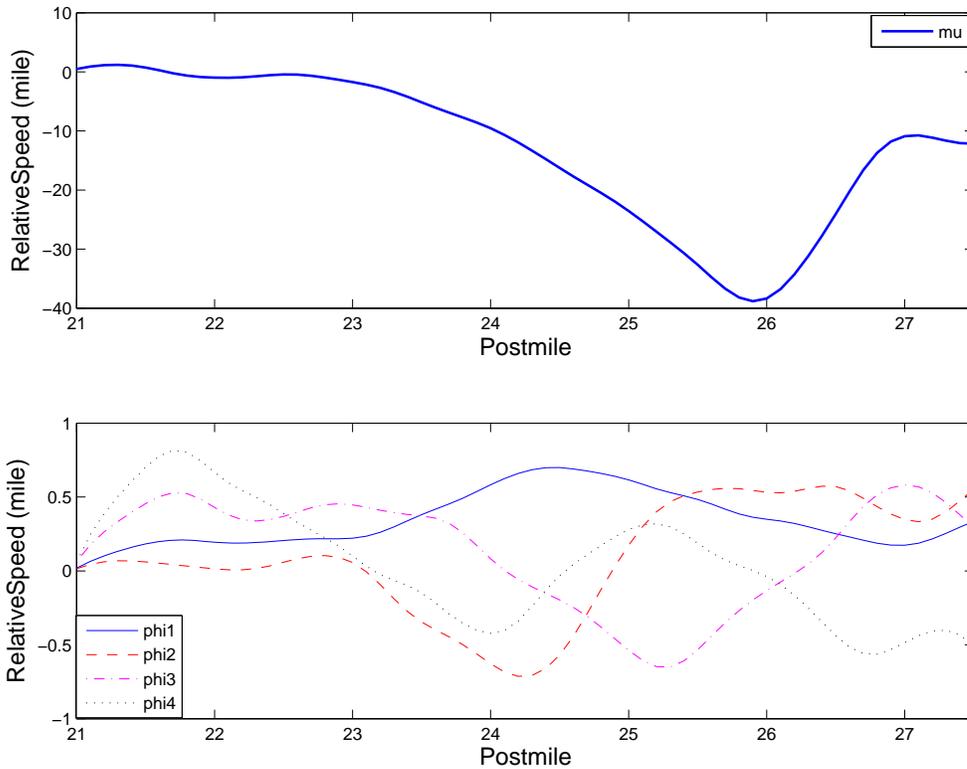


Figure 2: The estimated mean function  $\mu$  (top) and the first four eigenfunctions  $\phi_k$ ,  $k = 1, 2, 3, 4$  (bottom), for predictor processes  $X_i$ .

For the  $i$ th individual test car “trip” that enters the highway segment at time  $T_i$ , the

response function  $Y_i(t)$ ,  $t \in [0, 300]$  seconds, is defined as the relative speed curve over a future time period with length 300 seconds, where relative speed is obtained by subtracting the initial speed when the individual trip enters the highway segment from the actual subsequent vehicle speeds. Thus we reset  $Y_i(0) = 0$  for all  $i$ . Assembling the data from all individual trips that were run along the highway segment  $[21, 27.5]$ , with various entering times  $T_i$  between 10am and 6pm, results in a sample size of  $n = 126$ .

