

Short-Term Traffic Forecasting using the Local Linear Regression Model

UCI-ITS-TS-WP-02-5

Hongyu Sun ¹
Henry X. Liu ²
Heng Xiao ¹
Bin Ran ¹

¹ Department of Civil and Environmental Engineering
University of Wisconsin at Madison; Madison WI 53706, U.S.A.
sun@cae.wisc.edu, xiaoh@cae.wisc.edu, bran@cae.wisc.edu

² California PATH, ATMS Center, Institute of Transportation Studies
University of California, Irvine, hliu@translab.its.uci.edu

July 2002

Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697-3600, U.S.A.
<http://www.its.uci.edu>

Short Term Traffic Forecasting

Using the Local Linear Regression Model

Hongyu Sun

Department of Civil and Environmental Engineering
University of Wisconsin at Madison
1415 Engineering Drive, Madison, WI 53706
Tel: (608) 262-3967, Fax: (608) 262-5199
e-mail: sun@cae.wisc.edu

Henry X. Liu

California PATH
University of California at Irvine
Institute of Transportation Studies
523 Social Science Tower
Irvine, CA 92697-3600
Tel: (949) 824-2949, Fax: (949) 824-8385
e-mail: hliu@translab.its.uci.edu

Heng Xiao*

Bin Ran**

Department of Civil and Environmental Engineering
University of Wisconsin at Madison
1415 Engineering Drive, Madison, WI 53706
Tel: (608) 262-0052; (608) 262-5199
*e-mail: xiaoh@cae.wisc.edu
**e-mail: bran@engr.wisc.edu

Submitted to the 2003 TRB Annual Meeting

July 2002

Short Term Traffic Forecasting

Using the Local Linear Regression Model

Hongyu Sun¹, Henry X. Liu², Heng Xiao¹ and Bin Ran¹

¹ Department of Civil and Environmental Engineering,
University of Wisconsin at Madison, Madison, WI, U.S.A.

² California PATH, University of California at Irvine, Irvine, CA, U.S.A.

Abstract

Traffic data is highly nonlinear and also varies with times of day. It changes abruptly when entering or leaving a congestion hour. Therefore, the prediction of travel time requires accurate models. This leads to the problem of approximating nonlinear and time-variant functions. In this paper, we propose and apply local linear regression models to the short-term traffic prediction. The local linear regression is one type of the local weighted regression methods. They have been applied to many problems, including artificial intelligence, dynamic system identification, data mining, etc. It can be used for nonlinear time series prediction under certain mixing conditions. The performance of the proposed model is compared with the previous nonparametric approaches, such as nearest neighborhood and kernel methods using the 32 day traffic speed data collected on Houston US-290 Northwest Freeway. We found that the local linear methods consistently have better performance than the nearest neighborhood and the kernel smoothing method.

1 Introduction

The problem of short-term traffic forecasting is to determine the variables such as traffic volume, travel speed, and travel time in the next time window, usually in the range of five minutes to half an hour. The motivation for such forecasting is obvious, since the developments of many ITS systems such as route guidance systems, adaptive ramp metering, adaptive signal control, and variable message signs have become increasingly demanding for the accurate short-term forecasting. Depending on the data source used in the forecasting process, short-term traffic forecasting models can be mainly categorized into three groups: 1) those using only historical data, 2) those using only real-time data, and 3) those using both historical and real-time data. The key issue to the traffic forecasting problem is how to make use of these two sources of information (Lin, 2001).

A large body of past work is available on the short-term forecasting algorithms (Van Arem et al., 1997), for example, ones that based on the time-series model (Ahmed et al., 1979), Kalman filtering theory (Okutani et al., 1984; Chen et al., 2001), simulation models (Rathi et al., 1989; Van Vliet, 1980; Duncan et al., 1997), dynamic traffic assignment models (Ben-Akiva et al., 1995; Ran, 2000), neural-network models (Park et al., 1998), and nonparametric methods (Davis et al., 1991, Smith et al., 1999). The nonparametric approach has been employed in many research studies. However, most of the applications within the transportation field only investigated methods such as the nearest neighborhood (Smith et al., 1999) or kernel estimator (Faouzi, 1996). These methods belong to the approach of local constant regression in the family of the local polynomial regression models (Fan et al., 1996), the main types of the nonparametric models. Fan (Fan et al., 1996) showed that local linear regression has a high minimax efficiency among all possible estimators. This method also can handle a wide range of data distributions while avoiding boundary and cluster effects. Thus, the study of the approach to travel time prediction is expected to be rewarding.

