Space-time Multivariate Negative Binomial Regression for Urban Short-Term Traffic Volume Prediction

Yousef-Awwad Daraghmi*, Chih-Wei Yi*, and Tsun-Chieh Chiang†
*Department of Computer Science, National Chiao Tung University, 1001 University Road, Hsinchu City 30010, Taiwan.
Email:yousef@cs.nctu.edu.tw, cwyi.nctu@gmail.com.
†Division of Communication Systems, Information and Communication Research Lab, ITRI, Taiwan.
Email: tcchiang@itri.org.tw

***Abstract*—The accuracy of short-term traffic volume prediction in urban areas depends on the traffic volume characteristics and how prediction models address these characteristics. In this paper, we propose a space-time multivariate Negative Binomial (NB) regression for short-term traffic volume prediction in urban areas. The NB regression spatially correlates multiple overdispersed traffic volumes on multiple roads. We add the temporal correlation of volumes by allowing each volume to correlate with its values at previous time segments. Data consisting of traffic volumes collected in Taipei city are used to verify the model. The root mean square error is used to compare the proposed model with the Holt-Winters (HW) and Multivariate Structural Time series (MST) models. The results show that the proposed model is more accurate than the HW and MST models in all traffic conditions. The proposed model also determines causal interactions among spatial variables which assists in identifying roads affecting the prediction accuracy. Upstream roads are always significant, distant roads are always insignificant and downstream roads are significant during rush hours only.**

***Index Terms*—Autocorrelation, causal interactions, Negative Binomial, overdispersion, short-term prediction, spatial correlation, traffic volume.**

I. Introduction

One of the applications of the intelligent transportation system is congestion prevention in urban areas which requires continuous prediction of traffic volumes. A prediction for a short-term horizon, around 15 minutes, is sufficient to provide information regarding future traffic conditions and achieve efficient traffic management strategies [1]. In this context, the short-term prediction is to predict traffic volumes on certain roads or road segments of a route at short and sufficient time ahead. Different short-term prediction models have been proposed but their accuracy depends on how these models address traffic volume characteristics, and how they deal with relationships among roads. In this work, we study the problem of how to improve the accuracy of the traffic volume prediction in urban areas.

Traffic volumes in urban areas have spatial and temporal characteristics. Spatially, roads have causal interactions which play a major role in determining the evolution of traffic volumes over time. For example, a traffic volume on a road is affected by traffic flows on upstream roads and downstream roads. Consequently, traffic volume patterns and variations are correlated with the status of traffic flows in upstream and downstream roads. Temporally, each road's traffic volume is correlated with previous volumes, which is called temporal autocorrelation and is also affected by volumes on other roads. Additionally, vehicles often have high variant arriving and leaving rates which cause overdispersion in traffic volumes, i.e. the variance is always greater than the mean.

To address the above characteristics, we propose the space-time Negative Binomial (NB) regression. The NB regression is selected because it is a well-known method for modeling overdispersed counts, it allows including multiple variables, and can identify cause-effect relationships among variables [2]. We represent correlated roads in the space by the NB dependent variable and predictors. We also let each variable correlate with its values at previous times. The space-time NB regression handles overdispersion, captures temporal autocorrelations and identifies causal interactions among correlated roads. We use the L-BFGS-M optimization technique, which requires linear time and space complexity [3], to reduce the complexity of the NB regression.

Data from Taipei city are used to verify the proposed method. The residual error and the root mean square error (RMSE) are calculated to compare the proposed model with the Holt-Winters (HW) [4], [5] and the multivariate structural time series (MST) [6]. The HW and MST methods are chosen because they have the same complexity as our method since their coefficients can optimized using the L-BFGS-M algorithms [7], and to enable the comparison with one univariate method (HW) and one multivariate method (MST).

The contribution of our work is a space-time NB regression which handles traffic volume overdispersion, captures autocorrelations, has linear time and space complexity and outperforms other prediction methods used in this field such as the HW and MST. The model can be used for all traffic conditions such as low traffic, average traffic and rush hours. Using the proposed model, the causal interactions among correlated roads is identified. Upstream road segments are always significant, downstream segments are significant during rush hours, and distant segments are always insignificant.