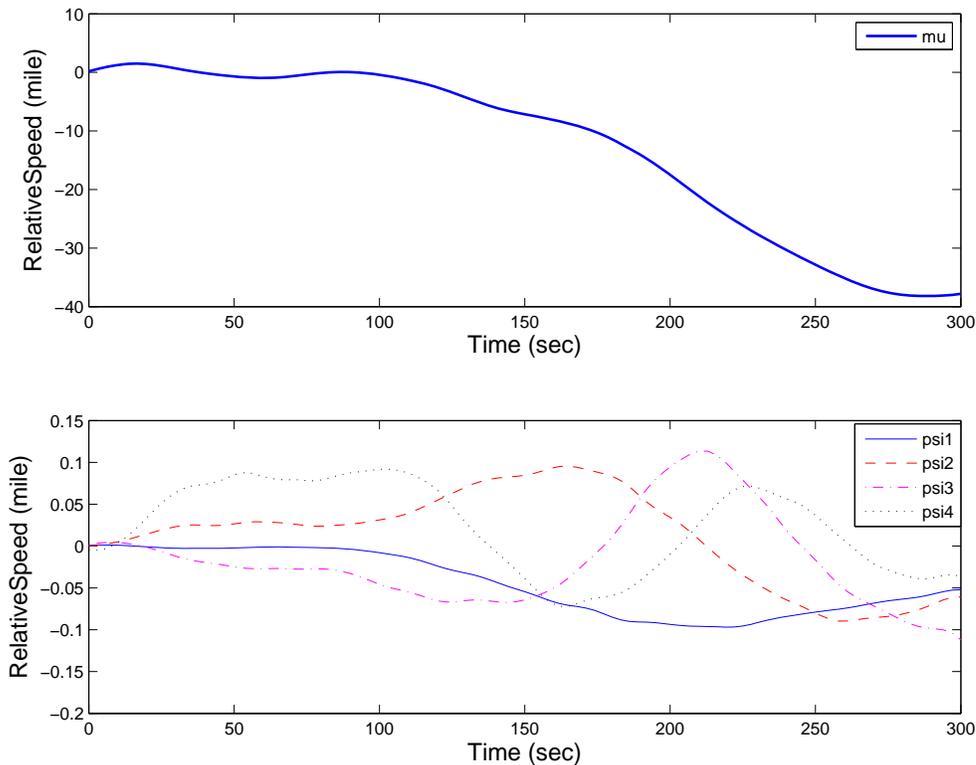


Figure 3: The estimated mean function (top) and the first four eigenfunctions  $\psi_j(s)$ ,  $j = 1, 2, 3, 4$  (bottom), for response processes  $Y_i$ .

Each trip  $i$  is represented by densely observed data  $(\{t_{ij}, V_{ij}\}, j = 1, \dots, N_i)_{i=1, \dots, n}$ , based on which one can estimate mean function  $\mu_Y$ , eigenfunctions  $\psi_j$ , and functional principal components  $\zeta_{ij}$ . The number of included components  $P$  is chosen to be 4, selected through FVE with threshold 0.9. Estimates of the mean function  $\mu_Y$  and of the first four eigenfunctions  $\psi_l$ ,  $l = 1, 2, 3, 4$ , are visualized in Figure 3, demonstrating that vehicle velocities systematically decline towards the right end of the highway segment. The variability as

reflected in the eigenfunctions is high just before this slowing down happens.

Following the procedure in Section 3, we estimated the mean and conditional covariance function of  $Y$ , conditional on  $X$ ; see eq. (9) and (13). Furthermore, given current road condition  $X$  at times  $T$ , we constructed the 90% prediction region for the future speed curve  $Y$  of a car that enters the freeway at postmile 21 at time  $T$ . These regions were constructed by plugging corresponding estimates into eq. (17) and (18). The constant  $\mathcal{C}$  was computed empirically by controlling the empirical average coverage to be 90%.