One approach with a close relationship to this local model is the time varying coefficient linear model. It was applied to traffic data by Rice et al. (2001), but the weighting

function is determined by the difference in the departure time instead of the distance of the covariates space and the covariate is only the current time instant data. At every time point, the relationship between future and current data is linear, although our local linear model does not assume any linearity.

A good survey on local weighted regression can be found in Atkeson et al.'s paper (1997). The local model can learn locally, as opposed to the global models such as a neural network. This is implemented easily by using weighted regression whereby the weights are the kernel functions with bandwidth defining the validity region of a local model. The local linear model can give good approximations of the nonlinear time variant functions with reasonable costs (Lewandowski et al., 2001).

The remainder of the paper is organized as follows. In Section 2, after reviewing the problem of short-term traffic forecasting, we will discuss the local linear regression model formulation. An empirical study using actual freeway data is devised to test the proposed model, and compared with the nearest neighbor approach and the kernel method in Section 3. Conclusive remarks and future research directions are given in the last section.

2 Methodology

2.1 Prediction Problem

The traffic prediction problem can be described in the following way: Given the observed traffic data, $TT(i)$, $i = 1, \dots, t$, the prediction is to generate an estimate of $TT(t + s)$, where s is the prediction horizon.

Bridging smoothing and time series studies led to the nonparametric approach to the time series prediction problem (Györfi, 1989). The basic difference between the parametric and nonparametric approaches lies in the fact that for the latter we have to compute a

whole function instead of a set of parameters. Typically such a smoothing function is estimated by a local average in covariates space.

Under some mixing conditions (Robinson, 1983), the prediction problem can be formulated as the nonparametric regression models. The covariate vector \mathbf{x} is a set of $\{TT(t-d+1), \dots, TT(t)\}$ or it may include other available data such as the historical average for the same time interval over past days. And the response variable $y = TT(t+s)$. Smith et al. (1999) chose four variables in the covariates: $\mathbf{x} = \{TT(t-1), TT(t), TT_{\text{hist}}(t), TT_{\text{hist}}(t+1)\}^T$, and $y = TT(t+1)$. In this paper, all the covariates are the time series of past observations, that is, $\mathbf{x} = \{TT(t-d+1), \dots, TT(t)\}$. The choice of d will be discussed in the next section.

Figure 1 presents a flow chart of the general settings of traffic prediction.

