An Online Change-Point-Based Model for Traffic Parameter Prediction

Gurcan Comert and Anton Bezuglov, *Member, IEEE*

***Abstract*—This paper develops a method for predicting traffic parameters under abrupt changes based on change point models. Traffic parameters such as speed, flow, and density are subject to shifts because of weather, accidents, driving characteristics, etc. An intuitive approach of employing the hidden Markov model (HMM) and the expectation-maximization (EM) algorithm as change point models at these shifts and accordingly adapting the autoregressive-integrated-moving-average (ARIMA) forecasting model is formulated. The model is fitted and tested using publicly available 1993 I-880 loop data. It is compared with basic and mean updating forecasting models. Detailed numerical experiments are given on several days of data to show the impact of using change point models for adaptive forecasting models.**

***Index Terms*—Change point models, hidden Markov model (HMM), time-series autoregressive integrated moving average (ARIMA), traffic prediction.**

I. Introduction

TRAFFIC parameter prediction is a critical component for many intelligent transportation systems (ITS) applications, such as Advanced Traveler Information Systems, real-time route guidance, and emergency response systems planning. ITS target safety, environment, and economic competitiveness with research in implementations of new technologies to transportation networks. For instance, accurate and robust traffic parameter prediction can be used in real-time routing to find the fastest way to a destination and to avoid congestion for on-time delivery. Ultimately, it can lead to reliable transportation networks where goods are delivered at a minimum cost of time, fuel, and emissions. For such applications, the accuracy of the predicted input traffic parameters is critical. Contrarily, traffic data can exhibit nonlinearities due to various factors such as traffic accidents, inclement weather conditions, and demand surges, which adversely affect the performance of the parameter prediction models. This paper proposes an online change-point-based (OCPB) model for traffic prediction that can effectively handle nonlinearities in the data by monitoring and switching the forecasting models [1].

In the literature, researchers have proposed a number of methods for short-term traffic parameter prediction. These methods can be broadly classified as parametric and nonparametric. Parametric models include time-series models [2]–[8] and Kalman filtering (KF) models [9]–[12]. Nonparametric models consist of nonparametric regression methods [2], [13] and neural networks (NNs) [14]–[21]. For detailed reviews of the various short-term forecasting models and their critical aspects, the reader is referred to the work by Vlahogianni *et al.* [22], Smith and Demetsky [23], Smith *et al.* [2], and Vlahogianni and Karlaftis [18]. In summary, these studies report that the nonparametric techniques have superior performance over simple time-series models. However, they require higher computational power and demand more data. Major advantages of the parametric approach are well-developed theory and ability to interpret the parameters. Generally, parametric prediction methods first train a model using historical data to estimate parameters and then test it on different data. The main assumption in this parametric modeling approach is unchanged process characteristics (e.g., mean and variance), which may affect the prediction accuracy. Since traffic is subject to occasional abrupt disturbances (e.g., incidents and weather) or congestion level (see [24] and [25]) that can potentially change the underlying dynamics of the data generation process, robust forecasting models are needed, which incorporate the changes when they occur.

Although adaptive models are available in the literature, there are many points to improve in existing methods [26], particularly when accidents happen and whenever traffic congestion sets in and dissolves [27], [28]. For example, Zhang and Rice [28] adapt the coefficient of the linear regression model to estimate freeway travel times. Tavana and Mahmassani [29] indicate that simple continuum models that use static equilibrium speed–density relationships cannot capture the dynamics of the system. Instead, they develop dynamic speed–density relations using bivariate time-series models. Huynh *et al.* [30] extend the work of Tavana and Mahmassani [29] by implementing an adaptive model where the model parameters are periodically updated online (e.g., every 5 min) based on the prevailing traffic conditions. The numerical experiments on synthetic data sets show that adaptive methods provide better estimates compared with nonadaptive methods. They also investigate the impact of the update frequency and the length of past observations (which are used in calculating the updates) on the accuracy. The numerical results suggest that an update frequency of

Manuscript received June 18, 2012; revised October 8, 2012; December 20, 2012; and March 27, 2013; accepted April 25, 2013. Date of publication May 30, 2013; date of current version August 28, 2013. This work was supported by the 2012 National Nuclear Security Administration Summer Faculty Research Fellowship through the Office of Research, Benedict College. The Associate Editor for this paper was L. Li.

G. Comert is with the Physics and Engineering Department, Benedict College, Columbia, SC 29204 USA (e-mail: comertg@benedict.edu).

A. Bezuglov is with the Mathematics and Computer Science Department, Benedict College, Columbia, SC 29204 USA (e-mail: bezuglova@benedict.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2013.2260540

10 min using 45 min of past data provide the lowest error. The results of Huynh *et al.* [30] are further corroborated by Qin and Mahmassani [31], where actual sensor data are used for the numerical evaluations.

More recently, hybrid methods have been used to obtain adaptive models. Cetin and Comert [32] combine autoregressive integrated moving average (ARIMA) model with expectation-maximization (EM) and cumulative summation (CUSUM) algorithms to update the intercept. The EM-ARIMA model in [32] is compared with the OCPB model in this paper. Qi [33] applies hidden Markov models (HMMs) for traffic condition prediction. Adaptive fuzzy logic and Bayesian adaptive credit assignment algorithms in [34] and [35] are applied to select the best NN flow predictor at a given period. Min and Wynter [35] present an extended time-series model that includes spatial and temporal correlations for full-time range-network level-flow forecast in urban networks. Tan *et al.* [36] offer an aggregate model that combines exponential smoothing, moving average (MA), and ARIMA based on an NN and calls for the best model for a given period. Kindzerske and Ni [37] use nonparametric regression with an NN for traffic condition forecasting. In addition to the mean level forecasting, there have been new studies on the volatility to generate confidence bounds [38], [39]. Karlaftis and Vlahogianni [40] develop a fractional integrated process to uncover the long-term structure of the time series and apply the model on traffic volume prediction. The study fits an autoregressive (AR) fractionally integrated MA with fractionally integrated volatility process (FIGARCH) model and states that the model better represents the transportation time series compared with the classical ARIMA, although a very large time series is needed for these models. Dimitiou *et al.* [34] give an adaptive fuzzy rule-based meta-optimized model for flow forecasting at urban networks. The model is provided with offline, online, univariate, and multivariate options. Model parameters are optimized by a genetic algorithm, and the results with classical ARIMA, KF, offline univariate, multivariate, online univariate, and multivariate models are compared. The best accuracy is obtained from the online multivariate model with the highest R^2 value.

Applications of change point models in other fields are vastly common particularly in finance and economics [41]–[46]. Mostly, models involve HMM-based change point models and consist of mainly level change assumption. Change point models are classified as offline (retrospective) and online (prospective) analyses. Offline analysis focuses on multiple change points and distributional properties. Online analysis looks for the detection of the first-time change point (see, for instance, Moreno *et al.* [47] and Giron *et al.* [48]). Although the offline version [47]–[49] has been extensively studied, Bayesian online/sequential change point models have also recently received attention with applications to computer science, climatic time series, deoxyribonucleic acid sequencing, and medical applications [49]–[53]. For example, Saatci *et al.* [54] and Turner *et al.* [55] utilize the Bayesian detection model that is offered by Adams and MacKay [56] with Gaussian processes to generate nonparametric time-series models that adapt to abrupt changes. Similarly, Garnett *et al.* [57] develop a sequential Bayesian prediction with nonstationary covariance functions in the Gaussian process that model change points.

