Final report of ITS Center project: Abnormal event traffic forecasting

A Research Project Report

For the Center for ITS Implementation Research
A U.S. DOT University Transportation Center

Short Term Speed Variance Forecasting Using Linear Stochastic Modeling of Univariate Traffic Speed Series

Principal Investigator
Dr. Brian L. Smith

Center for Transportation Studies
University of Virginia
351 McCormick Road, P.O. Box 400742
Charlottesville, VA 22904-4742
434.924.6362

July 16, 2007

Research Report No. UVACTS-15-17-10
Date: July 17, 2007

Short Term Speed Variance Forecasting Using Linear Stochastic Modeling of Univariate Traffic Speed Series

Jianhua Guo
Research Associate

Brian L. Smith
Associate Professor

University of Virginia
Center for Transportation Studies

July 16, 2007

A Research Project Report for the Intelligent Transportation Systems Implementation Center (ITS)
A U.S. DOT University Transportation Center

Dr. Brian L. Smith, PhD
Department of Civil Engineering
Email: bls2z@virginia.edu

Center for Transportation Studies at the University of Virginia produces outstanding transportation professionals, innovative research results and provides important public service. The Center for Transportation Studies is committed to academic excellence, multi-disciplinary research and to developing state-of-the-art facilities. Through a partnership with the Virginia Department of Transportation's (VDOT) Research Council (VTRC), CTS faculty hold joint appointments, VTRC research scientists teach specialized courses, and graduate student work is supported through a Graduate Research Assistantship Program. CTS receives substantial financial support from two federal University Transportation Center Grants: the Mid-Atlantic Universities Transportation Center (MAUTC), and through the National ITS Implementation Research Center (ITS Center). Other related research activities of the faculty include funding through FHWA, NSF, US Department of Transportation, VDOT, other governmental agencies and private companies.

CTS Website	Center for Transportation Studies
http://cts.virginia.edu	University of Virginia
	351 McCormick Road, P.O. Box 400742
	Charlottesville, VA 22904-4742
	434.924.6362

<table>
<tr><td>1. Report No.

UVACTS-15-17-10</td><td>2. Government Accession No.</td><td>3. Recipient's Catalog No.</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle
Short Term Speed Variance Forecasting Using Linear Stochastic Modeling of Univariate Traffic Speed Series</td><td>5. Report Date

July 17, 2007</td></tr>
<tr><td>6. Performing Organization Code</td></tr>
<tr><td colspan="2">7. Author(s)

Dr. Brian Smith, Jianhua Guo</td><td>8. Performing Organization Report No.</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization and Address

Center for Transportation Studies
University of Virginia
PO Box 400742
Charlottesville, VA 22904-7472</td><td>10. Work Unit No. (TRAIS)</td></tr>
<tr><td>11. Contract or Grant No.</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agencies' Name and Address
Office of University Programs, Research Innovation and Technology Administration
US Department of Transportation
400 Seventh Street, SW
Washington DC 20590-0001</td><td>13. Type of Report and Period Covered
Final Report</td></tr>
<tr><td>14. Sponsoring Agency Code</td></tr>
<tr><td colspan="3">15. Supplementary Notes</td></tr>
<tr><td colspan="3">16. Abstract</td></tr>
<tr><td colspan="2">17 Key Words
Speed Variance Forecasting, Univariate Traffic Speed Series</td><td>18. Distribution Statement
No restrictions. This document is available to the public.</td></tr>
</table>

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	21	N/A

1 Introduction

Intelligent transportation systems (ITS) offers the potential to address critical transportation needs. However, most ITS currently operates in a reactive mode. While this provides some level of benefit, "*the full benefits of ITS cannot be realized without an ability to anticipate traffic conditions in the* short-*term (less than one hour into future)*." (Smith and Oswald, 2003) Based on the anticipated traffic condition, proactive transportation management and comprehensive traveler information services are feasible. Therefore, traffic condition forecasting has been identified as one of the major challenges for ITS.

Considering the forecasting process as a state extrapolation process governed by certain regularity, the development of traffic condition forecasting methods demands a sound understanding of traffic condition dynamics. Volume, speed, and density (or occupancy for the widely-deployed inductive loop detectors) are three traffic variables that are most commonly used to characterize traffic conditions, and suitable traffic condition forecasting methods are expected to be built upon traffic condition dynamics in terms of these traffic variables.

Previous efforts addressing traffic flow forecasting at a higher aggregation interval, such as 15-minutes, indicate traffic conditions to be linear stochastic. The traffic flow forecasting methods can be roughly classified into nonlinear theory based methods and linear theory based methods. The former assumes traffic dynamics are nonlinear, and can be emulated through nonlinear operations. Typically, this category includes non-parametric regression, neural networks, kernel smoothing, and local linear regression. The latter assumes traffic dynamics are linear and can be emulated through linear operations. Typically, this category includes univariate Box-Jenkins approach, exponential smoothing, spectral analysis, and multivariate time series methods. Adaptive methods, such as Kalman filter and recursive least square, can be classified into linear methods due to their nature as sequential projection in linear space. Williams (1999), Smith et al. (2002), and Williams and Hoel (2003) showed that traffic flow forecasting method based on Seasonal Autoregressive Integrated Moving Average (SARIMA) process outperformed nonlinear theory based methods, supporting the adoption of SARIMA process to describe traffic flow dynamics. Guo (2005) further appended a Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model to describe the conditional variance for forecasting confidence interval construction. In addition, recent research on traffic state transition summarized in Daganzo (2002) revealed that traffic dynamics could be modeled using simple first order continuum theory, suggesting a linear regularity in traffic dynamics.

