Submitted to the Annals of Applied Statistics

DYNAMICAL FUNCTIONAL PREDICTION AND CLASSIFICATION, WITH APPLICATION TO TRAFFIC FLOW PREDICTION

By Jeng-Min Chiou

Academia Sinica

Motivated by the need for accurate traffic flow prediction in transportation management, we propose a functional data method to analyze traffic flow patterns and predict the future traffic flow for an up-to-date and partially observed flow trajectory. In this study, we approach the problem by sampling traffic flow trajectories from a mixture of stochastic processes, each reflecting a distinct pattern of daily traffic flow trajectories. The proposed *functional mixture prediction* approach combines functional prediction with probabilistic functional classification to take distinct traffic flow patterns into account. The probabilistic classification procedure, which incorporates functional clustering and discrimination, hinges on subspace projection that lays the groundwork for functional mixture prediction. The proposed methods not only assist in predicting traffic flow trajectories, but also identify distinct patterns in daily traffic flow of typical temporal trends and variabilities. Although motivated by traffic flow analysis and prediction, the proposed methodology is widely applicable in analysis and prediction of longitudinally recorded functional data.

1. Introduction. Traffic flow is an important macroscopic traffic characteristic in transportation systems. The measurement and forecasting of traffic flow are crucial in the design, planning and operations of highway facilities [Zhang and Ye (2008)]. Traffic flow can be measured automatically using various types of vehicle detectors such as the commonly used dual loop detectors, which are installed in certain roads at regular intervals. Real-time traffic flow information in conjunction with historical traffic flow records makes it possible to predict traffic flow in the short term. The importance of traffic prediction for intelligent transportation systems has long been recognized in many applications, including the development of traffic control strategies in advanced traffic management systems and real-time route guidance in advanced traveler information systems [Zheng, Lee and Shi (2006)]. However, dynamic features of traffic flow, along with unstable traffic conditions and unpredictable environmental factors, contribute to the challenge of pursuing accuracy in predictions.

Short-term traffic flow prediction has been intensively investigated for more than two decades and various types of methodologies have been developed. These include time series models [e.g. Williams and Hoel (2003); Stathopoulos and Kar-

Keywords and phrases: clustering, discrimination, functional regression, mixture model, subspace projection

laftis (2003)], Kalman filtering methods [e.g. Xie, Zhang and Ye (2007); Okutani and Stephanides (1984)], local linear regression models [Sun et al. (2003)], neural network based methods [e.g. Chen and Grant-Muller (2001); Zheng, Lee and Shi (2006); Çetiner, Sari and Borat (2010)] and fuzzy neural models and fuzzy logic system methods [Yin et al. (2002); Zhang and Ye (2008)], among others. In addition, there are many articles comparing parametric time series models, nonparametric regression models and neural networks in traffic prediction, such as in Kirby, Waston and Dougherty (1997), Smith and Demetsky (1997) and Smith, William and Oswald (2002). To the best of our knowledge, a functional data approach to predicting traffic flow has not yet been investigated in the literature.

1.1. *Illustration of traffic flow prediction and the proposed functional data method.* Motivated by a practical need for accurate traffic flow prediction, we develop a novel functional data method for predicting future, or unobserved, daily traffic flow for an up-to-date and partially observed traffic flow trajectory. Figure 1 illustrates a sample of daily traffic flow trajectories. The data were collected by a dual loop vehicle detector located near Shea-San Tunnel on National Highway 5 in Taiwan in 2009 and are based on the flow rates (vehicle count per min) over 15-min time intervals, a metric suggested in *Highway Capacity Manual 2000* for operational analyses [Zheng, Lee and Shi (2006)]. The trajectories sample 70 days as the training data, while the remaining 14 days are used as the test data to validate the prediction performance. The aim is to predict the unobserved traffic flow trajectory for a partial trajectory with updated flow information up to the "current

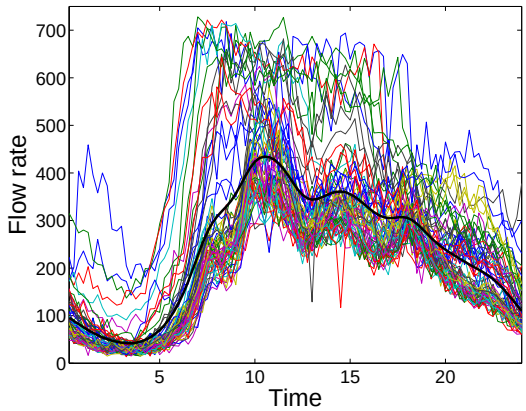
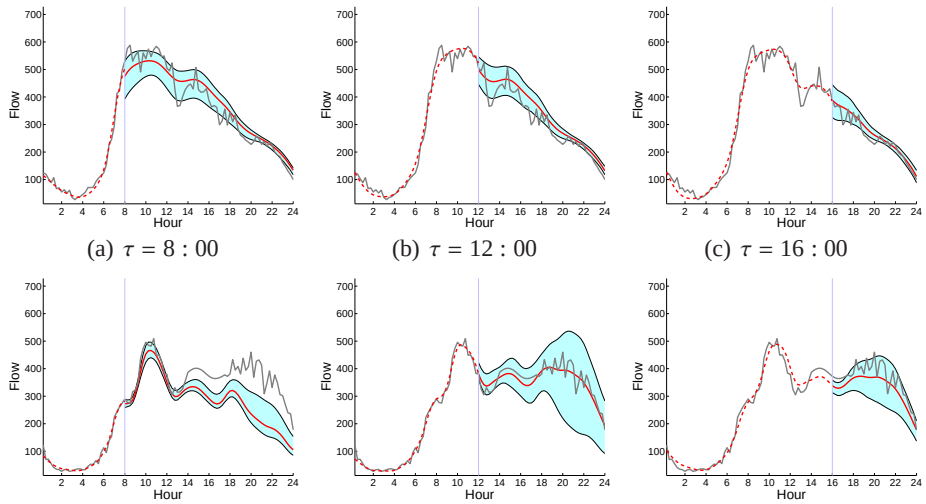


Fig 1: *Daily traffic flow trajectories (training data) with the estimated mean function superimposed on the observed trajectories.*



(a) $\tau = 8:00$ (b) $\tau = 12:00$ (c) $\tau = 16:00$

(d) $\tau = 8:00$ (e) $\tau = 12:00$ (f) $\tau = 16:00$

Fig 2: *Two samples of trajectories from the test data set ((a)–(c) for Test sample 1; (d)–(f) for Test sample 2). The fitted curves (dotted lines before times τ) and the predicted curves (solid curves after times τ) with 95% prediction intervals for a partially observed trajectory available up to times $\tau = 8$, 12 and 16:00, superimposed on the complete trajectory (gray line).*

time" τ, which is given as a time of day. In Figure 2, the raw trajectories (gray lines) before $\tau = 8$, 12 and 16:00 are observations from the test data, superimposed on the curves (dotted lines) fitted by functional principal component analysis. After the last observation time point τ, the predicted traffic flow trajectories (solid lines) are obtained by the proposed functional mixture prediction model coupled with the 95% bootstrap prediction intervals. The real data trajectories after times τ (gray lines) are unobserved in the prediction scenario and are displayed for comparative purposes. The prediction for the trajectory is dynamically updated as the "current time" τ progresses.

We note that the aforementioned methods in the literature were largely developed based on "short-term" or "next-step" traffic prediction modeling, where "short-term" refers to a forecast horizon of a time interval. A 15-min interval is commonly used as the forecast horizon for traveler-oriented applications and operational analysis [Zhang and Ye (2008)]. In contrast to the "next-interval" prediction methods in the existing literature, our functional data method is more flexible, as illustrated in Figure 2, allowing valid prediction periods extended from the "current time" to the end of the day and thus providing more information to relevant users.

Since future traffic conditions and temporal traffic flow patterns play a critical role in traffic prediction [e.g. Smith, William and Oswald (2002); Vlahogianni, Karlaftis and Golias (2008)], to take into account distinct daily traffic flow patterns we propose a *functional mixture prediction* approach that combines functional prediction with a probabilistic classification procedure. Specifically, we propose to implement functional cluster analysis of past traffic flow trajectories to obtain typical daily traffic flow patterns or clusters, followed by a probabilistic classification for the traffic flow trajectory observed thus far. Based on the traffic flow patterns or clusters identified by the proposed method, we can predict the unobserved traffic flow trajectories by a functional prediction model in conjunction with the estimated posterior membership probabilities of traffic flow clusters. Although motivated by traffic flow analysis and prediction, the proposed methodology is by no means restricted to this particular field and is generally applicable to a wide variety of longitudinally recorded functional data.

In many real applications, clustering of curve data can be challenging and misclassification of an up-to-date and partially observed trajectory can cause loss of prediction accuracy. Hence, a simple "prediction-after-classification" approach based on hard clustering results may not be the best approach. This will be illustrated in our numerical studies including the real data application and simulations. In contrast, the proposed functional mixture prediction framework addresses challenges related to prediction of complex functional data, such as those containing heterogeneous patterns and large variability over time. The proposed *functional mixture prediction* approach to traffic flow prediction has the following features:

- It is the first approach to employ functional data techniques to address traffic flow prediction applications, which are critical in many intelligent transportation systems.
- The functional data approach allows for interval prediction in contrast to "next-step" traffic flow prediction applications found in the literature.
- The proposed functional mixture prediction model plays a central role in predicting the future trajectory for an up-to-date and partially observed trajectory and takes distinct traffic flow patterns into account to improve prediction accuracy. This study extends the idea of the subspace projected functional clustering method of Chiou and Li (2007) to identify distinct patterns of daily traffic flow from the past data, coupled with the forward functional testing procedure of Li and Chiou (2011) to determine the number of clusters, which lays the groundwork for the functional mixture prediction approach.
- The probabilistic classification approach, including functional clustering and discrimination, is new. It allows for the prediction of posterior membership probabilities using partial information up to the current time, in contrast to clusters constructed by complete trajectories from past data.

- The predictive function in the functional mixture prediction model is built on an existing functional linear model, which is widely used in functional regression modeling [e.g. Ramsay and Dalzell (1991), Müller, Chiou and Leng (2008)] and is easy to implement.

1.2. *Literature review of relevant functional data methods.* Statistical tools for functional data analysis have been extensively developed during the past two decades to deal with data samples consisting of curves or other infinite-dimensional data objects. Systematic overviews of functional data analysis are provided in the monographs of Ramsay and Silverman (2002, 2005) and Ferraty and Vieu (2006) and in the review articles of Rice (2004), Zhao, Marron and Wells (2004) and Müller (2005, 2008). Functional data analysis provides a wide range of applications in many disciplines. These include biomedical and environmental studies [Di et al. (2009); Gao and Niemeier (2008)], analysis of time-course gene expression profiles [Müller, Chiou and Leng (2008); Coffey and Hinde (2011)], linguistic pitch analysis [Aston, Chiou and Evans (2010)] and demographic and mortality forecasting [Hyndman and Ullah (2007); Chiou and Müller (2009); D'Amato, Piscopo and Russolillo (2011)], among many others. In relation to functional data prediction, Müller and Zhang (2005) proposed a functional data approach to predicting remaining lifetime and age-at-death distributions from individual event histories observed up to the current time. More recently, Zhou, Serban and Gebraeel (2011) proposed a functional data approach to degradation modeling for the evolution of degradation signals and the remaining life distribution. These are relevant works that contain novel functional data techniques with interesting applications to the prediction of an unobserved event for a partial trajectory observed up to the current time.