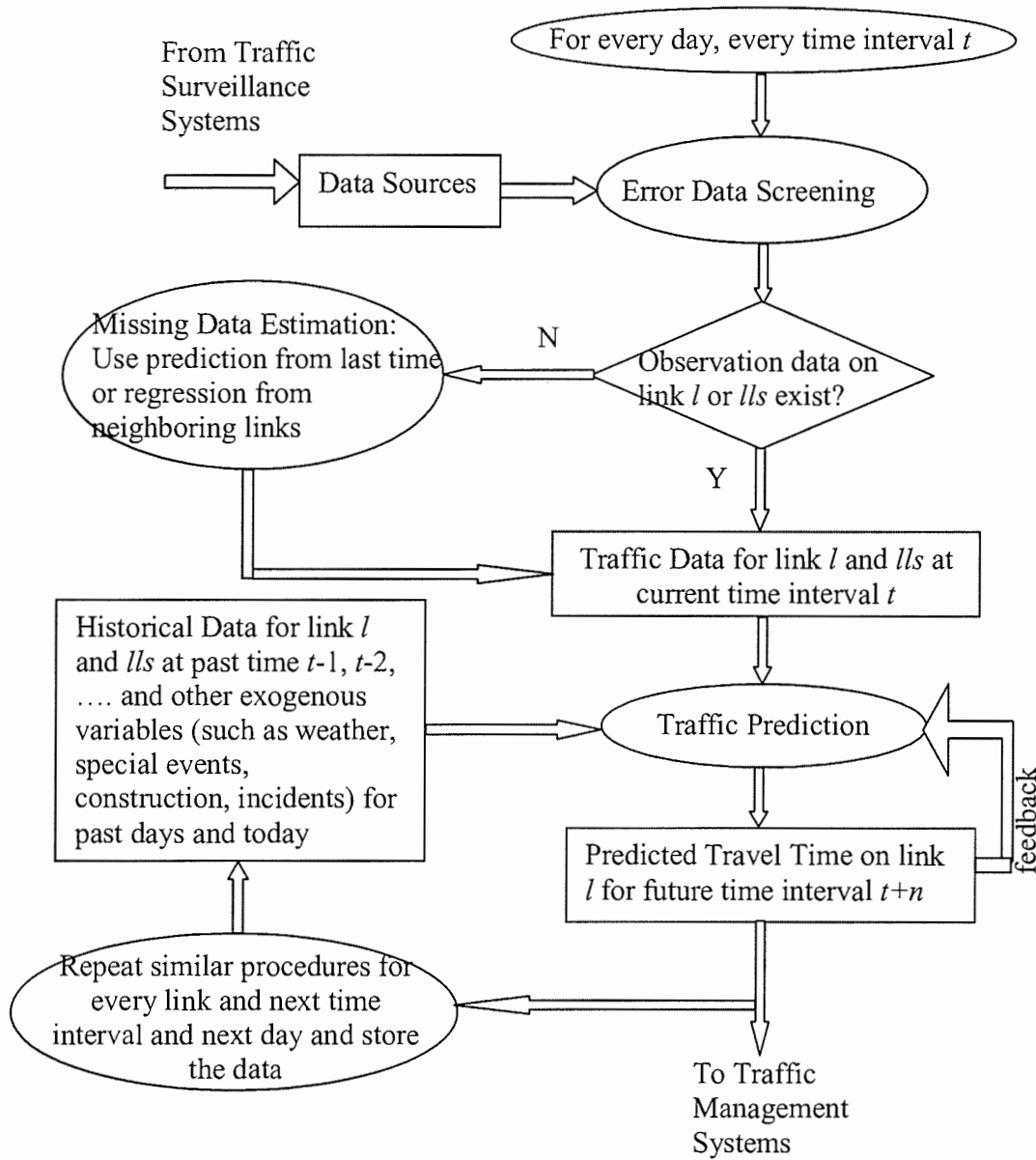


Figure 1. General Settings of Traffic Prediction

2.2 Local Linear Regression Model

We use local linear regression model to approximate the relationship of the future traffic with the past and current traffic data. Multivariate local linear regression estimator (Fan

et al., 1996) will be a main point of discussion and investigation in this paper. Given multivariate covariate \mathbf{X} and a univariate response Y , it is of interest to estimate the mean regression function, the prediction of travel time in this paper, $m(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$, where $\mathbf{x}^T = (x_1, \dots, x_d)$ is a point in \mathbb{R}^d . The $\beta = (\beta_0, \dots, \beta_d)^T$ to minimize

$$\sum_{i=1}^n \left\{ Y_i - \beta_0 - \sum_{j=1}^d \beta_j (X_{ij} - x_j) \right\}^2 K_B(\mathbf{X}_i - \mathbf{x})$$

$$\text{is } \hat{\beta} = (\hat{\beta}_0, \dots, \hat{\beta}_d)^T = (\mathbf{X}_d^T \mathbf{W} \mathbf{X}_d)^{-1} \mathbf{X}_d^T \mathbf{W} \mathbf{y}, \quad (1)$$

where the observations are given by $\{(\mathbf{X}_i^T, Y_i): i = 1, \dots, n\}$, and $\mathbf{X}_i = (X_{i1}, \dots, X_{id})^T$.

$$\mathbf{X}_d = \begin{pmatrix} 1 & X_{11} - x_1 & \cdots & X_{1d} - x_d \\ 1 & X_{21} - x_1 & \cdots & X_{2d} - x_d \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} - x_1 & \cdots & X_{nd} - x_d \end{pmatrix},$$

$$\mathbf{W} = \text{diag} \{K_B(\mathbf{X}_i - \mathbf{x})\}, \mathbf{y} = (Y_1, \dots, Y_n)^T$$

Here $K_B(\mathbf{u}) = \frac{1}{|B|} K(B^{-1}\mathbf{u})$, where K is a multivariate probability density function with mean zero and the covariance matrix of $\mu_2(K)\mathbf{I}_d$, with \mathbf{I}_d the $d \times d$ identity matrix. B is called bandwidth matrix and $|B|$ denotes its determinant. In this paper the weighting kernel K is chosen as Gaussian function and take $B = h\mathbf{I}_d$.

Thus $\hat{m}(\mathbf{x}) = \hat{\beta}_0$, and $(\frac{\partial \hat{m}}{\partial x_j})(\mathbf{x}) = \hat{\beta}_j, j = 1, \dots, d$. We use $\hat{y} = \hat{\beta}_0 + \sum_{j=1}^d \hat{\beta}_j (X_{ij} - x_j)$ as the prediction value (Atkeson et al., 1997).