Based on linear traffic dynamics, the traffic speed series is naturally expected to be described using linear stochastic theory. In this report, first, the conditional mean of traffic speed series will be investigated using ARIMA model. Second, the conditional variance of traffic speed series will be investigated using GARCH model. On one hand, forecast confidence intervals could be constructed using conditional variance computed from GARCH process, indicating the certainty associated with the forecasts (or predictability of traffic). On the other hand, conditional variance could be utilized in more broad ITS settings, one of which is to determine the sample size for probe vehicle based traffic surveillance systems. In summary, the combination of

ARIMA model and GARCH model will present a complete structure describing the first two conditional moments of traffic speed series, facilitating the development of an online forecasting system.

The rest of this report is organized as follows. Section 2 discusses the related literature; section 3 presents an overview of time series modeling; section 4 applies the model to the real traffic speed series from the US and the UK; finally, section 5 concludes this report with a discussion of the results.

2 Related Literature

Like the efforts on traffic flow forecasting that indicate traffic flow series can be best described as SARIMA process (Williams 1999; Smith et al., 2002; Williams and Hoel, 2003), the investigation of traffic speed series was largely contained within the development of traffic speed estimation and forecasting methods.

- Yang et al. (2004) provided an online recursive traffic speed forecasting method based on an arbitrary autoregressive process. The authors use Kalman filter to process this autoregressive model and provide forecasting confidence interval using prediction variance, which will converge to a constant eventually. Actually, the maximum difference between the lengths of confidence intervals (Figure 6 in their paper) is about 0.1 mph for confidence interval width around 8 mph.
- Based on nonlinear dynamics assumption, Sun et al. (2003) used local linear regression method to forecast 5-minute traffic speed data collected on the US-290 Northwest freeway in Houston. A comparative study showed that local linear regression outperformed k-nearest neighbor and kernel smoothing method. In addition, Sun et al. (2004) presented a method to construct prediction confidence intervals for predictions from local linear regression. Asymptotic prediction confidence intervals and prediction confidence intervals based on bootstrap method were constructed and the bootstrap method was recommended. Their results showed that a large proportion of real observations were outside of prediction confidence intervals given by the bootstrap method, which invalidated the goodness of their approach.
- Different from univariate analysis, Tavana et al. (2000) approached using transfer function model (bivariate time series model). In their work, based on the concept of dynamic speed-density relationship, the variation of speed around a static speed-density relation is captured by a transfer function model, in which density (estimated from occupancy) was used as forcing function and the deviation of speed from its static equilibrium value was used as system output. The static equilibrium speed-density relation was assumed to be linear and was estimated from speed-density data. The speed equilibrium value was then computed using density and the estimated static equilibrium speed-density relation. First-order differencing was performed on the input and output to introduce stationarity. The transfer model was identified using cross-correlation analysis after applying pre-whitening operation to both input and output series. The noise term after removing the effects of cross-correlation was modeled as MA(1). In essence, the transfer function model is a regression of the output (speed deviation) onto input (density) with MA(1) errors. The model parameters were estimated through an

optimization process minimizing the conditional sum square of errors. In forecasting speed for time t, the density at time t needs to be obtained first. Two options were suggested: prediction using univariate time series forecasting method and simulation. The latter was adopted in their work.

- Based on the transfer function model in Tavana et al. (2000), Huynh et al. (2002) proposed an adaptive process of estimating the model parameters. Results from simulation-based analysis showed that adaptive model outperformed non-adaptive model. Qin et al. (2004) proposed a calibration of the same model using real sensor data from Irvine, CA, and reaffirmed that adaptive model outperformed the non-adaptive model and also showed that the increase of number of links applied in the model improved speed estimation.
- Antoniou et al. (2005) approached using state space model. The speed-density relation was regarded as the measurement equation, and the parameters in the equation were regarded as system state. The transition of system state was formulated as an autoregressive process. The degree of the transition process, the transition matrix, and the errors were estimated off-line. Three variations of Kalman filter algorithm, including extended Kalman filter (EKF), Iterated EKF, and Unscented Kalman filter (UKF), were tested using real sensor data from Europe and the US. Numerical results showed that speed estimation using online calibrated speed-density relation was improved over speed estimated from off-line calibrated speed-density relation. As is in the transfer function model (Tavana et al. 2000, Huynh et al. 2002, and Qin et al. 2004), the prediction of the occupancy (density) has to be obtained before obtaining the forecast of speed.

3 Time Series Models

In this section, the theory of time series modeling is briefly reviewed for completeness (Fuller 1996). A time series defined as $\{X_t : t \in T\}$ can be strictly stationary or weakly stationary. Given $F_X(\cdot)$ as the probability distribution function and lag h, a time series is strictly stationary if

$$F_{X_{t_1},X_{t_2},\dots,X_{t_n}}\left(x_{t_1},x_{t_2},\dots,x_{t_n}\right)=F_{X_{t_1+h},X_{t_2+h},\dots,X_{t_n+h}}\left(x_{t_1},x_{t_2},\dots,x_{t_n}\right) \tag{1}$$

for all the possible sets of $t_1,t_2,\dots,t_n$ and $t_1+h,t_2+h,\dots,t_n+h$. This definition states that the joint probability of any given subset of a strictly stationary time series is independent of the lag. A time series is weakly stationary if

(1) the expected value of X_t is constant for all t; and

(2) the covariance matrix of $\left(X_{t_1},X_{t_2},\dots,X_{t_n}\right)$ is the same as that of $\left(X_{t_1+h},X_{t_2+h},\dots,X_{t_n+h}\right)$ for all finite sets $(t_1,t_2,\dots,t_n)$ and h.

