Fast Boundary Flow Prediction for Traffic Flow Models using Optimal Autoregressive Moving Average with Exogenous Inputs (ARMAX) Based Predictors

Paper number: TRB14-4645

Cheng-Ju Wu (corresponding author)
Department of Mechanical Engineering, University of California, Berkeley
Berkeley, California 94720-1742
chengju@berkeley.edu

Thomas Schreiter
California PATH, University of California, Berkeley
410 McLoughlin Hall, Berkeley, CA 94720
schreiter@berkeley.edu

Roberto Horowitz
Department of Mechanical Engineering University of California at Berkeley
Berkeley, California 94720-1742
horowitz@berkeley.edu

Gabriel Gomes
California PATH, University of California, Berkeley
410 McLoughlin Hall, Berkeley, CA 94720
gomes@path.berkeley.edu

Paper submitted to TRR 2014, 15-Mar-2014
4900 words + 6 figure(s) *250 + 2 table(s) *250 = 6900 'words'

ABSTRACT

Traffic Management Centers (TMC) want to improve the performance of road networks and reduce congestion by actively managing the infrastructure of a freeway corridor. A promising avenue for proactive traffic management is the prediction of the near-future traffic conditions in real-time by employing a traffic flow model. An important set of calibration parameters of such a model are the boundary flows, i.e., the amount of traffic that is expected to enter the network during the prediction horizon. In this paper, we propose a boundary flow prediction method that combines the most recent traffic data with historical traffic data. An autoregressive moving average with exogenous input (ARMAX) is estimated on-line based on the most recent vehicle detector station (VDS) data. An optimal multiple step ahead traffic demand predictor is obtained based on the estimated ARMAX model by solving a corresponding Bezout equation for each predictor. Results obtained using empirical freeway mainline and on-ramp data show that this method outperforms methods that rely only on the historical average of the data to perform a prediction, especially during days with unusual traffic flow demands, such as a Super Bowl Sunday. Due to its simplicity and robustness, this method is useful in practical applications.

[HERE FIGURE 1]

INTRODUCTION

Reducing congestion both under recurrent and non-recurrent conditions is a challenge for freeway traffic management centers (TMC). The Connected Corridors program at the University of California PATH program is currently developing Decision Support Systems (DSS) that are able to forecast future short-term traffic conditions and evaluate potential traffic management strategies to improve mobility and safety in freeway corridors. A schematic block diagram of the traffic flow prediction module of a DSS is depicted in Figure 1. Many traffic predictors, like [1], use a traffic flow model based on the Cell Transmission Model [2] to predict near-future traffic conditions in real-time, usually with a prediction horizon of approximately one hour. In order to accurately estimate the potential benefits of traffic management strategies such as ramp metering, traffic routing and detouring, and lane management, the traffic flow model has to be calibrated well so that its predictions closely match actual future traffic conditions. Accurate calibration of traffic flow models is still a major challenge and usually requires careful estimation of its parameter values. As Figure 1 shows, there are many parameters that affect the prediction, such as network modeling, the fundamental diagrams, estimation of the current traffic state, and the prediction of boundary flows and split ratios. This paper focuses on one set of these parameters, namely the prediction of the incoming flows at the boundaries of the road network.

Much research has been devoted to the prediction of traffic flow. The area of traffic investigation known as OD prediction aims at calculating the amount of traffic that is expected to travel between each origin (O) and each destination (D) of the road network. Examples include [3], [4] and [5]. An advantage of OD prediction is that the resulting flow information is detailed and permits the search for true optimal traffic management strategies. However, OD estimation is a difficult problem, which is due to the fact that the available loop detector data is generally insufficient to determine the OD flow uniquely. As a consequence, the implementation of optimal traffic management strategies based on OD predictions may not yield the expected performance gains if the true OD patterns cannot be determined. Employing Lagrangian mobile data may provide partial flow information for each OD pair, which may significantly improve the quality of OD prediction in the future.

[HERE FIGURE 2]

An alternative and somewhat coarser form of boundary flow prediction is to disregard the destination and to only estimate the *total* flow at each origin. (To complete the traffic assignment, one must also predict split fractions at each bifurcation of the network, which determine how much traffic is leaving the network at each off-ramp. Although similar in nature to boundary flow prediction, split ratio prediction is out of the scope of this paper.) As shown in [6] and [7], this information can be determined uniquely in freeways, using only historical mainline data

flows. In this paper we will assume that historical freeway mainline and ramp traffic flow data is available through data collection and repository sites such as the Performance Measurement System (PeMS) [8]. In the following, we briefly discuss existing boundary flows forecasting methods.

One naive forecasting methodology is the zero-order-hold predictor, which applies the latest flow measurement over the prediction horizon. Another simple method is to directly apply a historical average, such as the mean flow of the previous weeks for the same day and time of day.

A more sophisticated forecasting technique is based on autoregressive moving average (ARMA) models, which treat data as a time series and linearly combine previous flow measurements to predict the future flow. The forecasting technique proposed in [9] employs an ARMA model with an integral action (so-called ARIMA) to forecast flow. Other ARMA-like models include a seasonal term (called seasonal ARIMA, or SARIMA), which exploit the periodicity of traffic flow. Figure 2 shows the flow profile of a detector over three weeks. One can clearly identify two periods of repeating traffic patterns. Firstly, traffic is following a daily pattern, with high traffic during the day and low traffic during the night. Working days thereby tend to have more traffic and show two peak periods, whereby weekend days exhibit a lower flow with only one peak period. Secondly, a weekly pattern is visible repeating the pattern of five working days and two weekend days.

In many works, the flow of the previous week is combined with recent measurements, for example [10], [11] and [12]. These approaches assume that data from different locations are uncorrelated. To model spatio-temporal correlations between nearby locations, [13] and [14] proposed a spatio-temporal seasonal ARIMA model. Another forecasting technique originates from machine learning. Artificial neural networks were used to forecast boundary flows to capture nonlinear traffic flow phenomena in [15] and [16]. Support vector machines (SVM) are another machine learning technique employed to forecast boundary flows, such as by [17].

Since the literature of traffic flow forecasting is large, several authors have compared different approaches. The research in [18] compared statistical approaches such as ARMA-like models with machine learning approaches and provided many references in transportation beyond the forecast of flow. Recently, [19] compared many approaches experimentally, including simple forecasting methodologies, ARMA-like models, and machine learning models. They confirmed that models that combine historical data with recent data clearly outperform approaches that use only one or the other. For a model to be suitable for on-line traffic prediction applications, it must be both computationally efficient and accurate in its prediction. Simple non-model based prediction approaches are fast, but usually lack the accuracy of more sophisticated model-based or machine learning-based models. On the other hand, machine learning techniques can be slow and hard to calibrate. A special case is the seasonal ARIMA model: some authors use data only from one previous period (commonly the previous week), which leads to fast computation and good accuracy. The drawback is that if the previous week contained irregular data, for example

caused by an extraordinary event such as an accident, the accuracy of the forecast can diminish significantly.