The dimension d of the covariate vector and the bandwidth h are the parameters we need to select. The optimal bandwidth can be determined by cross-validation which is used in this paper. The basic idea of the cross-validation method (Fan et al., 1996) is to remove a sample $\{(\mathbf{X}_i^T, Y_i)\}$ from the training database, and use the remaining $(n-1)$ samples to build a regression function $\hat{m}_{h,-i}()$ and compute the $\hat{Y}_{i,-i} = \hat{m}_{h,-i}(\mathbf{X}_i)$, then the bandwidth is chosen to optimize an overall measure involving all values of \hat{Y}_i . In this paper, we use

the average square errors as the measure. That is, $h = \arg \min_h \left(n^{-1} \sum_{i=1}^n \{Y_i - \hat{Y}_{i,-i}^{(h)}\}^2 \right)$, where

$\hat{Y}_{i,-i}^{(h)}$ is the prediction computed with parameter h .

d can be chosen by a cross-validation-like procedure to minimize the prediction error (Faouzi, 1996). That is, $d_{opt} = \arg \min_d \left(d_x^{-1} \sum_{i=n-d_x+1}^n \{Y_i - \hat{Y}_{i,-i}^{(d)}\}^2 \right)$, where $\hat{Y}_{i,-i}^{(d)}$ is the prediction computed with d covariates, d_x is the length of the training sample used to get the optimal d . We choose $d_x = n$ in this paper.

Since we only use the time series as covariates, we could borrow the result in the nonlinear time series to choose d if the chaos exists in the nonlinear time series. d can be determined by using the time delay embedding theorem (Kantz et al., 1997), where we use delay 1 and d is the embedding dimension to be identified by false nearest neighbors or mutual information. But if the chaos is not evident or the data is too noisy, this method is not accurate.

When the matrix $\mathbf{X}_d^T \mathbf{W} \mathbf{X}_d$ is nearly singular, the ridging of the estimator was proposed (Seifert et al., 1996):

$$\mathbf{X}_d^T \mathbf{W} \mathbf{X}_d = \mathbf{X}_d^T \mathbf{W} \mathbf{X}_d + \text{diag}(\lambda_0, \dots, \lambda_d)$$

For local linear regression, ridge parameters λ_i can be calculated as $\sqrt{n_0}/3$, where n_0 is a mean number of observations in a smoothing interval. A default value of the ridge parameters is chosen in this paper for simplicity.

2.3 k -Nearest Neighborhood (k -NN) and Kernel Regression Models

The prediction of travel time by the k -NN (Fan et al., 1996) is given as

$$\hat{y} = (1/k) \sum_i^k I(\text{dis}(\mathbf{x} - \mathbf{X}_i)/h_k) Y_i,$$

where $I()$ is the indicator function, h_k is the neighbor range where k neighbor \mathbf{X}_i around \mathbf{x} can be found. In this paper the distance function is the Euclidean distance.

The prediction of travel time by the kernel regression was given (Faouzi, 1996) as

$$\hat{y} = \sum_{i=1}^n w_i Y_i, \text{ where } w_i = K_B(\mathbf{X}_i - \mathbf{x}) / \sum_{i=1}^n K_B(\mathbf{X}_i - \mathbf{x}), \text{ if the denominator is not zero;}$$

otherwise, $\hat{y} = 0$. Since the denominator becomes zero due to the sparse design when the database is small, this kernel regression formulation has obvious limitation by giving the prediction value of 0. In this paper, we reformulate it as a special case ($d=0$) of local linear regression (Fan et al., 1996) and so we can add ridging to it. That is, to replace \mathbf{X}_d in (1) with $[\underbrace{1, \dots, 1}_n]^T$, where T denotes the matrix transpose.

The relationship between kernel regression and the local weighted regression was discussed in several studies including (Fan et al., 1996). Local weighted regression is preferred over kernel regression for irregular data distributions (Hastie et al., 1993). For regular distributed data away from any boundary the local weighted regression and kernel regression are equivalent (Müller, 1987).

3 Results

One step and multiple steps prediction have been studied in this paper. We tested the cases of $s = 1, 2, \dots, 5$. Since the speed data is more comparable for different links than travel time data and it is convenient for the joint temporal spatial data mining in our future work. In the paper, speed data is used instead of travel time which can be directly computed from speed and distance. This study is based on Houston US-290 Northwest freeway eastbound traffic speed data collected from Feb to July 2002 every 5 minutes. We chose the same freeway to study as Park's research (Park et al., 1998) in order to compare the results in some way. The data is retrieved by software from the online real-time information provided to the public by the Houston TranStar Automatic Vehicle Identification (AVI) traffic monitoring system (Please refer to traffic.tamu.edu). The road

segment picked is from the cross street Sam Houston Tlwy to the cross street Fairbanks since this segment has the most usable data. The distance of this segment is 1.55 miles and the usual travel speed is about 68 mph, and the travel time is about 1:22 minutes.