II. RELATED WORK

The literature of prediction methods in the transportation field reveals that multivariate methods, which focus on modeling traffic parameters of different locations, are more accurate than univariate methods [1], [6], [8]. Limited variations of multivariate models for the short-term prediction in urban areas can be found in the literature. The most related models can be found in [1], [6], [8]–[10]. A multivariate time-series state space model based on the Autoregressive Integrated Moving Average (ARIMA) was used to model and predict traffic volumes [1]. The results showed that there is a need for prediction models that can capture traffic behavior during rush hours [1]. The used model suffers from high computational complexity due to the stationary requirements and the large number of estimated parameters [6]. Experiments showed that ARIMA based prediction models have time complexity of $O(n^2)$ where n is the size of data [11].

Another example of the used multivariate models is the Space-Time ARIMA (STARIMA) [8] which is also ARIMA-based and suffers from high computational complexity. In this study [8], the distance between two locations was assumed too long so that congestions would not affect the flow patterns, and distant roads were assumed having no effect on the prediction. The results showed that the traffic speed on a road is correlated with the speed on upstream roads [8]. However, the effect of downstream roads was not examined. The only study outlined the correlation between a road speed and its downstream road speed is [12] which focused on traffic speed on highways.

To reduce the computation complexity of the aforementioned models, a Multivariate Structural Time Series model (MST) that uses unrelated time series equation was proposed [6]. The MST model combines seasonal components such as trends and cycles with other explanatory variables to model and forecast 15-minute aggregated traffic volumes [6]. The results showed that including volumes of the nearest upstream junctions as explanatory variables improves the prediction accuracy while distant junctions do not affect it [6]. The model was accurate in normal traffic conditions and was not tested under congestion state. The effect of downstream roads on upstream roads was not examined in the model. The computation complexity was significantly reduced because the model does not require stationary data and many estimated parameters.

The aforementioned studies are fundamental in the transportation field as they addressed the temporal and the spatial correlations. However, the related studies did not provide any model that works in all traffic conditions. None of the studies handled traffic volume overdispersion and consider the causal interactions among roads. Compared to the related studies, the proposed model focuses on the temporal and the spatial correlations, handles traffic volume overdispersion and considers causal interactions among spatial variables.

III. THEORETICAL BACKGROUND

The NB regression is the most popular method for modeling overdispersed counts. It reveals the correlation and the cause-effect relationship among the variables [2]. Other count-modeling methods such as the Poisson models have been used to model traffic density [13]. The Poisson models assume equidispersion; the variance and the mean are equal which makes the Poisson models inappropriate when the variance is greater than the mean, i.e. overdispersion [2]. Unlike the Poisson models, the NB regression handles overdispersion by employing a parameter, φ, which makes the variance, σ^2, equal to $\mu + \varphi\mu^2$ where μ is the mean [2].

In the generalized linear model of the NB regression, the mean, μ, of a dependent variable, $\mathbf{Y}$, is related to an exponential function of a predictor, $\mathbf{X}$, by $\mu = e^{\mathbf{X}'\beta}$, where β is a regression coefficient and $\mathbf{X}'$ is the transpose of $\mathbf{X}$ [2]. When multiple predictors exist, the regression equation can be written as

$$\ln \mathbf{Y} = \alpha + \mathbb{X}\beta + \eta \tag{1}$$

where $\mathbb{X}$ is the predictors' matrix, α is the intercept, β is the regression coefficients vector, and η is the regression error vector where the error is independent of all covariates and distributed with mean $= 1$ and a variance$= 1/\varphi$. This model assumes no feedback from the dependent variable to the predictors which means that the predictors affect the dependent variable and not the opposite.