In this paper, we propose a technique that combines the strengths of a model that uses a large amount of historical data with an autoregressive moving average with exogenous input (ARMAX) model that is fast in computation. We choose an ARMAX model as the foundation of our method because it is a well-known time-series technique that is supported by well-established system identification theory [20]. Moreover, under relatively mild assumptions, optimal multi-step ahead output predictors can be determined in a computationally efficient manner based on ARMAX models [21]. In our proposed traffic flow prediction method, historical data are aggregated to flow profiles, which represent a meaningful average for each weekday. To reduce the effect of noise and obtain an accurate historical average, multiple months of data are aggregated to seven typical daily profiles for each detector. As shown in Figure 2 and mentioned above, traffic exhibits a daily and a weekly period. We therefore calculate flow profiles that represent a typical day, one for each day of the week. Since this computation involves a large amount of data, it can be computationally demanding. This, however is not a problem in practice, since these nominal historical profiles can be computed in an off-line fashion, so that no running time constraints apply. Using the nominal historical profile as a deterministic input and the actual traffic flow as the noise-contaminated measured output, we identify in a real-time recursive fashion the parameters of an autoregressive moving average with exogenous input (ARMAX) model that best describes this input/output relation. Using the results of the ARMAX identification process, we determine the optimal multi-step ahead predictor model of the traffic flow, which utilizes the flow measurements up to the current time and the nominal historical profile by solving a Bezout equation [21]. Due to its relatively low on-line computation time, this system can be applied in practice to predict traffic flow in real-time.

The remainder of this paper is organized as follows. The next section explains the methodology of the proposed algorithm. It consists of the aggregation of the historical data, the ARMAX model description, the estimation of its parameter values, and the actual forecast of the boundary flows. Then, the results from experiments are presented based on empirical mainline and on-ramp inductive-loop data from the I-15 freeway in California, where a one-hour flow forecast is compared against measured flow data. The paper concludes with a summary and a brief discussion concerning practical applications and further research.

[HERE FIGURE 3]

METHODOLOGY

In this section, we present how to determine the multi-step ahead optimal predictor of traffic flow, based on both historical and the most recent traffic data. The concept of flow prediction is illustrated in Figure 3 for a vehicle detector station (VDS). In this figure, profile $u(k)$ is the nominal historical flow data for one day, sampled with sampling rate Δt so that the profile

consists of $N = \frac{24h}{\Delta t}$ data points. Profile $y(k)$ is the recent flow data of the prediction day from midnight to the current time step L. Given $u(k)$ and $y(k)$, the proposed prediction algorithm computes the multi-step traffic flow prediction $y^{\mathrm{p}}(k+D|k)$ for a given prediction horizon N_p, where $D = 1, \dots, N_p$.

A brief description of the proposed prediction algorithm is as follows. First, the historical traffic flow data is categorized by the day of the week, so that a nominal flow profile $u(k)$, for $k = 1, \dots, N$ is obtained for each day of the week. Using the representative nominal flow profile $u(k)$ as the deterministic input, and the recent traffic flow $y(k)$ as the noise-contaminated output, the parameters of an ARMAX model are recursively estimated at each time step k. Based on the estimated ARMAX parameters, the Bezout equation is solved at every time step, which produces the parameters of the optimal D-step ahead output predictor $y^{\mathrm{p}}(k+D|k)$ given only output data up to time step L and the sequence of historical data $u(k)$. In the following, we present each step in detail.

Nominal Flow Profile for each Day of the Week

In the first step, nominal historical flow profiles are determined for each day of the week. The nominal profile is robust with respect to outliers, which are caused by, for example, traffic accidents or detector failures.

Let a VDS measure the flow over the course of S full days. These data are represented as a set

$$M = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_S\} \tag{1}$$

of historical traffic flow data, where each element

$$\mathbf{x}_i = [x_i(1), x_i(2), \cdots, x_i(N)]^{\mathrm{T}} \in \mathrm{R}_+^{\mathrm{N}} \tag{2}$$

is a traffic flow profile over day i, for $i = 1, \dots, S$, with $x_i(k)$ as the flow measured at time step k.

The data are then categorized by the day of the week. Define the nominal flow profile of a given day of week as $u_d(k)$, where $d \in \{\mathrm{Mon}, \mathrm{Tue}, \mathrm{Wed}, \mathrm{Thu}, \mathrm{Fri}, \mathrm{Sat}, \mathrm{Sun}\}$. The nominal flow profile of each day of the week is calculated as the median of each sampling time step. For example, the typical profile of Monday is calculated as

$$u_{\mathrm{Mon}}(k) = \mathrm{median}\{x_i(k)\text{: day } i \text{ is a Monday}\} \tag{3}$$

Aggregating many profiles to few typical profiles filters out some of the noise of the data, this leads to a robust estimation of typical profiles. Furthermore, by using the median, the effect of outliers is eliminated.

In this aggregation step, potentially a large amount of data is aggregated, which might require significant computational resources. However, this is not an issue for the on-line operation of the system, because the typical profile can be computed off-line.

For the on-line forecast procedure that will be described in the next section, the nominal profile of the respective day of the week is used, i.e. if the forecast day is a Monday, then $u(k) = u_{\mathrm{Mon}}(k)$ will be used.

ARMAX Model Definition and Parameter Estimation

In order to model the correlations of flow measurements in time, an autoregressive moving average model with exogenous inputs (ARMAX) model is employed in this research. Consider a recursive ARMAX model in every time step k described by the following linear stochastic difference equation

$$A(q^{-1})y(k) = B(q^{-1})u(k) + C(q^{-1})w(k), \tag{4}$$

where $y(k)$ is the measured flow of the VDS at time step k, $u(k)$ is the nominal historical flow data at time step k, and $w(k)$ is assumed to be a zero-mean innovation sequence, i.e. $E\{w(k)\} = 0$ and $E\{w(k)w(k-j)\} = 0$ for $0 < j \leq k$. $A(q^{-1})$, $B(q^{-1})$ and $C(q^{-1})$ are scalar polynomials in the backward shift operator q^{-1} $[q^{-1}\, y(k) \triangleq y(k-1)]$ of orders n_a, n_b and n_c, respectively, defined by

$$A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a}$$

$$B(q^{-1}) = b_0 + b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}$$

$$C(q^{-1}) = 1 + c_1 q^{-1} + \cdots + c_{n_c} q^{-n_c}. \tag{5}$$

The order of the polynomials n_a, n_b and n_c are design parameters.