The relative mean error (RME) is used as the performance index to compare the three models: nearest neighborhood, kernel smoother and local linear model.

$$\text{RME} = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|, \text{ where } Y_i \text{ is the observation value, } \hat{Y}_i \text{ is the predicted value.}$$

After deleting data from holidays and weekends, and screening out data with too many missing values, 32 days were left. For every day, the first two points are deleted since they are missing for some days, the rest of the missing data are replaced by the nearest recent time data in that day. Thus every day has 286 points. We use 30 days data as the training set and use 2 days as the testing set. Since 32 days database is not large, we leave different combinations of two days out to form the training set and the two days left out to form the testing set. Thus we ran the 992 combinations. We average all the error performance over all results from all runs to evaluate a method.

Since the eastbound is inbound, the morning peak hour has more occurrence than the afternoon peak hour. Descriptive statistics are used to summarize the distribution of our data. A plot with "error bars" reporting a mean value and a range of variation (1 error) is shown in Figure 2 for the measured speed during the morning period from 6am to 10am over the 32 days. The 1σ range would include about 68% of normal data where σ is a population standard deviation.

We find that the variation of the speed profiles from day to day in the morning peak hour is quite large. This is due to the random factors such as the construction, traffic incidents, special events, weather, vehicle types, driver types and so on. Our study doesn't assume such information is available.