Among the various settings in functional regression analysis [Müller (2005)], models with both the response and predictor variables as functions serve this study's purpose with regard to prediction. Functional regression models of this kind have been considered, for example, in Yao, Müller and Wang (2005b), Chiou and Müller (2007), Müller, Chiou and Leng (2008) and Antoch et al. (2010). Methods of functional data clustering that are found in the literature include the use of multivariate clustering algorithms on the finite-dimensional coefficients of basis function expansions [e.g. Abraham et al. (2003); Serban and Wasserman (2005)], model-based functional data clustering [e.g. James and Sugar (2003); Ma and Zhong (2008)], a general descending hierarchical algorithm [Chapter 9 of Ferraty and Vieu (2006)] and various depth-based classification methods [Cuevas et al. (2007); López-Pintado and Romo (2006)], among others. Of particular interest with regard to functional prediction models are the methods that define clusters via subspace projection [Chiou and Li (2007, 2008)]. The subspace projection method considers

cluster differences not only in mean functions, but also in eigenfunctions of covariance kernels that takes into account individual random process variations, making it suitable for interpreting the stochastic nature of traffic flow and suggesting a natural link with functional regression models.

This article is organized as follows. In Section 2, we represent traffic flow trajectories as a mixture of stochastic processes and discuss functional clustering and classification methods to take traffic flow patterns into account. Section 3 discusses the functional mixture prediction model, including the algorithm for implementing functional mixture prediction. Sections 4 and 5 illustrate the empirical analysis of traffic flow patterns and results of predicting traffic flow trajectories. Section 6 presents a simulation study to evaluate the performance of the functional mixture prediction in comparison with related methods. Concluding remarks and discussion are provided in Section 7. More information in selecting the number of clusters, the bootstrap prediction intervals and additional details in the simulation design and results are deferred to Supplementary Materials.

2. Modeling traffic flow trajectories and clustering traffic flow patterns. Previous studies in traffic flow prediction and modeling have revealed that traffic condition data is characteristically stochastic, as opposed to chaotic [Smith, William and Oswald (2002)]. The stochastic features of traffic flow trajectories are suggestive of a functional data approach. In the functional data framework, we adopt the notion that each daily traffic flow trajectory is a realization of a random function sampled from a mixture of stochastic processes. Let Z denote the random function for the daily traffic flow trajectory in the domain $\mathcal{U} = [0, T]$. Here, the random function Z is square integrable with the inner product of any two functions f_1 and f_2 defined as $\langle f_1, f_2\rangle_{\mathcal{U}} = \int_{\mathcal{U}} f_1(t) f_2(t) dt$ with the norm $\|f_1\|_{\mathcal{U}} = \langle f_1, f_1\rangle_{\mathcal{U}}^{1/2}$. It is assumed that the random function $Z(t)$ has a smooth mean function $EZ(t) = \mu_Z(t)$ and covariance function $\mathrm{cov}(Z(s), Z(t)) = G_Z(s, t)$, for s and t in $\mathcal{U}$.

2.1. *Functional clustering of historical traffic flow trajectories.* While temporal traffic flow patterns are critical in traffic prediction, the underlying traffic flow structures and number of typical patterns are unknown and remain to be explored. We assume the mixture process Z consists of K sub-processes, with each sub-process corresponding to a cluster. The random cluster variable C for each individual cluster membership is randomly distributed among clusters with label $c \in \{1, \ldots, K\}$. For each sub-process associated with cluster c, define the conditional mean function $E(Z(t) \mid C = c) = \mu^{(c)}(t)$ and covariance function $\mathrm{cov}(Z(s), Z(t) \mid C = c) = G^{(c)}(s, t)$, for $c \in \{1, \ldots, K\}$. Let $(\lambda_j^{(c)}, \varphi_j^{(c)})$ be the corresponding eigenvalue-eigenfunction pairs of the covariance kernel $G^{(c)}$, where $\lambda_j^{(c)}$ are in non-ascending order. Assume, under mild conditions, each sub-process possesses a Karhunen-Loève expansion for the daily traffic flow trajectory Z given

by

$$Z^{(c)}(t) = \mu^{(c)}(t) + \sum_{j=1}^{\infty} \xi_j^{(c)} \varphi_j^{(c)}(t), \tag{2.1}$$

where $\xi_j^{(c)} = \langle Z - \mu^{(c)}, \varphi_j^{(c)} \rangle_{\mathcal{U}}$ with $\langle \varphi_j^{(c)}, \varphi_l^{(c)} \rangle_{\mathcal{U}} = 1$ for $j = l$ and 0 otherwise. In practice, it is often the case that the representation only requires a small number of components to approximate the trajectories. In general, trajectories with simpler structure require fewer components as compared to more complex trajectories.

Following the conventional approach, the best cluster membership c^* given Z is determined by maximizing the posterior probability $P_{C|Z}(\cdot \mid \cdot)$ such that

$$c^*(Z) = \underset{c \in \{1,\ldots,K\}}{\arg\max}\, P_{C|Z}(c \mid Z). \tag{2.2}$$

We propose estimating the posterior membership probability $P(C = c \mid Z)$ using the so-called discriminative approach, as opposed to the generative approach [see e.g. Dawid (1976); Bishop and Lasserre (2008)]. While there is no general consensus for choosing between generative and discriminative approaches [Ng and Jordan (2002); Xue and Titterington (2008)], the former requires *a priori* knowledge on the class-conditional probability density functions, information that is difficult to justify incorporating for the traffic flow trajectories. It is easier to use the discriminative approach that directly estimates the class-membership probabilities without attempting to model the underlying probability distributions of the random functions. Following this line, the multiclass logit model is a popular method for estimating the posterior membership probabilities. We propose incorporating a distance measure between Z and its projection associated with each cluster as the covariate in the multiclass logit model.

Consider the relative L^2 distance as the distance measure based on cluster subspace projection as

$$d^{(c)} = \frac{\|Z - \tilde{Z}^{(c)}\|^2}{\sum_{k=1}^{K} \|Z - \tilde{Z}^{(k)}\|^2}, \tag{2.3}$$

where $\tilde{Z}^{(c)}(t) = \mu^{(c)}(t) + \sum_{l=1}^{M_c} \xi_l^{(c)} \varphi_l^{(c)}(t)$, with $\xi_l^{(c)} = \langle Z - \mu^{(c)}, \varphi_l^{(c)} \rangle_{\mathcal{U}}$. The value M_c is finite and is chosen data-adaptively so that Z is well approximated by $Z^{(c)}$ by the M_c components. Let $\mathbf{d} = (1, d^{(1)}, \ldots, d^{(K-1)})^\top$ and $\boldsymbol{\gamma}_c = (\gamma_{0c}, \gamma_{1c}, \ldots, \gamma_{(K-1)c})^\top$. Taking the vector $\mathbf{d}$ as the covariate, we can estimate the posterior cluster membership probability using the multiclass logit model.

$$P(C = c \mid Z) = \frac{\exp\{\boldsymbol{\gamma}_c^\top \mathbf{d}\}}{\sum_{k=1}^{K} \exp\{\boldsymbol{\gamma}_k^\top \mathbf{d}\}}, \tag{2.4}$$

for $c = 1, \ldots, K-1$ and $P(C = K \mid Z) = 1 - \sum_{c=1}^{K-1} P(C = c \mid Z)$ with the Kth cluster being the baseline. The vector of regression coefficients $\boldsymbol{\gamma}_c$ remains to be estimated.

Clusters defined by criterion (2.2) are based on subspace projection. Let $S_M^{(c)}$ be the linear span of the set of eigenfunctions $\{\varphi_1^{(c)}, \ldots, \varphi_{M_c}^{(c)}\}$, $c = 1, \ldots, K$. For identifiability, it is assumed that for any two clusters c and d the following two conditions do not hold simultaneously: (i) $S_M^{(c)}$ belongs to $S_M^{(d)}$, (ii) $\mu^{(c)} = \mu^{(d)}$, or $\mu^{(c)} \in S_M^{(d)}$ and $\mu^{(d)} \in S_M^{(c)}$. These conditions were derived in Theorem 1 of Chiou and Li (2007) for identifiability of clusters defined via subspace projection. Criterion (2.2) leads to clusters with similar curves that are embedded in the cluster subspace spanned by the cluster center components, the mean function and the eigenfunctions of the covariance kernel that represent the functional principal component subspace.

In addition, the number of clusters is unknown, and must be determined in practice. The method used to determine the number of clusters in this study is based on a sequence of tests on cluster structures done to ensure statistical significance in the difference between cluster types as proposed in Li and Chiou (2011). The number of clusters K is determined by testing a sequence of null hypotheses $H_{01} : \mu^{(c)} = \mu^{(d)}$ and $H_{02} : S_M^{(c)} = S_M^{(d)}$, for $1 \le c \ne d \le K$. The *forward functional testing* procedure aims to search for the maximum number of clusters while retaining differences with statistical significance among the clusters. The procedure is especially suitable for the subspace projected functional clustering method. The sequence of the functional hypothesis tests helps identify significant differences between cluster structures, and provides additional insight into further cluster analysis. Since the hypothesis tests are based on bootstrap resampling methods, it takes substantial computational time to construct the reference distribution. Details of the procedure are discussed in [Li and Chiou (2011)] and we have briefly summarized them in Supplementary Material A.

2.2. *Probabilistic functional classification of traffic flow patterns.* For the purpose of prediction, the time domain $\mathcal{U}$ of the process Z is decomposed into two exclusive time domains $\mathcal{S}(\tau) = [0, \tau]$ and $\mathcal{T}(\tau) = [\tau, T]$. Now, let Z^* be a newly observed trajectory of the process Z, denoted by $Z^*_{\mathcal{S}(\tau)}$ as observed up to time τ. We predict the cluster membership probability of the trajectory Z^* based on the known trajectory $Z^*_{\mathcal{S}(\tau)}$ observed until time τ, which will then be used to predict the unobserved trajectory $Z^*_{\mathcal{T}(\tau)}$.

