This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Real-Time Prediction of Seasonal Heteroscedasticity in Vehicular Traffic Flow Series

Wei Huang, Wenwen Jia, Jianhua Guo, Billy M. Williams, Guogang Shi, Yun Wei, and Jinde Cao, *Fellow, IEEE*

***Abstract*— Over the past decade, traffic heteroscedasticity has been investigated with the primary purpose of generating prediction intervals around point forecasts constructed usually by short-term traffic condition level forecasting models. However, despite considerable advancements, complete traffic patterns, in particular the seasonal effect, have not been adequately handled. Recently, an offline seasonal adjustment factor plus GARCH model was proposed in Shi *et al.* 2014 to model the seasonal heteroscedasticity in traffic flow series. However, this offline model cannot meet the real-time processing requirement proposed by real-world transportation management and control applications. Therefore, an online seasonal adjustment factors plus adaptive Kalman filter (OSAF+AKF) approach is proposed in this paper to predict in real time the seasonal heteroscedasticity in traffic flow series. In this approach, OSAF and AKF are combined within a cascading framework, and four types of online seasonal adjustment factors are developed considering the seasonal patterns in traffic flow series. Empirical results using real-world station-by-station traffic flow series showed that the proposed approach can generate workable prediction intervals in real time, indicating the acceptability of the proposed approach. In addition, compared with the offline model, the proposed online approach showed improved adaptability when traffic is highly volatile. These findings are important for developing real-time intelligent transportation system applications.**

***Index Terms*— Heteroscedasticity, seasonal adjustment factors, generalized autoregressive conditional heteroscedasticity (GARCH), adaptive Kalman filter.**

I. Introduction

TRAFFIC congestion has become a growing concern for transportation management agencies and traffic researchers around the world. Due to the limitation in constructing new roads in already crowded metropolitan areas, traffic management and control systems under the umbrella of intelligent transportation systems (ITS) are considered as one of the most effective methods to relieve traffic congestion. Compared to reactive systems that respond to currently observed traffic conditions, proactive systems call for an ability to anticipate traffic condition in the near future, yielding the so called short-term traffic forecasting problem.

Short term traffic condition forecasting includes level forecasting, i.e., point forecast generation, and quantification of uncertainty associated with the point forecast, i.e., prediction interval generation [1]. For traffic condition level forecasting, the variables used in the analysis are mainly the means or averages of the traffic condition characteristics, e.g., mean traffic speed, mean traffic flow rate, or mean traffic occupancy. Statistically, level forecasting is to analyze the first order conditional moment of the traffic condition series. Although traffic condition level forecasting is still a hot research topic with on-going debates found in the literature, time series models have received sustained attention for their outstanding forecasting performance and sound theoretical foundation [2]–[14]. The theoretical justification of the time series method for modeling the traffic condition series was firstly pointed out in [12] through recognizing the Wold Decomposition Theorem, and the seasonal autoregressive integrated moving sverage (SARIMA) model was proposed to model the 15-minute traffic flow series from M25 in London, in which weekly pattern is deemed as the primary seasonal effect and is handled through weekly differencing.

For traffic condition uncertainty quantification, the variables used are usually in terms of traffic condition variance, and statistically, the uncertainty analysis deals primarily with the second order conditional moment of the traffic condition series. Traffic uncertainty analysis can also be considered as heteroscedasticity modeling, and traffic conditional series has been shown to demonstrate heteroscedasticity in [15]. In this regard, generalized autoregressive conditional heteroscedasticity (GARCH) model and stochastic volatility (SV) model are two major types of methods. The GARCH models, proposed in [16] through generalizing the autoregressive conditional heteroscedasticity (ARCH) model proposed in [17], have been investigated in a number of studies. Using 15-minute traffic flow series as the representative, [18] applied the GARCH model for modeling the conditional variance of traffic flow series and an online algorithm based on Kalman filter was also designed to process the GARCH model, generating workable prediction intervals in term of kickoff percentage

Manuscript received March 6, 2017; revised August 7, 2017; accepted November 2, 2017. This work was supported by the National Natural Science Foundation of China under Grant 61573106. The Associate Editor for this paper was Y. Wang. *(Corresponding author: Wei Huang.)*

W. Huang, W. Jia, and J. Guo are with the Intelligent Transportation System Research Center, Southeast University, Nanjing 210096, China (e-mail: seuhwei@126.com; 879627553@qq.com; jg2nh@yahoo.com).

B. M. Williams is with the Department of Civil, Construction, and Environmental Engineering North Carolina State University, Raleigh, NC 27695 USA (e-mail: billy_williams@ncsu.edu).

G. Shi is with the Transportation College, Southeast University, Nanjing 210096, China (e-mail: kkbb0507@qq.com).

Y. Wei was with the Intelligent Transportation System Research Center, Southeast University, Nanjing 210096, China. He is now with the Beijing Urban Construction Design & Development Group Company Ltd., Beijing 100037, China (e-mail: luckyboy0309@163.com).

J. Cao is with the Department of Mathematics, Southeast University, Nanjing 210096, China, and also with the Research Center for Complex Systems and Network Sciences, Southeast University, Nanjing 210096, China (e-mail: jdcao@seu.edu.cn).

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.2017.2774289

and the interval width-to-level ratio; further in [19], the 15-minute time interval was extended to a spectrum of data collection time intervals. Reference [5] applied the GARCH model in modeling the relative velocity, and [20] applied the ARMA + GARCH model in forecasting link travel time variability with the generalized Pareto distribution (GPD) incorporated into the ARMA + GARCH model to obtain a more realistic percentile of travel time; however, no online algorithm was proposed in these studies. Reference [21] proposed to use the ARIMA + GARCH and the autoregressive fractionally integrated moving average plus fractionally integrated asymmetric power autoregressive conditional heteroscedasticity (ARFIMA + FIAPARCH) for forecasting traffic volatility in real time for urban networks with the online approach constructed through recursively performing the quasi-maximum likelihood estimation method provided in commercial software. Reference [22] applied the GARCH model to predict the conditional variance of traffic speed series using layered Kalman filters. Reference [23] applied the GARCH model for modeling the reliability of travel time forecasting. Reference [24] applied the adaptive Kalman filters to SARIMA + GARCH structure to update the process variance of the state space model. Reference [25] also applied ARIMA + GARCH model on modeling traffic flow series. Other GARCH type models include multivariate GARCH model [26] and asymmetric GARCH models [27], [28]. For the SV models, [29] proposed a discrete-time parametric stochastic volatility model based on a latent stochastic process to provide short-term forecasts of traffic speed variability. Reference [30] applied ARFIMA + SV model to predict the average and the prediction interval of travel time, and the experiment showed that the ARFIMA+SV model outperforms the ARIMA+GARCH model. However, in terms of evaluating the likelihood function, SV model is more difficult than GARCH model. In summary, though demonstrating considerable advancements, all these models are not able to account for the complete traffic patterns, especially, the seasonality pattern that has been shown to be significant in traffic condition series [14].

Several studies have been proposed in modeling seasonal heteroscedasticity in traffic condition series. Reference [31] proposed a seasonal adjustment factor approach for deseasonalizing the seasonal effect before applying the GARCH model, and using the real world traffic condition series, the deseasonalizing effect was shown to be effective and the overall heteroscedasticity modeling procedure was shown to outperform the conventional GARCH model. However, the analysis is an offline procedure, which cannot meet the online processing requirement. In addition, [32] proposed two component GARCH models to model trend and seasonal components in travel time through decomposition.

In this paper, targeting the requirement of real time processing proposed by real world transportation applications, an online seasonal adjustment factor plus adaptive Kalman filter (OSAF + AKF) model is proposed to predict in real time the seasonal heteroscedasticity in traffic flow series. The performance of the proposed OSAF + AKF model will be investigated and compared with those of the offline model. In addition, it is important to clarify that the input to the proposed model is the level removed residual series. The rest of the paper is organized as follows. First, necessary statistical background is briefly presented for SARIMA and GARCH. Then, the proposed OSAF+AKF model is presented. Afterward, data used in this study are described, followed by a presentation of empirical results, including the performance of the OSAF + AKF model, the comparisons between the online model and the offline model, the disaggregated prediction interval generation demonstration, and the computational efficiency of the proposed online models. Finally, the paper concludes with summaries and discussions.