TABLE I
EXAMPLE OF LANE BLOCKING ACCIDENTS IN THE DATA

Date	Start Time	End Time	Location	Duration
18-Feb	6 : 52 : 00 AM	6 : 59 : 00 AM	4	0 : 07 : 00
18-Feb	7 : 35 : 00 AM	7 : 44 : 00 AM	4	0 : 09 : 00
18-Feb	8 : 57 : 00 AM	9 : 16 : 00 AM	4	0 : 19 : 00
18-Feb	6 : 15 : 00 PM	6 : 19 : 00 PM	4	0 : 04 : 00
3-Mar	7 : 07 : 00 AM	8 : 21 : 00 AM	4	1 : 14 : 00
3-Mar	7 : 04 : 00 AM	9 : 28 : 00 AM	17	2 : 24 : 00

TABLE II
DATA USED IN NUMERICAL EXPERIMENTS

Day	Location	Usage	Changes
02-Mar AM/PM	Loop 04	ARIMA fit	No
03-Mar AM	Loop 04	HMM train/ARIMA fit	Major
03-Mar PM	Loop 17	HMM train/ARIMA fit	Major
16-Feb AM/PM	Loop 04	testing	No/Major
18-Feb AM/PM	Loop 04	testing	Major/Major
19-Feb AM/PM	Loop 04	testing	No/Major
23-Feb AM/PM	Loop 04	testing	Major/Major
24-Feb AM/PM	Loop 07	testing	No/Minor
25-Feb AM/PM	Loop 04	testing	No/Major
26-Feb AM/PM	Loop 12	testing	No/Major
01-Mar AM/PM	Loop 04	testing	No/Minor
05-Mar AM/PM	Loop 20	testing	No/Major
15-Mar AM/PM	Loop 04	testing	Major/No
16-Mar AM/PM	Loop 12	testing	No/Major
17-Mar AM/PM	Loop 05	testing	No/Major
18-Mar AM/PM	Loop 04	testing	Major/Minor
19-Mar AM/PM	Loop 04	testing	No/Major

II. PROBLEM STATEMENT AND DATA

The main objective in this paper is to treat the data with unknown break points. Between any two change points (i.e., parameter states), the characteristics of the process are assumed to stay the same. In such case, continuously updating parameters (or periodically) may not be necessary as the model parameters within a state are not statistically changing. This paper focuses on illustrating the value of the change point approach in short-term traffic parameter prediction. In this framework, change point detection/clustering algorithms of HMMs and the EM are employed to detect shifts, update the mean levels, and switch the forecasting models. Time-series ARIMA models are utilized as the main forecasting tool. For implementation in the field, recalibration intervals of both detection and forecasting model parameters are needed to be investigated. However, they are not dealt within this paper. For numerical experiments, publicly available loop data set collected by the California Partners for Advanced Transit and Highways program on I-880 for the Freeway Service Patrol Project (http://ipa.eecs.berkeley.edu/~pettyk/FSP/) is used. This database contains travel speeds, occupancies, and traffic volumes from 170 loop detectors at every 30 s collected over 46 weekdays. The loop data were collected in the morning (5:00 A.M.–10:00 A.M.) and afternoon (2:00 P.M.–8:00 P.M.) periods (see Table I). In addition to the loop data, detailed records of all incidents including incident type, location, severity, and duration were kept for the data collection period. Table II shows the total of 16 days of 1-min

TABLE III
SHAPIRO WILK TEST FOR NORMALITY

State(AM)	p-val	s size	State(PM)	p-val	s size
S1	0.019	55	S1	0.457	17
S2	0.001	81	S2	0.762	76
S3	0.002	38	S3	0.197	57
S4	0.002	120	S4	0.003	204

aggregated travel speed data that are used for model building and testing. Loop-detector numbers that are closest to the incidents are also shown in the table. The testing of all models is performed on days with no incident occurring near these loops and on days with accidents. Very broadly, changes that occurred on the testing days are also shown in the table. Although the changes do not show uniform effects on different days, to have a better idea about the performance of the proposed model, they are given considering the drop of average speed to state 2 or 3 (minor change), state 1 (major change), and no change. These states are defined next in Section III-A.

Major assumptions for the proposed model follow, i.e., constant mean and variance for the data generation process on a normal driving day without an incident, constant mean and variance within states on a day with incident, and underlying normal distribution for the detection algorithms (i.e., HMM and EM). The normality is tested with Shapiro Wilk in R for state-by-state A.M. and P.M. sections of the data where p values are given in Table III. More than half of the tests reject the null hypothesis of the filtered section that is normally distributed. The normal probability plots are also checked. Overall, the samples look approximately normal. Although the normality is partially supported by the tested samples in this paper, the assumption may not always hold in practice.

The rest of this paper is organized as follows: In Section III, the proposed OCPB model with its components is introduced. Section IV presents the detailed numerical analysis that evaluate the prediction performance of the OCPB model against the intercept adaptive (i.e., EM-ARIMA) and basic prediction models [i.e., random walk (RW), simple MA, ARIMA(1,0,2), and ARIMA(1,1,1)]. Finally, Section V summarizes the findings and addresses possible future research directions.

III. METHODOLOGY

Here, the OCPB traffic parameter prediction model is explained. The OCPB model is a hybrid model that consists of three major components, i.e., change point algorithm (HMM) that is used to detect shifts and switch forecasting models, mean level update (EM algorithm), and main forecasting tool (time-series ARIMA models). Each of these methods and their roles in the OCPB model are thoroughly described next.

A. HMM

The HMM is a fundamental model that is applicable to the problem of change point detection [33], [45], [46]. The HMM is a stochastic model that contains two variables X_t and Z_t that are related to each specific point in time t. X_t denotes an unobservable probabilistic state variable where Z_t denotes an

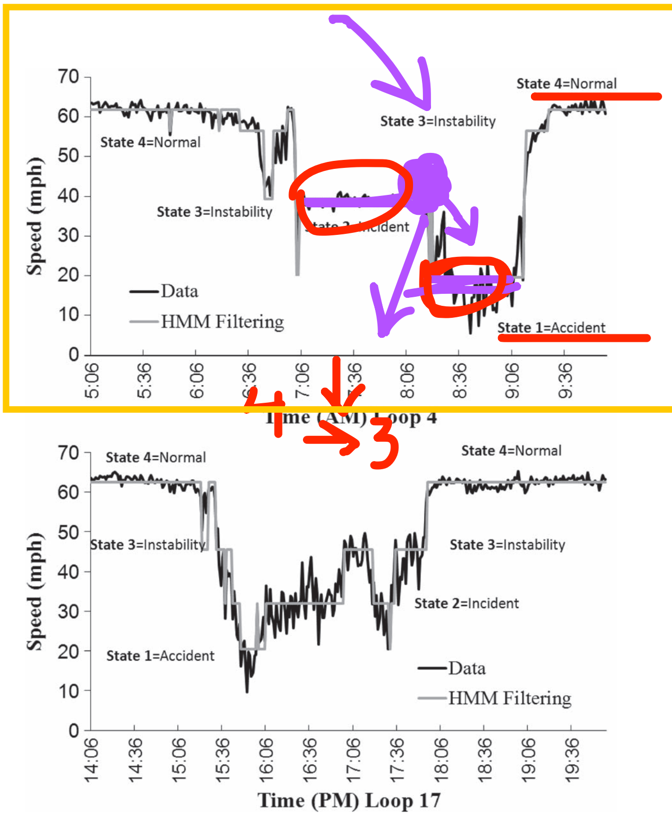

Fig. 1. HMM filtering of March 3 showing different states of the process.

observable variable. The state variable X_t can be discrete or continuous. When it is discrete, the number of distinct values (or the number of states in the model) can be manually or algorithmically determined [58]. The models typically demonstrate better accuracy when the number of states increase. However, models with too many states become overconfident about the training data set, and they are not efficient to be used on other data sets.