The best fit model is chosen by selecting predictors that have significant regression coefficients. A variable is significant when its coefficient's P-value is less than the significance level. Usually a significance level of 5% is used in stepwise forward or backward elimination algorithms to estimate regression coefficients and eliminate insignificant variables. The estimation also employs optimization methods such as Fisher's scoring method which estimates the coefficients based on maximizing a likelihood function [2]. However, these techniques may have high time complexity which may reach $O(n^3)$ [14]. To reduce the computational complexity of coefficients optimization, we use a different technique as will be shown in the proposed model. The log-likelihood function used in [2] performs well for few predictors. To enable including multiple predictors, we use the log-likelihood function derived in [15] and given by

$$f_{NB} = \sum_{i=1}^{N} \left(\begin{array}{c} \sum_{j=1}^{n} y_{ij} \ln \mu_{ij} + y_i \ln \varphi + \ln \Gamma(y_i + \frac{1}{\varphi}) \\ -(y_i + \frac{1}{\varphi}) \ln \Gamma(\varphi\mu_i + 1) - \ln \Gamma(\frac{1}{\varphi}) \end{array} \right) \tag{2}$$

where Γ is the gamma function, and y_{ij} is the the value of the variable i at time j [15]. The best fit model is also chosen based on the smallest value of Akaike Information Criterion (AIC) [2].

IV. METHODOLOGY

A. Data set

The traffic volume data were collected in 13 signalized arterial roads in Taipei city, Fig. 1. Microwave Radar Vehicle Detectors were used to record the data every two minutes for

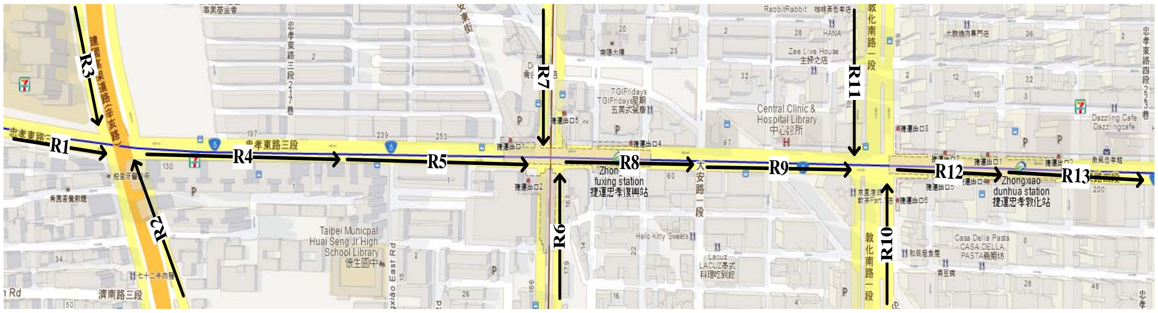

Fig. 1. Map of selected roads in Taipei city

19 days from January 21, 2008 at 00:00 to February 8, 2008 at 11:58.

Initially, the data were inspected and invalid records were found including missing volume values, negative volume values and zero volume values when speed is greater than zero. The invalid data resulted from detectors malfunction or failure. The number of the invalid records of each road was less than 20 records per day which is not large and does not affect the accuracy of modeling. To replace an invalid data record, we used an interpolation function that calculates the average of the preceding value and the following value of that record.

Here, we classify the road segments into a targeted road segment, upstream segments and downstream segments. The adjacent upstream/downstream segments to the targeted segment are called direct segments, the neighbors of the adjacent segments are called indirect segments and others are distant segments. The upstream segments are obviously the main source of traffic to the targeted segment. The downstream segments may affect the upstream during rush hours mainly when roads are short [10], [12]. This assumption is tested in this study.

B. Data characteristics

The volume time series data of a selected road for 19 days is plotted in Fig. 2. Fig. 2 illustrates how the data have weekly and daily seasonal patterns which are repeated periodically. The weekly pattern includes the work-day pattern of five days cycle length and the week-end pattern of two days cycle length. The daily pattern includes different variations categorized as: low traffic when the volume is less than the mean and often exists in the early morning from 00:00 to 6:00, rush hours when the volume is greater than the mean and covers time periods from 7:00 to 9:00 and 17:00 to 19:00, and average traffic when the volume is around the mean. The mean values of selected roads are shown in Table I. Within the above traffic categories, the mean and the variance are time dependent and they have different values.