II. Statistical Background

In this section, statistical backgrounds are briefly presented for the SARIMA model and the GARCH model. Interested readers are referred to [33], [34], and [36] for more statistical treatments of these models.

A. SARIMA

SARIMA is a member of Box-Jenkins family of time series models and has been extensively used in many disciplines such as industrial engineering, advanced automatic control, and financial engineering [33], [34]. This method assumes that a linear combination of past observations and forecasting errors can capture the autocorrelation structure of the time series, thereby satisfactory forecasts can be constructed [10], [12], [14]. For a discrete traffic flow series x_t, the SARIMA$(p, d, q)(P, D, Q)_s$ structure is defined as Eq.(1).

$$\varphi(B)\Phi(B^s)(1-B^s)^D(1-B)^d x_t = \theta(B)\Theta(B^s)\varepsilon_t, \quad (1)$$

where t = time index; B = backshift operator such that $Bx_t = x_{t-1}$; p = order of the short-term autoregressive polynomial; q = order of the short-term moving average polynomial; d = order of the short-term differencing; p = order of the seasonal autoregressive polynomial; Q = order of the seasonal moving average polynomial; D = order of the seasonal differencing; S = seasonal period; $(1-B^s)^D$ = seasonal differencing; $(1-B^s)^d$ = short term differencing; $\phi(B) = 1-\phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ = short-term autoregressive polynomial; $\theta\ (B) = 1-\theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ = short-term moving average polynomial; $\Phi\ (B^s) = 1 - \Phi_1(B^s) - \Phi_2(B^s)^2 - \cdots - \Phi_p(B^s)^p$ = seasonal autoregressive polynomial; $\Theta(B^s) = 1-\Theta_1(B^s) - \Theta_2(B^s)^2 - \cdots - \Theta_Q(B^s)^Q$ = seasonal moving average polynomial.

For the SARIMA model defined in Eq.(1), the roots of $\phi(B)$, $\theta\ (B)$, $\Theta\ (B)$, and $\Phi\ (B^s)$ are expected to be outside of the unit circle so as to ensure the causality and invertibility of the model, which indicates that the process can be expressed in terms of past innovations and past observations, respectively [33], [34], [36]. In addition, these polynomials have no common factors. $\varepsilon\ (t)$ is the residual series, conventionally assumed to be Gaussian white noise with mean zero and constant variance.

For the SARIMA structure, previous studies have shown that the SARIMA$(1, 0, 1)(0, 1, 1)_{672}$ model can be used to fit the traffic flow rate series aggregated at 15-minute interval [12], [14]. In this paper, the SARIMA$(1,0,1)(0,1,1)_{672}$ model is used to capture and hence remove the first-order autocorrelation structure of the traffic flow series.

B. GARCH

GARCH is a commonly applied approach of modeling the second order conditional moment of heteroscedastic series. This approach was first proposed in [17] in the form of ARCH model, and then the ARCH model was generalized into the GARCH model in [16]. Compared with the SARIMA model in which the residual series is assumed to be having constant variance, termed as homoscedasticity, the GARCH model assumes the residual series ε_t to be having non-constant variance. In time series modeling, the residual series ε_t is assumed to follow the GARCH(u, v) model, defined as follows:

$$\varepsilon_t = \sqrt{h_t} e_t, \tag{2}$$

$$h_t = \alpha_0 + \sum_{i=1}^{\upsilon} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{u} \beta_i h_{t-i}, \tag{3}$$

$$e_t \sim IN(0,1), \tag{4}$$

where t = time index; h_t = conditional variance at t, i.e., $\varepsilon_t/\Psi_{t-1} \sim N(0,h_1)$ with Ψ_{t-1} as the information up to $t-1$; u = autoregressive order of GARCH process with $u > 0$; υ = moving average order of GARCH process with $\upsilon > 0$; α_0 = positive constant coefficient; $\alpha_{i,i-1,\cdots,\upsilon}$ = non-negative coefficients of the lagged sample variance ε_{t-i}^2; $\beta_{i,i-1,\cdots,\upsilon}$ = non-negative coefficients of the lagged conditional variance h_{t-i}; e_t = independent normal variable with zero mean and unit variance, i.e., $IN(0,1)$. For the GARCH(u,v) process, the order of u and v can be identified theoretically using the conventional information criteria such as the Akaike information criterion (AIC), however, in practice, many analysts assume that GARCH(1,1) would be adequate for modeling the second order conditional variance and in [18], the GARCH(1,1) model was compared with other GARCH models with different orders, showing insignificant differences between these models in modeling real world heteroscedastic traffic flow series. Therefore, in this paper, the GARCH(1,1) model will be used for modeling the conditional variance of the traffic flow series. Finally, it is also important to mention that the error series e_t is assumed to be normal for facilitating the construction of prediction intervals.

III. Proposed Methodology

Considering the real time processing requirement in real world transportation applications, an online seasonal adjustment factor plus adaptive Kalman filter (OSAF + AKF) approach is proposed in this section based on previous work in [31], including online seasonal adjustment factor computation, adaptive Kalman filter development, and integrated seasonal heteroscedasticity modeling. For this approach, the seasonal heteroscedasticity effect will be first removed in an online fashion through applying the real time computed seasonal adjustment factor, and then the remaining heteroscedasticity will be modeled using the adaptive Kalman filter constructed based on the GARCH model.

A. Online Seasonal Adjustment Factor Computation

The purpose of this subsection is to introduce an online seasonal adjustment factor computation algorithm through implementing a rolling horizon based updating mechanism so that the seasonal heteroscedasticity effect can be first removed from the heteroscedastic traffic flow series in an online fashion. As proposed in [31], the four types of seasonal adjustment factors are defined as Eqs.(5-8), including daily adjustment factor (DF), weekly adjustment factor (WF), log daily adjustment factor (LNDF), and log weekly adjustment factor (LNWF).

$$DF_n^2 = \frac{1}{D}\sum_{i}^{D} \varepsilon_{i,n}^2, \tag{5}$$

$$LNDF_n^2 = exp[\frac{1}{D}\sum_{i}^{D} ln(\varepsilon_{i,n}^2)], \tag{6}$$

$$WF_{d,n}^2 = \frac{1}{N_d}\sum_{s\in S_d} \varepsilon_{i,n}^2, \tag{7}$$

$$LNWF_{d,n}^2 = exp[\frac{1}{N_d}\sum_{s\in S_d} ln(\varepsilon_{i,n}^2)], \tag{8}$$

where D = number of days in the traffic series; N = number of intraday intervals with $N = 96$ for 15-minute time interval; n = intraday time interval index, $N = 1,\cdots,N$; $\varepsilon_{i,n}$ = traffic data for time interval n at day i; d = day of week index, i.e., $d = 1$ for Sunday, $d = 2$ for Monday, $d = 3$ for Tuesday, $d = 4$ for Wednesday, $d = 5$ for Thursday, $d = 6$ for Friday, and $d = 7$ for Saturday in this study; S_d = 1 set of days with same day of week index d, e.g., all the Sunday in traffic data series for $d = 1$; $\varepsilon_{s,n}$ = traffic data for time interval n at the s^{th} day in S_d; N_d = number of $\varepsilon_{s,n}$ values found in S_d; DF_n and $LNDF_n$ = daily adjustment factors at intraday time interval n; $WF_{d,n}$ and $LNDF_{d,n}$ = weekly adjustment factors at intraday time interval n of day d of the week.