For this paper, to show the usefulness of the HMM in traffic parameter prediction, the authors only manually checked the impact of the number of states on the HMM performance and did not notice significant accuracy improvement with more than four states. Fig. 1 shows the output of the four-state HMM on the training data set, where the states are Major Accident, Minor Incident, Instability, and Normal Driving: $X_t = \{\text{Acc}, \text{Inc}, \text{Inst}, \text{Norm}\}$. Probability distribution $P(X_t|X_{t-1})$ determines the change points.

The observable variable Z_t represents aggregated 1-min average traffic speed at time t. Z_t is assumed to be normally distributed for each state $X_t : P(Z_t|X_t = i) = N(\mu_i, \sigma_i)$. Under these assumptions, two conditions need to be identified to apply the HMM to change point detection [59].

1) Given model λ and a sequence of observations $Z_{1\ldots t}$, maximize the likelihood of the assigned states $X_{1\ldots t}$.
2) Given a set of observations $Z^i_{1\ldots t}$, find an HMM with λ^* that will maximize the likelihood of set $L(Z;\lambda) = P(Z_{1\ldots t}|\lambda)$, i.e., training HMM.

The usual procedure for solving the aforementioned problems is to calculate forward $\alpha_t(i)$ and backward $\beta_t(i)$ variables using the forward–backward algorithm [60]. These variables represent the probabilities to finish (forward) or to start (backward) of an HMM at state i at time t given observations $Z_{1\ldots T}$, i.e.,

$$\alpha_t(i) = P(Z_{1\ldots t}, X_t = i|\lambda)$$
$$\beta_t(i) = P(Z_{t+1\ldots T}, X_t = i|\lambda).$$

The forward–backward algorithm suggests an iterative procedure to calculate the variables, i.e.,

$$\alpha_1(i) = \pi_i \cdot b_i(Z_1)$$
$$\alpha_{t+1}(i) = \left(\sum_{j=1}^{N} \alpha_t(j) \cdot a_{ij}\right) \cdot b_i(Z_{t+1})$$
$$\beta_T(i) = 1$$
$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} \cdot b_j(Z_{t+1}) \cdot \beta_{t+1}(j)$$

where π_i is the probability of the model to start at state i : $\pi_i = P(X_1 = i)$. In this algorithm, b_i is the probability of observing Z_1 at state $i : b_i = P(Z_t = Z_t|X_t = i)$. Finally, a_{ij} represents the switching behavior of the HMM, i.e., $a_{ij} = P(X_{t+1} = i|X_t = j)$. To solve the first problem and find the state sequence with the highest likelihood, variable $\gamma_t(i)$ (a normalized product of $\alpha_t(i)$ and $\beta_t(i)$) is introduced, i.e.,

$$\gamma_t(i) = \frac{\alpha_t(i) \cdot \beta_t(i)}{\sum_{j=1}^{N} \alpha_t(j) \cdot \beta_t(j)}. \quad (1)$$

As a result, the state sequence with the highest likelihood at each time t is $X_t = \arg, \max_{1\leq i\leq N} \gamma_t(i)$. In addition, the forward–backward algorithm is used to train HMM (finding the best fitting HMM) using $\xi_t(i, j) = P(X_t = i, X_{t+1} = j, Z|\lambda)$, which is calculated by

$$\xi_t(i, j) = \frac{\alpha_t(i) \cdot a_{ij} \cdot b_j(Z_{t+1}) \cdot \beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_t(i) \cdot a_{ij} \cdot b_j(Z_{t+1}) \cdot \beta_{t+1}(j)}. \quad (2)$$

Both variables $\gamma_t(i)$ and $\xi_t(i, j)$ allow calculating expected values for HMM parameters $\bar{\pi}_i = \gamma_1(i)$, $a_{\bar{i}j}$, and $b_{\bar{j}}(k)$ given the observed data. It has been demonstrated [61] that, when these expected values are used as the new HMM parameters, its likelihood improves. If this reestimation step is iteratively used (i.e., the Baum–Welch algorithm), then within several iterations, the likelihood reaches a local maximum and the HMM is considered as trained. This procedure is, in fact, a special application of the EM algorithm [62].

As with most local optimization algorithms, the key to the success is proper initialization. The initial HMM parameters in this paper are set to 10 mi/h for Major Accident, 20 mi/h for Minor Incident, 40 mi/h for Instability, and 60 mi/h for Normal Driving (see Table IV). The choice of the initial mean values in the table facilitates the interpretation of the model and enables relating the state definitions after the trained HMM.

In this paper, HMMs are separately fitted on A.M. and P.M. data. Data collected on March 3 at loops 4 and 17 contains all four states. Thus, March 3 data are used to train an A.M. HMM and a P.M. HMM (see Fig. 2). Table IV summarizes $P(Z_t|X_t)$ values for each of the two models. Since the change points occur relatively rare, both A.M. and P.M. models are conservative, which is evident in Fig. 3, which shows $P(X_t|X_{t-1})$ of the A.M. model. The probability of staying at the same state is very high for Major Accident, Minor Incident, and Normal Driving (96.3%, 4.6%, and 97.6%, respectively). It has the smallest

TABLE IV
HMM PARAMETERS BEFORE AND AFTER TRAINING

HMM	μ_1	σ_1	μ_2	σ_2	μ_3	σ_3	μ_4	σ_4
Initial	10.0	1.0	20.0	1.0	40.0	1.0	60.0	1.0
AM	19.6	53.5	39.3	2.7	56.5	13.0	61.8	2.6
PM	20.5	24.99	31.9	25.96	45.5	39.09	62.5	1.54

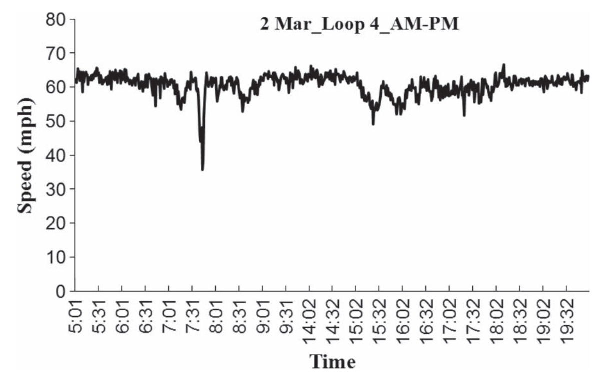

Fig. 2. One-minute aggregated speed of March 2, which is used for ARIMAs and EM-ARIMA.

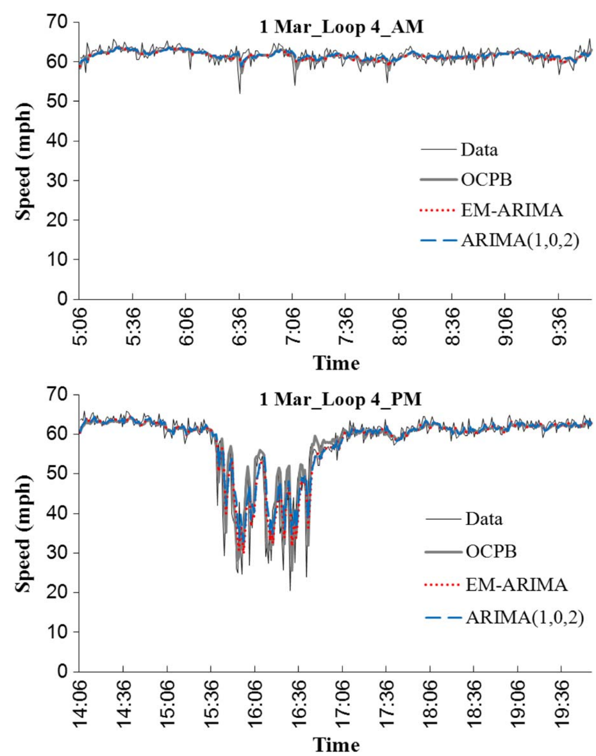

Fig. 3. Traffic speed predictions from Loop 4 on March 1 A.M.–P.M. with all three methods.