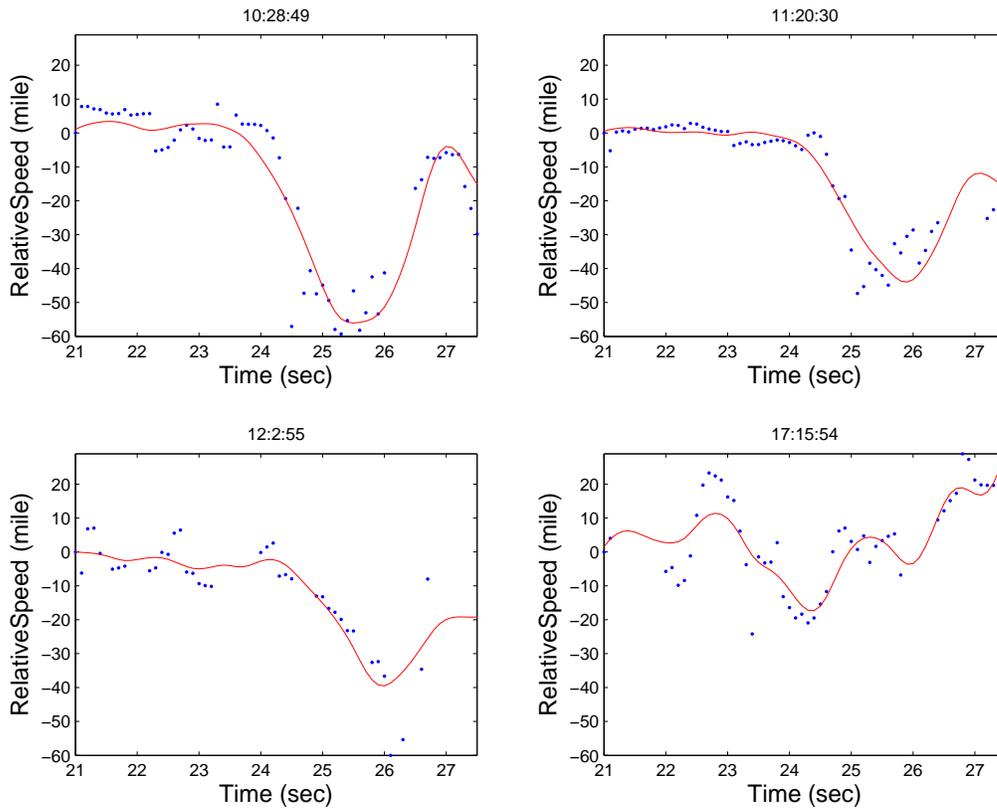


Figure 4: Noisy observations and recovered predictor functions  $\hat{X}_i$  (solid curves) for velocity trajectories  $X_i$ , representing the speed (relative to postmile 21) of the traffic flow on the highway segment  $[21, 27.5]$ , for four randomly selected entering times  $T_i$ , where data are combined from trips entering the freeway at times close to  $T_i$ .

For four randomly selected individual trips that enter the highway segment at different times  $T_i$ , we display the predictor functions  $X_i(s)$  and the relative speed on the highway

section [21, 27.5] at times  $T_i$  in Figure 4. It is noteworthy that velocity varies widely along this stretch of freeway. Adopting the empirical value of  $\mathcal{C}$ , the estimated functions of  $\mu_{Y|X}(t)$  and  $\text{var}(Y_X(t))$ , we can evaluate eq. (17) by plugging in the estimated functional principal components of  $X_i(s)$  at the four selected times  $T_i$ .