Based on above definitions, the online seasonal adjustment factor computation algorithms can be developed within a rolling horizon framework, in which a fixed period is selected and rolling forward across each time slice to compute the seasonal adjustment factors accordingly. Therefore, Eqs.(5-8) can be reformulated into recursive forms. Given $\varepsilon_{s,d,n}^2$ as as the traffic data for time interval n at the S^{th} day in S_d, the online daily adjustment factors, denoted by $ODF_{i,n}$ and $OLNDF_{i,n}$ for time interval n at day i, and the online weekly adjustment factors, denoted by $OWF_{s,d,n}$ and $OLNWF_{s,d,n}$ for time interval n at the s^{th} day in S_d are defined as

$$ODF_{i+1,n}^2 = ODF_{i,n}^2 + \frac{1}{D}(\varepsilon_{i+1,n}^2 - \varepsilon_{i,n}^2)], \tag{9}$$

$$\begin{aligned} OLNDF_{i+1,n}^2 &= OLNDF_{i,n}^2 + exp\{\frac{1}{D}[ln(\varepsilon_{i+1,n}^2) \\ &\quad - ln(\varepsilon_{i,n}^2)]\}, \end{aligned} \tag{10}$$

$$\begin{aligned} OWF_{s,d,n}^2 &= OWF_{s,d,n}^2 + \frac{1}{N_d}[ln(\varepsilon_{s+1,d,n}^2) \\ &\quad - ln(\varepsilon_{s,d,n}^2)], \end{aligned} \tag{11}$$

$$\begin{aligned} OLNWF_{s+1,d,n}^2 &= OLNWF_{s,d,n}^2 + exp\{\frac{1}{N_d}[ln(\varepsilon_{s+1,d,n}^2) \\ &\quad - ln(\varepsilon_{s,d,n}^2)]\}. \end{aligned} \tag{12}$$

B. Adaptive Kalman Filter Development

The purpose of this subsection is to introduce the adaptive Kalman filter that will be applied to model the remaining heteroscedasticity left after deseasonalizing the traffic condition series using the online computed seasonal adjustment factor. Note that there have been publications on constructing the adaptive Kalman filter based on the GARCH model. e.g., [24], and the following treatments are presented for the completeness of the proposed approach.

First, the GARCH(1,1) model is converted into the state space model. Taking $\eta_t = \varepsilon_t - h_t$ into Eq.(3), the GARCH process can be interpreted as an autoregressive moving average (ARMA) process in ε_t^2 as Eq.(13), i.e.

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{n}(\alpha_i + \beta_i)\varepsilon_{t-i}^2 - \sum_{i=1}^{n}\beta_i \eta_{t-i} + \eta_t, \qquad (13)$$

where η_t is serially uncorrelated with mean zero and $n = \max(u, v)$ [16]. Therefore, for the GARCH(1,1) model, we have Eq.(14) as

$$\varepsilon_t^2 = \alpha_0 + \alpha\varepsilon_{t-1}^2 + \beta\eta_{t-i} + \eta_t, \qquad (14)$$

where $\beta = -\beta_1$, $\alpha = \alpha_1 + \beta_1$. Then, the state transition and observation equations can be defined as Eqs.(15) and (16) as below, respectively.

$$x_t = \Phi x_{t-1} + z_t, \qquad (15)$$

$$Y_t = H_t x_t + \eta_t, \qquad (16)$$

where x_t = state variable $(\alpha_0, \alpha, \beta)^T$; Φ = state transition matrix defined as $diag\{\lambda^{-\frac{1}{2}}\}$, with λ defined as a forgetting factor; z_t = state noise series with state noise covariance matrix $Cov(z_t z_t^T) = Q_t$; Y_t = current observation defined as ε_t^2; H_t = time varying observation matrix defined as $(1\ \varepsilon_{t-1}^2\ \eta_{t-1})^T$; η_t = observation noise series with observation noise covariance matrix $Cov(\eta_t \eta_t^T) = R_t$.

Combined, Eqs.(15) and (16) constitute the state space representation for the GARCH(1,1) model. The reparameterized parameter vector $(\alpha_0\ \alpha\ \beta)^T$ is treated as the hidden state and the squared residual series after applying the SARIMA model is treated as the driving observation process.

Second, the state space model is solved through adaptive Kalman filter. Compared with the conventional Kalman filter that assumes constant process variance, adaptive Kalman filter can update the process variance in real time. In this paper, the process variance adaptation mechanism is adopted from the mechanism proposed in [39].

In summary, given the state space model defined in Eqs.(15) and (16) and assume the initial state variable estimate as $\hat{x}_0$, posterior state estimation error covariance estimation as $\hat{P}_0^+$, observation noise covariance as R_0, state noise covariance as Q_0, average observation errors as $\hat{\eta}_0$, and average system state estimation errors as $\hat{z}_0$, the adaptive Kalman filter is performed as below.

Step 1: State propagation and prior state estimation error covariance estimation

$$\hat{x}_{t|t-1}^- = \Phi\hat{x}_{t-1|t-1}^+, \qquad (17)$$

$$\hat{P}_{t|t-1}^- = \Phi\hat{P}_{t-1|t-1}^+\Phi^T + Q_{t-1}. \qquad (18)$$

Step 2: Update observation noise covariance matrix R_t,

$$\eta_t = Y_t - H_t\hat{x}_{t|t-1}^+, \qquad (19)$$

$$\hat{\eta}_t = \frac{1}{N}\sum_{i=1}^{N}\eta_{t-j+1}, \qquad (20)$$

$$R_t = \frac{1}{N}\sum_{i=1}^{N}\{(\eta_{t-j+1} - \hat{\eta}_t)(\eta_{t-j+1} - \hat{\eta}_t)^T - \frac{N-1}{N}H_{t-j+1}\hat{P}_{t-j+1|t-j}^- H_{t-j+1}^T\}. \qquad (21)$$

Step 3: Kalman gain computation

$$K_t = \frac{\hat{P}_{t|t-1}^- H_{t-1}}{H_{t-1}^T\hat{P}_{t|t-1}^- H_{t-1} + R_t}. \qquad (22)$$

Step 4: Posterior state estimation and posterior state estimation error covariance estimation

$$\hat{x}_{t|t}^+ = \hat{x}_{t|t-1}^- + K_t(Y_t - H_{t-1}^T\hat{x}_{t|t-1}^-), \qquad (23)$$

$$\hat{P}_{t|t}^+ = (I - K_t H_{t-1}^T)\hat{P}_{t|t-1}^-). \qquad (24)$$

Step 5: Update state noise covariance matrix Q_t

$$z_t = x_{t|t}^+ - \Phi x_{t-1|t-1}^+, \qquad (25)$$

$$\hat{z}_t = \frac{1}{N}\sum_{j=1}^{N} z_{t-j+1}, \qquad (26)$$

$$Q_t = \frac{1}{N}\sum_{j=1}^{N}\Big\{(z_{t-j+1} - \hat{z}_t)(z_{t-j+1} - \hat{z}_t)^T - \frac{N-1}{N}[\Phi\hat{P}_{t-j|t-j}^+\Phi^T - \hat{P}_{t-j+1|t-j+1}^+]\Big\}. \qquad (27)$$

where $\hat{P}_{t|t-1}^-$ = prior state estimation error covariance; $\hat{P}_{t|t}^+$ = posterior state estimation error covariance; K_t = Kalman gain at time t; η_t = observation errors; $\tilde{\eta}_t$ = average observation errors; z_t = system state estimation errors; $\hat{z}_t$ = average system state estimation errors; N = prescribed memory size of the adaptive Kalman filter.

C. Integrated Seasonal Heteroscedasticity Modeling

In this subsection, an overall real time procedure is proposed. First, the seasonal adjustment factors will be computed in real time and applied to deseasonalize the traffic series through dividing the traffic series using the corresponding factors. Then, the deseasonalized traffic series will be filtered using the adaptive Kalman filter to generate the heteroscedasticity or volatility estimation in the deseasonalized scale. Finally, the results will be converted back to the original scale through multiplying the estimated volatility by the online seasonal adjustment factors, based on which, the prediction intervals can be generated upon a normality assumption. In summary, the general steps of the proposed integrated real time seasonal heteroscedasticity modeling procedure are described as follows.