Weakly stationarity only restricts the first and second moment of a time series, and hence it is also termed as covariance stationarity. According to the Wold Decomposition Theorem, a weakly stationary time series Y_t defined on $T = \{0, \pm 1, \pm 2, ...\}$ can be represented as

$$Y_t = X_t + Z_t, \tag{2}$$

where Z_t is a deterministic component and X_t can be modeled as an autoregressive moving average (ARMA) process defined as

$$\phi(B) X_t = \theta(B) \varepsilon_t, \tag{3}$$

where

B : backshift operator such that $BX_t = X_{t-1}$;

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \Lambda - \phi_p B^p$: short-term AR polynomial with order p;

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \Lambda - \theta_q B^q$: short-term MA polynomial with order q;

For a non-stationary time series $\{X_t\}$, the non-stationarity can be handled by proper differencing, and the whole model, seasonal integrated autoregressive moving average-SARIMA $(p,d,q)(P,D,Q)_S$, is defined as

$$\phi(B) \Phi\left(B^S\right)\left(1 - B^S\right)^D (1 - B)^d X_t = \theta(B) \Theta\left(B^S\right) \varepsilon_t \tag{4}$$

where

$\left(1 - B^S\right)^D$: seasonal differencing with seasonality S and order D;

$(1 - B)^d$: normal differencing with order d;

$\Phi(B^S) = 1 - \Phi_1(B^S) - \Phi_2(B^S)^2 - \Lambda - \Phi_P(B^S)^P$: P^{th} order seasonal AR polynomial;

$\Theta(B^S) = 1 - \Theta_1(B^S) - \Theta_2(B^S)^2 - \Lambda - \Theta_Q(B^S)^Q$: Q^{th} order seasonal MA polynomial.

It is assumed that the roots of $\phi(B)$, $\theta(B)$, $\Phi(B^S)$, and $\Theta(B^S)$ are outside the unit circle, which enables the causality and invertibility of the process; also, these polynomials should have no common factors. $\{\varepsilon_t\}$ is assumed to be Gaussian white noise with mean zero and constant variance σ_ε^2; in addition, it is further assumed to follow GARCH process defined as below.

$$\varepsilon_t = \sqrt{h_t} e_t, \quad h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}, \; e_t \sim IN(0,1) \tag{5}$$

where

h_t: conditional variance at time t, i.e., $\varepsilon_t | \Psi_{t-1} \sim N(0, h_t)$ with Ψ_{t-1} as the information up to time $t-1$;

p, q: order of the GARCH process with $p \geq 0$ and $q > 0$;

α_0: positive constant coefficient;

$\alpha_{i, i=1,\ldots,q}$: non-negative coefficients of the lagged sample variance ε_{t-i}^2;

$\beta_{i, i=1,\ldots,p}$: non-negative coefficients of the lagged conditional variance h_{t-i};

If $p = 0$, GARCH(p,q) process is reduced to ARCH (q) process; if $p = q = 0$, the GARCH(p,q) process is simplified into Gaussian white noise, and the SARIMA+GARCH structure is reduced to SARIMA process. An alternative representation shows that GARCH process can be interpreted as ARMA process in ε_t^2 as:

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{n} (\alpha_i + \beta_i) \varepsilon_{t-i}^2 - \sum_{i=1}^{p} \beta_i \eta_{t-i} + \eta_t \tag{6}$$

where $\eta_t = \varepsilon_t^2 - h_t$ and $n = \max(p, q)$ (Bollerslev 1986), indicating that the modeling of GARCH model is similar to that of SARIMA in the sense of squared series. Box-Jenkins approach provided a complete framework of SARIMA modeling, including identification, model estimation, forecasting, and diagnostic check. Interested readers are referred to (Box and Jenkins 1994) for detail.

The combination of equation (4) and (5) presented a complete stochastic structure, SARIMA+GARCH. Using results in Engle (1982) and Bollerslev (1986), this structure can be processed separately without efficiency lost: (1) the processing of SARIMA, and (2) the processing of GARCH using the residuals from the SARIMA model.

4 Empirical Study

This section applied the SARIMA+GARCH structure to traffic speed series. First, the sensor data used in this analysis is discussed; second, using the sensor data, the model for speed levels is identified; third, heteroscedasticity tests and GARCH model selection are discussed; finally, the adequacy of the identified stochastic structure is demonstrated by investigating its performance.

4.1 Data

Taffic data collected from two highway systems, including the motorway system in the

United Kingdom and freeway systems in Richmond, Virginia in the US, are used in this study. In this study, these traffic speed data were aggregated into 15-minute data according to Edie (1963), i.e., flow weighted harmonic mean. Obvious erroneous data are excluded and treated as missing values. In total, 10 traffic speed series were compiled (see Table 1).

Table 1: Overview of Selected Traffic Speed Series

Region	Station	Total Length	Percent Missing(%)
RHMD	12040	6720	7.05
RHMD	12042	6666	5.55
RHMD	12043	6666	4.64
UK	2737a	30912	4.71
UK	2808b	20676	5.05
UK	4565a	6289	4.55
UK	4762a	35039	3.93
UK	4897a	30912	7.87
UK	6951a	35039	1.78
UK	6954b	35039	1.92