value for Instability (a still very high 81.3%), after which the state usually switches to Normal Driving (17.1%), i.e.,

$$P(X_t|X_{t-1}) = \begin{pmatrix} 0.963 & 0.037 & 0.0 & 0.0 \\ 0.014 & 0.946 & 0.04 & 0.0 \\ 0.001 & 0.015 & 0.813 & 0.171 \\ 0.0 & 0.0 & 0.024 & 0.976 \end{pmatrix}. \quad (3)$$

To find whether there is a switch point at time t, the trained HMM looks at five immediate preceding speed values $Z_{-1\ldots-5} = \{Z_{t-1}, Z_{t-2}, Z_{t-3}, Z_{t-4}, Z_{t-5}\}$ and applies the forward–backward algorithm [59] to find the posterior probability of which state the HMM would be at time t: $P(X_t|Z_{-1\ldots-5})$. State i maximizes the predicted state $P(X_t = i|Z_{-1\ldots-5})$, which is utilized to choose the proper ARIMA model. To account for the potential effect of the number of historical data points for change point detection, HMMs are also trained on 10 and 20 min of data with no significant difference in the outcomes.

B. EM Algorithm

In the OCPB model, the EM algorithm is utilized for updating the mean of the process. The EM algorithm is often used to estimate the parameters of mixture models or models with latent variables [62]. The notation for EM is adopted from [63]. Given N sample points that are generated by a mixture of two distributions as in (4), the EM algorithm can be applied to determine the parameters of these two distributions $\boldsymbol{\theta} = [\theta_1 = (\mu_1, \sigma_1), \theta_2 = (\mu_2, \sigma_2), \pi]$. The first step of the EM algorithm specifies initial values for the parameters. In the expectation step, the algorithm computes responsibilities γ_i (i.e., the probability of an observation belonging to Y_2) for each data point. Using the calculated responsibilities, it then computes the five parameters in the maximization step. The iterations continue until the likelihood function converges. The convergence of a basic EM algorithm is slow; however, there are techniques for accelerating the convergence [64]. Simple equations pertaining to the EM are given below. First, the probability density of Y is written as a mixture, i.e.,

$$Y = (1 - \Delta)Y_1 + \Delta Y_2 \quad (4)$$

where $Y_1 \sim N(\mu_1, \sigma_1^2)$, $Y_2 \sim N(\mu_2, \sigma_2^2)$, and $\Delta \in 0, 1$, with $P(\Delta = 1) = \pi$, and

$$g_Y(y) = (1 - \pi)\phi_{\theta_1}(y) + \pi\phi_{\theta_2}(y) \quad (5)$$

where $\phi_\theta(x)$ denotes normal density. For a data set of N points, the log-likelihood function can be written as follows:

$$l(\theta, Z) = \sum_{i=1}^{N} \ln\left[(1 - \pi)\phi_{\theta_1}(y_i) + \pi\phi_{\theta_2}(y_i)\right] \quad (6)$$

where $\boldsymbol{\theta} = [\theta_1 = (\mu_1, \sigma_1), \theta_2 = (\mu_2, \sigma_2), \text{and } \pi]$, and Z represents the data points. The analytical maximization of (6) is difficult; however, if the observation is known to belong to Y_2 (i.e., with latent variable $\Delta_i = 1$; otherwise, $\Delta_i = 0$), the log-likelihood can be written as in (7), and $\Delta_i = 1$ s can be estimated by (8), i.e.,

$$l(\theta; \Delta, Z) = \sum_{i=1}^{N} [(1 - \Delta_i)\ln(1 - \pi)\phi_{\theta_1}(y_i) + \Delta_i \ln \pi\phi_{\theta_2}(y_i)] \quad (7)$$

$$\gamma_i(\theta) = E(\Delta_i \mid \theta, Z) = P(\Delta_i = 1 \mid \theta, Z) \quad (8)$$

$$\hat{\mu}_{t+1} = \sum_{t-1}^{t} \left[\hat{\mu}_1(1 - \gamma_i) + \hat{\mu}_2\gamma_i\right]/3. \quad (9)$$

Given N data points that are assumed to be generated by mixture of two distributions (i.e., normal and abnormal speeds), the EM algorithm is applied to determine the distribution parameters and responsibilities. The number of mixtures could be decided based on another algorithm as in [65]. N data points constitute the main input to the algorithm. To see the impact of the sample size, prediction performances of the EM algorithm with $N = 10$ and $N = 20$ samples randomly selected from March 2 are checked. In particular, with accident data, EM with $N = 10$ is able to provide better mean values. Thus, ten sample normal-speed data points are used to represent the expected behavior, and the most recent three observations are used for the real-time status. Technically, the EM algorithm provides the real-time estimation of the process mean at each time point from (9), which is subsequently used in the OCPB and EM-ARIMA models for prediction.

C. ARIMA Forecasting

ARIMA models are commonly used forecasting tools in the literature [45], [66]. These models handle time-correlated forecasts by exploiting the autocorrelation structures in time-series data. ARIMA models and their extensions have been applied to the traffic data before [67]. The notation for a basic ARIMA(p, d, q) model can be written as follows:

$$\phi_p(B)(1 - B)^d Z_t = \theta_q(B)a_t \quad (10)$$

where

B	backshift operator defined by $B^j Z_t = Z_{t-j}$;
$\phi_p(B) = (1 - \phi_1 B - \ldots - \phi_p B^p)$	AR polynomial of order p;
$\theta_q(B) = (1 - \theta_1 B - \ldots - \theta_q B^q)$	MA polynomial of order q;
$(1 - B)^d$	polynomial for ordinary differencing;
Z_t	value of the series at time t;
a_t	white noise with zero mean and variance σ^2.

A typical ARIMA model fitting consists of plotting the data, transforming the data if needed, identifying potential models (i.e., determining p, d, and q), estimating model parameters, and checking diagnostics. In this paper, all of these steps are carried out in the statistical package R. After examining the autocorrelation function (ACF) and the partial ACF, a set of alternative models are identified with different model orders p, d, and q. To determine the best model among the alternatives, diagnostics and Akaike information criteria (AIC) are used. Finally, the model with good diagnostics and good AIC value is chosen.

TABLE V
BASIC ARIMA AND EM-ARIMA PARAMETERS

Day	Models	ϕ	θ_1	θ_2	Intercept
MAR-02 AM	ARIMA(1,0,2)	0.95	0.67	0.09	61.87
MAR-02 PM	ARIMA(1,0,2)	0.96	0.52	0.07	60.63
MAR-02 AM	ARIMA(1,1,1)	0.70	0.97	-	-
MAR-02 PM	ARIMA(1,1,1)	0.17	0.72	-	-

TABLE VI
STATE-BY-STATE PARAMETERS FOR THE OCPB MODEL

Day	State	Models	ϕ	θ_1	θ_2	Intercept
AM	1	ARIMA(1,0,2)	0.90	0.49	0.12	20.64
AM	2	ARIMA(1,0,2)	0.02	-0.20	-0.57	39.39
AM	3	ARIMA(1,0,2)	0.75	0.34	-0.17	55.21
AM	4	ARIMA(1,0,2)	0.94	0.70	0.04	61.90
PM	1	ARIMA(1,0,2)	0.71	0.06	0.36	19.52
PM	2	ARIMA(1,0,2)	0.81	0.51	0.14	31.64
PM	3	ARIMA(1,0,2)	0.71	0.18	0.05	45.39
PM	4	ARIMA(1,0,2)	0.95	0.46	0.22	62.50