Step 1: Use Eqs.(9-12) to obtain the online seasonal adjustment factors;

Step 2: Use Eq.(28) to deseasonalize the traffic series;

$$w_t = \varepsilon_t/(\text{online seasonal adjustment factor}). \qquad (28)$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

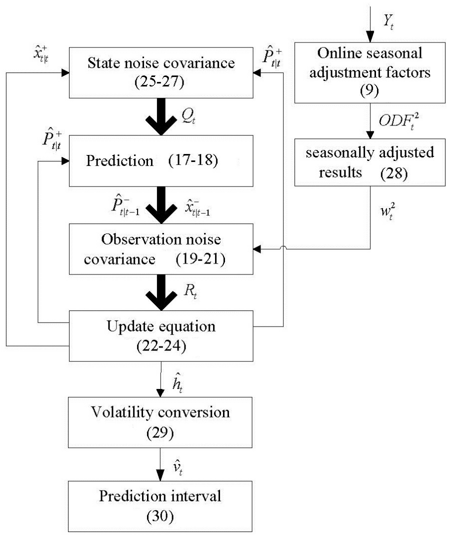


Fig. 1. Flow chart of the real time seasonal heteroscedasticity modeling approach.

Step 3: Use adaptive Kalman filter, i.e., Eqs.(17-27) to model the deseasonalized traffic series with w_t^2 as current observation and $(1\ w_{t-1}^2\ \eta_{t-1})^T$ as the time varying observation matrix H_t so as to estimate the traffic conditional volatility $\hat{h}_t$;

Step 4: Convert the estimated volatility back to original scale;

$$\hat{\upsilon}_t = \hat{h}_t \cdot (\text{online seasonal adjustment factor}). \tag{29}$$

Step 5: Generate prediction interval for time t, denoted by PI_t, at 95% significant level;

$$PI_t = [m_t - 1.96\hat{\upsilon}_t, m_t + 1.96\hat{\upsilon}_t]. \tag{30}$$

where ε_t = original traffic series before seasonal adjustment; w_t = deseasonalized traffic series; $\hat{h}_t$ = estimated traffic condition volatility from adaptive Kalman filter model for deseasonalized traffic series; $\hat{\upsilon}_t$ = traffic condition volatility estimated for unadjusted traffic series; m_t = mean traffic flow at time t generated by SARIMA$(1, 0, 1)(0, 1, 1)_{672}$.

According to the abovementioned general steps and taking the ODF+AKF as the example, the detailed flow chart of performing the real time seasonal heteroscedasticity modeling in time t is shown in Fig.1. According to Fig.1, first, at time t, the observation Y_t, i.e., the squared residual at time t denoted by ε_t^2, is used to compute the seasonal adjustment factor at time t, i.e., ODF_t^2. Second, ε_t^2 is converted into w_t^2 using the seasonal adjustment factor at time t. Then the volatility $\hat{h}_t$ is computed by adaptive Kalman filter. Finally, volatility in the original non-deseasonalized scale of $\hat{\upsilon}_t$ is computed, based on which the prediction interval at time t , i.e., PI_t is computed. Note that after initialization of the algorithm, at each time step t, the input of this process is Y_t, and the computation will then be carried on sequentially with respect to the flow chart.

IV. Data

In this section, the traffic data used in this paper will be described, including data collection, residual acquisition, and performance measures.

A. Data Collection

Real world traffic flow data collected from the motorway system in the United Kingdom and the metropolitan freeway systems of Minnesota in the United States are used in this study. Though originally archived at different time intervals, in this study, these data are aggregated into 15-minute time intervals according to [35]. In performing the aggregation, missing values were propagated upward, and simple screening procedures, i.e., a threshold test and a hang-on test, were applied to eliminate the obvious erroneous data. For the hang-on test, if the 15-minute traffic flow outputs from a certain detector stick to a single value for at least 2 hrs, i.e., at least eight consecutive equivalent traffic flow values, the eighth and all the following outputs with the same value are deemed suspicious and treated as missing. More information can be found in [18] on these screening procedures. In this study, 24 long (yearly or almost-yearly series) traffic flow series were selected as representatives for these two regions across different years with the aggregation time interval set to be 15-minute. In other words, for each series, one day of traffic data will have 96 data points, and one week of traffic data will have 672 data points. Note that the Minnesota data can be accessed publicly through visiting http://www.d.umn.edu/~tkwon/TMCdata/TMCarchive.html, while the UK data was obtained through contacting directly the Highway Agency of the United Kingdoms.

B. Residual Acquisition

Recall that the input series for the proposed OSAF + AKF approach are the residual series with the autocorrelation structure removed from the original traffic condition series, and in this subsection, the residual acquisition procedure is presented. In doing so, the SARIMA$(1, 0, 1)(0, 1, 1)_{672}$ model was applied to remove the autocorrelation structure from the original traffic condition series considering its established performance of modeling the 15-minute traffic flow series [14]. When applying the model, the conventional Box-Jenkins time series modeling procedure was followed to make sure the autocorrelation structure of the original traffic flow series has been adequately removed. Therefore, we will have 24 15-minute traffic flow residual series to be used in this study.

Furthermore, the 24 traffic flow residual series were further divided into two separate datasets, i.e., the initialization dataset and the evaluation dataset. The purpose of the initialization dataset is to start the proposed OSAF+AKF procedure. In this way, the length of the initialization dataset will be the rolling period for computing the four seasonal adjustment factors, and the proposed OSAF + AKF procedure will run into the evaluation dataset so that the performance of the proposed data can be computed. The evaluation dataset is also used for computing the performance for the conventional offline model so that the proposed model can be appropriately compared with the offline model. In summary, the overview of the 24 residual series is shown in Table 1.

C. Performance Measures

The performance of the prediction intervals generated by the proposed OSAF + AKF approach can be evaluated using

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
OVERVIEW OF THE RESIDUAL SERIES

				Initialization dataset			Evaluation dataset		
Region	Highway	Station	No. of Lanes	Start	End	Sample Size	Start	End	Sample Size
UK	M25	4565a	4	1/1/2002	9/30/2002	26,208	10/1/2002	12/31/2002	8832
UK	M25	4680b	4	1/1/2002	9/30/2002	26,208	10/1/2002	12/31/2002	8832
UK	M1	2737a	3	2/13/2002	9/30/2002	22,080	10/1/2002	12/31/2002	8832
UK	M1	2808b	3	2/13/2002	9/30/2002	22,080	10/1/2002	12/31/2002	8832
UK	M1	4897a	3	2/13/2002	9/30/2002	22,080	10/1/2002	12/31/2002	8832
UK	M6	6951a	3	1/1/2002	9/30/2002	26,208	10/1/2002	12/31/2002	8832
MN	I35W-NB	60	4	1/1/2000	9/30/2000	26,304	10/1/2002	12/31/2002	8832
MN	I35W-SB	578	3	1/1/2000	9/30/2000	26,304	10/1/2002	12/31/2002	8832
MN	I35E-NB	882	3	1/1/2000	9/30/2000	26,304	10/1/2002	12/31/2002	8832
MN	I35E-SB	890	3	1/1/2000	9/30/2000	26,304	10/1/2002	12/31/2002	8832
MN	169-NB	442	2	1/1/2000	9/30/2000	26,304	10/1/2002	12/31/2002	8832
MN	169-SB	737	2	1/1/2000	9/30/2000	26,304	10/1/2002	12/31/2002	8832
MN	I35W-NB	60	4	1/1/2004	9/30/2004	26,304	10/1/2002	12/31/2002	8832
MN	I35W-SB	578	3	1/1/2004	9/30/2004	26,304	10/1/2002	12/31/2004	8832
MN	I35E-NB	882	3	1/1/2004	9/30/2004	26,304	10/1/2002	12/31/2004	8832
MN	I35E-SB	890	3	1/1/2004	9/30/2004	26,304	10/1/2002	12/31/2004	8832
MN	169-NB	442	2	1/1/2004	9/30/2004	26,304	10/1/2002	12/31/2004	8832
MN	169-SB	737	2	1/1/2004	9/30/2004	26,304	10/1/2002	12/31/2004	8832
MN	I35W-NB	60	4	1/1/2015	9/30/2015	26,208	10/1/2015	12/31/2015	8832
MN	I35W-SB	578	3	1/1/2015	9/30/2015	26,208	10/1/2015	12/31/2015	8832
MN	I35E-NB	882	3	1/1/2015	9/30/2015	26,208	10/1/2015	12/31/2015	8832
MN	I35E-SB	890	3	1/1/2015	9/30/2015	26,208	10/1/2015	12/31/2015	8832
MN	169-NB	442	2	1/1/2015	9/30/2015	26,208	10/1/2015	12/31/2015	8832
MN	169-SB	737	2	1/1/2015	9/30/2015	26,208	10/1/2015	12/31/2015	8832