In addition to the seasonal patterns, the volume is overdispersed. As shown in Table I, the variance of the volume in each road is greater than the mean because of the volume variations within the seasonal patterns and the volume fluctuation due to traffic lights. The green light phase decreases the volume and the red light phase increases it, and the fluctuation is

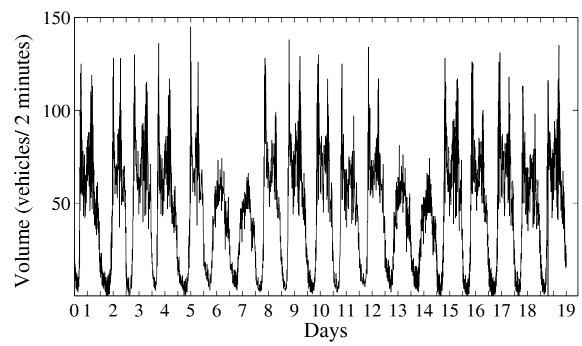

Fig. 2. Traffic volume of a selected road for 19 days

significant during rush hours which agrees with the results in [16]. The variations may be caused also by other effects such as changing weather conditions, road work, and driving behavior.

TABLE I
THE OVERDISPERSION OF VOLUME IN THE SELECTED ROADS

Roads	R1	R4	R5	R8	R9	R12	R13
Mean	45.4	39.3	54.9	47.2	48.7	60.9	62.9
Variance	865	830	853	1001	608	989	1270

C. The space-time Negative Binomial regression

Our goal is to find a multivariate prediction method that can address the traffic volume temporal and spatial correlation. In this section we formulate the space-time multivariate model and deal with important characteristics of the research data that are: overdispersion, autocorrelations and causal interactions.

We use the NB regression to handle the overdispersion and model spatially correlated roads. When traffic volume data come from N correlated roads $(\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_N)$, each road's volume is a spatial variable and is represented by a vector of size n as $\mathbf{Y}_i = (\mathbf{y}_{i,1}, \mathbf{y}_{i,2}, ..., \mathbf{y}_{i,n})$. One road's volume is considered a dependent variable, $\mathbf{Y}_d$, and the other $N-1$ roads' volumes are predictors.

The temporal correlation can be addressed by allowing the volume of each road to correlate with itself at previous time segments. A similar temporal correlation approach that focused on highway speed temporal correlations for limited time segments, $t-1$ and $t-2$, using fixed effect regression methods was introduced in [12]. In our approach, we focus on the whole temporal autocorrelation, $t-1, t-2, ..., t-n$, for all spatially correlated road segments. We use subscript i to denote the road number and a subscript j to indicate that road i has a data vector at a previous time j, *e.g.,* $\mathbf{Y}_{3,2}$ is the traffic volume vector for road three two time segments ago, $t-2$. The multivariate space-time NB regression model can be written compactly as

$$\ln y_{d,t} = \alpha + \sum_{i=1}^{N}\sum_{j=1}^{n}\beta_{i,j} y_{i,j} + \eta \tag{3}$$

where $y_{d,t}$ is the traffic volume value of the dependent road at time t, $\beta_{i,j}$ is the regression coefficient corresponding to road i at time j and $1 \leq d \leq N$. The model in equation 3 contains, within the summation, the temporal autocorrelation part of the dependent road and the predictors. The coefficient β corresponds to the spatial and temporal correlation. This model handles the overdispersion by making the variance, σ^2, equal to $\mu + \varphi\mu^2$, and capture the temporal autocorrelation by referring to previous time segments, and identify causal interactions among roads since any change in the predictors causes a change in the dependent variable.