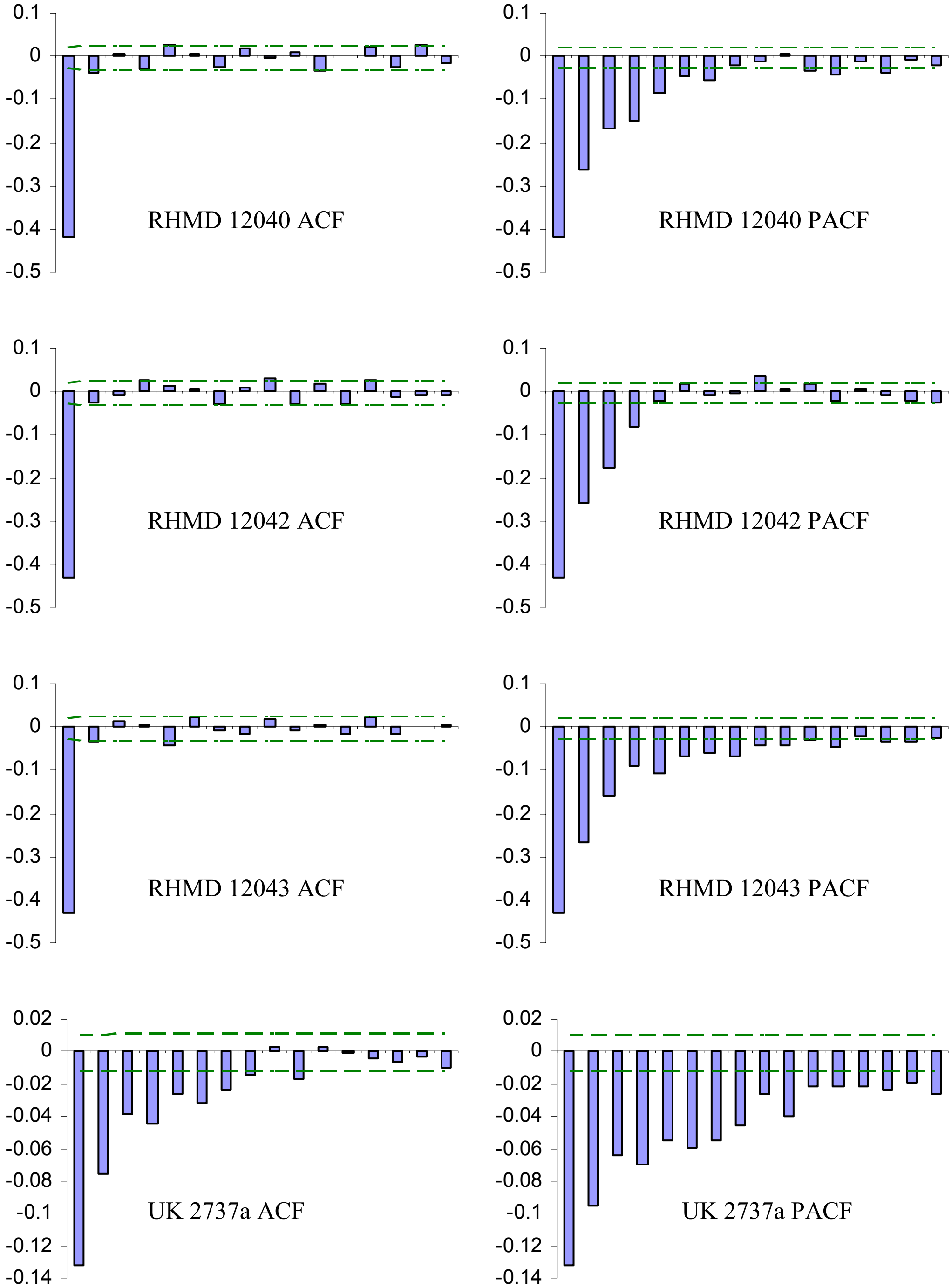
Note: UK: the United Kingdoms; RHMD: Richmond, VA.