We define the relative L^2 distance in a manner similar to (2.3) via cluster subspace projection, but it is based on the partially observed $Z^*_{\mathcal{S}(\tau)}$ rather than the entire Z^* since the part $Z^*_{\mathcal{T}(\tau)}$ is not yet observed. Suppose that the cluster subspaces $\mu^{(c)}$ and $\{\varphi^{(c)}_{\mathcal{S}(\tau),j}\}$, $c = 1 \ldots, K$, are being identified as in Section 2.1. Then, the relative

L^2 distance is defined as

$$d^{*(c)}_{\mathcal{S}(\tau)} = \frac{\|Z^*_{\mathcal{S}(\tau)} - \tilde{Z}^{*(c)}_{\mathcal{S}(\tau)}\|^2}{\sum_{k=1}^K \|Z^*_{\mathcal{S}(\tau)} - \tilde{Z}^{(k)}_{\mathcal{S}(\tau)}\|^2}, \tag{2.5}$$

where $\tilde{Z}^{*(c)}_{\mathcal{S}(\tau)}(s) = \mu^{(c)}(s) + \sum_{l=1}^{M_c} \xi^{*(c)}_{\mathcal{S}(\tau),l} \varphi^{(c)}_{\mathcal{S}(\tau),l}(s)$, with $\xi^{*(c)}_{\mathcal{S}(\tau),l} = \langle Z^*_{\mathcal{S}(\tau)} - \mu^{(c)}, \varphi^{(c)}_{\mathcal{S}(\tau),l} \rangle_{\mathcal{S}(\tau)}$. Here, the set of eigenfunctions $\{\varphi^{(c)}_{\mathcal{S}(\tau),l}\}$ corresponds to the covariance kernel $G^{(c)}_{\mathcal{S}(\tau)}$ of the random process $Z_{\mathcal{S}(\tau)}$. Taking $\mathbf{d}^*_{\mathcal{S}(\tau)} = (1, d^{*(1)}_{\mathcal{S}(\tau)}, \ldots, d^{*(K-1)}_{\mathcal{S}(\tau)})^\top$ as the covariate, we can predict the cluster membership probability based on the newly observed $Z^*_{\mathcal{S}(\tau)}$ using the multiclass logit model

$$P(C = c \mid Z^*_{\mathcal{S}(\tau)}) = \frac{\exp\{\boldsymbol{\gamma}_c^\top \mathbf{d}^*_{\mathcal{S}(\tau)}\}}{\sum_{k=1}^K \exp\{\boldsymbol{\gamma}_k^\top \mathbf{d}^*_{\mathcal{S}(\tau)}\}}, \tag{2.6}$$

for $c = 1, \ldots, K-1$, and $P(C = K \mid Z^*_{\mathcal{S}(\tau)}) = 1 - \sum_{c=1}^{K-1} P(C = c \mid Z^*_{\mathcal{S}(\tau)})$ with the Kth cluster being baseline. We note that the vector of coefficients $\boldsymbol{\gamma}_c$ here is the same as that in (2.4) based on the historical or training data.

2.3. *Estimation for probabilistic functional classification.* In practice, the observed trajectories may be contaminated with random measurement errors. Let $Y_i(t_{ij})$ be the jth observation of the ith individual flow trajectory from the underlying process $Z_i^{(c)}$ of cluster c observed at time t_{ij}, $0 \le t_{ij} \le T$, such that $Y_i(t_{ij}) = Z_i^{(c)}(t_{ij}) + \epsilon_{ij}$, where $Z_i^{(c)}$ is the underlying random function such that $Z_i^{(c)}(t) = \mu^{(c)}(t) + \sum_{k=1}^\infty \xi_{ki}^{(c)} \varphi_k^{(c)}(t)$, and the random measurement errors ϵ_{ij} are independent of $\xi_{ki}^{(c)}$ with mean zero and variance σ^2.

To identify the structures of cluster subspaces, $\{\hat{\mu}^{(c)}, \{\hat{\varphi}_k^{(c)}\}_{k=1,\ldots,M_c,}\}$, $c = 1, \ldots, K$, we follow the idea of defining clusters via subspace projection and apply the proposed subspace-projected functional clustering procedure to the training data set. In the initial step, since cluster membership is unknown, the clustering is based on functional principal component scores of an overall single random process. Details of the initial clustering refer to Section 2.2.1 of Chiou and Li (2007). In the iterative updating step, cluster membership is determined by criterion (2.2) in a hard clustering manner. The clustering procedure is implemented iteratively in identifying between (i) cluster subspaces and (ii) cluster memberships until convergence.

Cluster subspaces. Given the observations $\{(t_{ij}, Y_i(t_{ij})), i = 1, \ldots, n, j = 1, \ldots, m\}$, from the historical or training data, and the cluster memberships of the trajectories, using the observations belonging to cluster c, the mean function $\mu^{(c)}$ can be estimated by applying the locally weighted least squares method while the estimates

of the components $\varphi_k^{(c)}$ and $\xi_{ik}^{(c)}$ rely on the covariance estimate $\hat{G}^{(c)}$ by applying the smoothing scatterplot data $(Y_{ij} - \hat{\mu}^{(c)}(t_{ij}))(Y_{il} - \hat{\mu}^{(c)}(t_{il}))$ to fit a local linear plane. Details of this estimation can be found in Chiou, Müller and Wang (2003) and Yao, Müller and Wang (2005a), for example. The smoothing parameters in the mean and covariance estimation steps are chosen data-adaptively via the 10-fold cross-validation method. An estimate of $\xi_{ik}^{(c)}$ can be obtained by the conditional expectation approach of Yao et al. (2005a) for the case of sparse designs. Here, we simply obtain the estimate $\hat{\xi}_{ik}^{(c)} = \langle Y_i - \hat{\mu}^{(c)}, \hat{\varphi}_k(c)\rangle_{\mathcal{U}}$ by numerical approximation for the case of dense designs of traffic flow recording. The value M_c is selected as the minimum that reaches a certain level of the proportion of total variance explained by the M_c leading components such that

$$M_c = \underset{L:\ L\geq 1}{\arg\min}\left\{\sum_{l=1}^{L}\hat{\lambda}_l^{(c)}\Big/\sum_{l=1}^{\infty}\hat{\lambda}_l^{(c)}1_{\{\hat{\lambda}_l^{(c)}>0\}} \geq \delta\right\}, \tag{2.7}$$

where $\delta = 90\%$ in this study. These cluster structure estimates are used in turn to estimate the vector of regression coefficients $\boldsymbol{\gamma}_k$ in (2.8) below.

Cluster memberships. Given the structure of each cluster based on the training data, we then use the discriminative approach to fit the posterior probabilities of cluster membership $P(C = c \mid Y_i)$ such that

$$\hat{P}(C = c \mid Y_i) = \frac{\exp\{\hat{\boldsymbol{\gamma}}_c^\top \hat{\mathbf{d}}_i\}}{\sum_{k=1}^{K}\exp\{\hat{\boldsymbol{\gamma}}_k^\top \hat{\mathbf{d}}_i\}},\ c = 1, \dots, K-1, \tag{2.8}$$

and $\hat{P}(C = K \mid Y_i) = 1 - \sum_{k=1}^{K-1}\hat{P}(C = k \mid Y_i)$, taking the Kth cluster as the baseline. Here, the relative L^2 distance vector $\hat{\mathbf{d}}_i = (1, \hat{d}_i^{(1)}, \dots, \hat{d}_i^{(K-1)})^\top$ serves as the predictor variable and is calculated by

$$\hat{d}_i^{(c)} = \frac{\|Y_i - \hat{Z}_i^{(c)}\|^2}{\sum_{k=1}^{K}\|Y_i - \hat{Z}_i^{(k)}\|^2},\ c = 1, \dots, K-1, \tag{2.9}$$

where $\hat{Z}_i^{(c)}(s) = \hat{\mu}^{(c)}(s) + \sum_{j=1}^{M_c}\hat{\xi}_{ij}^{(c)}\hat{\varphi}_j^{(c)}(s)$, with $\hat{\xi}_{ij}^{(c)}$ defined above. The coefficient estimates $\hat{\boldsymbol{\gamma}}_c = (\gamma_{0c}, \gamma_{1c}, \dots, \gamma_{(K-1)c})$, are obtained by the conventional iterated reweighted least squares method [McCullagh and Nelder (1989)]. The resulting estimate (2.8) is used to determine the cluster membership according to (2.2).

Now, given a newly observed trajectory Y^* up to time τ, denoted by $Y^*_{\mathcal{S}(\tau)}$, we obtain the covariate vector $\hat{\mathbf{d}}^*_{\mathcal{S}(\tau)} = (1, \hat{d}^{*(1)}_{\mathcal{S}(\tau)}, \dots, \hat{d}^{*(K-1)}_{\mathcal{S}(\tau)})^\top$, where

$$\hat{d}^{*(c)}_{\mathcal{S}(\tau)} = \frac{\|Y^*_{\mathcal{S}(\tau)} - \hat{Z}^{*(c)}_{\mathcal{S}(\tau)}\|^2}{\sum_{k=1}^{K}\|Y^*_{\mathcal{S}(\tau)} - \hat{Z}^{(k)}_{\mathcal{S}(\tau)}\|^2},\ c = 1, \dots, K-1. \tag{2.10}$$

Here, $\hat{Z}^{*(c)}_{\mathcal{S}(\tau)}(s) = \hat{\mu}^{(c)}(s) + \sum_{j=1}^{M_c} \hat{\xi}^{*(c)}_{\mathcal{S}(\tau),j}\ \hat{\varphi}^{(c)}_{\mathcal{S}(\tau),j}(s)$, and $\hat{\xi}^{*(c)}_{\mathcal{S}(\tau),j}$ can be obtained by a numerical approximation to $\langle Z^*_{\mathcal{S}(\tau)} - \hat{\mu}^{(c)}_{\mathcal{S}(\tau)}, \hat{\varphi}^{(c)}_{\mathcal{S}(\tau),j}\rangle_{\mathcal{S}(\tau)}$. To obtain $\{\hat{\varphi}^{(c)}_{\mathcal{S}(\tau),j}\}$ we simply decompose the covariance estimate $\hat{C}^{(c)}$ into blocks corresponding to the time domains $\mathcal{S}(\tau)$ and $\mathcal{T}(\tau)$ without re-estimating the covariance function, making the dynamical prediction step easy to implement for any given τ. We predict the cluster membership for the newly observed $Y^*_{\mathcal{S}(\tau)}$ by the posterior probability

$$\hat{P}(C = c \mid Y^*_{\mathcal{S}(\tau)}) = \frac{\exp\{\hat{\boldsymbol{\gamma}}_c^\top\, \hat{\mathbf{d}}^*_{\mathcal{S}(\tau)}\}}{\sum_{k=1}^{K} \exp\{\hat{\boldsymbol{\gamma}}_k^\top\, \hat{\mathbf{d}}^*_{\mathcal{S}(\tau)}\}},\ c = 1, \ldots, K-1, \tag{2.11}$$

and $\hat{P}(C = K \mid Y^*_{\mathcal{S}(\tau)}) = 1 - \sum_{c=1}^{K-1} \hat{P}(C = c \mid Y^*_{\mathcal{S}(\tau)})$.