two measures, namely the kickoff percentage and the width to flow ratio [18], [19], [22]. Note that [38] proposed a measure of prediction interval coverage probability (PIPC) that is similar to the selected measure of kickoff percentage in this paper. In fact, these two measures are complementary in theory.

The kickoff percentage is defined as the total number of traffic flow observations falling outside of the corresponding prediction intervals divided by the total number of traffic flow observations, as shown in Eq.(31).

$$P_{kf} = N_{kf}/N * 100/\%, \tag{31}$$

where P_{kf} = kickoff percentage; N_{kf} = total number of traffic flow observations falling outside of the corresponding prediction intervals; N = total number of traffic flow observations.

The width to flow ratio is defined as the average of width to flow ratios for all the prediction intervals, as shown in Eq.(32).

$$r = \frac{c}{f}, \tag{32}$$

where r = width to flow ratio; c = prediction interval width; f = corresponding mean traffic level.

In this study, 95% significant level is adopted when interpreting these measures. Statistically, for the 95% significant level, the kickoff percentage is expected to be close to 5%, and under this condition, the width to flow ratio is expected to be small for yielding more precise quantification on the uncertainty associated with the point forecasts. In order to show the detailed performance in terms of the two measures, the evaluating traffic condition data are grouped by traffic flow levels and time of day as defined in Tables 2 and 3, respectively. In this way, the performance measures are computed for each traffic series with respect to each group, and then the average measures for all the traffic data series can be obtained through averaging the measures of each group across all the traffic series with respect to group definition.

In the end, the average computation time of each online model processing each traffic data observation will be computed for each data series in order to show the computational efficiency of these online models. Note this is critical for supporting the real time applications that are common for intelligent transportation systems.

TABLE II
GROUPS BY TRAFFIC LEVEL

Group	Group description
TTL	For all traffic observations
L1	$\geq$ 0 and $<$500 veh/h/ln
L2	$\geq$ 500 and $<$1000 veh/h/ln
L3	$\geq$ 1000 and $<$1500 veh/h/ln
L4	$\geq$ 1500 and $<$2000 veh/h/ln
L5	$\geq$ 2000 veh/h/ln

TABLE III
GROUPS BY TIME OF DAY

Group	Group description
T1	[00:00-04:00)
T2	[04:00-06:00)
T3	[06:00-07:00)
T4	[07:00-08:00)
T5	[08:00-09:00)
T6	[09:00-10:00)
T7	[10:00-12:00)
T8	[12:00-14:00)
T9	[14:00-16:00)
T10	[16:00-17:00)
T11	[17:00-18:00)
T12	[18:00-19:00)
T13	[19:00-20:00)
T14	[20:00-22:00)
T15	[22:00-24:00)

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

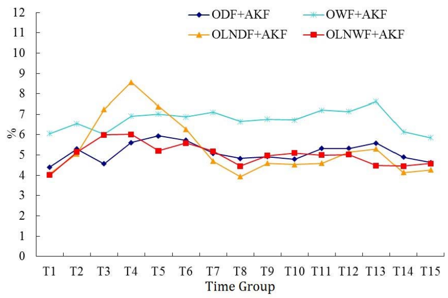


Fig. 2. Kickoff percentage with respect to time of day.

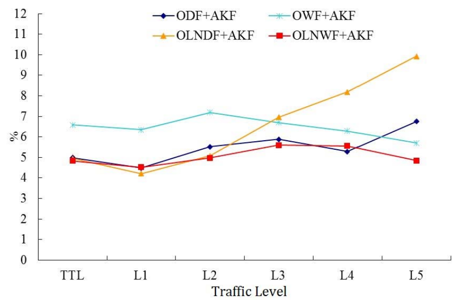


Fig. 3. Kickoff percentage with respect to traffic levels.

V. Empirical Results

Empirical test results are presented in this section. Note that in the rest of the paper, the acronym of BATCH is adopted for denoting the conventional offline approach of the seasonal adjustment factor plus GARCH model.

A. Aggregated OSAF + AKF Performance Investigation

The purpose of this test is to show the performances of the proposed OSAF + AKF algorithm and to find which specific online model generates the best performance. First, the performances of the proposed approaches in terms of kickoff percentage with respect to time of day and traffic levels are presented in Figs.2 and 3, respectively. On observing Figs.2 and 3, it is clear that all the proposed OSAF + AKF models can generate the kickoff percentage around 5%, indicating that workable prediction intervals can be generated using the proposed online approach. This is desirable in showing that the proposed online computing mechanisms are acceptable for modeling the seasonal heteroscedasticity in the traffic condition series. In addition, on comparing the performance of the four OSAF + AKF models, OLNWF + AKF can be seen to generate in general better prediction intervals in terms of kickoff percentage than the rest three models. Specifically, on observing Fig.3, OLNWF+AKF and ODF+AKF are showing comparable performance in general for all groups except for L5, for which OLNWF+AKF clearly outperforms ODF+AKF, while in Fig.2, OLNWF + AKF outperforms ODF+AKF when traffic is running for afternoon peak hour groups T9, T10, T11, and T12, and for group T13 as well. The above observations can be substantiated by looking at the quantitative performance for OLNWF + AKF and ODF+AKF shown in Table 4 and Table 5 with respect to time of day and traffic levels, respectively. This is consistent with the findings in previous studies on supporting the weekly pattern in traffic condition series [12], [14], [19], and this is also in agreement with the findings in [31] for supporting the log weekly adjustment factor.

Second, the performance of OSAF + AKF in terms of prediction interval width-to-flow ratio with respect to different time of day and traffic levels are shown in Figs.4 and 5, respectively. On observing Figs.4 and 5, it can be seen that

TABLE IV

Kickoff Percentage With Respect to Time of Day

Time Group	ODF+AKF	OLNWF+AKF
T1	4.40	4.01
T2	5.30	5.14
T3	4.57	5.99
T4	5.61	6.01
T5	5.94	5.21
T6	5.72	5.59
T7	5.08	5.18
T8	4.83	4.45
T9	4.92	4.97
T10	4.79	5.10
T11	5.32	4.99
T12	5.32	5.03
T13	5.59	4.48
T13	5.59	4.48
T14	4.90	4.45
T15	4.63	4.58

TABLE V

Kickoff Percentage With Respect to Traffic Levels

Traffic Level	ODF+AKF	OLNWF+AKF
TTL	5.00	4.83
L1	4.49	4.52
L2	5.51	4.98
L3	5.88	5.60
L4	5.28	5.56
L5	6.75	4.85

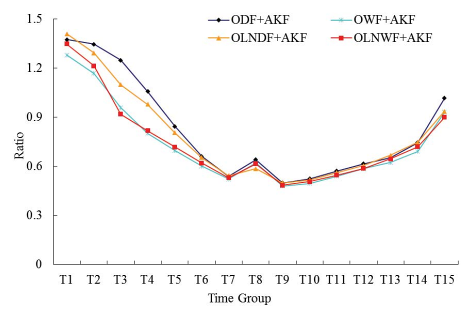


Fig. 4. Width-to-flow ratio with respect to time of day.