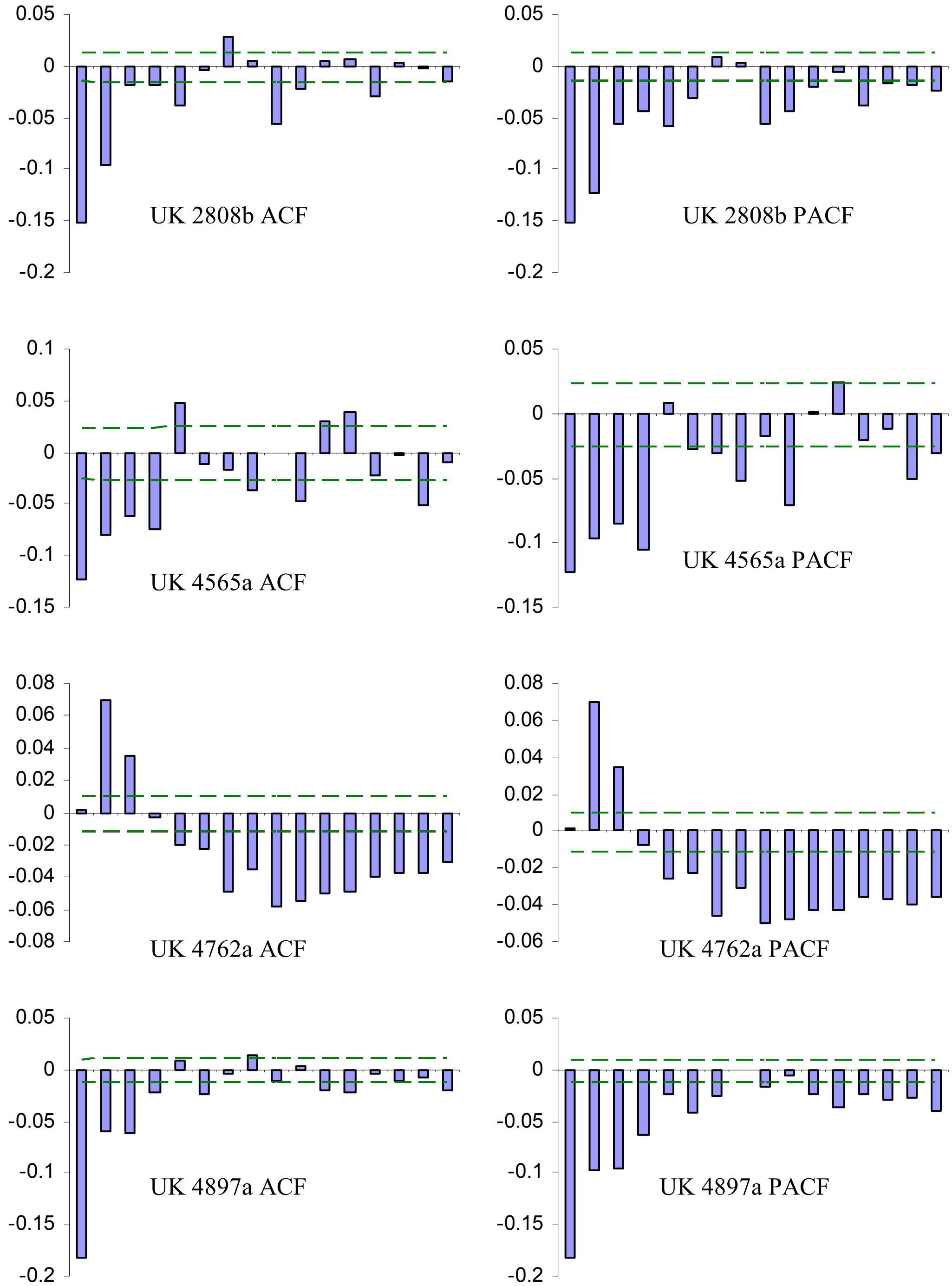
4.2 ARIMA Model Selection

In this section, the order of ARIMA model is determined. Unlike traffic flow series in which strong daily or weekly pattern exists (Williams 1999), for speed series, the daily/weekly seasonality is not significant. According to previous research, the *speed remains nearly constant even at quite high flow rates* (Revised Traffic Flow Theory Monograph, 1994). The same argument is also reflected in Highway Capacity Manual 2000 (See Exhibit 8-26 and Exhibit 8-27 for speed and volume variations across weekdays and Saturday, and Exhibit 13-2 for the speed-flow relationships for the basic freeway segment). All these evidences lead to the fact that the seasonal factor of the SARIMA process could be eliminated in modeling traffic speed series.

Although speed series will roughly maintain at a level for most of the time, speeds do oscillate due to disturbances like incident and weather, causing non-stationarity in the series. In this study, the speed series will be differenced once to introduce stationarity, which is in accordance with previous efforts in (Tavana et al. 2000, Huynh et al. 2002, and Qin et al. 2004).

After the first order differencing, and ignoring the seasonality, an autoregressive moving average (ARMA) model could be constructed. The sample ACF and sample PACF from lag 1 to lag 16 for the differenced series are computed using SAS ®. See Figure 1 below.


RHMD 12040 ACF
RHMD 12040 PACF
RHMD 12042 ACF
RHMD 12042 PACF
RHMD 12043 ACF
RHMD 12043 PACF
UK 2737a ACF
UK 2737a PACF


UK 2808b ACF
UK 2808b PACF
UK 4565a ACF
UK 4565a PACF
UK 4762a ACF
UK 4762a PACF
UK 4897a ACF
UK 4897a PACF

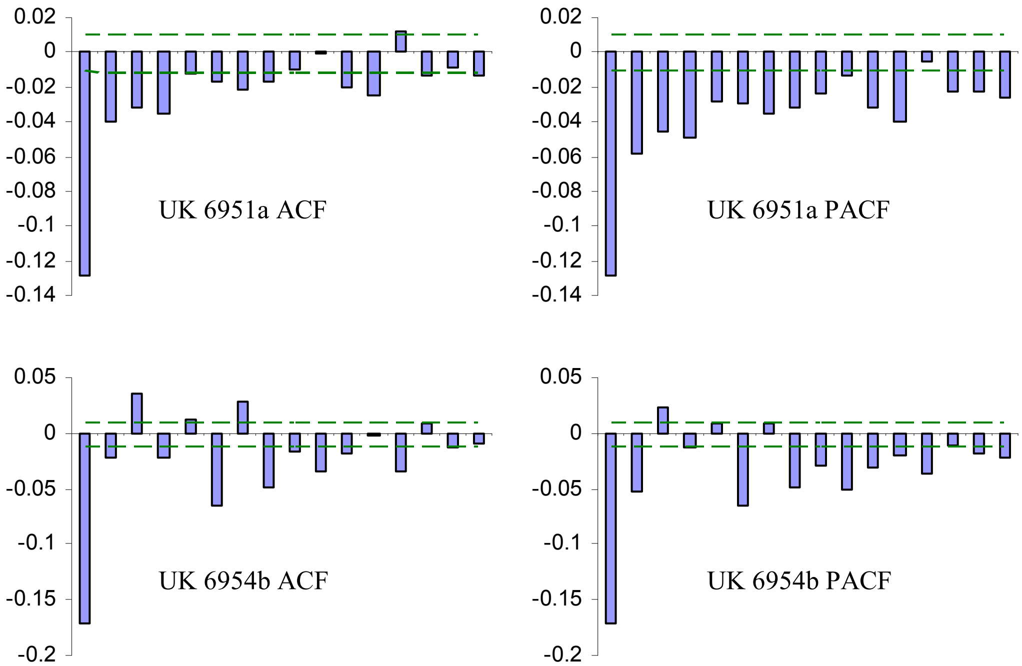


Figure 1: Sample ACF and Sample PACF of First-Order Differenced Speed Series (x-axis: lag; y-axis: ACF or PACF; Dashed line around x-axis: 95% confidence limits)

The ACF (cutoffs at lag one) and PACF (tails off) for Richmond stations indicates the series can be modeled as MA(1) process. The ACF and PACF for UK stations are quite volatile, and no single AR model or MA model could be used. However, one common point is that the magnitude of the ACF and PACF of these stations is small. Under this circumstance, combing the consideration of Richmond station, in this study, we use ARMA(1,1) to model the first-order differenced speed series for all the stations. To summarize, the speed level process is modeled as ARIMA(1,1,1). It is worthwhile to note that the model is meant to describe the general trend of speed process in the long run for building an online forecasting model, and this model will not necessarily be the exact model for each series. The discrepancy will be compensated in the online forecasting algorithm that can adapt to local variations.