A total of 12 (i.e., A.M.–P.M.) ARIMA models are obtained for forecasting. Two ARIMA(1,0,2) (or ARMA(1,2)) models are fitted to March 2 A.M.–P.M.-Loop 4, i.e., a normal day data. Fig. 2 shows the speed measurements on March 2, which are found to be stationary; therefore, no differencing is needed. In order to compare the performance of the OCPB model with integrated ARMA forecasts, two ARIMA(1,1,1) models with good diagnostics are also fitted to March 2 A.M.–P.M. sections. The estimated parameters for these ARIMA models are shown in Table V. The remaining eight ARIMA(1,0,2) models are fitted to each state in the OCPB model. Since it represents all four states (see Fig. 1), March 3 (i.e., A.M.-Loop 4 and P.M.-Loop 17) is used for the ARIMA parameter estimation for the OCPB model. All of the state-by-state models that are fitted to ARIMA(1,0,2) give pretty good diagnostics and AIC values, which is desirable as the model orders remain the same when parameters are changing. For state-by-state ARIMA fitting, segments of the same states are linked together. The impact of mixing on the model fitting is kept relatively limited as March 3 data consists of long runs of particular states, and only one day of data is used for model building. Estimated parameter values for the OCPB model are given in Table VI.

For ARIMA forecasts, AR(∞) representation in (11) is adopted. Using (11), forecasting formulas for ARIMA(1,0,2) and ARIMA(1,1,1) can be derived as in (12) and (13), respectively. Intercept μ and model parameters (e.g., ϕ, θ) are kept constant in these ARIMA models. EM-ARIMA predictions are produced using (12), as well as (9), to update μ, i.e.,

$$\psi(B)(Z_t-\mu)=a_t\psi(B)=\phi(B)(1-B)^d/\theta(B) \quad (11)$$

$$\hat{Z}_{t+1}=\left(1-\sum_{i=1}^{\infty}\psi_i\mu\right)+Z_t\sum_{i=1}^{\infty}B\psi_i \quad (12)$$

where $\psi_0=1$, $\psi_1=\phi-\theta_1$, $\psi_2=\psi_1\theta_1-\theta_2$, and $\psi_i=\psi_{i-1}\theta_1-\psi_{i-2}\theta_2$ for $i>2$, and

$$\hat{Z}_{t+1}=Z_t\sum_{i=1}^{\infty}B\psi_i \quad (13)$$

where $\psi_0=1$, $\psi_1=1+\phi-\theta_1$, $\psi_2=\psi_1\theta_1-\phi$, and $\psi_i=\psi_{i-1}\theta_1$ for $i>2$.

OCPB predictions are calculated from (14) where the conditional probability of each state given five previous speed observations weighs the forecasts from corresponding states obtained from (12). Model intercepts are updated based on (9) in the OCPB model as well, i.e.,

$$\hat{Z}_{t+1}=\sum_{k=1}^{4}P(X_t=k|Z_{-1\ldots-5})\hat{Z}_{t+1},k. \quad (14)$$

IV. NUMERICAL ANALYSIS

Here, numerical results are presented for the performance of the OCPB model in (14). The model is compared with a mean-adapting EM-ARIMA (12) and (9), a basic ARIMA(1,0,2) (12), a model with differencing ARIMA(1,1,1) (13), and two naive models of RW (15) and simple MA (16) with five observations. All of the six models are used to generate 1-, 5-, and 10-lag ahead forecasts for the 14 days of speed data that are given in Table II, i.e.,

$$\hat{Z}_{t+1}=Z_t \quad (15)$$

$$\hat{Z}_{t+1}=\sum_{i=0}^{4}Z_{t-i}/5. \quad (16)$$

On the basis of the two error measures that are reported (i.e., mean square error (MSE) $=\sum_{t=1}^{n}[Z_t-\hat{Z}_t]^2/n$ and mean absolute error (MAE) $=\sum_{t=1}^{n}|Z_t-\hat{Z}_t|/n$), Table VII shows the summary of one-step-ahead prediction errors and maximum percent improvements from compared models for various days and loop-detector locations (i.e., closest loop to the incident; see Fig. 4). Fig. 5 shows the percent MSE improvements of the OCPB model over EM-ARIMA, ARIMA(1,0,2), and ARIMA(1,1,1) models for 1-lag ahead forecasts for different change types (i.e., no change, minor change, and major change). Moreover, Table VIII shows the overall average percent improvements of the OCPB model over the five compared models for 1-lag, 5-lag, and 10-lag ahead predictions. Based on the data that are analyzed in this paper, the following observations can be made.

1) When there is an accident (major or minor), errors of ARIMA and naive models are higher compared with the errors when there is no accident. For instance, the MSE of ARIMA(1,0,2) is 44.70 on February 18 A.M., a day with two major incidents (see Table I), whereas it is 2.87 for March 1 A.M., a day without a major incident. Fig. 3 shows 1-lag predictions on March 1 A.M.–P.M. From the figure in the A.M. section, no significant change occurs, and in general, all of the forecasting models similarly perform. In the afternoon, an incident causes a change with state-2 and state-3 oscillations (listed as minor change in Table II). The impact on the process is captured by the OCPB model with an MSE of 16.89 versus the closest MSE of 20.12 (i.e., 16% difference) that is achieved by the ARIMA(1,1,1) model. A similar process behavior on March 5 P.M. with a change to even lower speed state 1 (i.e., major change) results in an MSE difference of 40% from the next best model.