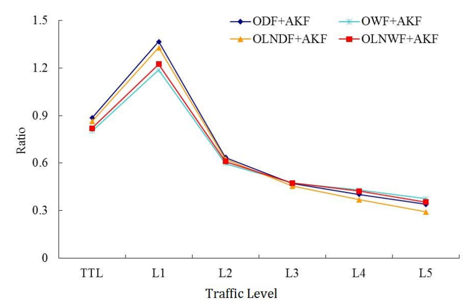


Fig. 5. Width-to-flow ratio with respect to traffic levels.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

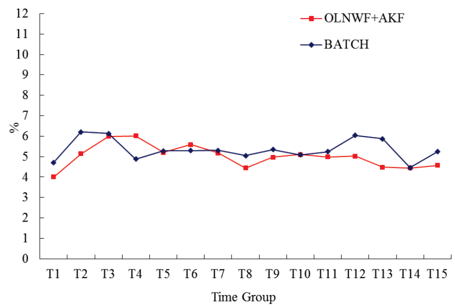


Fig. 6. Kickoff percentage with respect to time of day.

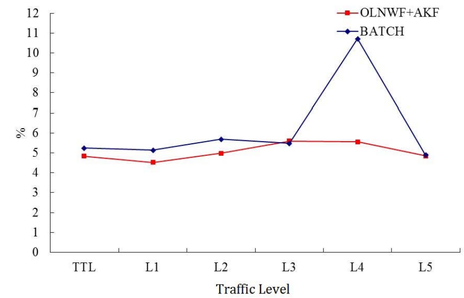


Fig. 7. Kickoff percentage with respect to traffic levels.

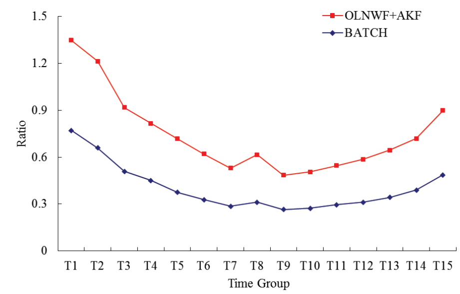


Fig. 8. Width-to-flow ratio with respect to time of day.

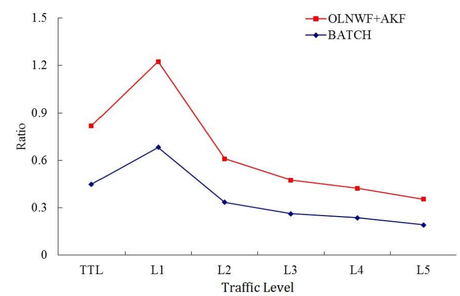


Fig. 9. Width-to-flow ratio with respect to traffic level.

all the four proposed OSAF + AKF models are showing comparable performance in terms of width to flow ratio. Specifically, the ratio is around 0.5 for groups corresponding to high level traffic, i.e., the prediction intervals are around half of the traffic flows, and the ratio will lift up to around 0.6-1.4 when the traffic is running at low level. This is not surprising in that the width to flow ratio is likely to be inflated due to the low level of the denominator. In addition, on a close look into the performance, in Fig.4, OWF+AKF and OLNWF + AKF consistently generate smaller width to flow ratio than ODF+AKF and OLNDF+AKF for all traffic levels and time of day except for T8, and in Fig.5, OWF+AKF and OLNWF + AKF outperform ODF+AKF and OLNDF+AKF for groups of TTL, L1, and L2, and all the four models are showing close performances for groups of L3, L4, and L5. This is also consistent with previous findings that the weekly model is more preferable than the daily models.

In summary, empirical results shows that the proposed OSAF+AKF model can generate workable prediction intervals in term of both kickoff percentage and width to flow ratio, and therefore, the proposed model can be used to predict in real time the seasonal heteroscedasticity in traffic flow series. In addition, the log weekly model is found to generate better prediction intervals, which is in alignment with previous studies and also the findings in [31], therefore, the OLNWF+AKF model will be selected for showing further the strength of the proposed approach.

B. Aggregated OLNWF + AKF and Batch Performance Comparison

The purpose of this test is to show the OLNWF + AKF model will have comparable performances with the corresponding BATCH model, i.e., the log weekly adjustment factor based BATCH model, which is important to justify the validity of the online model. First, the kickoff percentages of OLNWF + AKF and BATCH are shown with respect to time of day and traffic levels in Figs.6 and 7, respectively. On observing Figs.6 and 7, it is clear that both models yield kickoff percentages close to 5% for all time groups and traffic levels groups, which indicates that the proposed online computation mechanism is acceptable in generating prediction interval in real time. In addition, by further looking into the performance of OLNWF + AKF and BATCH, it can be seen that OLNWF + AKF works better than BATCH for high level traffic groups. Specifically, for Fig.6, OLNWF + AKF outperforms BATCH during morning and afternoon peak hours, i.e., for groups of T2, T3, T5, T7, T9, T11, T12, and T13, by yielding kickoff percentage closer to 5%, and for Fig.7, OLNWF + AKF outperforms BATCH for high level group L4 while the two approaches show comparable performance for the rest groups. This finding indicates the online OLWF+AKF model outperforms BATCH model during the peak hours when traffic is more volatile at high levels. On reflection, this finding is not surprising since a fixed set of parameters is used in BATCH for all traffic conditions whereas the OLNWF+AKF model can adapt to the changing of traffic conditions. This is desirable since most traffic management and control operations are targeting volatile and changing traffic conditions.

Second, the width to flow ratios of OLNWF + AKF and BATCH are shown with respect to time of day and traffic levels in Figs. 8 and 9, respectively. On observing Figs.8 and 9, it is clear that BATCH consistently outperforms OLNWF + AKF for all time groups and traffic level groups; however, for volatile traffic, e.g., for traffic level group L4 in Fig.7, BATCH has higher kickoff percentages than OLNWF + AKF, which offsets the added value of narrower prediction

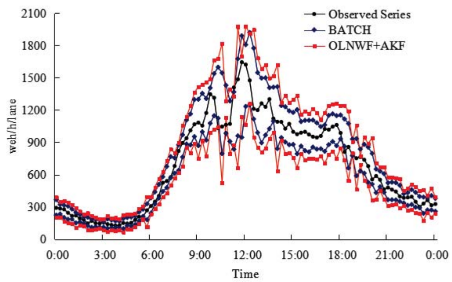


Fig. 10. Disaggregated prediction intervals generated by OLNWF + AKF and BATCH.

intervals generated by BATCH. In addition, on a close look into Figs. 8 and 9, for volatile traffic conditions, the mean width to flow ratio for OLNWF+AKF varies within the range from 0.35 to 0.48, indicating workable prediction intervals generated by OLNWF + AKF for these traffic conditions.

In summary, empirical results show that the performance of OLNWF + AKF and BATCH are mixed in term of kickoff percentage and width to flow ratio. However, considering the requirements of traffic management and control systems working under volatile traffic conditions, it is acceptable for the proposed OLNWF + AKF to model the seasonal heteroscedastic traffic condition series in real time.

C. Disaggregated OLNWF + AKF and Batch Performance Comparison

Previous investigations showed the acceptability of OSAF+ AKF in term of collective performances. In this subsection, the behavior of the OSAF + AKF approach will be demonstrated in a disaggregated fashion over a typical day together with the BATCH model. Based on previous investigations, OLNWF + AKF is used for this demonstration, with the disaggregated performance shown in Fig.10 for UK station 2737a on October 12, 2002, in which the prediction intervals generated from OLNWF + AKF and BATCH are shown directly around the observed traffic flow series.