3. Functional mixture prediction of future traffic flow trajectories. To accurately predict traffic flow trajectories under various traffic conditions, we combine the functional prediction model with functional clustering and classification methods. Given a newly observed trajectory $Z^*_{\mathcal{S}(\tau)}$ of the process Z as observed up to time τ, we propose a *functional mixture prediction model* to predict the trajectory of Z^* on the time interval $\mathcal{T}(\tau) = [\tau, T]$, denoted by $Z^*_{\mathcal{T}(\tau)}$ as,

$$E(Z^*_{\mathcal{T}(\tau)}(t) \mid Z^*_{\mathcal{S}(\tau)}) = \sum_{c=1}^{K} P(C = c \mid Z^*_{\mathcal{S}(\tau)})\, \tilde{Z}^{*(c)}_{\mathcal{T}(\tau)}(t)\,, \tag{3.1}$$

where $\tilde{Z}^{*(c)}_{\mathcal{T}(\tau)}(t) = E(Z^*_{\mathcal{T}(\tau)}(t) \mid Z^*_{\mathcal{S}(\tau)}, C = c)$ is the predictive function conditional on cluster $C = c$, and $P(C = c \mid Z^*_{\mathcal{S}(\tau)})$ is the posterior probability of cluster membership given the newly observed trajectory $Z^*_{\mathcal{S}(\tau)}$ up to time τ. The functional mixture prediction model (3.1), obtained by the law of iterated expectation on the random cluster membership variable C, $E\{E(Z^*_{\mathcal{T}(\tau)}(t) \mid Z^*_{\mathcal{S}(\tau)}, C)\}$, minimizes the expected risk, $E\left\{\mathcal{L}(Z^*_{\mathcal{T}(\tau)}, \widetilde{Z}^C_{\mathcal{T}(\tau)})\right\}$, where $\widetilde{Z}^C_{\mathcal{T}(\tau)}(t) = E(Z^*_{\mathcal{T}(\tau)}(t) \mid Z^*_{\mathcal{S}(\tau)}, C)$ and the loss function is defined as $\mathcal{L}(Z^*_{\mathcal{T}(\tau)}, \widetilde{Z}^c_{\mathcal{T}(\tau)}) = \int_{\mathcal{T}(\tau)} \left\{Z^*_{\mathcal{T}(\tau)}(t) - \tilde{Z}^{*(c)}_{\mathcal{T}(\tau)}(t)\right\}^2 dt$.

3.1. *Functional linear regression of traffic flow trajectories.* In a regression setting, the process $Z(s)$, for $s \in \mathcal{S}(\tau)$ denoted by $Z_{\mathcal{S}(\tau)}$, serves as the predictor function and the process $Z(t)$, for $t \in \mathcal{T}(\tau)$ denoted by $Z_{\mathcal{S}(\tau)}$, is the response function. The subspace projected functional clustering method developed above is well suited to identifying clusters in conjunction with functional prediction. For each cluster subspace, $Z^{(c)}$ is decomposed into $Z^{(c)}_{\mathcal{S}(\tau)}(s)$ and $Z^{(c)}_{\mathcal{T}(\tau)}(t)$ whose Karhunen-Loève expansions can be obtained such that $Z^{(c)}_{\mathcal{S}(\tau)}(s) = \mu^{(c)}(s) + \sum_{j=1}^{\infty} \xi^{(c)}_{\mathcal{S}(\tau),j}\, \varphi^{(c)}_{\mathcal{S}(\tau),j}(s)$ and $Z^{(c)}_{\mathcal{T}(\tau)}(t) = \mu^{(c)}(t) + \sum_{j=1}^{\infty} \xi^{(c)}_{\mathcal{T}(\tau),j}\, \varphi^{(c)}_{\mathcal{T}(\tau),j}(t)$, where the notations $\xi^{(c)}_{\mathcal{S}(\tau),j}$, $\varphi^{(c)}_{\mathcal{S}(\tau),j}$,

$\xi^{(c)}_{\mathcal{T}(\tau),j}$ and $\varphi^{(c)}_{\mathcal{T}(\tau),j}$ are defined analogously to those on the entire domain $\mathcal{U}$, but they correspond to the sub-domains $\mathcal{S}(\tau)$ and $\mathcal{T}(\tau)$.

We consider a functional linear regression model [e.g. Ramsay and Dalzell (1991), Müller, Chiou and Leng (2008)] conditional on cluster membership,

$$E(Z_{\mathcal{T}(\tau)}(t) \mid Z_{\mathcal{S}(\tau)}, C = c) = \mu^{(c)}(t) + \int_{\mathcal{S}(\tau)} \beta^{(c)}_{\tau}(s,t)\{Z_{\mathcal{S}(\tau)}(s) - \mu^{(c)}(s)\}ds, \tag{3.2}$$

for all $t \in \mathcal{T}(\tau)$. Here, given a fixed value of τ, assume the bivariate regression function $\beta^{(c)}_{\tau}(s,t)$ is smooth and square integrable, that is, $\int_{\mathcal{T}(\tau)}\int_{\mathcal{S}(\tau)} \beta^{(c)}_{\tau}(s,t)dsdt < \infty$. Under the smoothness assumption on the underlying random process, we further assume that the bivariate regression function $\beta^{(c)}_{\tau}(s,t)$ is a smooth function of τ for all s and t. Using the eigenbasis expansion for the regression coefficient function such that $\beta^{(c)}_{\tau}(s,t) = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \beta^{(c)}_{\tau,kj}\, \varphi^{(c)}_{\mathcal{S}(\tau),j}(s)\, \varphi^{(c)}_{\mathcal{T}(\tau),k}(t)$, model (3.2) can be expressed as

$$E(Z_{\mathcal{T}(\tau)}(t) \mid Z_{\mathcal{S}(\tau)}, C = c) = \mu^{(c)}(t) + \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \beta^{(c)}_{\tau,kj}\, \xi^{(c)}_{\mathcal{S}(\tau),j}\, \varphi^{(c)}_{\mathcal{T}(\tau),k}(t) \tag{3.3}$$

where $\xi^{(c)}_{\mathcal{S}(\tau),j} = \langle Z_{\mathcal{S}(\tau)} - \mu^{(c)}, \varphi^{(c)}_{\mathcal{S}(\tau),j}\rangle_{\mathcal{S}(\tau)}$ and $\beta^{(c)}_{\tau,kj} = E(\xi^{(c)}_{\mathcal{S}(\tau),j}\xi^{(c)}_{\mathcal{T}(\tau),k})/E\{(\xi^{(c)}_{\mathcal{S}(\tau),j})^2\}$ are the regression parameters to be estimated. Under the smoothness assumption on $\beta^{(c)}_{\tau}(s,t)$ along with τ, it follows that $\beta^{(c)}_{\tau,kj}$ is also smooth in τ for all k and j.

3.2. *Functional linear prediction model for future traffic flow.* Given $Z^*_{\mathcal{S}(\tau)}$, we aim to predict the values of $Z^*_{\mathcal{T}(\tau)}$. Suppose that the cluster structures $\mu^{(c)}$ and $\{\varphi^{(c)}_{\mathcal{S}(\tau),j}\}$ and the regression coefficients $\beta^{(c)}_{\tau,kj}$ are given. In practice, these estimates can be obtained from the functional clustering and the functional regression analysis using the historical or training data as described in Section 2. Then, the functional prediction model below is used to predict the unobserved trajectory conditional on a specific cluster:

$$E(Z^*_{\mathcal{T}(\tau)}(t) \mid Z^*_{\mathcal{S}(\tau)}, C = c) = \mu^{(c)}(t) + \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \beta^{(c)}_{\tau,kj}\, \xi^{*(c)}_{\mathcal{S}(\tau),j}\, \varphi^{(c)}_{\mathcal{T}(\tau),k}(t) \tag{3.4}$$

for all $t \in \mathcal{T}(\tau)$, where $\xi^{*(c)}_{\mathcal{S}(\tau),j} = \langle Z^*_{\mathcal{S}(\tau)} - \mu^{(c)}, \varphi^{(c)}_{\mathcal{S}(\tau),j}\rangle_{\mathcal{S}(\tau)}$ and will be obtained by numerical approximation.

Finally, given a partially observed trajectory $Z^*_{\mathcal{S}(\tau)}$, the unobserved trajectory $Z^*_{\mathcal{T}(\tau)}$ can be predicted by the functional mixture prediction model (3.1) using the results of the functional prediction model (3.4) in conjunction with the multiclass logit model (2.6). However, the components in these models remain to be estimated. The estimation procedure for the functional linear model is briefly summarized below.

3.3. *Estimation for functional mixture prediction models.* We note that the estimation of $\beta_{\tau,kj}$ in (3.3) and (3.4) can further be simplified using a simple linear regression approach [Müller, Chiou and Leng (2008)], such that

$$E(\xi^{(c)}_{\mathcal{T}(\tau),k} \mid \xi^{(c)}_{\mathcal{S}(\tau),j}) = \beta^{(c)}_{\tau,kj}\, \xi^{(c)}_{\mathcal{S}(\tau),j}\,,$$

for all pairs of (k, j). Therefore, functional linear regression can be decomposed into a series of simple linear regressions of functional principal component scores of the response processes in relation to those of the predictor processes.

For our predictions, given the cluster membership information and the subspace structure of each cluster, we estimate $\beta^{(c)}_{\tau}(s,t)$ in the functional linear regression model (3.2) based on the training data. Given the estimated principal component functions $\hat{\varphi}^{(c)}_{\mathcal{S}(\tau),j}(t)$ and $\hat{\varphi}^{(c)}_{\mathcal{T}(\tau),k}(t)$ and the principal component scores $\hat{\xi}^{(c)}_{\mathcal{S}(\tau),j}$ and $\hat{\xi}^{(c)}_{\mathcal{T}(\tau),k}$, the estimate of $\beta^{(c)}_{\tau,kj}$ can be obtained by

$$\tilde{\beta}^{(c)}_{\tau,kj} = \left\{(n_c - 1)\lambda^{(c)}_{\mathcal{S}(\tau),j}\right\}^{-1} \sum_{i=1}^{n_c} \left(\hat{\xi}^{(c)}_{\mathcal{S}(\tau),i,j} - \overline{\xi}^{(c)}_{\mathcal{S}(\tau),j}\right)\left(\hat{\xi}^{(c)}_{\mathcal{T}(\tau),i,k} - \overline{\xi}^{(c)}_{\mathcal{T}(\tau),k}\right), \tag{3.5}$$

where $\overline{\xi}^{(c)}_{\mathcal{S}(\tau),j}$ and $\overline{\xi}^{(c)}_{\mathcal{T}(\tau),k}$ are sample averages of $\hat{\xi}^{(c)}_{\mathcal{S}(\tau),i,j}$ and $\hat{\xi}^{(c)}_{\mathcal{T}(\tau),i,k}$, respectively.