As described above, the exogenous known input $u(k)$ in Eq. (4) is calculated as the median of a large number of historical flows for a unique day of the week (e.g. all Monday flows in a given month). Therefore it can be considered as a deterministic input in this model. Day to day variability in the flow $y(k)$ from the median $u(k)$ can be due to both deterministic factors (e.g. the flow on a particular day could be consistently 5% higher than the median) or stochastic effects, such as random colored noise. A fairly general model that includes many of these effects is a state space model with input and measurement white noises

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}u(k) + \mathbf{B}_w m(k) \tag{6}$$

$$y(k) = \mathbf{C}\mathbf{x}(k) + \mathrm{D}u(k) + v(k)$$

where the input noise $m(k)$ and the measurement noise $v(k)$ are both zero mean white noises, $\mathbf{x}(k) \in \mathrm{R}^n$ is the state and $\mathbf{A}$, $\mathbf{B}$, $\mathbf{B}_w$, and $\mathbf{C}$ are matrices of the appropriate dimensions, and D is a scalar. A well-known result in stochastic systems [21] is that, under fairly mild assumptions and stationarity, the system in Eq. (6) can also be described by the ARMAX model in Eq. (4), where in this case the zero mean innovation signal $w(k)$ is equal to the Kalman filter residual for the system in Eq. (6) and the roots of the polynomial $\bar{C}(z) = z^{n_c}C(z^{-1})$ are the closed loop poles of the Kalman filter. Similar results exist for time-varying matrices $\mathbf{A}$(k), $\mathbf{B}$(k), $\mathbf{B}_w(k)$, $\mathbf{C}$(k) and gain $\mathrm{D}(k)$ and/or non-stationarity conditions. However in these cases, the coefficients of the polynomials $A(q^{-1})$, $B(q^{-1})$ and $C(q^{-1})$ in Eq. (5) are time-varying.

We now assume that the flow $y(k)$ can be adequately described in terms of the ARMAX model (4), using the known exogenous input $u(k)$, but will not assume that the (possibly time-varying) coefficient of the polynomials $A(q^{-1})$, $B(q^{-1})$ and $C(q^{-1})$ in Eq. (5) are known, nor the innovations signal $w(k)$ is measurable. The coefficients of the polynomials in (5) are estimated using a Recursive Least Squares (RLS) Parameter Adaptation Algorithm (PAA) [21] with forgetting factor and covariance resetting [22]:

$$\hat{A}(q^{-1}) = 1 + \hat{a}_1(k)q^{-1} + \cdots + \hat{a}_{n_a}(k)q^{-n_a}$$

$$\hat{B}(q^{-1}) = \hat{b}_0(k) + \hat{b}_1(k)q^{-1} + \cdots + \hat{b}_{n_b}(k)q^{-n_b}$$

$$\hat{C}(q^{-1}) = 1 + \hat{c}_1(k)q^{-1} + \cdots + \hat{c}_{n_c}(k)q^{-n_c}. \tag{7}$$

The a-posteriori estimation output $\hat{y}(k)$ is given by

$$\hat{y}(k) = -\hat{A}^*(q^{-1})y(k) + \hat{B}(q^{-1})u(k) + \hat{C}^*(q^{-1})e(k) \tag{8}$$

$$e(k) = y(k) - \hat{y}(k)\,, \tag{9}$$

with $\hat{A}^*(q^{-1}) = \hat{A}(q^{-1}) - 1$and $\hat{C}^*(q^{-1}) = \hat{C}(q^{-1}) - 1$. The parameters are updated in order to make the residual $e(k)$ in (9) converge to an innovation sequence. Details regarding the implementation and convergence properties of the recursive parameter estimation algorithm used in this paper can be found in many adaptive signal processing and control texts, such as [20, 21]. In essence, under stationary assumption and constant coefficients, it can be shown that the output estimation error $e(k) = y(k) - \hat{y}(k)$ converges to an innovation sequence and, under further assumptions regarding $C(q^{-1})$, the coefficient estimates converge to the true coefficients in (5).

The algorithm in [22] uses two design parameters, the forgetting factor and the regularization factor, to improve the parameter estimation performance and allow tracking of slowly time-varying coefficients. The forgetting factor is a real number greater than zero and less than one (typically very close to one), which enhances the ability of the PAA to track parameter variations. The smaller the forgetting factor is, the smaller the contribution of older data is to the RLS PAA. The regularization factor is another RLS PAA design parameter. As detailed in [22], the

regularization procedure prevents the RLS gain matrix (also known as the covariance matrix) from becoming unbounded under lack of persistence of excitation and prevents the estimated parameter-bursting phenomenon. Estimated parameter bursting (i.e. parameters suddenly become large in magnitude) may lead to large overshoots in the optimal predictor.

[HERE FIGURE 4]

Optimal Traffic Flow Predictor

In this section, we describe the optimal prediction method that is employed in this paper. Assume that the RLS ARMAX parameter estimates in (7) converge such that $e(k) = y(k) - \hat{y}(k)$ converges to an innovation sequence and, in particular, the estimated noise polynomial $\hat{C}(q^{-1})$ is asymptotically stable (i.e. all roots of the polynomial $\hat{C}(z)$ lie outside the unit circle). Given the median historical flow data $u(k)$ for $k = 1, \dots, N$ and recent flow data $[y(1), \dots, y(L)]$, the optimal (minimum variance) D-step ahead predictor of $y(L+D)$, denoted as $y^{\mathrm{p}}(L+D|L)$, satisfies the following difference equation [21],

$$\hat{C}(q^{-1})y^{\mathrm{p}}(k+D|k) = \hat{G}(q^{-1})y(k) + \hat{F}(q^{-1})\hat{B}(q^{-1})u(k+D). \tag{10}$$

The polynomials $\hat{G}(q^{-1})$ and $\hat{F}(q^{-1})$

$$\hat{F}(q^{-1}) = 1 + \hat{f}_1 q^{-1} + \cdots + \hat{f}_{D-1} q^{-(D-1)} \tag{11}$$

$$\hat{G}(q^{-1}) = \hat{g}_0 + \hat{g}_1 q^{-1} + \cdots + \hat{g}_{n_a-1}(k) q^{-n_g} \tag{12}$$

with $n_g = \max(n_c - D, n_a - 1)$, are the unique solution of the Bezout equation,

$$\hat{C}(q^{-1}) = \hat{F}(q^{-1})\hat{A}(q^{-1}) + q^{-D}\hat{G}(q^{-1}). \tag{13}$$

Note that $\hat{F}(q^{-1})$ and $\hat{G}(q^{-1})$ depend on k and D. We omitted the indexes for legibility.