On observing Fig.10, it can be seen that both OLNWF + AKF and BATCH can generate time varying prediction intervals to adapt to the changing traffic conditions, which is important for the traffic management and control systems to respond accordingly when the traffic condition changes. In addition, compared with the BATCH model, the proposed OLNWF + AKF model can generate comparable prediction intervals around the observed traffic flow series, which indicates the online processing mechanism proposed in the OLNWF+AKF model is acceptable in converting the conventional offline approach of BATCH into an online one. This is also critical for supporting the traffic management and control systems that are usually running in real time. Therefore, in summary, the proposed OLNWF+AKF model is acceptable in modeling seasonal heteroscedastic traffic flow series in real time.

D. Computational Efficiency

The average computation time for processing each traffic observations is presented in Table 6 to show the computational efficiency of the proposed online models. As shown in Table 6, for all the four proposed online models, all the average computation time is less than 2 ms. Compared with the 15-min interval used for these models, these computation time are trivial so that online processing can be successfully achieved for these online models. This is important for fulfilling the real time requirement for ITS applications.

TABLE VI
AVERAGE COMPUTATION TIME (UNIT: ms)

Region	Highway	Station	Year	ODF +AKF	OWF +AKF	OLNDF +AKF	OLNWF +AKF
UK	M25	4565a	2002	1.74	1.80	1.57	1.48
UK	M25	4680b	2002	1.80	1.81	1.60	1.48
UK	M1	2737a	2002	1.59	1.94	1.36	1.36
UK	M1	2808b	2002	1.52	1.59	1.39	1.42
UK	M1	4897a	2002	1.65	1.62	1.36	1.33
UK	M6	6951a	2002	1.74	1.40	1.60	1.51
MN	I35W-NB	60	2000	1.74	1.74	1.57	1.68
MN	I35W-SB	578	2000	1.82	1.79	1.62	1.68
MN	I35E-NB	882	2000	1.76	1.88	1.57	1.57
MN	I35E-SB	890	2000	1.76	1.62	1.57	1.59
MN	169-NB	442	2000	1.79	1.68	1.59	1.54
MN	169-SB	737	2000	1.82	1.74	1.57	1.57
MN	I35W-NB	60	2004	1.82	1.62	1.65	1.51
MN	I35W-SB	578	2004	1.71	1.65	1.65	1.54
MN	I35E-NB	882	2004	1.65	1.59	1.65	1.57
MN	I35E-SB	890	2004	1.68	1.59	1.59	1.54
MN	169-NB	442	2004	1.65	1.71	1.65	1.68
MN	169-SB	737	2004	1.62	1.59	1.68	1.71
MN	I35W-NB	60	2015	1.51	1.63	1.54	1.63
MN	I35W-SB	578	2015	1.60	1.60	1.54	1.63
MN	I35E-NB	882	2015	1.60	1.60	1.34	1.66
MN	I35E-SB	890	2015	1.60	1.57	1.46	1.68
MN	169-NB	442	2015	1.68	1.60	1.46	1.63
MN	169-SB	737	2015	1.57	1.66	1.46	1.66

VI. DISCUSSION AND CONCLUSION

Over the past few decades, short term traffic condition forecasting has been studied extensively for supporting the development of proactive traffic management and control systems. In this regard, apart from many published studies on generating point forecast, prediction interval generation has also received growing attention to quantify the uncertainty associated with point forecasts. More importantly, due to the close relationship between the prediction interval generation and the properties of the second order moment structure of traffic condition series, a relatively independent field of second order moment modeling is emerging from the conventional short term traffic condition prediction field. In this end, a number of studies have been published over the past several years, and a recent study is on combining a seasonal adjustment factor together with the conventional GARCH model as in [31]. However, the model proposed in [31] is an offline model, which cannot meet the real time processing requirement by many transportation applications. Therefore, this paper proposed an online OSAF + AKF approach of predicting in real time the seasonal heteroscedasticity in traffic flow series.

Using real world traffic flow data, the proposed OSAF + AKF model is implemented and validated through investigations into the performance of the proposed model and the comparisons of the proposed model with the conventional offline model. Empirical results show that the proposed model can process the seasonal heteroscedastic traffic flow series in real time with comparable performance when compared with the offline model, which indicates that the proposed online model is acceptable for supporting the traffic management and control systems. In addition, empirical results also show that the proposed model demonstrates improved performance over the offline model when traffic is highly volatile, which demonstrates the added adaptive ability imparted into the proposed approach through the online adjustment mechanism.

Some remarks are provided as follows. First, uniform performance measures are urgently needed for evaluating the overall performance of prediction interval generation, in particular, when the two selected measures are showing diverting performances. Second, more deseasonalizing methods can be explored. The seasonal adjustment factor approach is a classical time domain approach, and other methods can be investigated such as the Fourier or harmonic analysis approach in frequency domain. Third, in addition to the GARCH model, refined heteroscedasticity modeling approach might be developed or discovered for supporting traffic management and control systems. Moreover, more adaptation mechanisms can be investigated. This is important for the transportation systems to adapt to the changing traffic patterns. Finally, in addition to the 15-minute interval used in this paper, the proposed algorithm can be investigated in terms of other time intervals so as to extend its applicability to meet the requirements of more ITS applications, and more data sets can be applied to test the proposed online algorithm.

Acknowledgements

The authors would like to thank the Minnesota Department of Transportation and the United Kingdoms Highways Agency for providing the data used in this study.

References

[1] C. Chatfield, "Calculating interval forecasts," *J. Bus. Econ. Stat.*, vol. 11, no. 2, pp. 121–135, 1993.

[2] M. S. Ahmed and A. R. Cook, "Analysis of freeway traffic time-series data by using Box-Jenkins techniques," *Transp. Res. Rec.*, no. 722, pp. 1–9, 1979.

[3] P. Cortez, M. Rio, M. Rocha, and P. Sousa, "Multi-scale Internet traffic forecasting using neural networks and time series methods," *Expert Syst*, vol. 29, no. 2, pp. 143–155, 2012.

[4] M. M. Hamed, H. R. Al-Masaeid, and Z. M. B. Said, "Short-term prediction of traffic volume in urban arterials," *J. Transp. Eng.*, vol. 12, no. 3, pp. 249–254, 1995.

[5] Y. Kamarianakis, A. Kanas, and P. Prastacos, "Modeling traffic volatility dynamics in an urban network," *Transp. Res. Rec.*, no. 1923, pp. 18–27, 2005.

[6] M. Levin and Y.-D. Tsao, "On forecasting freeway occupancies and 743 volumes," *Transp. Res. Rec.*, no. 773, pp. 47–49, 1980.

[7] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 871–882, Jun. 2013.

[8] 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, 2011.

[9] N. Nihan and K. Holersland, "Use of the Box and Jenkins time series technique in traffic forecasting," *Transportation*, vol. 9, no. 2, pp. 125–143, 1980.

[10] B. L. Smith, B. M. Williams, and R. K. Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transport. Res. C, Emerg. Technol.*, vol. 10, no. 4, pp. 303–321, 2002.

[11] M.-C. 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.

[12] B. Williams, "Modeling and forecasting vehicular traffic flow as a seasonal stochastic time series process," Ph.D. dissertation, Dept. Civil Eng., Univ. Virginia, Charlottesville, VA, USA, 1999.

[13] B. Williams, "Multivariate vehicular traffic flow prediction: Evaluation of ARIMAX modeling," *Transp. Res. Rec.*, no. 1776, pp. 194–200, 2001.

[14] 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, 2003.

[15] J. Guo, W. Huang, and B. M. Williams, "Integrated heteroscedasticity test for vehicular traffic condition series," *J. Transp. Eng.*, vol. 138, no. 9, pp. 1161–1170, 2012.

[16] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *J. Econometrics*, vol. 31, no. 3, pp. 307–327, 1986.

[17] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.

[18] J. Guo, "Adaptive estimation and prediction of univariate vehicular traffic condition series," Ph.D. dissertation, Dept. Civil, Construction Environ. Eng., North Carolina State Univ., Raleigh, NC, USA, 2005.

[19] J. Guo, B. Williams, and B. Smith, "Data collection time intervals for stochastic short-term traffic flow forecasting," *Transp. Res. Rec.*, no. 2024, pp. 18–26, 2008.