4.3 Heteroscedasticity Test and GARCH Model Selection

Upon the Selection of ARIMA(1,1,1), the modeling of the second conditional moments is described as follows. First, missing values were replaced by the one-step-ahead forecasts using ARIMA(1,1,1) model in an iterative procedure. The iteration will stop when the model standard error stops decreasing or the maximum number of iterations has been reached. Based on the missing value imputed series, ARIMA(1,1,1) model was estimated and the forecasting residuals were computed. In this process, the model parameters were estimated using maximum likelihood method. Then, using the forecasting residuals, heteroscedasticity test using Langrage multiplier test and portmanteau test was performed to justify the addition of the GARCH process. To account for test inflation due to series length, four strategies, i.e., test by length, by month, by week, and by day, are applied. In test by length, the entire residual series is tested; for the other three tests, the residual series is first broken into sub-series and then tested individually. According to the length of series in each test, the testing power is decreasing in the order of test by length, month, week, and day. To be conservative, for each series or sub-series, statistical significant heteroscedasticity will be claimed only when both tests (Langrage multiplier test and portmanteau test) are significant at <0.001 level for lags from 1 to 12. For test by month, week, and day, the percentages of the significant sub-series will be computed to show the degree of variance change for each station.

The results for test by month/week/day are presented in Table 2. For test by length, the results are uniformly significant across stations. It can be seen that the process variance is changing across time, and the percentage of significance for each test decreases with the decreasing of testing power.

Table 2: Heteroscedasticity Test Results

Region	Station	Test by Month			Test by Week			Test by Day		
		A	B	%	A	B	%	A	B	%
RHMD	12040	3	2	67	10	2	20	70	2	3
RHMD	12042	3	3	100	10	2	20	70	0	0
RHMD	12043	3	3	100	10	10	100	70	2	3
UK	2737a	11	10	91	46	19	41	322	10	3
UK	2808b	8	7	88	31	22	71	216	7	3
UK	4565a	3	3	100	10	9	90	66	3	5
UK	4762a	12	12	100	53	45	85	365	56	15
UK	4897a	11	11	100	46	36	78	322	17	5
UK	6951a	12	11	92	53	41	77	365	13	4
UK	6954b	12	12	100	53	42	79	365	20	5

Note: A – total number of month/week/day; B – number of significant month/week/day.

According to equation (6), GARCH process can be regarded as ARMA process in terms of squared series; therefore, the identification can be performed through investigating the sample ACF and PACF of squared residual series. In this study, GARCH(1,1) is selected, and as the case for ARMA(1,1) for first-order differenced speed series, GARCH(1,1) is used to describe the general trend of the changing variance across stations and the discrepancy can be compensated for when developing the online algorithm.

4.4 Model Adequacy Evaluation

Summarizing section 4.2 and 4.3, the model for traffic speed series was identified as ARIMA(1,1,1)+GARCH(1,1), which is presented below.

$$(1-\phi_1 B)(1-B)X_t = (1-\theta_1 B)\varepsilon_t \tag{7}$$

$$\varepsilon_t = \sqrt{h_t}\,e_t,\ \ h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1 h_{t-1},\ e_t \sim IN(0,1) \tag{8}$$

In this subsection, the adequacy of this structure will be demonstrated through investigating its performance.

4.4.1 Performance Measures

The performance measures include the measures of forecasting accuracy and prediction confidence intervals. Given X_t as observations, $\hat{X}_t$ as forecasts, and n as the total number of observations, the forecasting accuracy measures are defined as: (1) Mean Absolute Error: $MAE = \frac{1}{n}\sum_{t=1}^{n}\left|X_t - \hat{X}_t\right|$; (2) Mean Absolute Percentage Error: $MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{X_t - \hat{X}_t}{X_t}\right|$; (3) Root Mean Square Error: $RMSE = \frac{1}{n}\sqrt{n\sum_{t=1}^{n}\left(X_t - \hat{X}_t\right)^2}$. The evaluation of confidence intervals is complicated with multiple criteria. In this study, following two intuitions, i.e., (1) real observations are expected to fall within the prediction confidence intervals, and (2) narrower prediction confidence intervals are more informative than wider ones, two summary statistics, e.g., prediction confidence interval kickoff percentage and average prediction confidence interval width to speed ratio are constructed. A prediction confidence interval kickoff is identified as an observation falling outside of the corresponding prediction confidence interval, and a percentage can be calculated as the total number of kickoffs divided by the total number of observations. Discussed in (Kang and Schmeiser 1990; Schmeiser and Yeh 2002), suitable prediction confidence intervals will have a kickoff percentage close to the nominal value, i.e., 5% for 95% confidence interval, and under this condition, the narrower the interval, the better. In computing performance measures, only non-missing speeds and corresponding forecasts and prediction confidence intervals are used. The imputed speeds are excluded from the computation.