Let $\hat{H}(q^{-1})$ be the product of the two polynomial $\hat{F}(q^{-1})$ and $\hat{B}(q^{-1})$:

$$\hat{H}(q^{-1}) = \hat{F}(q^{-1})\hat{B}(q^{-1}) = \hat{b}_0 q^{-1} + (\hat{b}_1 + \hat{f}_1\hat{b}_0)q^{-2} + \cdots + \hat{f}_{D-1}\hat{b}_{n_b} q^{-(D-1+n_b)}. \tag{14}$$

Then, the optimal predictor in (10) can be written in transfer function form as

$$y^{\mathrm{p}}(k+D|k) = \frac{\hat{G}(q^{-1})}{\hat{C}(q^{-1})} y(k) + \frac{\hat{H}(q^{-1})}{\hat{C}(q^{-1})} u(k+D). \tag{15}$$

The block diagram of the optimal predictor in (15) is shown in Figure 4. The predictor is a two input one output system. Given the current data $y(k)$ and nominal historical flow data $u(k+D)$ as inputs, the system generates the D-step ahead prediction of $y(k)$, which is the output $y^{\mathrm{p}}(k+D|k)$. The polynomials in the transfer functions of the predictor are updated at each time step, allowing it to track time-varying changes in the parameters in Eq. (5) using both current

flow and historical flow data. This adaptive property means that the proposed predictor is potentially able to predict traffic flow under irregular and incidental traffic conditions, as the following experiments will show.

EXPERIMENTAL SETUP

In this section, we describe the experimental environment used to validate the proposed traffic flow prediction algorithm based on empirical data from a Californian freeway.

Data

Traffic flow data for six vehicle detector stations (VDS) along the freeway corridor I-15 North, near San Diego, California, were obtained through the Performance Measurement System (PeMS) [8]. Three of the six VDS are located on the mainline with identification numbers and post mileage (PM) 1108595 (at PM 24.065), 1108562 (at PM 26.249), and 1108767 (at PM 27.138). The three other VDS are located on on-ramps with identification numbers 1108596 (at PM 24.065), 1108563 (at PM 26.249), and 1108768 (at PM 27.138). Five-minute historical flow data were collected from August 1, 2012 to December 31, 2012. The collected flow data (288 observations per day) were aggregated to $\Delta t = 15$ min, which results in $N = 96$ observations per day. The five-minute historical flow data is aggregated into 15-min data to reduce the historical flow data points per day and for filtering the noise in five-minute historical data. Therefore, 15-min data were used to compute the nominal typical flow profiles $u_d(k)$ per day of the week. In order to study the influence of a special event on the traffic flow prediction method proposed in this paper, we used traffic data collected during the week when the Super Bowl XLVII game took place, which is a large event in the United States. The normal traffic flow pattern on a regular weekday is shown in Figure 5 (solid line, $y(k)$). The unusual traffic flow pattern due to the Super Bowl event is shown in Figure 6 (solid line, $y(k)$). On Super Bowl day, people tended to stay at home or to visit friends to watch the game which began at 15:30. As a consequence, the traffic flow pattern in the period from 15:30 to 18:30 is significantly lower than the historical flow. An increase in flow that can be observed after 20:00 by the end of the game, which is probably caused by people returning to their home after watching the game. The proposed method is used to predict the flow in a rolling-horizon fashion from Monday January 28 to Sunday February 3, 2013 between 06:00 and 22:00.

ARMAX Estimation and Prediction Parameter Values

The ARMAX parameters used for the experiment are as follows. The order of ARMAX model was selected to be $(n_a, n_b, n_c) = (2,1,2)$. The selection of the order of the ARMAX model was based on the following three reasons. First, increasing the order of these polynomials in ARMAX model leads to overfitting and reduces the computation efficiency. Second, we chose the order of

the polynomial $A(q^{-1})$ and $C(q^{-1})$ to be the same. Third, we noticed that making the orders of $A(q^{-1})$ and $C(q^{-1})$ larger than two, resulted in little or no prediction improvement. The forgetting factor was selected as 0.97, and the regularization factor was selected as 0.01 after a brief trial an error period. Both of these factors are design parameters for the RLS PAA algorithm in [22]. Since the sampling rate Δt is 15 min and the prediction horizon is one hour ($N_p = 4$), the prediction step D is varied from 1 to 4, in order to make 15 min, 30 min, 45 min and 60 min ahead predictions.

Prediction Accuracy Performance

Given the real flow data $y(k)$ and the predicted flow data $y^{\mathrm{p}}(k|k-D)$ at time step k and prediction step D, the performance was evaluated based on mean absolute percentage error (MAPE) of prediction step D for a VDS as follows

$$\mathrm{MAPE}(D) = \frac{1}{n}\sum_{k=k_0}^{k_1} \frac{|y(k)-y^{\mathrm{p}}(k|k-D)|}{y(k)}. \tag{16}$$

The starting and end time steps are $k_0 = \frac{06{:}00}{\Delta t}$ and $k_1 = \frac{22{:}00}{\Delta t}$, respectively, which leads to $n = k_1 - k_0 + 1 = 65$ time steps for each day where a prediction is started. This error (16) is further averaged over the detector type (mainline/on-ramp).

RESULTS AND DISCUSSION

In this section, the traffic flow prediction results of the method proposed are presented and compared against a simple predictor based only on the nominal historical profile. We first show the prediction of a regular day, and then of the Super Bowl day, where irregular traffic flows occurred.

Flow Prediction on a Regular Day

[HERE FIGURE 5]

The flow prediction result of a mainline VDS (1108595, 5 lanes) of a weekday (Thursday 31 January 2013) is shown in the top part of Figure 5. The flow measurements $y(k)$ are indicated by the solid line. The nominal historical profile $u(k)$ is shown by the dashed line. The prediction, indicated by the remaining four lines, is close to the actual measurements so that only very small prediction error occurs. Because the regular is close to the nominal historical profile, the prediction is very close to the actual data.

The flow prediction result of an on-ramp VDS (1108563) on a weekday (Wednesday 30 January 2013) is shown in the bottom part of Figure 5. The difference between the predicted and the actual flows is small.

Flow Prediction on Super Bowl Sunday

[HERE FIGURE 6]

The flow prediction of a mainline VDS (1108595) for the Super Bowl day (Sunday 3 February 2013) is shown in the top part of Figure 6. Notice that flow leading up to the beginning of the game at 15:30 is drastically lower than the nominal flow $u(k)$. Also, the top part of Figure 6 shows that the prediction result $y^{\mathrm{p}}(k+D|k)$ for $D = 1, \ldots, 4$ more closely follows the measured flow $y(k)$ than the historical flow $u(k)$.

Using only the nominal historical profile as forecast leads to very large errors, since the drop of the traffic flow at 15:30 is completely ignored. The proposed prediction algorithm therefore outperforms the historical predictor. Moreover, the result shows that the one step prediction $y^{\mathrm{p}}(k+1|k)$ has better performance than larger steps. As the prediction step increases, the prediction deviates from the measurement flow, especially in the case of special event. The prediction error increases when the prediction step D increases.