[20] K. Sohn and D. Kim, "Statistical model for forecasting link travel time variability," *J. Transp. Eng.*, vol. 135, no. 7, pp. 440–453, 2009.

[21] T. Tsekeris and A. Stathopoulos, "Real-time traffic volatility forecasting in urban arterial networks," *Transp. Res. Rec.*, no. 1964, pp. 146–156, 2006.

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

[23] M. Yang, Y. Liu, and Z. You, "The reliability of travel time forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 1, pp. 162–171, Mar. 2010.

[24] J. Guo, W. Huang, and B. M. Williams, "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 50–64, Jun. 2014.

[25] C. Chen, J. Hu, Q. Meng, and Y. Zhang, "Short-time traffic flow prediction with ARIMA-GARCH model," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 607–612.

[26] J. Xia, Q. Nie, W. Huang, and Z. Qian, "Reliable short-term traffic flow forecasting for urban roads using multivariate GARCH model," *Transp. Res. Rec.*, no. 2343, pp. 77–85, 2013.

[27] Y. Zhang, A. Haghani, and X. Zeng, "Univariate volatility-based models for improving quality of travel time reliability forecasting," *Transp. Res. Rec.*, no. 2365, pp. 73–81, 2013.

[28] Y. Zhang, A. Haghani, and R. Sun, "Stochastic volatility modeling approach that accounts for uncertainties in travel time reliability forecasting," *Transp. Res. Rec.*, no. 2442, pp. 62–70, 2014.

[29] T. Tsekeris and A. Stathopoulos, "Short-term prediction of urban traffic variability: Stochastic volatility modeling approach," *J. Transp. Eng.*, vol. 136, no. 7, pp. 606–613, 2010.

[30] Y. Zhang, Y. Zhang, and A. Haghani, "A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 65–78, Jun. 2014.

[31] G. Shi, J. Guo, W. Huang, and B. M. Williams, "Modeling seasonal heteroscedasticity in vehicular traffic condition series using a seasonal adjustment approach," *J. Transp. Eng.*, vol. 140, no. 5, p. 04014012, 2014.

[32] Y. Zhang, A. Haghani, and X. Zeng, "Component GARCH models to account for seasonal patterns and uncertainties in travel-time prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 719–729, Apr. 2015.

[33] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1994.

[34] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 4th ed. Hoboken, NJ, USA: Wiley, 2008.

[35] L. C. Edie, "Discussion of traffic stream measurements and definitions," in *Proc. 2nd Int. Symp. Theory Traffic Flow (OECD)*, Paris, France, 1963, pp. 139–154.

[36] W. Fuller, *Introduction to Statistical Time Series*, 2nd ed. Hoboken, NJ, USA: Wiley, 1996.

[37] Y. Kamarianakis and P. Prastacos, "Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches," *Transp. Res. Rec.*, no. 1857, pp. 74–84, 2003.

[38] E. Mazloumi, G. Rose, G. Currie, and S. Moridpour, "Prediction intervals to account for uncertainties in neural network predictions: Methodology and application in bus travel time prediction," *Eng. Appl. Artif. Intel.*, vol. 24, no. 3, pp. 534–542, 2011.

[39] K. Myers and B. Tapley, "Adaptive sequential estimation with unknown noise statistics," *IEEE Trans. Autom. Control*, vol. AC-21, no. 4, pp. 520–523, Aug. 1976.

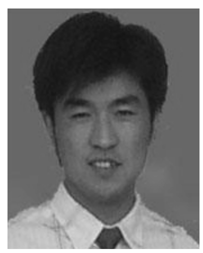
Wei Huang is currently a Distinguished Professor in civil engineering at the Intelligent Transportation System Research Center, Southeast University, Nanjing, China. He is a member of the Chinese Academy of Engineering. He enjoys the State Council special allowance and receives support from the New Century Talent Program, the National Outstanding Mid-aged Experts Program, the National Talents Engineering Program, and the Yangtze Scholar Program from various agencies and organizations. He is one of the forerunners in the research fields of long-span steel bridge pavement and intelligent transportation systems of China. As the leading awardee, he received 26 awards from both the national and provincial level. He has authored or co-authored 13 books.

Wenwen Jia received the bachelor's degree in mathematics from Chang'an University, Xi'an, China. She is currently pursuing the master's degree with the Intelligent Transportation System Research Center, Southeast University, Nanjing, China. She is currently involved in short-term traffic condition forecasting, especially on the second-order conditional moment modeling for vehicular traffic condition series.

Jianhua Guo is currently a Professor in transportation engineering at the Intelligent Transportation System Research Center, Southeast University, Nanjing, China. Prior to joining Southeast University, he was with University of Virginia and Federal Highway Administration. He received research funds from the Chinese Natural Science Foundation, the Ministry of Science and Technology of China, and various provincial agencies. His major research fields include intelligent transportation system applications, traffic simulation and control, statistical time series analysis, and discrete choice modeling. He serves as a reviewer for Chinese National Science Foundation and international journals. He received the Outstanding Reviewer Award for *ASCE Journal of Transportation Engineering* of the year 2012, and the Outstanding Reviewer Award for *ASCE Journal of Computing in Civil Engineering* of the years 2012, 2013, and 2014, consecutively.

Billy M. Williams is currently a Professor with the Department of Civil, Construction, and Environmental Engineering, North Carolina State University. Previously, he was an Assistant Professor with the School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA, USA. Before beginning his academic career, he spent over five years as a consulting engineer with the firm of Kimley-Horn and Associates and four years as a commissioned officer in the U.S. Navy Civil Engineer Corps. He is a recognized expert in the areas of analytical and simulation modeling of traffic operations and transportation networks, intelligent transport systems, and the application of rigorous statistical methods to a broad range of transportation modeling applications, including traffic condition forecasting and models of fundamental traffic flow characteristics. Funding for his research program has come from a variety of sources, including NSF, NCHRP, SHRP2, and the North Carolina Department of Transportation.

Guogang Shi received the Ph.D. degree from Southeast University, Nanjing, China, majoring in transport and transportation engineering. He is with the Highway Bureau of Jiangsu province, primarily in the fields of highway system maintenance, highway and bridge construction quality surveillance and control, highway traffic management and regulation, and intelligent highway systems. He is the Leading Researcher in various provincial funded research projects, and receives the Jiangsu Province Excellent Science and Technology Workers Award in 2012.



Yun Wei received the B.S. degree from the Nanjing University of Aeronautics and Astronautics in 2005, and the M.S. and Ph.D. degrees from Southeast University in 2008 and 2013, respectively. He is the Vice Director of the Urban Railway Green and Safe Construction National Engineering Laboratory. He is majoring in traffic information engineering and control and his research field includes intelligent vision analysis and pattern recognition. He is a member of the Comprehensive Intelligent Transportation Systems Technical Committee of the Chinese Association of Automation. He received the Beijing Science and Technology New Star Award and the Exceptional Talent Award of Beijing.

Jinde Cao (M'07–SM'07–F'16) was with Yunnan University from 1989 to 2000. In 2000, he joined the Department of Mathematics, Southeast University, Nanjing, China. From 2001 to 2002, he was a Post-Doctoral Research Fellow at the Department of Automation and Computer-Aided Engineering, The Chinese University of Hong Kong, Hong Kong. He is currently a Distinguished Professor, the Dean of the Department of Mathematics, and the Director of the Research Center for Complex Systems and Network Sciences at Southeast University. He is a member of the Academy of Europe. He was an Associate Editor of IEEE TRANSACTIONS ON NEURAL NETWORKS, *Journal of the Franklin Institute*, and *Neurocomputing*. He is an Associate Editor of IEEE TRANSACTIONS ON CYBERNETICS, *Differential Equations and Dynamical Systems*, *Mathematics and Computers in Simulation*, and *Neural Networks*. He has been named as Highly-Cited Researcher in mathematics, computer science, and engineering by Thomson Reuters.