TABLE VII
PREDICTION ERRORS OF ALL SIX FORECASTING MODELS

Day	OCPB		EM-ARIMA		ARMA		ARIMA		MA		RW		Max%Imp	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
FEB-16 AM L4	5.24	1.72	5.45	1.73	5.95	1.75	7.19	2.05	6.06	1.75	6.64	1.98	27	16
FEB-16 PM L4	16.08	2.80	24.55	3.26	36.00	4.19	30.09	3.48	39.47	3.82	25.24	3.42	59	33
FEB-18 AM L4	15.67	2.52	18.42	2.84	44.70	4.53	27.75	3.93	28.71	3.51	24.90	3.21	65	44
FEB-18 PM L4	19.27	3.17	28.33	3.66	42.82	4.72	33.26	3.92	40.60	4.28	35.58	4.20	55	33
FEB-19 AM L4	7.63	1.98	8.46	2.02	12.32	2.25	11.23	2.41	14.80	2.53	11.92	2.40	48	22
FEB-19 PM L4	15.62	2.90	21.32	3.28	30.78	3.80	23.67	3.35	27.00	3.59	29.57	3.81	49	24
FEB-23 AM L4	12.38	2.23	16.83	2.48	32.35	3.33	19.23	2.97	27.68	2.84	13.52	2.37	62	33
FEB-23 PM L4	10.54	2.37	17.11	2.69	21.79	2.94	19.84	2.78	23.93	2.93	19.59	2.95	56	20
FEB-24 AM L7	2.12	1.08	2.07	1.06	2.39	1.11	3.76	1.45	2.34	1.13	3.35	1.34	44	25
FEB-24 PM L7	6.92	1.92	14.91	2.27	17.95	2.33	18.03	2.36	27.00	2.69	12.46	2.24	74	29
FEB-25 AM L4	5.79	1.65	6.00	1.69	6.58	1.72	7.60	2.04	8.04	1.89	7.41	1.91	28	19
FEB-25 PM L4	11.09	2.49	16.39	2.96	30.27	4.20	18.89	3.09	23.63	3.44	19.29	3.23	63	41
FEB-26 AM L12	3.34	1.26	3.79	1.32	4.51	1.37	4.82	1.59	4.99	1.43	4.19	1.40	33	21
FEB-26 PM L12	12.38	2.34	16.77	2.56	22.46	2.98	19.51	2.69	24.64	3.00	20.26	2.78	50	22
MAR-1 AM L4	2.91	1.29	2.90	1.28	2.87	1.29	4.72	1.66	3.29	1.34	4.39	1.58	38	22
MAR-1 PM L4	16.89	2.32	31.37	2.76	32.55	2.74	20.12	2.60	25.72	2.84	23.90	2.74	48	18
MAR-5 AM L20	2.33	1.19	2.23	1.17	2.40	1.20	4.05	1.55	2.50	1.22	3.64	1.47	43	23
MAR-5 PM L20	11.47	2.38	19.07	2.66	23.15	2.79	22.03	2.77	26.03	2.91	23.22	2.90	56	18
MAR-15 AM L4	11.41	2.23	13.20	2.36	27.45	3.28	20.97	3.19	17.71	2.58	17.81	2.79	58	32
MAR-15 PM L4	4.66	1.61	5.39	1.65	5.00	1.61	4.99	1.61	5.58	1.71	6.59	1.84	29	13
MAR-16 AM L12	11.43	2.21	11.15	2.15	14.92	2.41	16.09	2.70	16.10	2.44	19.30	2.79	41	21
MAR-16 PM L12	16.67	3.16	18.59	3.19	25.28	3.66	20.19	3.31	26.23	3.66	19.72	3.30	36	14
MAR-17 AM L5	6.00	1.79	6.10	1.82	7.15	1.92	7.42	2.06	7.40	2.00	7.83	2.05	23	13
MAR-17 PM L5	10.79	2.46	21.32	3.21	27.17	3.49	26.57	3.48	37.49	4.04	20.92	3.22	71	39
MAR-18 AM L4	11.27	2.37	14.09	2.57	28.18	3.51	18.39	3.11	21.95	3.07	17.20	2.97	60	32
MAR-18 PM L4	14.04	2.94	21.76	3.44	26.95	3.70	24.71	3.57	32.99	4.01	25.46	3.71	57	27
MAR-19 AM L4	6.87	2.04	6.82	1.95	9.03	2.13	10.15	2.49	9.41	2.26	11.84	2.63	42	22
MAR-19 PM L4	19.97	3.46	29.53	3.97	37.50	4.30	34.54	4.21	43.02	4.63	37.06	4.58	54	25

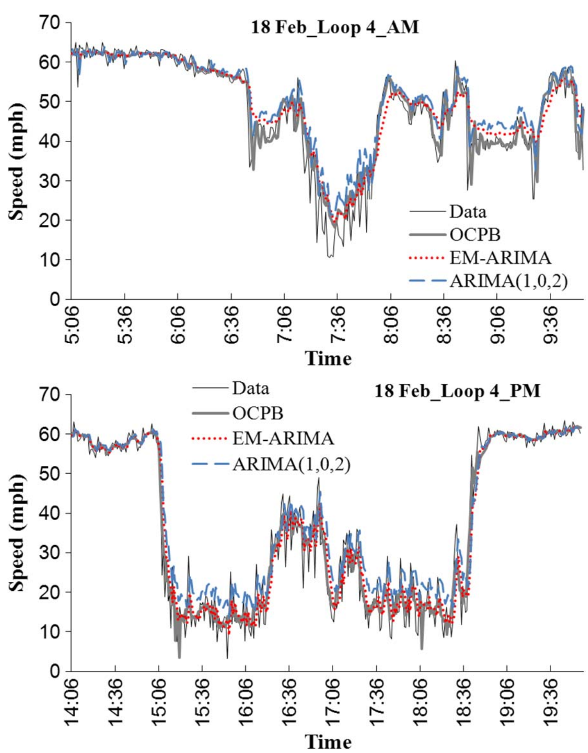

Fig. 4. Traffic speed predictions by Loop 4 on February 18 A.M.–P.M. with all three methods.

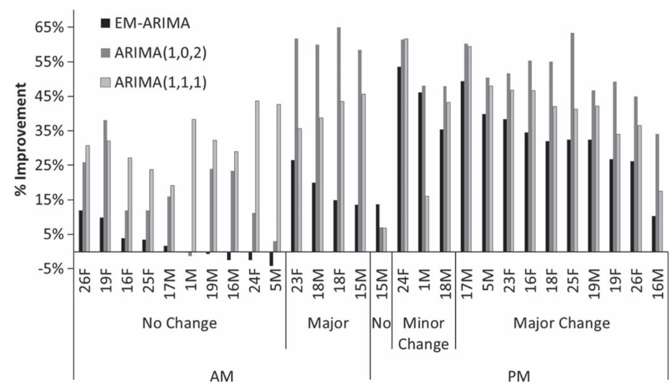

Fig. 5. Percent improvements in MSEs over EM-ARIMA, ARIMA(1,0,2), and ARIMA(1,1,1) models for 1-lag ahead forecasts.

2) ARIMA and naive models consistently overestimate speeds when the mean level drops, while the adaptive methods (i.e., OCPB and EM-ARIMA) capture the changes in the mean level. Fig. 4 reveals this phenomenon as the OCPB model follows the observed data closest. Other forecasting models lag behind as changes occur. In the February 18 A.M. section, many change points with less durations lead to less accuracy for nonadaptive models. The OCPB model's MSE of 15.67 is followed by EM-ARIMA with an MSE of 18.42 that has 15% difference. The P.M. section of February 18 shows longer state durations. The difference from EM-ARIMA has an MSE is 32%. In the March 17 P.M. section, similar multiple changes with long durations result in up to 50% accuracy improvement.
3) When there is no significant change in the process, the accuracy of the OCPB model is more or less equal to the compared models. Fig. 5 demonstrates the detailed prediction performances for no incident days. Clearly, from the figure, lower accuracy improvements are observed compared with the days with change points.

TABLE VIII
SUMMARY OF PERCENT IMPROVEMENTS

Interval	% Imp from EM-ARIMA MSE	% Imp from EM-ARIMA MAE	% Imp from ARMA MSE	% Imp from ARMA MAE	% Imp from ARIMA MSE	% Imp from ARIMA MAE	% Imp from MA MSE	% Imp from MA MAE	% Imp from RW MSE	% Imp from RW MAE
1-lag	20%	7%	39%	18%	37%	19%	40%	18%	35%	18%
5-lag	4%	1%	10%	5%	17%	10%	17%	7%	25%	14%
10-lag	0%	0%	0%	1%	17%	9%	9%	5%	19%	10%

4) Overall, both adaptive methods generate significantly better results than the MA, RW, and ARIMA models. The OCPB model outperforms nonadaptive models with average improvements reaching up to 39% in MSE and 19% in MAE, and surpasses EM-ARIMA performance with improvements reaching up to 20% in MSE and 7% in MAE for 1-lag ahead predictions. In Fig. 5, it is shown that the improvements are mainly positive. There are only a few instances of negative improvements, which are less than 5% on the days with no significant changes. The prediction accuracy gains reach up to 55% from EM-ARIMA, up to 65% from ARIMA(1,0,2), and 60% from ARIMA(1,1,1) when minor or major changes occur. No distinction is observed between minor and major changes; however, the incidents are only classified based on the impact reaching to state 1 (major) or 2 and 3 (minor). Certainly, multiple minor changes can also lead to higher improvements. For example, February 24 P.M. has multiple minor changes with short durations, and March 18 P.M. has multiple minor changes with high variance and long duration impacts. On these two days, the OCPB model provides up to 62% improvement in MSE.
5) The value of using the OCPB model over EM-ARIMA and nonadaptive models particularly decreases after 5-lag ahead forecasts as it contains no incident prediction mechanism. The OCPB also becomes late to the changes for multiple lag predictions. Percent improvements reduce to, on average, less than 5% from EM-ARIMA. From nonadaptive ARIMA, MA, and RW models, although the impact diminishes, it still stays higher than 10%.