The traffic flow prediction of an on-ramp VDS (1108563) on the Super Bowl day is shown in the bottom part of Figure 6. The influence of the Super Bowl game is visible at this location as well. The measured traffic flow $y(k)$ behaves normally until 15:30, then the flow decreases between 15:30 and 18:30. Before 15:30, the proposed prediction result $y^{\mathrm{p}}(k+D|k)$ for $D = 1, \ldots, 4$ and the historical flow $u(k)$ matches the measured traffic flow data $y(k)$. However, the prediction starts to deviate from the measured flow at 15:30. After 15:30, the prediction overestimates the traffic flow until 18:30, because the measured flow data has a significant difference with historical flow data.

Evaluation of Prediction Performance using the MAPE Criterion

[HERE TABLE 1]

The MAPE criterion defined in (16) are averaged for three mainline and three on-ramp VDS, which is shown in Tables 1 and 2, respectively. In addition, the MAPE criterion evaluates the prediction performance for the entire day rather than a specific time period of a day because the predictor is using the historical pattern of a whole day to make prediction.

[HERE TABLE 2]

In the mainline case, both of the MAPE values for the proposed method and the historical predictor are approximately 7% for the normal days (Monday to Saturday), which indicates that the MAPE error does not change significantly across normal days. However, due to the influence of the special event on Sunday, the MAPE values are higher than other days in the week. On Sunday (Super Bowl day), the error of the proposed method is 10.3% for the $D = 1$ step prediction. The error increases with the length of the prediction horizon. At $D = 4$, the prediction

error is 16.6%, while the MAPE error using historical data for prediction is 21.5%. The proposed method based on the ARMAX model therefore outperforms using only historical data to predict the traffic flow.

In the on-ramp case, the MAPE value for both the proposed method and the historical predictor are distributed from 20% to 29% for the normal days (Monday to Saturday), which is higher than normal days in the mainline case. On Sunday (Super Bowl day), the error of the proposed method is 31.3% for the $D = 1$ step prediction. At $D = 4$, the prediction error is 34.3%, while the MAPE error using historical data for prediction is 38.7%. The proposed method based on the ARMAX model still outperforms using only historical data to predict the traffic flow in the on-ramp case.

The predictions at mainline show a systematically lower error than those at on-ramps. The MAPE values in Tables 1 and 2 show an error for mainline flow prediction between 6% and 8%, while the on-ramp flow prediction error ranges from 20% to 33%, for a regular day. The reason is that mainline traffic is significantly higher than that of on-ramps and is, in fact, the aggregation of many single on-ramps flows, namely those that are located upstream. This aggregation leads to a systematic reduction of signal noise. Since the signal is less corrupted by noise, the estimation of the ARMAX parameters is more precise and consequently the prediction is more precise, i.e. its errors are lower.

CONCLUSIONS

We have developed a simple and fast, yet accurate, method to forecast traffic flow based on data from vehicle detector stations. An optimal multi-step ahead predictor combines the most recent measured flow data with a nominal historical flow profile. The parameters of this predictor are estimated by an autoregressive moving average model with exogenous input (ARMAX). This method predicts traffic flow under irregular conditions significantly better than a simple predictor based on the nominal historical profile alone. We showed by the example of the Super Bowl day, where traffic patterns differ drastically from regular Sundays, that the prediction error is reduced by up to ten percentage points.

The nominal historical profile is computed as the median over months of historical data, stratified by the day of the week. Using the median ensures that outliers caused by failing detectors or irregular days do not affect the nominal profile. The nominal profiles are computed off-line so that the on-line forecast only consists of the ARMAX estimation and prediction method, which is computationally very fast. This combination of a robust estimation of the nominal profile, the accurate prediction of the near-future traffic flow, and the fast computation make this method suitable for practical applications of boundary flow forecasting and its use in traffic management systems.

Further research will focus on the prediction of an error bound and the minimization of the prediction error, which provides information to the traffic flow forecast engine about the reliability of the boundary forecast. This in turn can be used to make the control decisions of the traffic manager more robust. Also, to incorporate route choice of travelers before they enter the network, we will investigate spatial correlations of neighboring on-ramp detectors, which is expected to increase the overall accuracy and decrease the uncertainty of the prediction. In addition, to predict boundary demand rather than flow, the inclusion of information about the traffic regime (free flow vs. congestion), measurements like density or occupancy will be necessary. Finally, since the literature already provides a large amount of flow forecasting algorithms, a comparison of different methods both in terms of prediction accuracy and running time would support scientists and practitioners in their choice of an appropriate boundary flow prediction model.

ACKNOWLEDGMENT

This work is supported by California Department of Transportation under the Connected Corridors program and by the National Science Foundation (NSF) through grant CDI-0941326.

REFERENCES

[1] Schreiter, T., J. W. C. Van Lint, and S. P. Hoogendoorn, Vehicle-class Specific Control of Freeway Traffic. In Transportation Research Board 92nd Annual Meeting, 2013, 13-0585.

[2] Daganzo, C. F. The Cell Transmission Model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. Transportation Research Part B: Methodological, Vol. 28, No. 4, 1994, pp. 269 – 287.

[3] Ashok, K., Estimation and prediction of time-dependent origin-destination flows. Ph.D. thesis, Massachusetts Institute of Technology, 1996.

[4] Zhou, X. and H. S. Mahmassani. A structural state space model for real-time traffic origin destination demand estimation and prediction in a day-to-day learning framework. Transportation Research Part B: Methodological, Vol. 41, No. 8, 2007, pp. 823 – 840.

[5] Djukic, T., J. W. C. Van Lint, and S. P. Hoogendoorn, Exploring Application Perspectives of Principal Component Analysis in Predicting Dynamic Origin-Destination Matrices. In Transportation Research Board 91st Annual Meeting, 2012, 12-2702.

[6] Muralidharan, A. and R. Horowitz. Imputation of ramp flow data for freeway traffic simulation. Transportation Research Record: Journal of the Transportation Research Board, Vol. 2099, No. 1, Trans Res Board, 2009, pp. 58–64.

[7] Muralidharan, A., Tools for modeling and control of freeway networks. Ph.D. thesis, University of California, Berkeley, 2012.

[8] PeMS Website http://pems.dot.ca.gov, accessed 1-Sep-2013.

[9] Lee, S. and D. B. Fambro. Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting. Transportation Research Record: Journal of the Transportation Research Board, Vol. 1678, No. 1, Trans Res Board, 1999, pp. 179–188.

[10] Williams, B. and L. Hoel. Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results. Journal of Transportation Engineering, Vol. 129, No. 6, 2003, pp. 664–672.