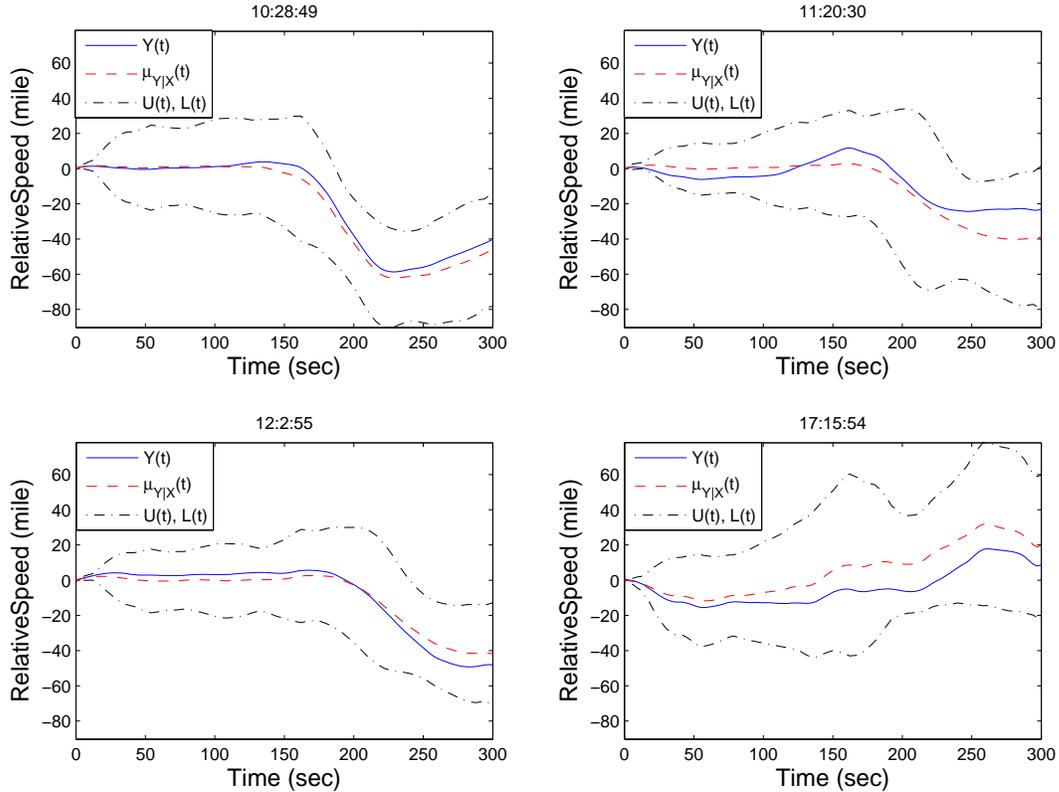


Figure 5: Relative speed functions  $Y_i(t)$  (solid) for four randomly selected test vehicles that enter the highway segment of interest at times  $T_i$ , over a future period of 5 minutes, along with the predicted conditional mean curve (dashed) and the prediction regions (dash-dotted), based on only the traffic flow  $X_i(s)$  at the entering times  $T_i$ . Note that in the construction of predicted mean speed functions and the prediction bands we do not use the observed speed curves.

In Figure 5, we plot the relative speed curve  $Y_i(t)$ ,  $t \in [0, 300]$ , overlaid with the predicted conditional mean speed function and the 90% prediction region, conditional on the corresponding predictor curve  $X_i$ . These regions reflect the conditional distribution of the responses and provide a notion of the level of uncertainty that is associated with predictions

obtained from functional mean regression implementations. As the figure indicates, there is a sizable degree of uncertainty, which increases as time progresses, but nevertheless, the overall trends of slowing traffic towards the end of the segment are unmistakable and are confirmed by the prediction bands.

For this particular traffic data prediction problem, only data for one day were available. Large batches of additional data are expected to become available in future follow-up projects and through many other studies worldwide. The proposed additive functional conditional distribution model will be of interest for the analysis of the larger data sets that these studies will generate, as well as for other traffic prediction problems.

## 5 Simulation

To illustrate our methods for prediction region construction, especially to assess the out of sample coverage for non-Gaussian cases, pairs of predictor and response curves  $(X_i, Y_i)$ ,  $i = 1, \dots, 400$ , were generated, and of these, 300 pairs were used for training the model and the remaining 100 pairs for testing. Based on the training sample, we obtain the functional principal components for  $X_i$  and  $Y_i$  as described in Section 2; then functions  $\hat{\mu}_{Y|X}(t)$  and  $\hat{\text{var}}(Y_X(t))$  are estimated as described in Section 3.1 and 3.2. For a given new predictor function  $X_i$ , one then obtains  $L_i(t)$  and  $U_i(t)$  by evaluating functions  $f_{jk}(\cdot)$  and  $g_{jk}(\cdot)$  at the functional principal components  $\xi_{ik}$ ,  $k = 1, \dots, K$  of predictor processes  $X_i(s)$ , where we substitute estimates for these quantities. For this simulation,  $\mathcal{C}_\alpha$  is chosen by controlling the empirical leave-one-out coverage for the training sample to be  $1 - \alpha$ . The coverages of the testing sample are recorded for nominal levels  $\alpha = \{0.5, 0.75, 0.9\}$ . We repeated the simulation 100 times.