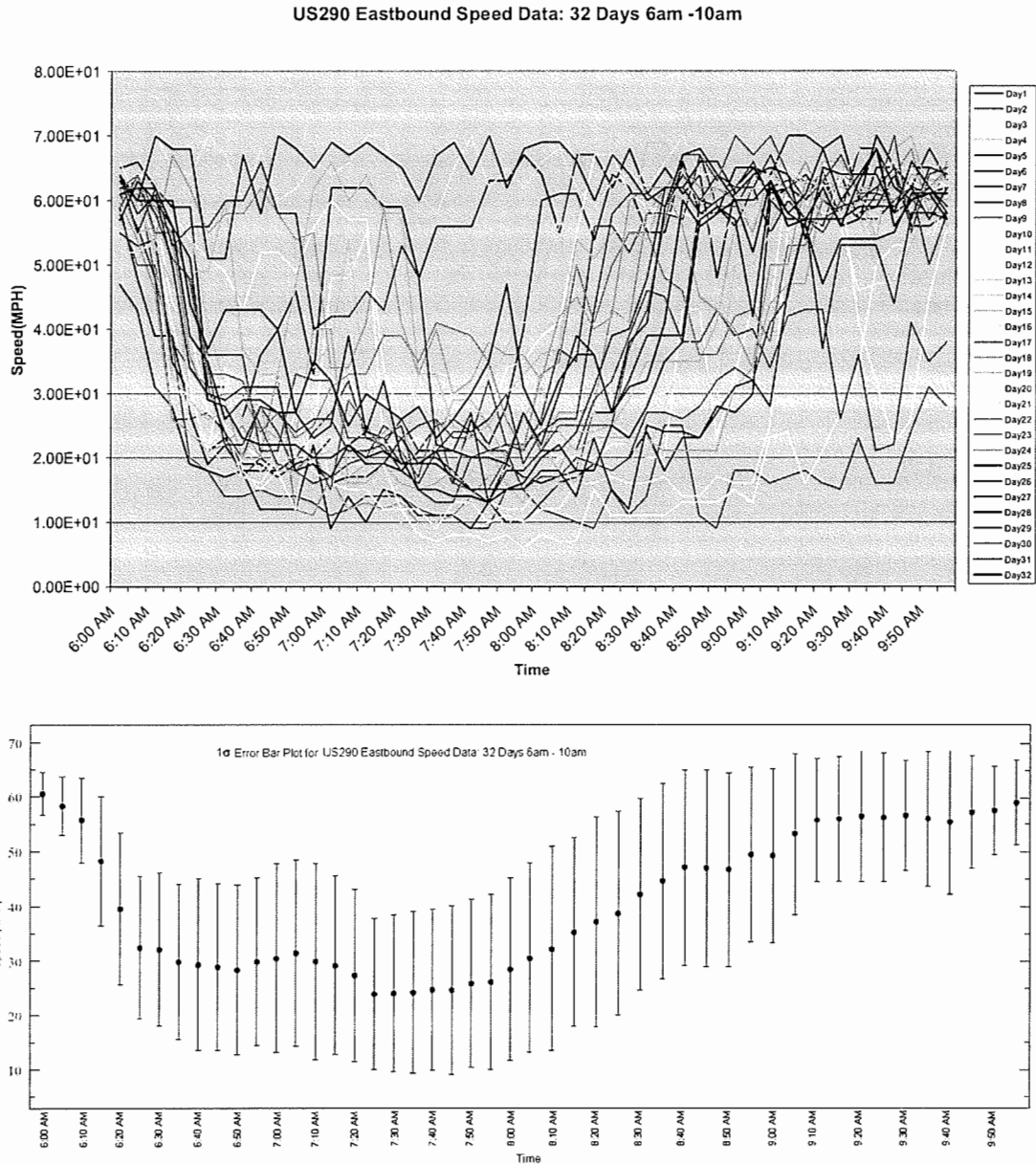


Figure 2. The US-290 Speed Data from 6am to 10am for 32 Days Selected from Feb 18 to July 10, 2002

To determine the number of covariates, that is, the number of previous and current time instants data to regress on, we use the cross-validation-like methods described in section 2 to test the prediction error performance with different d . The prediction performance with $d=1$ is not as good as $d=2$. And that with $d=4, 5$ and 11 is worse than $d=2$ also.

The d with lowest average prediction error is 3. We also tried the false neighborhood method in time delay embedding theorem (Kantz et al., 1997). Figure 3 shows the percentage of false neighbors versus the embedding dimension for the speed time series of the 32 days. We can see, at $d = 3$, the percentage of false neighborhood is about 35%. Thus, if we use 35% as the threshold to determine d , the method of time delay embedding can give the same choice of d parameter with the method of cross validation. But the real life data is noisy; the threshold may not be easy to obtain (Nair et al.). Although the cross validation result shows that $d = 3$ has a lower mean error than $d = 2$, after its trade-off with the higher computation complexity induced from the higher dimension, $d = 2$ is used in algorithm comparison. That is, the total number of the previous and current time instant data regressed on is 2.

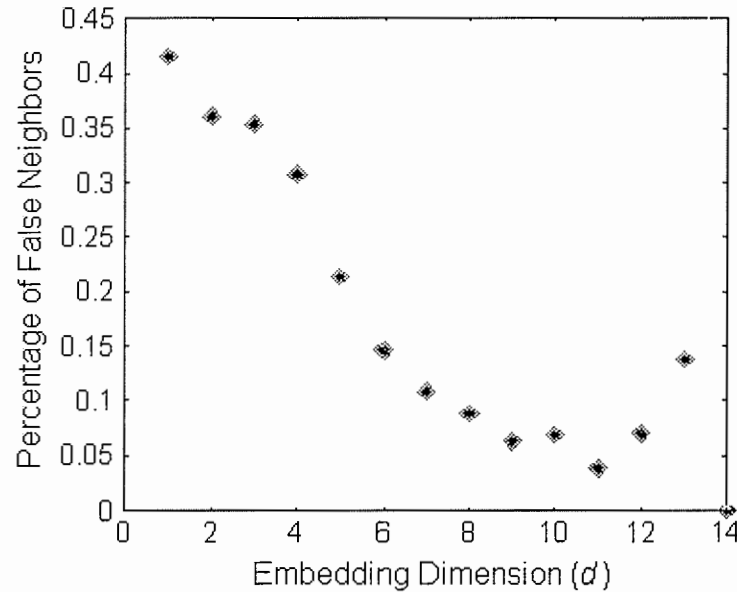


Figure 3. Plot of the Percentage of False Neighbors versus the Embedding Dimension

We tried two different ridge parameters 0.1 and 1e-10. The result shows that 0.1 is better, maybe because this value is closer to the ridge parameter formula outlined in section 2.2.

Table 1 and Figure 4 summarize the results of the average RME over 992 runs using 30 training days.

Models RME(%)	k -NN ($k=3$)	Kernel Regression		Local Linear
		No ridging	Ridging	
Prediction step =1	10.27	15.39	8.91	8.46
Prediction step =2	11.2	15.7	9.61	9.05
Prediction step =3	11.37	16.2	10.26	9.82
Prediction step =4	11.58	16.9	10.84	10.77
Prediction step =5	12.12	17.8	11.93	11.38