To take advantage of smoothness in prediction as the value τ progresses, we further smooth the estimates $\{\tilde{\beta}^{(c)}_{\tau,kj}, \tau = \tau_1, \ldots, \tau_Q\}$ over τ to obtain the smooth estimates $\hat{\beta}^{(c)}_{\tau,kj}$, where Q is the number of time points at which predicting the future trajectory is of interest. Here, we use the local linear smoothing method with cross-validated bandwidth [see, e.g. Fan and Gijbels (1996)]. Accordingly, using $\hat{\beta}^{(c)}_{\tau,kj}$ and the estimates $\hat{\mu}^{(c)}(t)$ and $\hat{\varphi}^{(c)}_{\mathcal{T}(\tau),k}(t)$, we obtain the predicted trajectory conditional on cluster c by

$$\hat{Z}^{*(c)}_{\mathcal{T}(\tau)}(t) = \hat{E}(Z^*_{\mathcal{T}(\tau)}(t) \mid Y^*_{\mathcal{S}(\tau)}\,, C = c) = \hat{\mu}^{(c)}(t) + \sum_{j=1}^{M_c}\sum_{k=1}^{M_c} \hat{\beta}^{(c)}_{\tau,kj}\, \hat{\xi}^{*(c)}_{\mathcal{S}(\tau),j}\, \hat{\varphi}^{(c)}_{\mathcal{T}(\tau),k}(t) \tag{3.6}$$

for all $t \in \mathcal{T}(\tau)$. Here, M_c is determined by (2.7). Finally, combining the results of (3.6) with (2.11), we obtain the predicted unobserved traffic flow trajectory

$$\hat{Z}^*_{\mathcal{T}(\tau)}(t) = \hat{E}(Z^*_{\mathcal{T}(\tau)}(t) \mid Y^*_{\mathcal{S}(\tau)}) = \sum_{c=1}^{K} \hat{Z}^{*(c)}_{\mathcal{T}(\tau)}(t)\, \hat{P}(C = c \mid Y^*_{\mathcal{S}(\tau)})\,. \tag{3.7}$$

3.4. *Implementation algorithm of functional mixture predictions.* Suppose there is a newly observed trajectory $\{(t^*_j, Y^*(t^*_j)); t^*_j < \tau\}$, denoted by $Y^*_{\mathcal{S}(\tau)}$ for short. The algorithm for functional mixture prediction that combines the functional classification procedure with the functional prediction model is summarized as follows.

Step 1 *Identification of cluster subspaces.* Perform the subspace-projected functional clustering procedure according to criterion (2.2) to identify cluster subspaces, $\{\hat{\mu}^{(c)}, \{\hat{\varphi}_k^{(c)}\}_{k=1,\ldots,M_c}\}$, $c = 1, \ldots, K$, based on the training data set as discussed in Sections 2.1 and 2.3.

Step 2 *Model fitting based on the historical or training data.*

(i) *Obtain the multiclass logit model for cluster membership distributions.* Obtain from Step 1 the regression coefficient estimates $\hat{\boldsymbol{\gamma}}_c$ in (2.8).

(ii) *Fit the functional linear regression model.* Fit the cluster-specific functional linear regression models and obtain the regression coefficient estimates $\hat{\beta}_{\tau,kj}^{(c)}$ as a smoothed version of (3.5).

Step 3 *Prediction of the future traffic flow trajectory for a new and partially observed $Y_{\mathcal{S}(\tau)}^*$ conditional on clusters.*

(i) *Predict the posterior membership probability of $Y_{\mathcal{S}(\tau)}^*$ associated with each cluster.* Calculate the relative L^2 distances $d_{\mathcal{S}(\tau)}^{*(c)}$ for the given $Y_{\mathcal{S}(\tau)}^*$ in (2.10) and obtain the posterior probability $\hat{Pr}(C = c \mid Y_{\mathcal{S}(\tau)}^*)$ in (2.11).

(ii) *Predict the unobserved trajectory $Y_{\mathcal{T}(\tau)}^*$ conditioning on each of the clusters.* Obtain the cluster-specific functional prediction model fit $\hat{E}(Y_{\mathcal{T}(\tau)}^* \mid Y_{\mathcal{S}(\tau)}^*, C = c)$ in (3.6).

Step 4 *Prediction of traffic flow trajectory by the functional mixture prediction model.* Calculate the predicted trajectory $\hat{E}(Y_{\mathcal{T}(\tau)}^* \mid Y_{\mathcal{S}(\tau)}^*)$ in (3.7) using the results of $\hat{Pr}(C = c \mid Y_{\mathcal{S}(\tau)}^*)$ and $\hat{E}(Y_{\mathcal{T}(\tau)}^* \mid Y_{\mathcal{S}(\tau)}^*, C = c)$ and obtain the bootstrap prediction intervals.

Details in constructing the bootstrap prediction intervals in Step 4 are provided in Supplementary Material B.

4. Analysis of traffic flow patterns. The sample data set of daily traffic flow trajectories from Section 1.1 is divided into a training data set (70 days) and a test data set (14 days) to examine the predictive performance of our model. Clusters of the traffic flow patterns from the training data are identified based on subspace projection using the proposed subspace-projected functional clustering method according to criterion (2.2). The implementation of the *functional forward testing procedure* of Li and Chiou (2011) leads to the choice of 3 clusters. Table 1 summarizes the empirical probabilities of rejecting the null hypotheses for $K = 2, 3, 4$, based on 200 bootstrap samples. The p-values with reference to the pre-determined level of significance 0.05, adjusted for multiple comparisons, indicate that when $K = 2$ and 3, the clusters are all significantly distinct, while Clusters 1 and 4

when $K = 4$ are not significantly different in terms of the mean functions and the eigenspaces.

The cluster memberships displayed in Figure 3 show that Cluster 1 contains mostly weekends (left panel), with 90% being holidays including weekends (right panel). Cluster 2 completely comprises weekdays including Mondays through Thursdays. Cluster 3 comprises mostly weekdays, especially Fridays (left panel). The mean functions of the three clusters and the overall trajectories are displayed in Figure 4. While Cluster 1 has a higher mean traffic flow rate than the other two clusters, Clusters 2 and 3 have relatively close mean flow rates in terms of shape and magnitude until 11:00, and they diverge thereafter with a higher mean flow

TABLE 1

Empirical probabilities of rejecting the null hypotheses H_{01} and H_{02}, respectively, based on 200 bootstrap samples.

Number of clusters	Clusters	$H_{01} : \mu^{(c)} = \mu^{(d)}$	$H_{02} : S^{(c)} = S^{(d)}$
2	1 vs. 2	0.000	0.010
3	1 vs. 2	0.000	0.005
	1 vs. 3	0.000	0.005
	2 vs. 3	0.005	0.015
4	1 vs. 2	0.000	0.055
	1 vs. 3	0.000	0.030
	1 vs. 4	0.155	0.025
	2 vs. 3	0.000	0.160
	2 vs. 4	0.000	0.035
	3 vs. 4	0.000	0.000

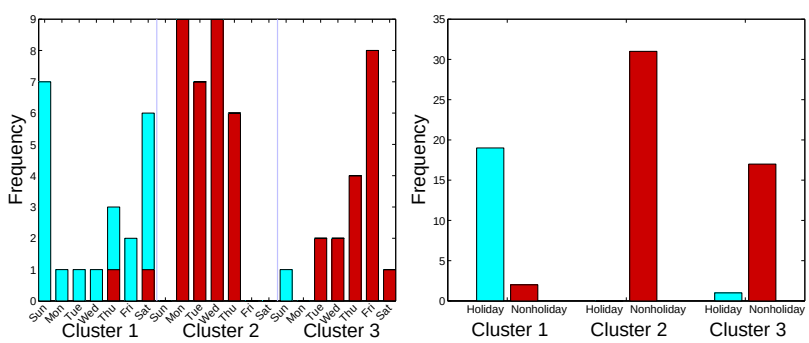


Fig 3: *Frequency plots of cluster labels by days of the week (left panel) and by non- and holidays (right panel) for Clusters 1, 2 and 3 (left, middle and right groups).*

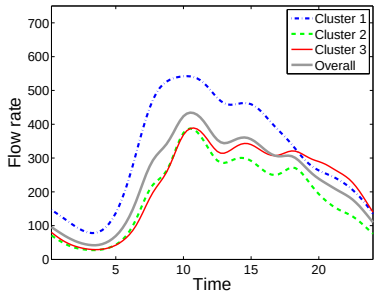


Fig 4: *Overall and cluster-specific mean functions of the training data of daily traffic flow rates.*

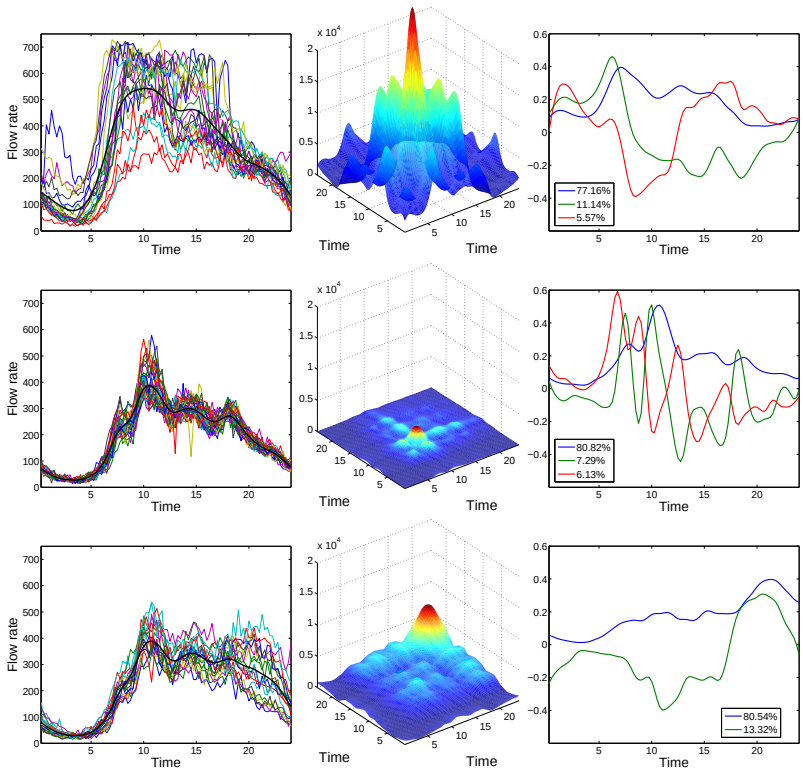
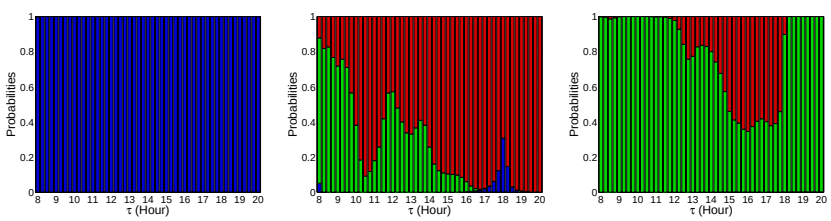


Fig 5: *Estimated mean functions (left column) superimposed on the observed trajectories, covariance functions (middle column) and the corresponding eigenfunctions (right column) of Clusters 1–3 (from top to bottom) based on the training data of daily traffic flow trajectories.*

rate in Cluster 3. The observed trajectories along with the corresponding covariance functions and leading eigenfunctions are shown in Figure 5. The variability of Cluster 1 is higher than the other two clusters while Cluster 2 has the lowest variability. The peak flow rate in Cluster 1 lasts from 07:00 to 17:00 and the three leading principal component functions explain 77.16%, 11.14% and 5.57% of total variability. The trajectories in Cluster 2 have a relatively uniform pattern with the major peak flow rate at around 11:00. Cluster 3 indicates a high variability of flow rates occurring after 18:00. The mean integrated prediction errors are defined as $n_c^{-1} \sum_{i=1}^{n_c} T^{-1} \int_0^T \left(\hat{Z}_i^{(c)}(t) - Y_i(t)\right)^2 dt$, where $\hat{Z}_i^{(c)}(t) = \hat{\mu}^{(c)}(t) + \sum_{j=1}^{M_c} \hat{\xi}_{ij}^{(c)} \hat{\varphi}_j^{(c)}(t)$, $Y_i(t)$ is the observed trajectory and n_c is the number of trajectories in Cluster c. These are 327.3, 78.8 and 122.4 for Clusters 1-3, respectively. Prediction using the overall trajectories without clustering, in contrast, returned an error of 300.5, indicating that there is a huge reduction in prediction errors when heterogeneity of cluster patterns are taken into account.