4.4.2 Forecasting Accuracy

The forecasting accuracy measures are listed in Table 3. It can be seen the forecasting accuracy measures of all the stations (except for station RHMD 12043) are satisfactory: the forecasting error generally varies around 1.5-2 mph, with relative error around 2.5-4%. The high forecasting performance is in accordance with the speed series dynamics. As discussed previously, the speed will generally maintain at a fixed level except for occasional period when traffic is subjected to external disturbances. For RHMD 12043, the forecasting measures indicate that speed series for this station is not as predictable as other stations. This argument will be further substantiated when investigating the forecasting confidence interval.

Table 3: Forecasting Accuracy Measures

Region	Station	MAE	MAPE	RMSE
RHMD	12040	2.17	4.07	3.59
RHMD	12042	1.63	3.00	2.55
RHMD	12043	7.56	22.97	10.58
UK	2737a	1.38	2.23	2.20
UK	2808b	1.65	3.04	3.04
UK	4565a	1.50	2.96	2.89
UK	4762a	2.32	4.85	4.36
UK	4897a	1.55	2.75	2.69
UK	6951a	1.46	2.75	2.52
UK	6954b	1.40	2.71	2.87

4.4.3 Prediction Confidence Interval

The performance measures of prediction confidence intervals are listed in Table 4. It can be seen that the prediction confidence interval kickoff percentage for all the stations varies around 3-5%, indicating the proposed prediction confidence interval construction method based on GARCH model is workable although the results is slightly conservative in yielding an high coverage. In the mean time, the average prediction confidence interval width to speed ratio are around 0.15-0.25, which can be translated to 9-15 mph using an assumed speed level at 60 mph. This is approximately ±5 –±8 mph around traffic speed level. The average prediction confidence interval width to speed ratio for RHMD 12043 is 0.93, indicating a wide confidence interval for the forecasts for this station and thus a low predictability. This is in accordance with the observation on the forecasting accuracy measures for this same station in Table 3.

Table 4: Prediction Confidence Interval Performance Measures

Region	Station	CI Kickoff Percentage	Average CI Width to Speed Ratio
RHMD	12040	2.51	0.23
RHMD	12042	2.46	0.17
RHMD	12043	5.63	0.93
UK	2737a	3.89	0.12
UK	2808b	3.51	0.16
UK	4565a	2.90	0.18
UK	4762a	4.50	0.28
UK	4897a	3.10	0.15
UK	6951a	3.57	0.15
UK	6954b	3.03	0.16

4.4.4 Results Illustration

The structure performance is illustrated in Figure 2. It can be seen that the proposed structure can produce forecast confidence intervals that can cover the real speed series, and the forecasts (average of the upper and lower confidence interval that are omitted in the diagrams) can track the evolution of speed series. For station RHMD 12043, the forecasting confidence interval can also cover the real speed observations, while the confidence intervals are very wide. Actually, the wide confidence interval indicates the reduced predictability of speed at this station, reflected by the huge oscillations of speeds at this station. In addition, the confidence interval will be enlarged when traffic is oscillating and the confidence intervals will shrink when traffic is stabilized.