[11] Shekhar, S. and B. M. Williams. Adaptive seasonal time series models for forecasting short-term traffic flow. Transportation Research Record: Journal of the Transportation Research Board, Vol. 2024, No. 1, Trans Res Board, 2008, pp. 116–125.

[12] Smith, B. L., B. M.Williams, and R. K. Oswald. Comparison of parametric and nonparametric models for traffic flow forecasting. Transportation Research Part C: Emerging Technologies, Vol. 10, No. 4, 2002, pp. 303 – 321.

[13] Kamarianakis, Y., W. Shen, and L. Wynter. Real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO. Applied Stochastic Models in Business and Industry, Vol. 28, No. 4, John Wiley & Sons, Ltd, 2012, pp. 297–315.

[14] Kamarianakis, Y. and P. Prastacos. Space-time modeling of traffic flow. Computers & Geosciences, Vol. 31, No. 2, 2005, pp. 119 – 133.

[15] Vlahogianni, E. I., M. G. Karlaftis, and J. C. Golias. Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach. Transportation Research Part C: Emerging Technologies, Vol. 13, No. 3, 2005, pp. 211 – 234.

[16] Jiang, X. and H. Adeli. Dynamic Wavelet Neural Network Model for Traffic Flow Forecasting. Journal of Transportation Engineering, Vol. 131, No. 10, 2005, pp. 771–779.

[17] Zhang, Y. and Y. Xie. Forecasting of short-term freeway volume with v-support vector machines. Transportation Research Record: Journal of the Transportation Research Board, Vol. 2024, No. 1, Trans Res Board, 2008, pp. 92–99.

[18] Karlaftis, M. and E. Vlahogianni. Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. Transportation Research Part C: Emerging Technologies, Vol. 19, No. 3, 2011, pp. 387 – 399.

[19] Lippi, M., M. Bertini, and P. Frasconi. Short-Term Traffic Flow Forecasting: An Experimental Comparison of Time-Series Analysis and Supervised Learning. Intelligent Transportation Systems, IEEE Transactions on, Vol. 14, No. 2, 2013, pp. 871–882.

[20] Ljung, L. System identification - Theory for the User. Prentice-Hall, 1999.

[21] Goodwin, G. C. and K. S. Sin. Adaptive Filtering Prediction and Control. Dover Publications, 2009.

[22] Gunnarsson, S., Combining tracking and regularization in recursive least squares identification. In Decision and Control, 1996., Proceedings of the 35th IEEE Conference on, 1996, Vol. 3, pp. 2551–2552 vol.3.

List of table titles and figure captions

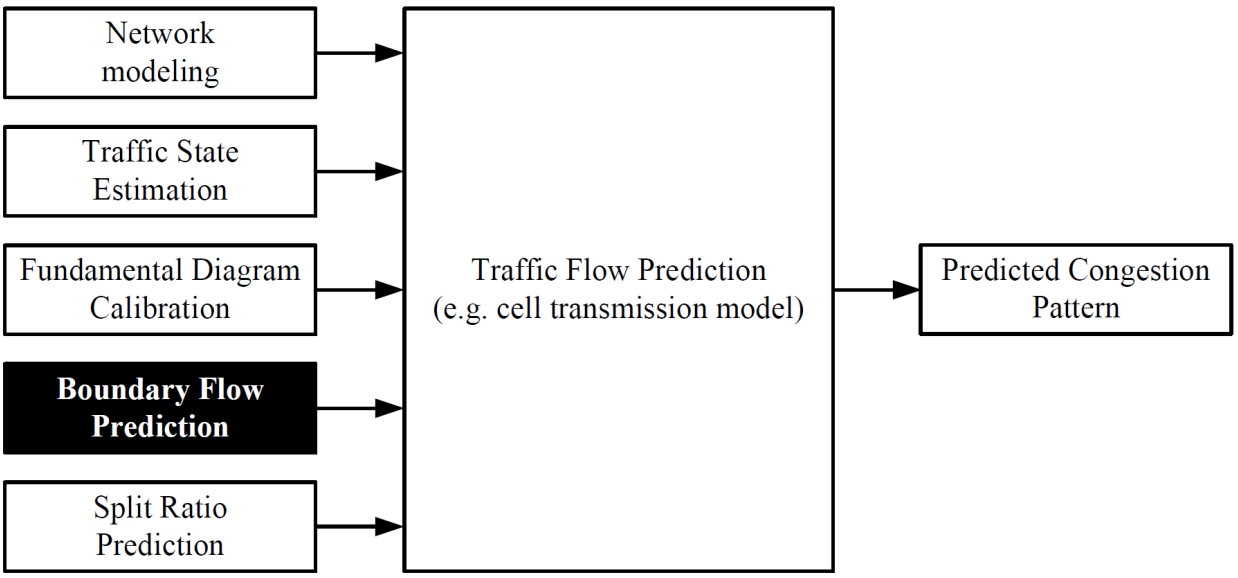

FIGURE 1 Components for on-line traffic flow prediction

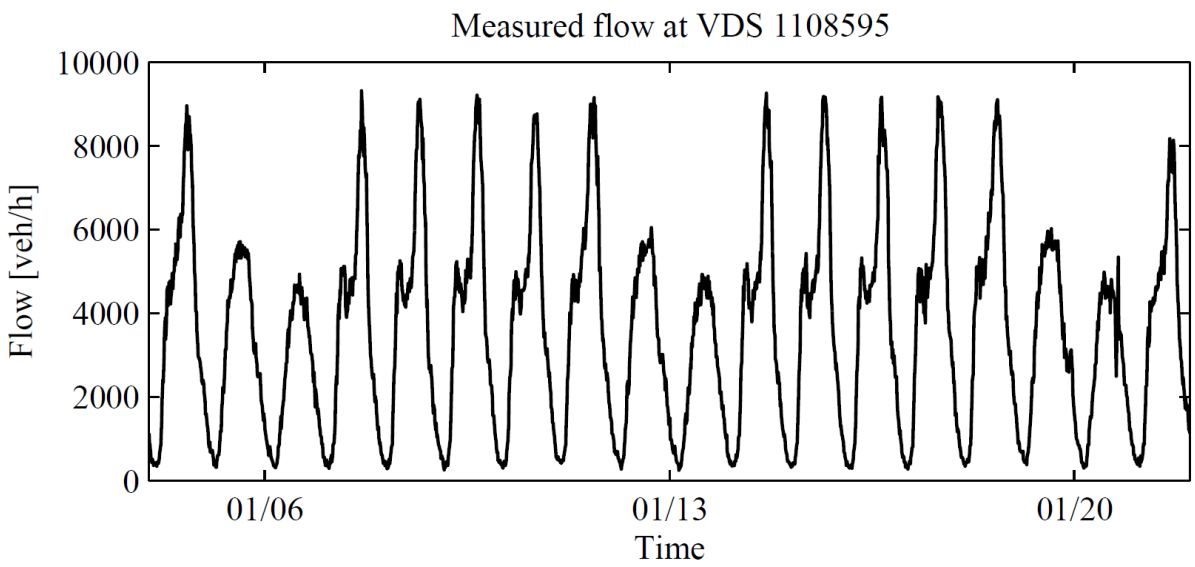

FIGURE 2 Flow measurements over the course of multiple weeks. A daily and a weekly period are clearly visible. (Data from mainline detector station 1108595 between 4 January and 21 January 2013, data source: [8])