V. CONCLUSION

This paper has developed a change-point-based short-term traffic parameter prediction model and has demonstrated its effectiveness with an extensive numerical evaluation. Based on the data utilized, the two adaptive techniques OCPB and EM-ARIMA provide more accurate results when the data generation process is not stable. The OCPB model is computationally more demanding because the HMM and the EM algorithm involve iterative solutions but significantly improves the errors. Although the approach seems promising, a number of questions need to be investigated. Among them, the retraining of HMM and the reestimation of ARIMA model parameters for different locations on the network can be listed as the most critical ones. Moreover, the recalibration interval of the model parameters can be studied for a specific location as new data received. An additional traffic parameter can be also observed to supply more information for the state probability calculations. Finally, the duration of the incidents can be incorporated into change point models for better detection.

ACKNOWLEDGMENT

The authors also would like to thank the anonymous reviewers for their insightful comments that significantly improved the content of this paper.

REFERENCES

[1] M. P. Clements, P. H. Franses, and N. R. Swanson, "Forecasting economic and financial time-series with non-linear models," *Int. J. Forecast.*, vol. 20, no. 2, pp. 169–183, Apr.–Jun. 2004.
[2] B. L. Smith, B. M. Williams, and R. K. Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 10, no. 4, pp. 303–321, Aug. 2002.
[3] A. Stathopoulos and M. G. Karlaftis, "A multivariate state space approach for urban traffic flow modeling and prediction," *Transp. Res. C, Emerg. Technol.*, vol. 11, no. 2, pp. 121–135, Apr. 2003.
[4] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, Nov. 2003.
[5] B. Ghosh, B. Basu, and M. O'Mahony, "Multivariate short-term traffic flow forecasting using time-series analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 2, pp. 246–254, Jun. 2009.
[6] T. Thomas, W. Weijermars, and E. van Berkum, "Predictions of urban volumes in single time series," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 1, pp. 71–80, Mar. 2010.
[7] M. T. Lee and B. Friedrich, "Comparison of univariate and multivariate time series models in short-term freeway traffic flow forecasting," in *Proc. TRB 89th Annu. Meeting Compend. Papers DVD*, Washington, DC, USA, 2010, vol. 1, pp. 1–13.
[8] T. T. Tchrakian, B. Basu, and M. O'Mahony, "Real-time traffic flow forecasting using spectral analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 519–526, Jun. 2012.
[9] J. Whittaker, S. Garside, and K. Lindveld, "Tracking and predicting a network traffic process," *Int. J. Forecast.*, vol. 13, no. 1, pp. 51–61, Mar. 1997.
[10] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended Kalman filter: A general approach," *Transp. Res. B, Methodol.*, vol. 39, no. 2, pp. 141–167, Feb. 2005.
[11] F. Yang, Z. Yin, H. X. Liu, and B. Ran, "An on-line recursive short-term traffic prediction algorithm," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1879, no. 1, pp. 1–8, Jan. 2004.
[12] S. Huang and W. A. Sadek, "A novel forecasting approach inspired by human memory: The example of short-term traffic volume forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 17, no. 5, pp. 510–525, Oct. 2009.
[13] I. Okutani and Y. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transp. Res. B, Methodol.*, vol. 18, no. 1, pp. 1–11, Feb. 1984.
[14] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach," *Transp. Res. C, Emerg. Technol.*, vol. 13, no. 3, pp. 211–234, Jun. 2005.
[15] L. Shen and M. Hadi, "Freeway travel time prediction with dynamic neural networks," in *Proc. TRB 89th Annu. Meeting Compend. Papers DVD*, Washington, DC, USA, Jan. 2010, vol. 1, pp. 1–15.
[16] K. Hamad, M. T. Shourijeh, E. Lee, and A. Faghri, "Near-term travel speed prediction utilizing Hilbert–Huang transform," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 24, no. 8, pp. 551–576, Nov. 2009.
[17] Y. Zhang and H. Ge, "Intelligent intersection traffic flow prediction based on fuzzy neutral network," in *Proc. TRB 89th Annu. Meeting Compend. Papers DVD*, Washington, DC, USA, Jan. 2010, vol. 1, pp. 1–19.
[18] E. I. Vlahogianni and M. G. Karlaftis, "Statistical methods versus neural networks in transportation research: Differences, similarities, and some

insights," *Transp. Res. C, Emerg. Technol.*, vol. 19, no. 3, pp. 387–399, Jun. 2011.

[19] W. Zheng, D. H. Lee, and Q. Shi, "Short-term freeway traffic flow prediction: Bayesian combined neural network approach," *J. Transp. Eng.*, vol. 132, no. 2, pp. 114–121, Feb. 2006.

[20] K. Y. Chan, T. S. Dillon, J. Singh, and E. Chang, "Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and Levenberg Marquardt algorithm," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 644–654, Jun. 2012.

[21] E. Castillo, M. Nogal, J. M. Menndez, S. Snchez-Cambronero, and P. Jimnez, "Stochastic demand dynamic traffic models using generalized beta-Gaussian Bayesian networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 565–581, Jun. 2012.

[22] E. I. Vlahogianni, J. C. Golias, and M. G. Karlaftis, "Short-term traffic forecasting: Overview of objectives and methods," *Transp. Rev.*, vol. 24, no. 5, pp. 533–557, Sep. 2004.

[23] B. L. Smith and M. J. Demetsky, "Traffic flow forecasting: Comparison of modeling approaches," *J. Transp. Eng.*, vol. 123, no. 4, pp. 261–266, Jul. 1997.

[24] L. Sun and J. Zhou, "Developing multi-regime speed–density relationships using cluster analysis," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1934, no. 1, pp. 64–71, Jan. 2005.

[25] Y. Kamarianakis, H. O. Gao, and P. Prastacos, "Characterizing regimes in daily cycles of urban traffic using smooth-transition regressions," *Transp. Res. C, Emerg. Technol.*, vol. 18, no. 5, pp. 821–840, Oct. 2010.

[26] J. W. C. V. Lint, S. P. Hoogendoorn, and H. J. van Zuylen, "Freeway travel time prediction with state-space neural networks modeling state-space dynamics with recurrent neural networks," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1811, no. 1, pp. 30–39, Jan. 2002.

[27] E. V. Zwet and J. Rice, "A simple and effective method for predicting travel times on freeways," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 3, pp. 200–207, Sep. 2004.

[28] X. Zhang and J. A. Rice, "Short term travel time predicting using a time-varying coefficient linear model," *Transp. Res. C, Emerg. Technol.*, vol. 11, no. 3/4, pp. 187–210, Jun./Aug. 2003.

[29] H. Tavana and H. Mahmassani, "Estimation and application of dynamic speed–density relations by using transfer function models," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1710, no. 1, pp. 47–57, Jan. 2000.

[30] N. Huynh, H. S. Mahmassani, and H. Tavana, "Adaptive speed estimation using transfer function models for real-time dynamic traffic assignment operation," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1783, no. 1, pp. 55–65, Jan. 2002.

[31] X. Qin and H. S. Mahmassani, "Adaptive calibration of dynamic speed–density relations for online network traffic estimation and prediction applications," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1876, no. 1, pp. 82–89, Jan. 2004.

[32] M. Cetin and G. Comert, "Short-term traffic flow prediction with regime switching models," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1965, no. 1, pp. 23–31, Jan. 2006.