We refer to the model as the space-time NB regression. The regression coefficients are estimated as in the NB binomial regression. The complexity of the model depends on the optimization algorithm which is the L-BFGS-B method. The L-BFGS-B algorithm is an extension of the limited-memory BFGS (Broyden–Fletcher–Goldfarb–Shanno (BFGS)). The L-BFGS-B algorithm requires a function to be minimized, a lower bound and an upper bound. The function is the negative of the log-likelihood function because the log-likelihood function should be maximized not minimized, and the lower and the upper bounds of the coefficients are -1 and 1, respectively. The time and space complexity of the L-BFGS-B algorithm can be kept to a linear order, i.e. $O(n)$ where n is the size of data [3]. Therefore, the space-time NB regression requires little running time and memory to estimate the parameters and predict future values. An example of using the BFGS method to estimate regression coefficients is shown in [17]. The proposed model is verified by using it to model and predict the traffic volume on selected roads in Taipei city.

V. Results and discussion

In this section we show the results of using the space-time NB regression for modeling and predicting data. We also compare the proposed model with the Holt-Winters (HW) method [4], [5] and the multivariate structural time series (MST) [6]. To train the models, we use the data recorded from January 21, 2008, at 00:00 to February 1, 2008, at 23:58, a total of 10 days, excluding weekends because they have different patterns. To test the prediction accuracy, we use the data recorded from February 4, 2008, at 00:00 to February 8, 2008, at 23:58 and we test the prediction models during low traffic, average traffic and rush hours.

The space-time NB regression coefficients are estimated for road segments whose upstream and downstream data are available (R4, R5, R8, R9 and R12). We examine different spatial variables such as upstream roads, downstream roads and distant roads during different traffic conditions. Initially, we target R8 which is spatially correlated with three direct upstream segments (R5, R6 and R7), one direct downstream segment (R9), one indirect upstream segment (R4), indirect downstream segment (R12) and distant segments (R1, R2, R3, and R13), as shown in Fig. 1.

The three traffic conditions cover different day times and have different lengths. If a single NB regression model is used for the whole day, the traffic condition with larger period will be given more weight than the others. This causes the prediction to be accurate in the large period only. To overcome this problem, we model each period separately by training the models using the low traffic period from 00:00 to 6:00, average traffic period, e.g. from 10:00-12:00, and the high traffic period, e.g. from 7:00 to 9:00. To model the temporal correlation, we use the stepwise forward elimination algorithm. We start by traffic volumes of roads at time segments $t-1$, then $t-2$, until $t-5$. we stop at $t-5$ because, at this time, the temporal correlations become insignificant. For R8, Table II presents the detailed spatial correlation and regression coefficients. The insignificant coefficients are represented by (-) in the Table.

Table II shows the coefficients related to space segments and time segments. Segments with P-values greater than the significance level (0.05) are not significant, and the significance increases when the P-value decreases. This is true for distance segments in space such as R1 and R13, and for distant segments in time such as $t-4$. Further, the coefficients values and degree of significance, in Table II, help to identify the effect of the spatial variables. The direct and indirect upstream segments are significant during all traffic conditions due to the usual traffic flow. The distant spatial variables R1 and R13 are insignificant and do not affect the targeted road. Additionally, the downstream segments are significant during the rush hour due to the large volume and the possible backward spread of congestions mainly when the distance between segments is not long. The length of road segments in our case ranges from $200m$ to $400m$ which is not enough to prevent the backward spread of congestions in crowded and busy areas. The overdispersion parameter is greater than zero in all traffic periods, which justifies the use of the NB regression.

We use the coefficients shown in Table II to produce prediction for several steps ahead, from one step to twenty steps and each step is two minutes. For each traffic condition, we use a different model because the coefficients are different. We compare the space-time NB with the HW and MST for a 20-minute prediction horizon. We use a single model for the HW and the MST methods during all traffic conditions. Table III presents the root mean square error (RMSE) of the models for selected roads during three traffic conditions for a 20-minute horizon. Figure 3 shows the three models during the three traffic periods.