The model fits of the multiclass logistic regression listed in Supplementary Material C are used in predicting the unobserved trajectory for an up-to-date and partially observed trajectory. Given a newly observed trajectory from the test data up to the current time τ, to consider different flow patterns, we predict the posterior probabilities for each of the associated clusters by functional classification based on the multiclass logistic regression model in (2.11), using the fitted regression coefficients in (2.8) with the relative L^2 distances (2.10) as the covariate. The posterior probabilities for some test samples are illustrated in Figure 6 with the values of τ progressing from 08:00 to 20:00 by 15-min intervals. In Figure 6(a), the predicted membership probabilities of *Test sample 1* degenerate to one for all values of τ. The associated predicted trajectories are shown in Figures 2(a)-2(c). The time-varying prediction intervals, which are wider closer to τ and taper off toward the



(a) Test sample 1 (b) Test sample 2 (c) Test sample 3

Fig 6: *The predicted cluster membership distribution for Clusters 1–3 (indicated in blue, green and red) as a function of the 'current time' τ (per 15 min) for samples from the test data based on the trajectories observed up to τ.*

end, depend on Cluster 1's variability pattern as illustrated in Figure 5 (top panels). In contrast, Figure 6(b) for *Test sample 2* indicates a more complex situation where the predicted membership distributions change with τ, with the associated predicted trajectories shown in Figures 2(d)-2(f). In this case, using the early trajectory information up to $\tau = 8:00$ may lead to misclassification, which makes it difficult to predict its future trajectory accurately. This issue is resolved as τ moves onwards. The wider prediction bands with τ at 12:00 and 16:00, in comparison with that at 8:00, reflect the fact that the variability of Cluster 3's traffic flow trajectories is larger in the afternoon as illustrated in Figure 5 (bottom panels). Figure 6(c) indicates that the posterior cluster membership probabilities of *Test sample 3* alternates between Clusters 2 and 3, owing to certain similarities in these two cluster patterns, and the predicted cluster membership remains with Cluster 3 after around 18:00. Given that the actual cluster memberships are unknown, the accuracy of functional classification for the up-to-date and partially observed trajectories in the test data will be investigated via a simulation study in Section 6.

5. Traffic flow prediction. In predicting the unobserved traffic flow trajectories, we also investigate the effects on the prediction performance of the interval length prior to time τ and the future interval length after it. To this end, we define $\mathcal{S}(\tau;\omega) = [\max(0,\tau-\omega),\tau]$, where ω is the length of the known interval to be used in prediction calculations and $\mathcal{T}(\tau;\kappa) = [\tau, \min(\tau+\kappa, T)]$, where κ is the length of the unknown interval to be predicted from time τ onwards. In the test data, given a sample Y_i^* observed up to time τ, denoted by $Y^*_{i,\mathcal{S}(\tau)}$, we define the mean integrated prediction error ($MIPE$) as the the performance measure of predicting $Y^*_{i,\mathcal{T}(\tau)}$. This is expressed as $MIPE(\tau,\omega,\kappa) = m_p^{-1}\sum_{i=1}^{m_p}\kappa^{-1}\int_0^\kappa\left\{\widehat{Z}^*_{i,\mathcal{T}(\tau)}(t) - Y^*_{i,\mathcal{T}(\tau)}(t)\right\}^2 dt$, where $Y^*_{i,\mathcal{T}(\tau)}(t) = Y_i^*(\tau+t)$, $\widehat{Z}^*_{i,\mathcal{T}(\tau)}(t)$ is obtained by (3.7) and m_p is the number of trajectories in the test data. For ease of comparisons across different values of τ, let $\tau_s = \max(0,\tau-\omega)$ and $\tau_e = \min(\tau+\kappa, T)$, for $\omega > 0$ and $\kappa > 0$. We define the total mean integrated prediction error ($TMIPE$) for the overall prediction performance by

$$TMIPE(\omega,\kappa) = \int_{\tau_s}^{\tau_e} MIPE(\tau,\omega,\kappa)\, d\tau\,, \tag{5.1}$$

where τ_s and τ_e are the smallest and the largest values, respectively, selected with respect to the times, τ, on the domain $[0,T]$. In this study, $T = 24$ (hours) and we set $\tau_s = 8$ and $\tau_e = 20$. For notational convenience, we let $\kappa^* = 24-\tau$ and $\omega^* = \tau$, $\tau_s \le \tau \le \tau_e$, such that κ^* denotes the interval length from the current time to the end of the day and ω^* denotes the maximal length of the past trajectory information available for prediction.

5.1. *Results and comparisons of traffic flow prediction.* In this study, we investigate the prediction performance by comparing the following methods:

- FP: Functional prediction based on functional linear regression using the same setting described in Section 3.1 but without considering clusters of different traffic flow patterns;
- FMP_H: Functional prediction using the proposed functional mixture prediction model except that the posterior membership probabilities (2.6) degenerate to zero or one such that $\sum_{c=1}^{K} P(C = c \mid Z^*_{\mathcal{S}(\tau)}) = 1$ (where the subscript H reflects the so-called *hard* classification);
- FMP_S: Functional prediction using the proposed functional mixture prediction model (where the subscript S reflects the so-called *soft* classification or probabilistic classification).

To examine prediction performance under various situations, we consider a wide range of values τ along with various values of ω and κ as defined in $\mathcal{S}(\tau;\omega)$ and $\mathcal{T}(\tau;\kappa)$. Table 2 indicates that the proposed FMP_S is robust, generally outperforming the other two (FP and FMP_H) under various values of ω and κ. Figure 7 (left panel) indicates that small values of ω (1,2,4) give a similar performance and it is not surprising that the performance for larger values of κ is worse. For fixed values of κ (right panel), TMIPE as a function of ω generally shows a positive slope as it moves away from the origin, with a minimum at $\omega = 5$ (for $\kappa = 4, 8, \omega^*$) or $\omega = 6$ (for $\kappa = 1$) from Table 2. The trend is relatively flat, but becomes steeper when $\omega = \omega^*$. This discrepancy is more pronounced with increasing κ. A possible explanation is that the flow trajectory patterns in Clusters 2 and 3 are close in shape and

TABLE 2
Performance comparisons for FP, FMP_H and FMP_S based on TMIPE ($\times 10^3$) under various values of κ and ω.

	κ \ ω	1	2	3	4	5	6	ω^*
FP	1	4.12	4.92	4.84	4.92	5.30	5.68	5.82
	4	7.42	7.71	7.69	8.10	8.60	9.06	8.95
	8	9.92	10.26	10.35	10.79	11.27	11.71	11.65
	κ^*	12.34	12.79	12.94	13.41	13.91	14.36	14.30
FMP_H (3 clusters)	1	3.48	3.22	3.20	3.33	3.53	3.52	3.14
	4	5.00	4.62	4.73	4.91	5.20	5.26	4.79
	8	8.88	8.44	8.48	8.68	8.97	9.07	8.81
	κ^*	12.24	11.82	11.87	12.06	12.36	12.47	12.33
FMP_S (3 clusters)	1	3.31	3.26	3.07	2.93	2.87	2.80	2.81
	4	4.37	4.14	4.05	3.98	3.80	3.86	4.18
	8	5.80	5.64	5.54	5.49	5.41	5.61	6.68
	κ^*	7.97	7.94	7.84	7.86	7.71	7.89	9.95

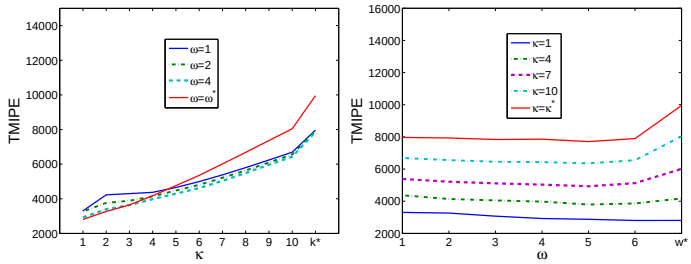


Fig 7: *Performance comparisons for FMP_S, based on $TMIPE$ (5.1), displayed as a function of κ (left) with ω fixed at 1, 2, 4, and ω^* and as a function of ω (right) with κ fixed at 1, 4, 7, 10 and κ^*.*

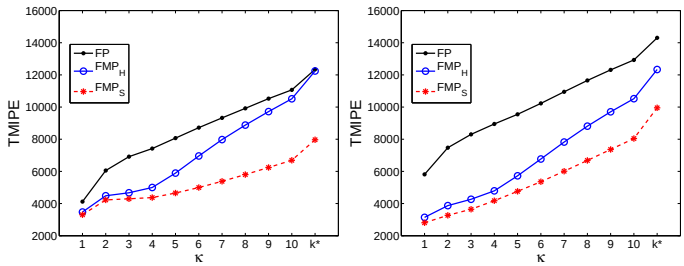


Fig 8: *Performance comparisons for FP, FMP_H and FMP_S, based on $TMIPE$ (5.1), displayed as a function of κ, with $\omega = 1$ (left) and $\omega = \omega^*$ (right).*

magnitude until noon and diverge thereafter and thus, using larger ω with more past information may not significantly improve the overall prediction accuracy. In the literature, Sentürk and Müller (2010) considered the length of past data to be used for prediction and suggested the optimal length using a data-adaptive criterion that minimizes the absolute prediction error. Our empirical results also suggest the use of a data-adaptive criterion to choose the length of past data. Additionally, comparisons for the prediction performance between the methods with fixed values of ω as illustrated in Figure 8 (with $\omega = 1$ on the left and $\omega = \omega^*$ on the right) reinforce the conclusion that FMP_S outperforms FP and FMP_H. In addition to the 3-cluster prediction performance illustrated above, results of the 2- and 4-cluster prediction performances are illustrated in Supplementary Material C for comparisons. These results also support our choice of the 3-cluster model, which outperforms the 2- and 4-cluster models.