In each simulation run, predictor trajectories  $X_i(s)$  were generated to mimic the traffic data. For each  $s_j$  equally distributed on  $[21, 27.5]$ , contaminated observations of  $X_i$  were generated as  $X_i(s_j) = \mu_X(s_j) + \sum_{k=1}^K \xi_{ik} \phi_k(s_j) + \epsilon_{ij}$ , where we substituted for  $\mu_X$ ,  $\phi_k$  the estimated functions that were obtained for the traffic data, as described in the previous section. Functional principal components  $\xi_{ik}$  were generated from a Gaussian distribution with mean zero and variances  $\lambda_k$ , which were chosen to equal the estimated eigenvalues for the traffic data. The additive contaminating errors  $\epsilon_{ij}$  were generated from a standard

Gaussian distribution.

For a given predictor function  $X_i$ , the corresponding response  $Y_i$  was generated as follows: For 101 equally distributed  $t_l \in [0, 300]$  we constructed noise-corrupted observations of response processes  $Y_X$  as  $Y_i(t_l) = \mu_{Y|X_i}(t_l) + \sum_{j=1}^P \zeta_j(X_i)\psi_j(t_l) + \varepsilon_{il}$ , where  $\zeta_j(X_i)$  are functional principal components with mean zero and variance depending on a function  $\gamma_j(X_i)$ , and  $\varepsilon_{il}$  are i.i.d additive errors corresponding to standard Gaussians. To mimic the traffic data, we used the estimated versions of  $\mu_{Y|X_i}$ ,  $\psi_j$  and  $\gamma_j(X_i)$  obtained for the traffic data, with detailed descriptions in the previous section. For the generation of the functional principal components  $\zeta_j(X_i)$  of response processes  $Y_X$ , we studied three scenarios:

- (i) *Gaussian case.* The response components  $\zeta_j(X_i)$  are generated as Gaussians  $\mathcal{N}(0, \gamma_j(X_i))$ .
- (ii) *Gaussian mixture.* The response components  $\zeta_j(X_i)$  are generated as a mixture distribution,  $B\mathcal{N}(-2, \gamma_j(X_i)/5) + (1 - B)\mathcal{N}(2, \gamma_j(X_i)/5)$ , where  $B$  is a Bernoulli variable with  $E(B) = 0.5$  that is independent of all other random variables and does not depend on  $X_i$ .
- (iii) *Gamma distribution.* The response components  $\zeta_j(X_i)$  are generated as  $\zeta_j(X_i) = \sqrt{\gamma_j(X_i)}(Z_{ij} - 4)/2$ , where  $Z_{ij}$  is a *Gamma*(4, 1) variable.

The results for mean coverage and mean values of  $\mathcal{C}_\alpha$  based on 100 simulation runs for each simulation scenario and levels  $\alpha \in (0.5, 0.75, 0.9)$  are in Table 1. They clearly show that the proposed method works well to keep the empirical level of the confidence bands close to the nominal levels for both Gaussian and non-Gaussian cases in scenarios that resemble the settings of the traffic data.

Table 1: Simulation results for mean coverage, and average values of empirically chosen  $\mathcal{C}_\alpha$ , based on 100 simulation runs, for levels  $1 - \alpha = \{0.5, 0.75, 0.9\}$ , where the functional principal components of the conditional process  $Y_X(t)$  are generated from Gaussian, Gaussian mixture, and transformed Gamma distributions.

distribution	$1 - \alpha = 0.5$		$1 - \alpha = 0.75$		$1 - \alpha = 0.9$	
	$\mathcal{C}_\alpha$	coverage	$\mathcal{C}_\alpha$	coverage	$\mathcal{C}_\alpha$	Coverage
Gaussian	<b>1.276</b>	<b>0.517</b>	<b>1.832</b>	<b>0.763</b>	<b>2.562</b>	<b>0.906</b>
Gaussian Mixture	<b>1.356</b>	<b>0.502</b>	<b>1.796</b>	<b>0.762</b>	<b>2.419</b>	<b>0.900</b>
Gamma	<b>1.238</b>	<b>0.496</b>	<b>1.784</b>	<b>0.745</b>	<b>2.663</b>	<b>0.900</b>

## 6 Assumptions and Discussion

In the following we list and discuss four key assumptions for the proposed approaches.