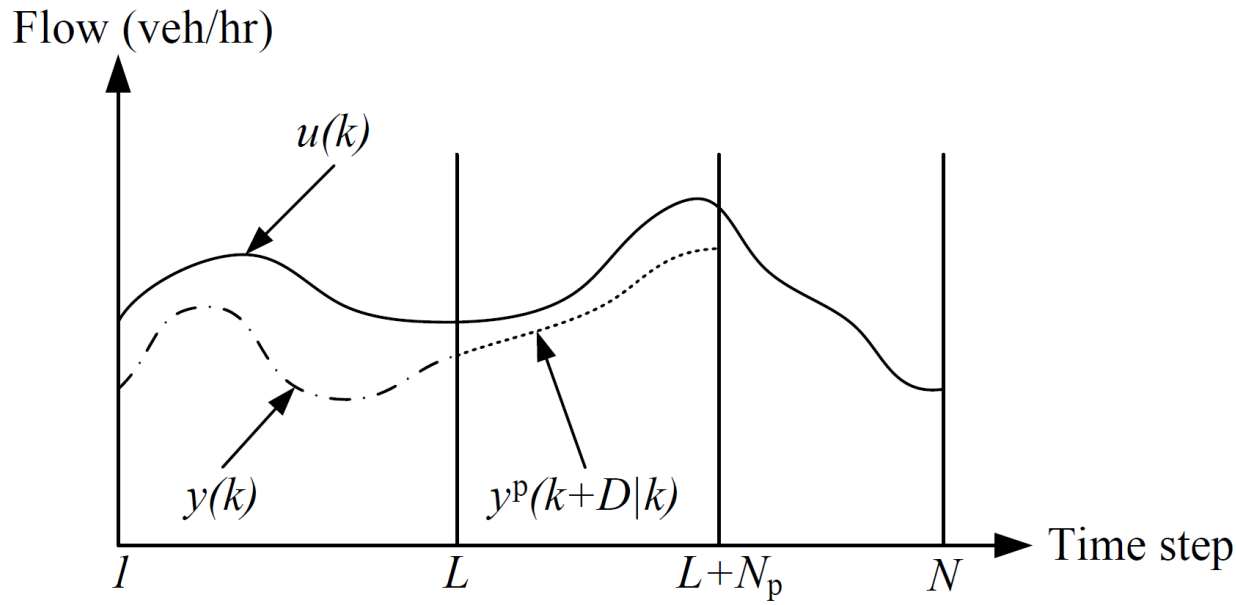

FIGURE 3 Concept of traffic flow prediction at time step L; dashed line: predicted traffic flow with horizon N_p; dash-dotted line: recent traffic flow measurements; solid line: nominal historical profile