5.2. *Comparisons with other methods.* We compare the functional mixture prediction approach with an existing method that could also be fitted into our func-

tional mixture prediction framework. One possible approach is to treat the unobserved future trajectory for a partially observed trajectory as missing in the entire trajectory. That is, we may replace (3.6) in Step 3 (ii) with the model

$$\hat{Z}^{*(c)}_{\mathcal{T}(\tau)}(t) = \hat{E}(Z^{*}_{\mathcal{T}(\tau)}(t) \mid Y^{*}_{\mathcal{S}(\tau)}, C = c) = \hat{\mu}^{*(c)}(t) + \sum_{j=1}^{M_c} \hat{\xi}^{*(c)}_j \, \hat{\varphi}^{*(c)}_j(t) \tag{5.2}$$

for all $t \in \mathcal{T}(\tau)$, where the estimated mean function $\hat{\mu}^{*(c)}(t)$ and the estimated eigenfunctions of the covariance kernel $\hat{\varphi}^{*(c)}_j(t)$ of cluster c are obtained by the training data set in the clustering step with the corresponding domain $\mathcal{T}(\tau)$. The key step is to estimate the functional principal component scores, $\hat{\xi}^{*(c)}_j$, which cannot be obtained easily since the trajectory is only partially observed. An existing method that can deal with this situation makes use of the expectation of the posterior distribution in Proposition 1 of Zhou, Serban and Gebraeel (2011), assuming that the prior distribution of the scores is Gaussian, for an application to degradation modeling. This formula coincides with the conditional expectation approach in equation (4) of Yao, Müller and Wang (2005) under Gaussian assumptions, although they are different in terms of statistical inference. We term this method the Functional Principal Component Prediction (FPCP) approach. We apply FPCP to the proposed functional mixture prediction algorithm including the cases with and without clustering/classification considerations for comparisons, including FPCP, FPCP_H and FPCP_S that are parallel to FP, FMP_H and FMP_S. The results shown in Table 3, in comparison with the results in Table 2, indicate that the functional mixture prediction approach in conjunction with functional linear regression outperforms FPCP

TABLE 3

Performance comparisons for FPCP, FPCP_H and FPCP_S based on TMIPE ($\times 10^3$) under various values of κ and ω.

	κ \ ω	1	2	3	4	5	6	ω^*
FPCP	1	14.30	13.94	13.63	12.62	10.49	8.48	7.69
	4	17.43	16.70	15.37	13.57	11.90	11.11	10.77
	8	17.97	16.98	15.96	14.93	14.10	13.73	13.30
	κ^*	19.27	18.59	17.88	17.13	16.50	16.19	15.66
FPCP_H (3 clusters)	1	7.83	8.62	8.90	8.90	10.57	10.67	2.85
	4	10.72	11.35	11.99	11.63	11.21	9.75	4.80
	8	12.57	12.62	12.28	11.94	12.05	11.65	9.19
	κ^*	14.35	14.36	14.24	14.13	14.14	13.94	12.48
FPCP_S (3 clusters)	1	6.26	7.29	7.54	7.73	9.41	9.44	3.07
	4	8.73	9.84	10.79	10.86	10.61	9.86	4.87
	8	10.13	11.16	11.66	11.65	11.60	11.28	8.80
	κ^*	12.12	12.95	13.23	13.12	13.14	13.14	12.37

approach. Additional results for the 2- and the 4-cluster models are also provided in Supplementary Material C for comparison.

6. Simulation. We implement a Monte Carlo simulation to evaluate the performance of the functional clustering and classification procedures as well as the functional prediction accuracy. We simulate the scenario of the real traffic flow trajectories analyzed in the previous sections. We generate a training data set and a test data set for each simulation run using the estimated results of the 3-cluster traffic flow trajectories as the true models with a total of 100 simulation replicates. The numbers of curves n_c are 21, 31 and 18 for Clusters 1–3 in each training data set and are 3, 8 and 3 in each test data as in the previous analysis. The synthetic curves of cluster c are generated by the model $y_i^{(c)}(t_j) = \mu^{(c)}(t_j) + \sum_{j=1}^{\widetilde{M}_c} \xi_{ij}^{(c)} \phi_j^{(c)}(t_j) + \varepsilon_{ij}^{(c)}$, for $i = 1, \dots, n_c$, where $\xi_{ij}^{(c)}$ are normal random variates with a mean of zero and variance $\lambda^{(c)}$ and the random measurement errors $\varepsilon_{ij}^{(c)}$ are independent and follow a normal distribution with a mean of zero and variance $\sigma_{(c)}^2$. The recording times $t_j = j/4$ for $j = 1, \dots, 96$, mimic the 15-min recording time interval. The quantities $\mu^{(c)}$, $\phi_j^{(c)}$, $\lambda_j^{(c)}$ and $\sigma_{(c)}^2$ use the model estimates of our real traffic flow data analysis. The numbers of components $\widetilde{M}_c$ are determined by the numbers of the estimated $\lambda_j^{(c)}$ that are strictly positive. Further details of the simulated models regarding the underlying functions $\mu^{(c)}$ and $\phi_j^{(c)}$, along with a sample of synthetic trajectories are displayed in Supplementary Material D. The clustering results of this simulated sample including the estimated mean function and the eigenfunctions along with the covariance functions, are also illustrated.

The average clustering error rates are 6.48% (with standard error 1.28%), 1.68% (0.45%) and 8.39% (2.01%) for Clusters 1–3 based on the 100 simulated training data sets. The accuracy of classification for the future trajectory to be predicted for a partially observed trajectory in the test data depends on the values τ, the "current" time observed thus far. The average classification error rate decreases with τ, ranging from 27.5% at 8:00 to 7.7% at 20:00, implying that prediction accuracy increases with τ. Additional details regarding accuracy of clustering and classification are compiled in Supplementary Material D.

The prediction performances based on the proposed functional mixture prediction (FMP) approach are summarized in Table 4. The method FMP_S^* is the same as FMP_S, apart from that FMP_S^* assumes the cluster/classification memberships are known, serving as the gold standard for prediction performance comparisons. The results clearly demonstrate that FMP_S outperforms FP and FMP_H, with relatively smaller values of TMIPE and the associated standard errors indicating that FMP_S has better prediction accuracy and is quite robust. In FMP_S, the prediction errors using $\omega = 1$ and $\omega = 2$ are close to each other and perform the best under

TABLE 4
Average TMIPE ($\times 10^3$) (with s.e. in parentheses) for FP, FMP_H, FMP_S and FMP_S^* under various values of κ and ω based on 100 simulation replicates.

	κ \ ω	1	2	4	6	ω^*
FP	1	2.99 (0.03)	3.44 (0.05)	3.90 (0.07)	4.40 (0.08)	4.87 (0.11)
	4	5.53 (0.09)	5.72 (0.10)	6.18 (0.12)	6.70 (0.14)	7.18 (0.17)
	8	6.77 (0.13)	6.92 (0.15)	7.33 (0.17)	7.80 (0.20)	8.26 (0.21)
	κ^*	7.50 (0.18)	7.67 (0.20)	8.08 (0.22)	8.52 (0.24)	8.97 (0.25)
FMP_H	1	2.88 (0.04)	3.17 (0.05)	3.42 (0.07)	3.58 (0.08)	3.61 (0.09)
	4	4.98 (0.11)	5.14 (0.12)	5.44 (0.15)	5.65 (0.17)	5.77 (0.18)
	8	6.00 (0.15)	6.12 (0.16)	6.42 (0.19)	6.66 (0.22)	6.89 (0.24)
	κ^*	7.18 (0.27)	7.47 (0.33)	8.02 (0.50)	8.41 (0.57)	8.80 (0.53)
FMP_S	1	2.80 (0.04)	3.09 (0.05)	3.30 (0.07)	3.49 (0.07)	3.49 (0.07)
	4	4.88 (0.09)	4.91 (0.10)	5.26 (0.11)	5.44 (0.13)	5.42 (0.13)
	8	5.90 (0.14)	5.99 (0.14)	6.34 (0.16)	6.38 (0.17)	6.59 (0.18)
	κ^*	6.90 (0.19)	7.07 (0.19)	7.25 (0.22)	7.58 (0.23)	7.86 (0.25)
FMP_S^*	1	2.60 (0.03)	2.81 (0.04)	3.03 (0.05)	3.13 (0.06)	3.13 (0.06)
	4	4.11 (0.06)	4.18 (0.07)	4.40 (0.09)	4.48 (0.10)	4.46 (0.11)
	8	4.57 (0.09)	4.59 (0.09)	4.74 (0.11)	4.78 (0.12)	4.73 (0.13)
	κ^*	4.58 (0.10)	4.57 (0.10)	4.65 (0.12)	4.68 (0.13)	4.62 (0.13)

various values of κ. The optimal selection of ω appears to be different from those obtained from our real traffic flow data. Although the generated data based on the model estimates may reach a high level of realism to traffic flow data, they may not be able to capture the entire data features such as outlying curves that could influence the prediction performance. In addition, classification errors of the partially observed trajectory may also play a role in prediction. We also compare the FMP method with the FPCP approach. The simulation results are listed in Supplementary Material D. The results suggest larger values of ω for minimal prediction errors. The intuition behind these results is that FPCP treats the partial trajectory to be predicted as missing values, especially when the data are more homogeneous within clusters and contain less outlying curves as in the simulated data. Overall, the results demonstrate that the proposed FMP approach outperforms the FPCP approach.

7. Concluding remarks and discussions. This study presents a methodological framework for uncovering traffic flow patterns and predicting traffic flow. The proposed functional data approaches, including classification and prediction, identify clusters with similar traffic flow patterns, facilitating accurate prediction of daily traffic flow. Although motivated by the subject of traffic flow prediction, the proposed methodology is generally applicable and transferable to the analysis and prediction of any longitudinally collected functional data, such as city electric-

ity usage or degradation studies in manufacturing systems. The empirical results demonstrate that our proposed method, functional mixture prediction, which combine functional prediction with probabilistic functional classification, can work reasonably well to predict traffic flow. We conclude that taking traffic flow patterns into account can greatly improve prediction performance as long as the traffic flow patterns can be satisfactorily identified.