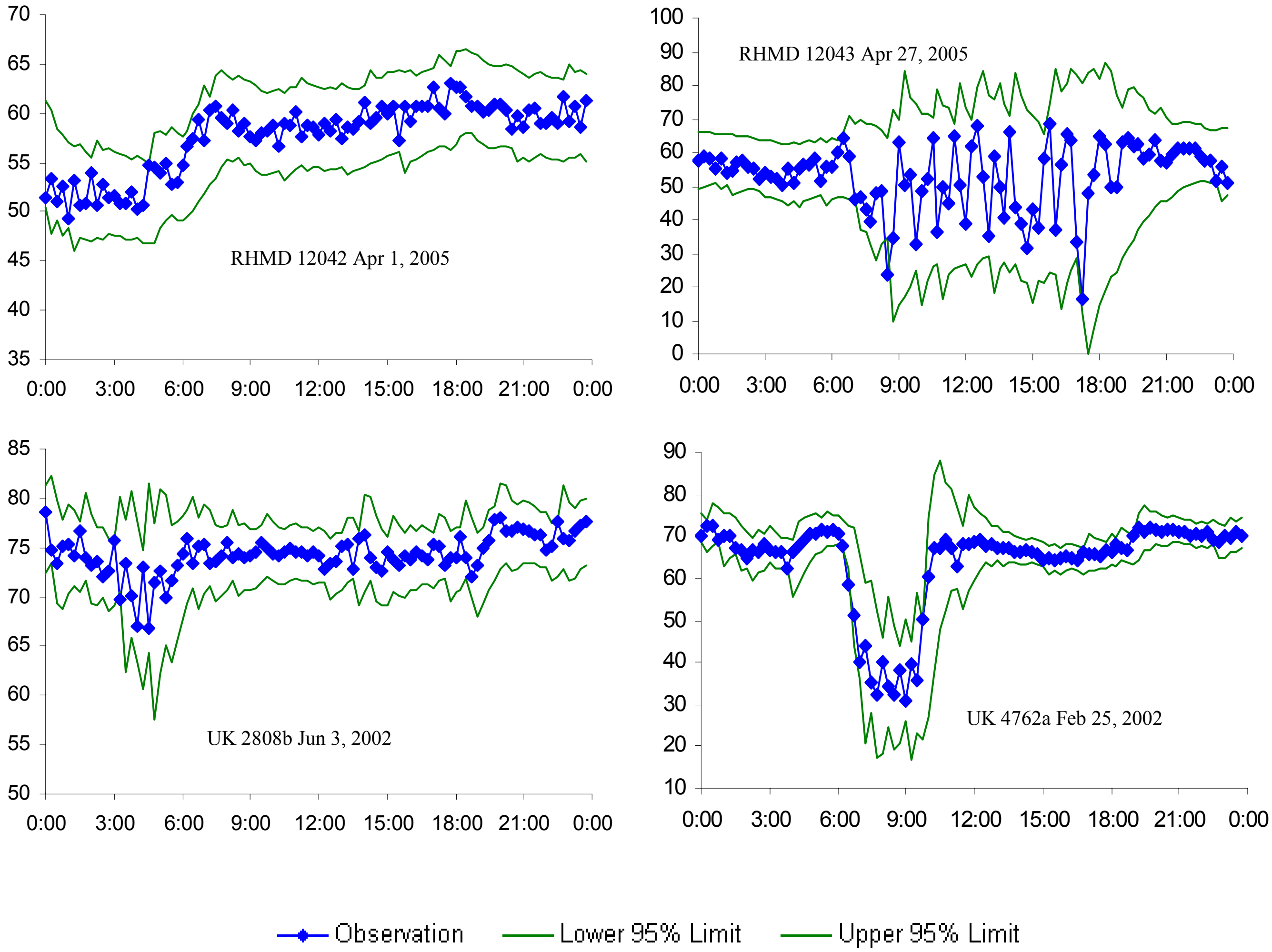


Figure 2: Structure Adequacy Illustration (x-axis: time; y-axis: speed in mph)

5 Conclusion

In this report, a stochastic structure, ARIMA(1,1,1)+GARCH(1,1) is proposed for modeling the traffic speed series for the purpose of developing an online forecasting system. The ARIMA(1,1,1) describes the evolution of traffic speed levels (conditional mean), and the GARCH(1,1) describes the evolution of traffic speed volatility (conditional variance). Based on these two components, speed forecasts and associate confidence interval can be generated. The empirical analysis using real sensor data from the US and the UK validates this structure through investigating speed forecasting accuracy and confidence intervals.

An important point to emphasize is that the proposed structure is meant to describe the general trend of traffic speed series, indicating the speed series dynamics in the long run. Actually, this structure does not rule out the possible variations across time and across stations as well, and further adaptive ability should be introduced in developing an online forecasting procedure using this structure.

An succeeding research effort will be to develop an online forecasting method based on this structure. Development of time series theory reveals that ARMA model can be formulated into state space model, based on which an online forecasting algorithm can be developed.

References

Antoniou, Constantinos, Moshe Ben-Akiva, and Haris N. Koutsopoulos (2005) On-line Calibration of traffic prediction models, Submitted for presentation at the 84th Annual meeting of the Transportation Research Board, Washington D.C.

Bollerslev, T. (1986) "Generalized Autoregressive Conditional Heteroscedasticity", *Journal of Econometrics*, 31, pp.307-327

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel, (1994) *Time Series Analysis: Forecasting and Control 3ed*, Prentice-Hall, Inc.

Daganzo, C. F. (2002) "A Behavioral Theory of Multi-Lane Traffic Flow. Part 1: Long Homogeneous Freeway Sections", *Transportation Research B*, 36, pp.131-158

Edie, L. C. (1963) "Discussion on Traffic Stream Measurements and Definitions", *Proceedings of 2nd International Symposium on the Theory of Traffic Flow*, pp.139-154, OECD, Paris, France

Engle, R. F. (1982) "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation", *Econometrica*, 50, pp.987-1008

Fuller, W. A. (1996) *Introduction to Statistical Time Series 2ed*, John Wiley & Sons

Guo, Jianhua (2005) "Adaptive Estimation and Prediction of Univariate Traffic Condition Series", Ph.D. dissertation, North Carolina State University, Raleigh, NC

Highway Capacity Manual, TRB, National Research Council, Washington D.C., 2000

Revised Traffic Flow Theory Monograph, FHWA, http://www.tfhrc.gov/its/tft/tft.htm

Huynh, Nhan, Hani S. Mahmassani, and Hossein Tavana (2002) "Adaptive speed estimation using transfer function models for real-time dynamic traffic assignment operation", Transportation Research Record, 1783, pp.55-65

Kang, K. and B. Schmeiser, (1990) "Graphical Methods for Evaluating and Comparing Confidence-Interval Procedures", *Operations Research*, 38, pp.546-553

Qin, Xiao and Hani S. Mahmassani (2004) "Adaptive calibration of dynamic speed-density relations for online network traffic estimation and prediction applications", Transportation Research Record, 1876, pp.82-89

Schmeiser, B. and Y. Yeh, "On Choosing A Single Criterion for Confidence-Interval Procedures", *Proceedings of the 2002 Winter Simulation Conference*, E. Yücesan, C.-H. Chen, J.L. Snowdon, and J.M. Charnes, eds., 2002

Smith, Brian L. and R. Keith Oswald (2003) "Meeting Real-Time traffic flow forecasting requirements with imprecise computations", Computer-Aided and Infrastructure Engineering, 18, pp.201-213

Smith, B. L., Williams, B. M., and Oswald, R. K., (2002), "Comparison of Parametric and Non-parametric Models for Traffic Flow Forecasting", *Transportation Research Part C,* Vol.10, pp.303-321

Tavana, Hossein and Hani S. Mahmassani (2000) "Estimation and application of dynamic speed-density relations by using transfer function models", Transportation Research Record, 1710, pp.47-57

Williams, B. M., (1999), Modeling and Forecasting Vehicular Traffic Flow as a Seasonal Stochastic Time Series Process, PhD dissertation, University of Virginia

Williams, B. M. and Hoel, L. A., (2003), “Modeling and Forecasting Vehicular Traffic Flow as A Seasonal ARIMA: Theoretical Basis and Empirical Results”, *Journal of Transportation Engineering*, Vol. 129, No.6, pp.664-672