In Table III, the RMSE values of the space-time NB

TABLE II
THE SPACE-TIME NB REGRESSION COEFFICIENTS WHEN R8 IS THE DEPENDENT ROAD

Variables	Coefficients β					
	low traffic		average traffic		rush hour	
	estimate	P-value	estimate	P-value	estimate	P-value
Intercept	2.13	2×10^{-16}	2.33	3.5×10^{-16}	2.75	5.1×10^{-16}
autocorrelation R_8, $t-1$	0.0830	0.62×10^{-16}	0.0852	1.4×10^{-16}	0.0892	4.1×10^{-16}
autocorrelation R_8, $t-2$	0.0680	3.3×10^{-10}	0.00753	4.7×10^{-11}	0.0795	3.4×10^{-12}
autocorrelation R_8, $t-3$	0.0461	1.2×10^{-5}	0.0512	3.8×10^{-6}	0.0563	9.1×10^{-7}
autocorrelation R_8, $t-4$	-	0.063	-	0.059	-	0.068
direct upstream R_5, $t-1$	0.0161	6.1×10^{-9}	0.0177	2.3×10^{-9}	0.0262	1.8×10^{-10}
direct upstream R_5, $t-2$	0.0061	2.0×10^{-9}	0.0082	5.4×10^{-9}	0.0158	1.3×10^{-9}
direct upstream R_5, $t-3$	0.0012	7.5×10^{-6}	0.0025	1.7×10^{-6}	0.0047	2.9×10^{-7}
direct upstream R_6, $t-1$	0.0123	1.1×10^{-9}	0.0145	2.4×10^{-9}	0.0280	1.6×10^{-10}
direct upstream R_6, $t-2$	0.0042	2.4×10^{-7}	0.0071	4.2×10^{-9}	0.0104	2.6×10^{-9}
direct upstream R_6, $t-3$	0.0017	4.1×10^{-6}	0.0037	9.2×10^{-6}	0.0055	3.5×10^{-7}
direct upstream R_7, $t-1$	0.0091	2.7×10^{-9}	0.0127	2.4×10^{-9}	0.0151	1.7×10^{-9}
direct upstream R_7, $t-2$	0.0024	1.4×10^{-8}	0.0035	4.5×10^{-8}	0.0064	2.4×10^{-9}
direct upstream R_7, $t-3$	0.0009	4.1×10^{-4}	0.0015	3.1×10^{-5}	0.0033	5.3×10^{-6}
indirect upstream R_4, $t-1$	0.0037	2.6×10^{-3}	0.0045	5.1×10^{-4}	0.0062	4.8×10^{-4}
indirect upstream R_4, $t-2$	0.0008	0.006	0.0014	0.003	0.0042	0.001
indirect upstream R_4, $t-3$	-	0.61	-	0.3	-	0.21
direct downstream R_9, $t-1$	-	0.49	0.0001	0.044	0.0013	0.008
direct downstream R_9, $t-2$	-	0.33	-	0.72	0.0011	0.023
indirect downstream R_{12}, $t-1$	-	0.12	-	0.24	0.0017	0.02
distant R1, $t-1$	-	0.23	-	0.24	-	0.37
distant R1, $t-2$	-	0.15	-	0.58	-	0.52
distant R1, $t-3$	-	0.45	-	0.71	-	0.82
# of observations	1800		600		600	
overdipsersion φ	2.33		2.64		2.07	

regression are less than the HW and the MST. The space-time NB regression performs better than the HW because the latter neither accounts for the spatial correlation nor the overdispersion. The MST model is multivariate and can include several spatial road segments but its inability to handle the overdispersion makes it less accurate than the space-time NB regression. During the rush hour, the HW and MST suffer from low accuracy since the RMSE values increase significantly. The plots in fig. 3 also show that the space-time NB regression performs better than the HW and the MST. However, the space-time NB regression accuracy starts to deteriorate before the MST during the low and average traffic conditions, as the space-time deteriorates at 20-minute horizon while the MST deteriorates at 24-minute horizon. During the rush hour, the MST accuracy deteriorates before the space-time NB regression. The space-time NB regression maintains the same deterioration time during all traffic conditions. Other prediction models such as the ARIMA based models, e.g. models in [1] and [8] also suffer from low accuracy during rush hours. Clearly, the space-time NB regression is more accurate than the HW and the MST methods during all traffic conditions