In the literature of intelligent transportation systems, conditional expectation is commonly used as the measure of traffic flow prediction/forecast of a future trajectory at a future time point or short period. However, it may be interesting to consider probabilistic forecasts [Gneiting (2010)], which take the form of probability distributions over future trajectories. A probabilistic forecast may engender a new way of thinking about traffic flow prediction, which may give a better account of uncertainty in potential flow trajectories. In this study, our focus was on predicting a future trajectory in the form of conditional expectation for an up-to-date and partially observed trajectory. Under the functional mixture prediction framework, a mixture of predictive distributions of the potential trajectories could instead serve as an ensemble for probabilistic forecasting. However, substantial efforts would be needed to accomplish the goal of probabilistic forecasting for traffic flow trajectories.

In addition to predictive accuracy, the real-time feature of traffic flow information is important in traffic management. Given that the components of our proposed model are estimated based on historical data, as in the training data, the proposed method also serves as a real-time prediction approach for predicting the future unobserved traffic flow trajectory for a partially observed flow trajectory. The fact that real-time information is quickly and easily updated will facilitate the establishment of effective reporting systems for traffic flow prediction. Furthermore, this article discussed single-detector traffic prediction, a category crucial in supporting demand forecasting as required in practice by operational network models. Future research might extend to multiple-detector traffic prediction and will be important in working toward the goal of better road network management.

Acknowledgments. The author would like to thank Mr. Y.-C. Zhang for computational assistance and Dr. J.A.D. Aston for carefully reading the manuscript. The author would also like to express his gratitude to the Editor, Associate Editor and referees, whose comments and suggestions helped improve the paper.

SUPPLEMENTARY MATERIAL

This supplement contains a PDF (FMPsupp.pdf) which is divided into four sections. **Supplement A**: Selection of the number of clusters; **Supplement B**: Boot-

strap prediction intervals; **Supplement C**: Additional results for traffic flow prediction; **Supplement D**: Additional simulation details and results.

REFERENCES

ABRAHAM, C., CORNILLON, P. A., MATZNER-LØBER, E., and MOLINARI, N. (2003). Unsupervised curve clustering using B-splines. *Scand. J. Statist.* **30** 581–595.

ANTOCH, J., PRCHAL, L., DE ROSA, M. and SARDA, P. (2010). Electricity consumption prediction with functional linear regression using spline estimators. *J. Appl. Statist.* **37** 2027–2041.

ASTON, J. A. D., CHIOU, J.-M. and EVANS, J. P. (2010). Linguistic pitch analysis using functional principal component mixed effect models. *J. Roy. Statist. Soc. Ser. C* **59** 297–317.

BISHOP, C. M. and LASSERRE, J. (2008). Generative or discriminative? Getting the best of both worlds. *Bayesian Statist.* **8** 3–24.

ÇETINER, B. G., SARI, M. and BORAT, O. (2010). A neural network based traffic-flow prediction model. *Math. and Comput. Appl.* **15** 269–278.

CHEN, H. and GRANT-MULLER, S. (2001). Use of sequential learning for short-term traffic flow forecasting. *Transportation Research Part C: Emerging Technologies* **9** 319–336.

CHIOU, J.-M. and LI, P.-L. (2007). Functional clustering and identifying substructures of longitudinal data. *J. Roy. Statist. Soc. Ser. B* **69** 679–699.

CHIOU, J.-M. and LI, P.-L. (2008). Correlation-based functional clustering via subspace projection. *J. Amer. Statist. Assoc.* **103** 1684–1692.

CHIOU, J.-M. and MÜLLER, H. G. (2007). Diagnostics for functional regression via residual processes. *Comput. Statist. Data Anal.* **51** 4849–4863.

CHIOU, J.-M. and MÜLLER, H. G. (2009). Modeling hazard rates as functional data for the analysis of cohort lifetables and mortality forecasting. *J. Amer. Statist. Assoc.* **104** 572–585.

CHIOU, J.-M., MÜLLER, H. G. and WANG, J. L. (2003). Functional quasi-likelihood regression models with smooth random effects. *J. Roy. Statist. Soc. Ser. B* **65** 405–423.

COFFEY, N. and HINDE, J. (2011). Analyzing time-course microarray data using functional data analysis - A review. *Statist. Appl. in Genetics and Molecular Biol.* **10** no. 23.

CUEVAS, A., FEBRERO, M. and FRAIMAN, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Comput. Statist.* **22** 481–496.

DAWID, A. P. (1976). Properties of diagnostic data distributions. *Biometrics* **32** 647–658.

D'AMATO, V. , PISCOPO, G. and RUSSOLILLO, M. (2011) The mortality of the Italian population: Smoothing techniques on the Lee-Carter Model. *Ann. Appl. Statist.* **5** 705–724.

DI, C-Z, CRAINICEANU, C. M., CAFFO, B. S. and PUNJABI, N. M. (2009). Multilevel functional principal component analysis. *Ann. Appl. Statist.* **3** 458–488.

FAN, J. and GIJBELS, I. (1996), *Local Polynomial Modelling and its Applications.* London: Chapman and Hall.

FERRATY, F. and VIEU, P. (2006) *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, New York.

GAO, H. O. and NIEMEIER, D. A. (2008). Using functional data analysis of diurnal ozone and NOx cycles to inform transportation emissions control. *Transportation Research Part D: Transport and Environment* **13** 221–238.

GNEITING, T. (2008). Editorial: Probabilistic forecasting. *J. Roy. Statist. Soc. Ser. A* **171** 319–321.

HYNDMAN, R. J. and ULLAH, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Comput. Statist. Data Anal.* **51** 4942–4956.

JAMES, G. M. and SUGAR, C. A. (2003). Clustering for sparsely sampled functional data. *J. Amer. Statist. Assoc.* **98** 397–408.

KIRBY, H. R., WASTON, S. M. and DOUGHERTY, M. S. (1997). Should we use neural networks or statistical models for short-term motorway traffic forecasting? *Int. J. Forecasting* **13** 43–50.

LI, P.-L. and CHIOU, J.-M. (2011). Identifying cluster number for subspace projected functional data clustering. *Comput. Statist. Data Anal.* **55** 2090–2103.

LÓPEZ-PINTADO, S. and ROMO, J. (2006). Depth-based classification for functional data. In *DIMACS Ser. Math. Theo. Comput. Sci.* (R. Liu, R., Serfling, and D.L. Souvaine, eds.) **72** 103–119.

MA, P. and ZHONG, W. (2008). Penalized clustering of large scale functional data with multiple covariates. *J. Amer. Statist. Assoc.* **103** 625–636.

MCCULLAGH, P. and NELDER, J. A. (1989). Generalized Linear Models. Chapman and Hall, London.

MÜLLER, H. G. (2005), Functional modelling and classification of longitudinal data (with discussion). *Scand. J. Statist.* **32** 223–246.

MÜLLER, H. G. (2008). Functional modeling of longitudinal data. In *Longitudinal Data Analysis* (G. Fitzmaurice, M. Davidian, G. Verbeke and G. Molenberghs, eds.) Wiley & Sons, Inc, 223–252.

MÜLLER, H. G., CHIOU, J. M. and LENG, X. (2008). Inferring gene expression dynamics via functional regression. *BMC Bioinformatics* **9** 60.

MÜLLER, H. G. and ZHANG, Y. (2005). Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories. *Biometrics* **61** 1064–1075.

NG, A. Y. and JORDAN, M. I. (2002). On discriminative vs. generative classifier: A comparison of logistic regression and naive Bayes. *Neural Information Processing System* **2** 841–848.

OKUTANI, I. and STEPHANIDES, Y. J. (1984). Dynamic prediction of traffic volume through Kalman filtering theory. *Transportation Research Part B: Methodological* **18** 1–11.

RAMSAY, J. O. and DALZELL, C. J. (1991). Some tools for functional data analysis (with Discussion). *J. Roy. Statist. Soc. Ser. B 53* 539-572.

RAMSAY, J. and SILVERMAN, B. (2002), *Applied Functional Data Analysis*. Springer, New York.

RAMSAY, J. and SILVERMAN, B. (2005), *Functional Data Analysis*, 2nd ed. Springer, New York.

RICE, J. (2004). Functional and longitudinal data analysis: perspectives on smoothing. *Statist. Sinica* **14** 631–647.

SERBAN, N. and WASSERMAN, L. (2005). CATS: Clustering after transformation and smoothing. *J. Amer. Statist. Assoc.* **100** 990–999.

SMITH, B. L. and DEMETSKY, M. J. (1997). Short-term traffic flow prediction: Neural network approach. *Transportation Research Record: Journal of the Transportation Research Board* **1453** 98–104.

SMITH, B. L., WILLIAM, B. M. and OSWALD, R. K. (2002). Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies* **10** 303–321.

STATHOPOULOS, A. and KARLAFTIS, M. G. (2003). A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies* **11** 121–135.

SUN, H., LIU, H. X., XIAO, H., HE, R. R. and RAN, B. (2003). Use of local linear regression model for short-term traffic forecasting. *Transportation Research Record* **1836** 143–150.

VLAHOGIANNI, E. I., KARLAFTIS, M. G. and GOLIAS, J. C. (2008) Temporal evolution of short-term urban traffic flow: a nonlinear dynamic approach. *Computer-Aided Civil and Infrastructure Engineering* **22** 326–334.

WILLIAMS, B. M. and HOEL, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of Transportation Engineering* **129** 664–672.

XIE, Y., ZHANG, Y. and YE, Z. (2007). Short-term traffic volume forecasting using Kalman filter with discrete wavelet decomposition. *Computer-Aided Civil and Infrastructure Engineering* **22** 326–334.

XUE, J. H. and TITTERINGTON, D. M. (2008). Comment on "On discriminative vs. generative classifier: A comparison of logistic regression and naive Bayes." *Neural Process Letters* **28** 169–187.

YAO, F., MÜLLER, H. G. and WANG, J. L. (2005a). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590.

YAO, F., MÜLLER, H. G. and WANG, J. L. (2005b). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33** 2873–2903.

YIN, H., WONG, S.C., XU, J. and WONG, C. K. (2002). Urban traffic flow prediction using fuzzy-neural approach. *Transportation Research Part C: Emerging Technologies* **10** 85–98.

ZHANG, Y. and YE, Z. (2008). Short-term traffic flow forecasting using fuzzy logic system methods. *Journal of Intelligent Transportation System* **12** 102–112.

ZHAO, X., MARRON, J. S. and WELLS, M. T. (2004). The functional data analysis view of longitudinal data. *Statist. Sinica* **14** 789–808.

ZHENG, W., LEE, D.-H. and SHI, Q. (2006). Short-term freeway traffic prediction: Bayesian combined neural network approach. *Journal of Transportation Engineering* **132** 114–121.

ZHOU, R. R., SERBAN, N. and GEBRAEEL, N. (2011). Degradation modeling applied to residual lifetime prediction using functional data analysis. *Annals of Applied Statistics* **5** 1586–1610.

INSTITUTE OF STATISTICAL SCIENCE
ACADEMIA SINICA
128 SEC. 2 ACADEMIA ROAD
NANKANG, TAIPEI 11529
TAIWAN
E-MAIL: jmchiou@stat.sinica.edu.tw