*Assumption 1. Predictor processes  $X$  and response processes  $Y$  are square integrable, and belong to the Hilbert space  $(X, Y) \in L_2 \times L_2$ .*

This is a standard assumption in functional data analysis and is not restrictive.

*Assumption 2. The conditional response processes, conditioning on the predictor processes, are Gaussian.*

This assumption is needed to simplify the very complex task of modeling conditional distributions, especially for the case that we consider here, where predictors are also functional. Gaussianity implies that the functional principal components for  $Y$  given  $X$  are independent, and crucially, that conditional distributions of response processes only depend on conditional mean functions and conditional covariance surfaces. Without this assumption, the characterization of the conditional distribution of a random process, given functional predictors, would require extremely complex modeling for conditional higher moments and cross-moments of response processes and thus unrealistically large sample sizes.

The proposed methods for the construction of prediction regions are motivated by the Gaussian case, and we indicate how to extend these to the non-Gaussian case, by focusing on coverage in an average sense, in section 3.3. As the simulations and Table 1 clearly indicate, the extension of these ideas to non-Gaussian situations works overall reasonably well.

*Assumption 3. Additive modeling approaches are used for the conditional mean function and conditional covariance function.*

These assumptions are needed to achieve a reasonable balance between model flexibility and feasibility. Modeling conditional mean functions corresponds to the usual functional regression problem with functional responses, where one targets conditional means. The modeling of conditional means for the case of functional responses has been extensively studied, the additive model being among the most general such approaches. Therefore it makes sense to borrow this existing and very flexible approach and to extend it to the conditional covariance modeling. Additive modeling is not only flexible but also structurally stable and is well known to yield excellent results in many regression settings.

Functional linear regression can also be considered as an alternative for conditional mean modeling. Modeling conditional smooth covariance functions is new as far as we know, and to develop an additive model for this modeling task as described in section 3.2 is quite natural, given the success and flexibility of additive models for conditional mean regression. A further assumption of *common functional principal components* is made implicitly in the modeling of the conditional covariances. This assumption is necessary in order to keep the complexity of the resulting models manageable and has been successfully employed before (Benko et al., 2009).

*Assumption 4. The functional principal component scores of the predictors  $X$  are independent.*

This assumption is needed for the straightforward practical implementation of the additive modeling for mean and covariance estimation, and is always satisfied in case the predictor processes are Gaussian (but not necessarily jointly Gaussian with the response processes). This assumption was shown to work well for functional additive modeling in previous work (Müller and Yao, 2008, 2010). If this assumption is violated, one can adopt more elaborate backfitting procedures (Mammen and Park, 2005) for the fitting of the two additive models for conditional mean functions and conditional covariance surfaces.

The problem to describe and quantify conditional distributions for functional responses has not been considered before as far as we know, even for the case of scalar or vector predictors. When the predictors are also functions, as in the traffic application, this adds substantial complexity to this problem. Previous models for function to function regression based on bivariate functional observations  $(X, Y) \in L^2 \times L^2$  exclusively focused on modeling the mean regression function  $E(Y|X = x)$ , while our interest here is the modeling of the entire conditional distribution  $\mathcal{L}(Y|X = x)$ , then applying these models to the construction of conditional prediction regions. The difficulty is that the distribution and especially the conditional distribution of functional data is hard to quantify and is unwieldy for statistical practice. This motivates to introduce simplifying structural assumptions as described above, which serve to reduce the intractable general problem to a series of feasible estimation problem. These assumptions lead to a practical and still sufficiently flexible solution and are not overly restrictive.

For the analysis of time-dynamic traffic monitoring data, functional methods convey the advantages of being highly flexible, while they are also consistent and statistically well motivated. Especially valuable for traffic monitoring are functional regression models, where response curves are related to predictors. Besides the commonly used but somewhat restrictive linear functional model, additive models and some other flexible models have been developed in recent years. Especially the flexible yet stable functional additive model has proved successful and we adopt it for the mean regression. A major extension that we propose in this paper is the additive modeling of conditional covariance functions.