The prediction using the space-time NB regression depends on the real time values of the dependent variable and the predictors, as well as the values associated with the temporal correlations at previous time segments. The prediction using the NB regression is one step ahead which may be very short, and the prediction for a longer time requires knowing the values of the predictors at that time. However, when the NB regression is used during one traffic condition, where data are in the same level, the prediction horizon can be extended to several steps assuming that the future values of the dependent variable and the predictors will follow the real values and the temporal correlated values. When data transit from a level to another level, i.e. from one traffic condition to another, the NB regression accuracy decreases because the temporal correlation becomes invalid. In traffic contexts, data have seasonal patterns, meaning that traffic conditions are repeated daily or weekly. A single NB regression can be used during one season only and can not capture all seasons.

TABLE III
COMPARISON BETWEEN PREDICTION MODELS FOR DIFFERENT ROADS IN DIFFERENT TRAFFIC PERIODS.

Traffic condition	Method	RMSE			
		R4	R5	R8	R9
Low traffic	Space-time NB	3.13	2.26	1.79	2.45
	HW	3.35	2.72	2.16	3.42
	MST	2.96	2.73	2.74	2.29
Average traffic	Space-time NB	3.53	4.54	4.59	4.18
	HW	5.86	6.62	5.95	7.61
	MST	3.76	3.10	3.52	3.64
Rush hour	Space-time NB	6.5	5.73	5.47	4.97
	HW	10.9	11.2	9.42	9.62
	MST	6.42	6.35	5.87	6.20

VI. CONCLUSION AND FUTURE WORK

Based on the results of previous studies which stated that the multivariate prediction approach is more accurate than the univariate approach, we have proposed a space-time multivariate NB regression model for short-term traffic volume prediction

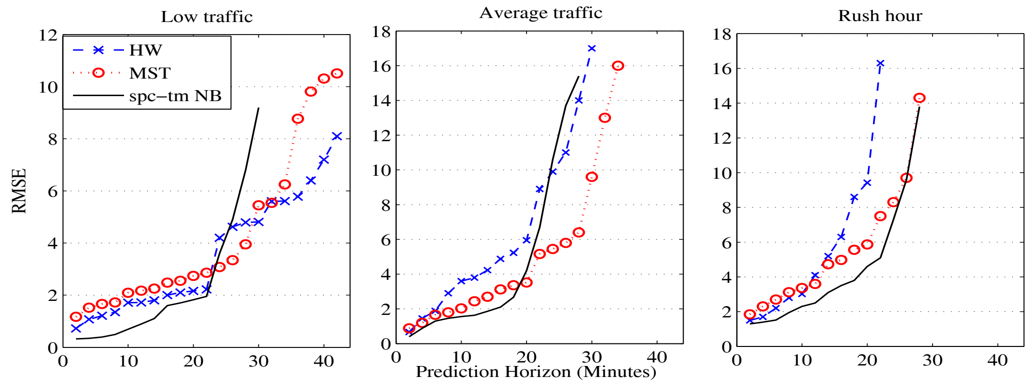

Fig. 3. RMSE against prediction horizons during the three traffic periods

in urban areas where volumes are autocorrelated and overdispersed. The space-time NB regression has linear time and space complexity and outperforms the HW and MST methods. Causal interactions among roads can be identified using the proposed model and consequently spatial variables influencing the accuracy of urban traffic volume prediction can be known. In agreement with [6] and [8], the direct/indirect upstream road segments are significant during all traffic conditions while distant segments are always insignificant. In addition to the results found in related studies, the direct/indirect downstream segments are significant only during rush hours.

Our future work will include some improvements of the proposed model so that it can capture traffic volumes seasonal patterns. We will combine the NB regression with models that deal with autocorrelation and seasonal patterns such as the HW [4], [5], double cycle HW [18] and the multiple cycle HW [19].