Table 1. Comparison of the Overall Average Relative Error for the Predictors

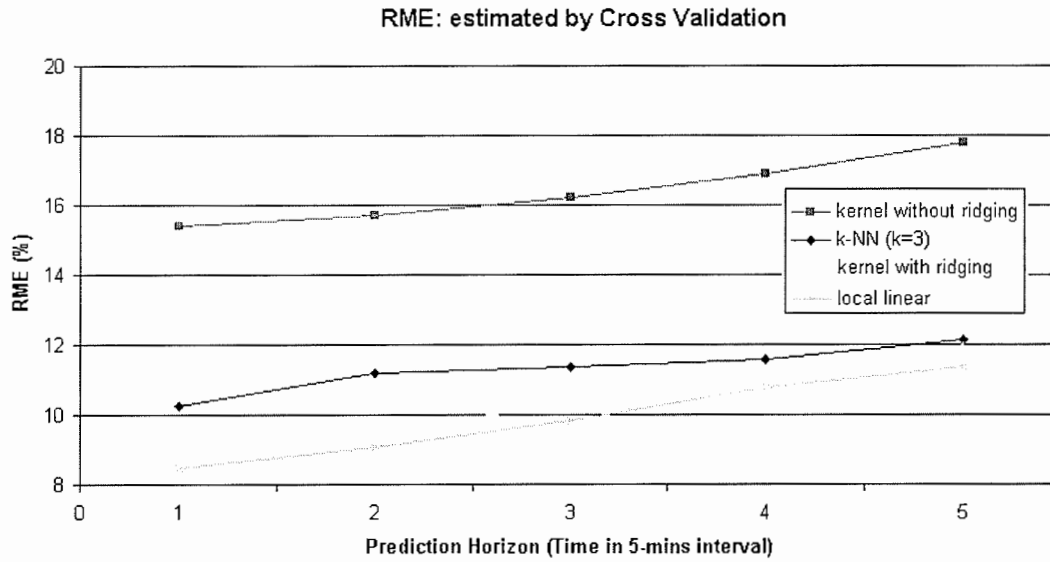


Figure 4. Relative Mean Error versus the Prediction Horizon for Different Models

The k -NN method uses $k = 3$. Since when $k = 4$ and the prediction step = 1, the RME of k -NN is 10.23%, which doesn't have much advantage than that (10.27%) under $k = 3$. Thus, we only use $k = 3$ in the comparison study.

Figure 5 illustrates the one step prediction result for a single day using local linear predictor.

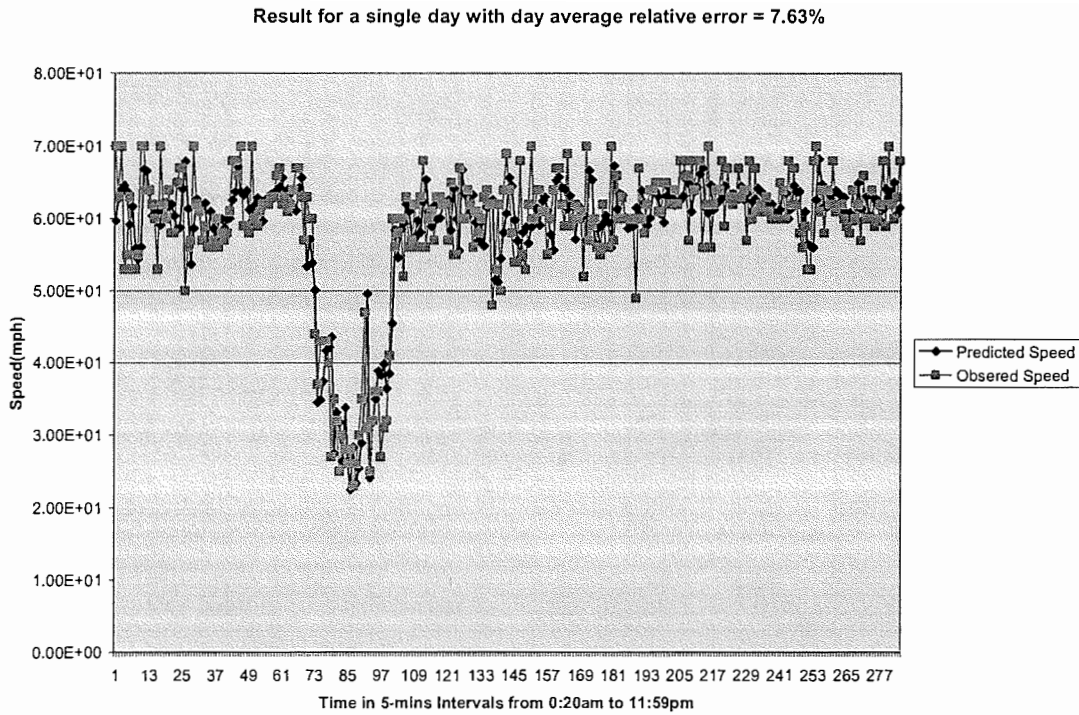


Figure 5. One Step Prediction Result for a Single Day Using Local Linear Predictor