When evaluating predicted future speeds in response to current traffic flows, it is clearly of interest not only to obtain point estimates of conditional speeds, but also simultaneous prediction regions around these estimates, as these characterize the uncertainty in the conditional mean response functions, irrespective of which function to function regression model is applied. Thus, one is not only interested in predicting the response curve, given predictors, but also in the level of confidence one could reasonably have regarding the accuracy of such predictions. We quantify this uncertainty by constructing conditional prediction regions.

While in our application to GPS-enabled mobile phone traffic monitoring we only considered one stretch of highway extending over a few miles, for wide spread applicability one would need to extend this approach to larger networks of roads and dynamic prediction. Such more complex applications can be handled by breaking the prediction down into prediction problems for a number of smaller road segments, for each of which the method can be applied as described, thus implementing a “divide-and conquer” strategy.

We expect that the proposed functional data analysis approaches, especially functional additive regression and the construction of conditional covariance models, will prove useful for time-dynamic traffic modeling and monitoring as well as for the same-day prediction of various traffic characteristics.

## References

- Benko, M., Härdle, W., and Kneip, A. (2009), “Common functional principal components,” *The Annals of Statistics*, 37, 1–34.
- Bickel, P. and Li, B. (2007), “Local Polynomial Regression On Unknown Manifolds,” *Com-*

- plex Datasets And Inverse Problems: Tomography, Networks And Beyond*, ser. *IMS Lecture Notes-Monograph Series.*, 54, 177–186.
- Cardot, H. (2007), “Conditional Functional Principal Components Analysis,” *Scandinavian Journal of Statistics*, 34, 317–335.
- Cardot, H., Crambes, C., and Sarda, P. (2005), “Quantile regression when the covariates are functions,” *Journal of Nonparametric Statistics*, 17, 841–856.
- Chen, K. and Müller, H.-G. (2011), “Conditional quantile analysis when covariates are functions, with application to growth data,” *Journal of the Royal Statistical Society: Series B*, 73, 000—000.
- Chiou, J.-M. and Müller, H.-G. (2009), “Modeling hazard rates as functional data for the analysis of cohort lifetables and mortality forecasting,” *Journal of the American Statistical Association*, 104, 572–585.
- Chiou, J.-M., Müller, H.-G., and Wang, J.-L. (2004), “Functional response models,” *Statistica Sinica*, 14, 675–693.
- Chu, L., Oh, S., and Recker, W. (2005), “Adaptive Kalman filter based freeway travel time estimation,” in *84th TRB Annual Meeting, Washington DC*.
- Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and its Applications*, London: Chapman & Hall.
- Faraway, J. J. (1997), “Regression analysis for a functional response,” *Technometrics*, 39, 254–261.
- Flury, B. (1988), *Common Principal Components and Related Multivariate Models*, John Wiley and Sons, Inc.
- Hall, P. and Hosseini-Nasab, M. (2006), “On properties of functional principal components analysis,” *Journal of the Royal Statistical Society: Series B*, 68, 109–126.
- Hall, P., Müller, H.-G., and Yao, F. (2008), “Modeling sparse generalized longitudinal observations with latent Gaussian processes,” *Journal of the Royal Statistical Society: Series B*, 70, 730–723.