[33] Y. Qi, "Probabilistic Models for Short Term Traffic Conditions Prediction," Ph.D. dissertation, Civil and Environ. Eng., Louisiana State Univ., Baton Rouge, LA, USA, 2010.

[34] L. Dimitriou, T. Tsekeris, and A. Stathopoulos, "Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow," *Transp. Res. C, Emerg. Technol.*, vol. 16, no. 5, pp. 554–573, Oct. 2008.

[35] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transp. Res. C, Emerg. Technol.*, vol. 19, no. 4, pp. 606–616, Aug. 2011.

[36] M. Tan, S. C. Wong, J. M. Xu, Z. R. Guan, and P. Zhang, "An aggregation approach to short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 1, pp. 60–69, Mar. 2009.

[37] M. D. Kindzerske and D. Ni, "Composite nearest neighbor nonparametric regression to improve traffic prediction," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1993, no. 1, pp. 30–35, Jan. 2007.

[38] Y. A. Kamarianakis and P. P. Kanas, "Modeling traffic volatility dynamics in an urban network," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1923, no. 1, pp. 18–27, Jan. 2005.

[39] J. Guo and B. M. Williams, "Real-time short term traffic speed level forecasting and uncertainty quantification using layered Kalman filters," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2175, no. 1, pp. 28–37, Dec. 2010.

[40] M. G. Karlaftis and E. I. Vlahogianni, "Memory properties and fractional integration in transportation time-series," *Transp. Res. C, Emerg. Technol.*, vol. 17, no. 4, pp. 444–453, Aug. 2009.

[41] M. P. Clements and D. F. Hendry, "Intercept correction and structural change," *J. Appl. Econom.*, vol. 11, no. 5, pp. 475–494, Sep. 1996.

[42] M. P. Clements and D. F. Hendry, "Forecasting from misspecified models in the presence of unanticipated location shifts," in *The Oxford Handbook of Economic Forecasting*, M. P. Clements and D. F. Hendry, Eds. London, U.K.: Oxford Univ. Press, 2011, ch. 11, pp. 315–355.

[43] J. L. Castle, N. W. Fawcett, and D. F. Hendry, "Forecasting breaks and forecasting during breaks," in *The Oxford Handbook of Economic Forecasting*, M. P. Clements and D. F. Hendry, Eds. London, U.K.: Oxford Univ. Press, 2011, ch. 10, pp. 271–315.

[44] M. Aiolfi, C. Capistran, and A. Timmermann, "Forecast combinations," in *The Oxford Handbook of Economic Forecasting*, M. P. Clements and D. F. Hendry, Eds. London, U.K.: Oxford Univ. Press, 2011, ch. 12, pp. 355–391.

[45] J. D. Hamilton, "Calling recessions in real time," *Int. J. Forecast.*, vol. 27, no. 4, pp. 1006–1026, Oct.–Dec. 2010.

[46] J. D. Hamilton, *Regime-Switching Models*, 2nd ed. Basingstoke, U.K.: Palgrave McMillan, 2008.

[47] E. Moreno, G. Casella, and A. Garcia-Ferrer, "An objective Bayesian analysis of the change point problem," *Stoch. Environ. Res. Risk Assess.*, vol. 19, no. 3, pp. 191–204, Aug. 2005.

[48] F. J. Giron, E. Moreno, and G. Casella, "Objective Bayesian analysis of multiple change points for linear models," in *Proc. Bayesian Stats.*, 2007, vol. 8, pp. 1–27.

[49] P. Fearnhead, D. Wyncoll, and J. Tawn, "A sequential smoothing algorithm with linear computational cost," *Biometrika*, vol. 97, no. 2, pp. 447–464, Jun. 2010.

[50] R. Lund and J. Reeves, "Detection of undocumented change points: A revision of the two-phase regression model," *J. Climate*, vol. 15, no. 17, pp. 2547–2554, Sep. 2002.

[51] Y. Ritov, A. Raz, and H. Bergman, "Detection of onset of neuronal activity by allowing for heterogeneity in the change points," *J. Neurosci. Methods*, vol. 122, no. 1, pp. 25–42, Dec. 2002.

[52] C. Beaulieu, T. B. M. J. Ouarda, and O. Seidou, "A review of homogenization techniques for climate data and their applicability to precipitation series," *Hydrol. Sci. J.–J. Sci. Hyrolog.*, vol. 52, pp. 18–37, Feb. 2007.

[53] R. Turner, "Bayesian change point detection for satellite fault prediction," in *Proc. IGC*, Cambridge, U.K., 2010, vol. 1, pp. 213–221.

[54] Y. Saatci, R. Turner, and C. E. Rasmussen, "Gaussian process change point models," in *Proc. Int. Conf. Mach. Learn.*, Haifa, Israel, Jun. 2010, vol. 1, pp. 927–934.

[55] R. Turner, Y. Saatci, and C. E. Rasmussen, "Adaptive sequential Bayesian change point detection," in *Proc. Temp. Segment. Workshop NIPS*, Whistler, BC, Canada, Dec. 2009, vol. 1, pp. 1–4.

[56] R. P. Adams and D. J. MacKay, "Bayesian Online Change-Point Detection," Univ. of Cambridge, Cambridge, U.K., Tech. Rep., 2007.

[57] R. Garnett, M. Osborne, S. Reece, A. Rogers, and S. Roberts, "Sequential Bayesian prediction in the presence of change points and faults," *Comput. J.*, vol. 53, no. 9, pp. 1430–1446, Nov. 2010.

[58] G. Celeux and J. B. Durand, "Selecting hidden Markov model state number with cross-validated likelihood," *Comput. Stat.*, vol. 23, no. 4, pp. 541–564, Oct. 2008.

[59] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[60] L. Baum and J. Eagon, "An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology," *Bull. Amer. Math. Soc.*, vol. 73, no. 3, pp. 360–363, May 1967.

[61] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Stat.*, vol. 41, no. 1, pp. 164–171, Feb. 1970.

[62] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc., B*, vol. 39, no. 1, pp. 1–38, Jan. 1977.

[63] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer-Verlag, 2009.

[64] G. McLachlan and D. Peel, *Finite Mixture Models*. Hoboken, NJ, USA: Wiley, 2000.

[65] V. Melnykov and I. Melnykov, "Initializing the EM algorithm in Gaussian mixture models with an unknown number of components," *Comput. Stat. Data Anal.*, vol. 56, no. 6, pp. 1381–1395, Jun. 2012.

[66] D. Brook and D. A. Evans, "An approach to the probability distribution of CUSUM run length," *Biometrika*, vol. 59, no. 3, pp. 539–549, Dec. 1972.

[67] B. M. Williams, "Multivariate vehicular traffic flow prediction: Evaluation of ARMAX modeling," in *Proc. Transp. Res. Rec.*, 2001, vol. 1776, pp. 194–200.

Gurcan Comert received the B.Sc. and M.Sc. degrees in industrial engineering from Fatih University, Istanbul, Turkey, in 2003 and 2005, respectively, and the Ph.D. degree in civil engineering from the University of South Carolina, Columbia, SC, USA, in 2008.

He is currently with the Physics and Engineering Department, Benedict College, Columbia. His research interests include applications of statistical models to transportation problems, traffic parameter prediction, and stochastic models.

Anton Bezuglov (M'06) received the B.Sc. and M.Sc. degrees in computer science from Taganrog State University, Taganrog, Russia, in 1997 and 1999, respectively, and the Ph.D. degree in computer science and engineering from the University of South Carolina, Columbia, SC, USA, in 2006.

He is currently with the Mathematics and Computer Science Department, Benedict College, Columbia. His research interests involve decision support systems and Bayesian networks.