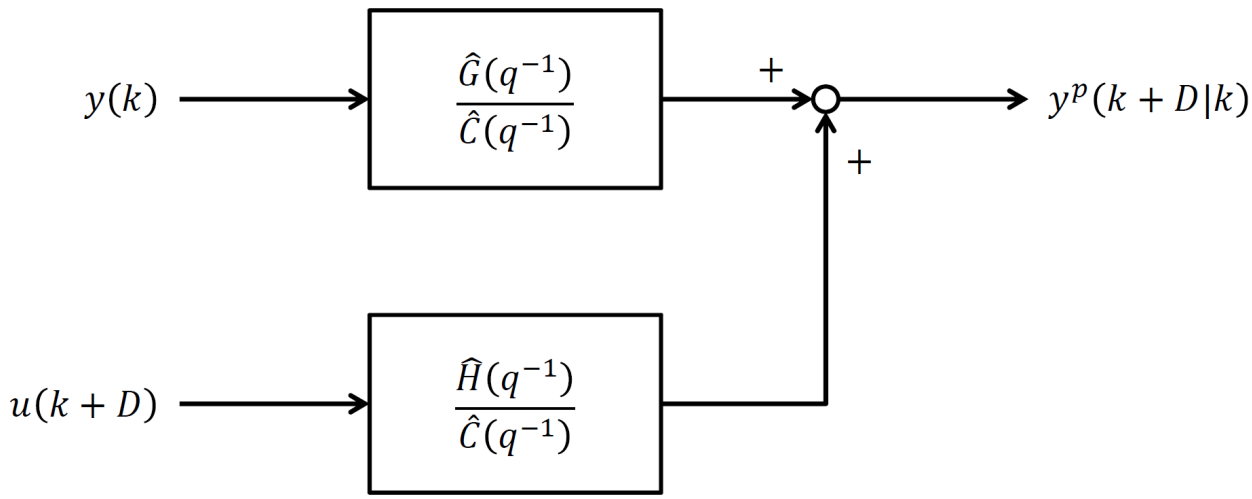

FIGURE 4 The optimal traffic flow predictor

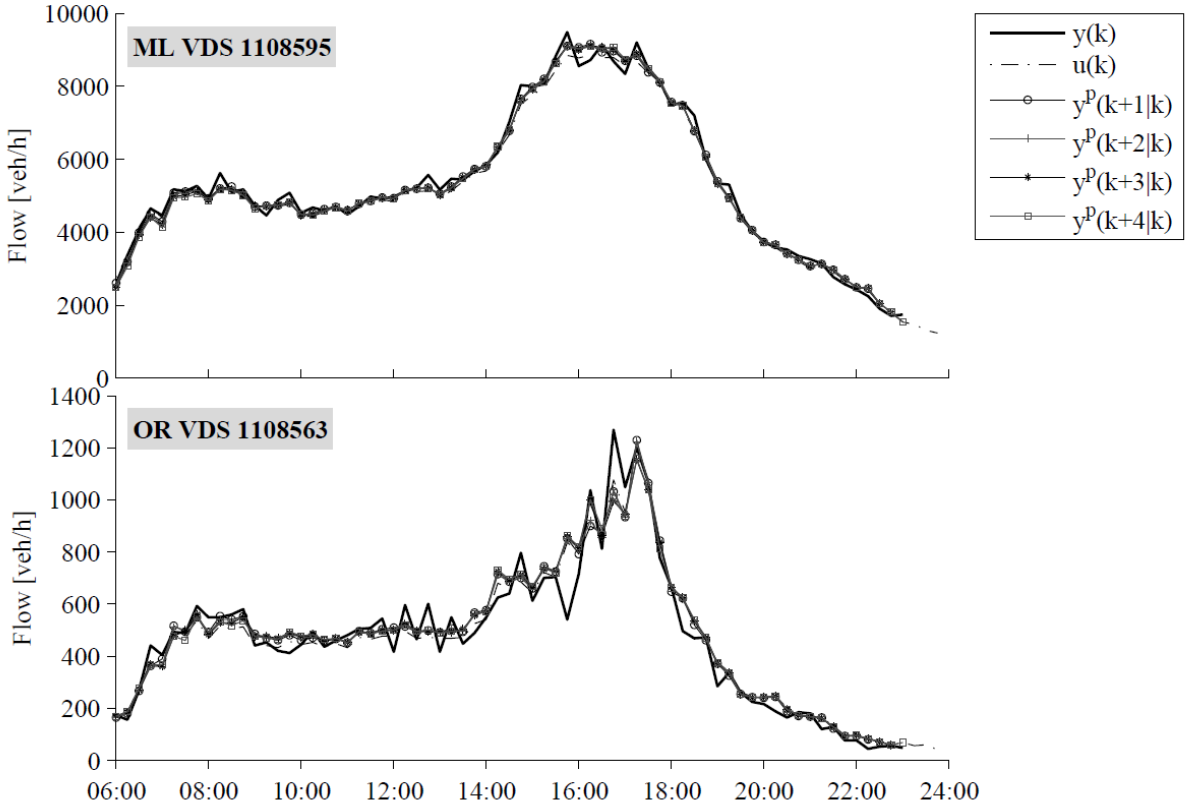

FIGURE 5 Flow prediction in weekday; top: mainline; bottom: on-ramp

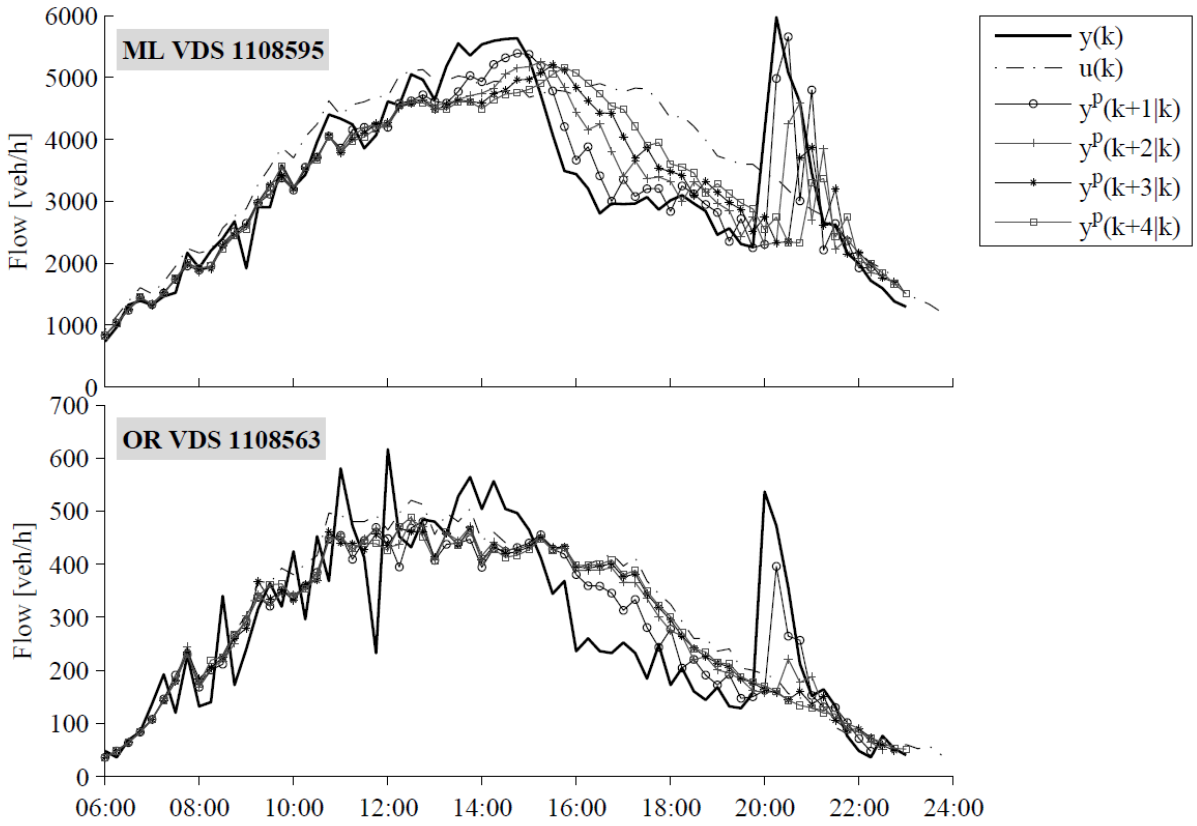

FIGURE 6 Flow prediction of Super Bowl Sunday; top: mainline; bottom: on-ramp

	Mainline MAPE [%]						
Predictor	Mon	Tue	Wed	Thu	Fri	Sat	Sun*
$y^{\mathrm{p}}(k+1\|k)$	6.4	5.9	6.2	5.1	6.7	6.3	10.3
$y^{\mathrm{p}}(k+2\|k)$	7.4	5.9	6.4	5.3	7.0	7.1	12.4
$y^{\mathrm{p}}(k+3\|k)$	7.8	5.9	6.7	5.5	7.0	7.2	14.5
$y^{\mathrm{p}}(k+4\|k)$	7.8	6.1	7.0	5.5	7.0	7.6	16.6
$u(k)$	8.0	7.0	7.6	5.9	7.4	7.9	21.5

*Super Bowl day

TABLE 1 Comparison of Proposed Predictor and Historical Predictor for Mainline VDS

	On-ramp MAPE [%]						
Predictor	Mon	Tue	Wed	Thu	Fri	Sat	Sun*
$y^{\mathrm{p}}(k+1\|k)$	22.1	21.9	23.1	21.9	28.1	31.7	31.3
$y^{\mathrm{p}}(k+2\|k)$	21.9	21.6	23.1	21.4	27.9	32.1	32.9
$y^{\mathrm{p}}(k+3\|k)$	21.8	22.3	22.7	21.6	27.0	31.6	33.6
$y^{\mathrm{p}}(k+4\|k)$	20.6	20.6	22.8	21.1	26.4	31.8	34.3
$u(k)$	22.8	21.9	24.0	21.9	28.8	32.3	38.7

*Super Bowl day

TABLE 2 Comparison of Proposed Predictor and Historical Predictor for On-ramp VDS