- (2009), “Estimation of functional derivatives,” *Annals of Statistics*, 37, 3307–3329.
- Herrera, J. and Bayen, A. (2010a), “Incorporation of Lagrangian measurements in freeway traffic state estimation,” *Transportation Research Part B: Methodological*, 44, 460–481.
- (2010b), “Incorporation of Lagrangian measurements in freeway traffic state estimation,” *Transportation Research Part B: Methodological*, 44, 460–481.
- Herrera, J., Work, D., Herring, R., Ban, X., Jacobson, Q., and Bayen, A. (2010), “Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment,” 18, 568–583.
- Herring, R., Hofleitner, A., Amin, S., Nasr, T., Khalek, A., Abbeel, P., and Bayen, A. (2010), “Using mobile phones to forecast arterial traffic through statistical learning,” in *89th Transportation Research Board Annual Meeting, Washington DC*.
- Izadpanah, P., Hellinga, B., and Fu, L. (2011), “Real-Time Freeway Travel Time Prediction Using Vehicle Trajectory Data,” in *Transportation Research Board 90th Annual Meeting*, no. 11-2826.
- Jiang, C. and Wang, J. (2010), “Covariate adjusted functional principal component analysis,” *Annals of Statistics*, 38, 1194–1226.
- Mammen, E. and Park, B. U. (2005), “Bandwidth selection for smooth backfitting in additive models,” *Annals of Statistics*, 33, 1260–1294.
- Mazare, P.-E., Tossavainen, O.-P., Bayen, A., and Work, D. (2012), “Trade-offs between inductive loops and GPS probe vehicles for travel time estimation: A Mobile Century case study,” in *Transportation Research Board 91st Annual Meeting (TRB’12)*.
- Mihaylova, L., Boel, R., and Hegyi, A. (2007), “Freeway traffic estimation within recursive Bayesian framework,” *Automatica*, 43, 290–300.
- Müller, H.-G., Chiou, J.-M., and Leng, X. (2008), “Inferring gene expression dynamics via functional regression analysis,” *BMC Bioinformatics*, 9, 60.

- Müller, H.-G. and Yao, F. (2006), “Regressing longitudinal response trajectories on a covariate,” in *Frontiers in Statistics. Festschrift for Peter Bickel*, eds. Fan, J. and Kou, H. L., London: Imperial College Press, pp. 305–324.
- (2008), “Functional additive models,” *Journal of the American Statistical Association*, 103, 1534–1544.
- (2010), “Additive modeling of functional gradients,” *Biometrika*, 97, 791–806.
- Nanthawichit, C., Nakatsuji, T., and Suzuki, H. (2003), “Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway,” *Transportation Research Record: Journal of the Transportation Research Board*, 1855, 49–59.
- Ramsay, J. O. and Dalzell, C. J. (1991), “Some tools for functional data analysis,” *Journal of the Royal Statistical Society: Series B*, 53, 539–572.
- Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*, Springer Series in Statistics, New York: Springer, 2nd ed.
- Rice, J. A. (2004), “Functional and longitudinal data analysis: Perspectives on smoothing,” *Statistica Sinica*, 631–647.
- Shen, Q. and Faraway, J. (2004), “An F test for linear models with functional responses,” *Statistica Sinica*, 14, 1239–1257.
- Tao, S., Manolopoulos, V., Rodriguez, S., and Rusu, A. (2012), “Real-Time Urban Traffic State Estimation with A-GPS Mobile Phones as Probes,” *Journal of Transportation Technologies*, 2, 22–31.
- Wang, Y. and Nihan, N. (2000), “Freeway traffic speed estimation with single-loop outputs,” *Transportation Research Record: Journal of the Transportation Research Board*, 1727, 120–126.
- Wang, Y. and Papageorgiou, M. (2005), “Real-time freeway traffic state estimation based on extended Kalman filter: a general approach,” *Transportation Research Part B: Methodological*, 39, 141–167.

- Wang, Y., Papageorgiou, M., and Messmer, A. (2008), “Real-time freeway traffic state estimation based on extended Kalman filter: Adaptive capabilities and real data testing,” *Transportation Research Part A: Policy and Practice*, 42, 1340–1358.
- Work, D., Tossavainen, O., Blandin, S., Bayen, A., Iwuchukwu, T., and Tracton, K. (2008), “An ensemble Kalman filtering approach to highway traffic estimation using GPS enabled mobile devices,” in *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on, IEEE*, pp. 5062–5068.
- Yao, F., Müller, H.-G., and Wang, J.-L. (2005), “Functional data analysis for sparse longitudinal data,” *Journal of the American Statistical Association*, 100, 577–590.