Acknowledgment

This work of C.-W. Yi was supported in part by NSC under Grant No. NSC100-2218-E-009-002 and NSC101- 2628-E-009-014-MY3, by Sinica under Grant No. AS-102- TP-A06, by ITRI under Grant No. 101-EC-17-A-05-01-1111 and 101-EC-17-A-03-01-0619, by the MoE ATU plan, and by D-Link, Inc..

References

[1] A. Stathopoulos and M. Karlaftis, "A multivariate state space approach for urban traffic flow modeling and prediction," *Transportation Research Part C*, vol. 11, no. 2, pp. 121–135, April 2003.

[2] J. M. Hilbe, *Negative Binomial Regression*, 2nd ed. Cambridge: Cambridge University Press, 2011.

[3] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM Journal on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, Septmeber 1995.

[4] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 664–672, November 2003.

[5] B. Basu, B. Ghosh, and M. M. O'Mahony, "Time-series modelling for forecasting vehicular traffic flow in Dublin," in *Proceedings of the 84th Transportation Research Board Annual Meeting*, Washington D.C., USA, January 2005, pp. 1–22.

[6] B. Ghosh, B. Basu, and M. OMahony, "Multivariate short-term traffic flow forecasting using time-series analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 2, pp. 246–254, June 2009.

[7] J. Durbin and S. J. Koopman, *Time Series Analysis by State Space Methods*. Oxford University Press, 2001.

[8] Y. Kamarianakis and P. Prastacos, "Space-time modelling of traffic flow," *Computers and Geosciences*, vol. 31, no. 2, pp. 119–133, March 2005.

[9] S. Sun and X. Xu, "Variational inference for infinite mixtures of gaussian processes with applications to traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 466–475, May 2011.

[10] Y. Kamarianakis and P. Prastacos, "Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches," in *Proceedings of the 82nd Transportation Research Board Annual Meeting*, Washington D.C., USA, January 2003, pp. 1–25.

[11] M. Krunz and A. Miskowski, "A source model for VBR video traffic based on $M/G/\infty$ input processes," in *Proceedings of the 17th Annual Joint Conference of the IEEE Computer and Communications Societies INFOCOM 98*, no. 3, San Francisco, USA, March 1998, pp. 1441–1448.

[12] J. Hu, I. Kaparias, and M. Bell, "Spatial econometrics models for congestion prediction with in-vehicle route guidance," *IET Intelligent Transportation Systems*, vol. 3, no. 2, pp. 159–167, June 2009.

[13] S. Panichpapiboon and W. Pattara-atikom, "Evaluation of a neighbor-based vehicle density estimation scheme," in *8th International Conference on ITS Telecommunications*, Phuket, Thailand, 24 October 2008.

[14] R. Ksantini, D. Ziou, B. Colin, and F. Dubeau, "Weighted pseudometric discriminatory power improvement using a bayesian logistic regression model based on a variational method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 253–266, February 2008.

[15] I. L. Solis-Trapala1 and V. T. Farewell, "Regression analysis of overdispersed correlated count data with subject specific covariates," *Statistics in Medicine*, vol. 24, no. 16, pp. 2557–2575, August 2005.

[16] T. Thomas, W. Weijermars, and E. van Berkum, "Predictions of urban volumes in single time series," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 1, pp. 71–80, March 2010.

[17] A. Charpentier, "Multivariate probit regression using (direct) maximum likelihood estimators," May 2011. [Online]. Available: http://freakonometrics.blog.free.fr/index.php?post/2011/05/11/bivprobit

[18] J. W. Taylor, "Exponentially weighted methods for forecasting intraday time series with multiple seasonal cycles," *International Journal of Forecasting*, vol. 26, pp. 627–646, 2010.

[19] P. G. Gould, A. B. Koehler, J. K. Ord, R. D. Snyder, R. J. Hyndman, and F. Vahid-Araghi, "Forecasting time series with multiple seasonal patterns," *European Journal of Operational Research*, vol. 191, no. 1, p. 207222, November 2008.