4 Conclusions and Future Work

The results have shown that the local linear method has a better performance than the nearest neighborhood and the kernel smoothing method for our traffic speed data. This is consistent with the proof in the statistical perspective by Fan (1996) and others. Besides its statistical justification, a tentative explanation can be that both the nearest neighborhood and the kernel method use the historical data average while only using the real time (current) data for computing weight function, but the local linear model uses the current data explicitly in the prediction value. So the local linear model makes better use of both historical and real time data than the previous methods. Also, notice that the kernel method defined in (Faouzi, 1996) did not perform well due to giving zero prediction when the denominator is zero. After ridging, the performance turns out normal, that is, it is supposed to perform better than the nearest neighborhood method.

The kernel method has a close performance with the local linear method. This may be because the irregular distribution in our data under study doesn't have high proportions. If we use less free flow data or use only peak hour data, the difference between these two models may be more evident.

The multiple steps prediction and one-step prediction results are shown. All methods have decreased performance when prediction horizon increases.

Note that the nonparametric methods strongly depend on the database. When the database is larger, it may store more patterns. Then the possibilities increase for the current situation (affected by all kinds of random factors) to find well-matched patterns in the database. Thus more accurate predictions can be made. Therefore, we used cross validation methods to get the average error so as to evaluate a model in a more general sense. Also, the results could be better if we have a larger database.

Further work includes optimizing the parameters such as ridge parameters and tuning parameters such as the bandwidth adaptively. The short term prediction has a major difference relative to the long term one, that is, the feedback arrives quickly. This characteristic can be utilized by using adaptive algorithms. The gain in the ratio of the performance and computation cost is expected. Also, the missing data issue may need to be further addressed. In addition, we tried to incorporate upstream and downstream links data into the covariates, but the result needs to be improved. Therefore, it is not presented here. This could also be included in future work.

References

Ahmed, M. & Cook, A. (1979). Analysis of Freeway Traffic Time Series Data by Using Box-Jenkins Techniques." *Transportation Research Board*, No. 722, pp.1-9.

Atkeson, C., Moore, A. & Schaal, S. (1997). "Locally Weighted Learning." *Artificial Intelligence Review*, 11, 1-5, pp.11-73.

Ben-Akiva, M., Cascetta, E. & Gunn, H. (1995). "An on-line dynamic traffic prediction model for an inter-urban motorway network." *Urban traffic networks : dynamic flow modeling and control*. Berlin : Springer-Verlag, 1995.

Chen, M. & Chien, S. (2001). "Dynamic Freeway Travel Time Prediction Using Probe Vehicle Data: Link-based vs. Path-based." TRB Paper No. 01-2887.

Davis, G. & Nihan, N. (1991). "Nonparametric Regression and Short-Term Freeway Traffic Forecasting." *Journal of Transportation Engineering*, 1991, p178-188.

Duncan, G. & Littlejohn, J. (1997). "High Performance Microscopic Simulation for Traffic Forecasting." Computing and Control Division Colloquium on Strategic Control of Inter-Urban Road Networks, London, England: Strategic control of inter-urban road networks: Computing and Control Division colloquium. London: Institution of Electrical Engineers, 1997.

Fan, J. & Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. London: Chapman & Hall.

Faouzi, E. (1996). "Nonparametric Traffic Flow Prediction Using Kernel Estimator." *Transportation and Traffic Theory: Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, Lyon, France, 24-26, July, 19, 1996.

Hastie, T. & Loader, C. (1993). "Local Regression: Automatic Kernel Carpentry." *Statistical Science*, 8(2): 120-143.

Kantz, H. & Schreiber, T. (1997). *Nonlinear Time Series Analysis*, Cambridge University Press.

- Györfi, L. (1989). *Nonparametric Curve Estimation from Time Series*, Springer-Verlag.
- Lewandowski, A. & Protzel, P. (2001). "Predicting Time-Varying Functions with Local Models." *IDA2001*, LNCS 2189, pp.44-52, 2001.
- Lin, W. (2001). "A Gaussian Maximum Likelihood Formulation for Short-Term Forecasting of Traffic Flow." 2001 IEEE Intelligent Transportation Systems Conference Proceedings, 150-155, Oakland, CA, USA, August 25-29, 2001.
- Müller, H.-G. (1987). "Weighted Local Regression and Kernel Methods for Nonparametric Curve Fitting." *Journal of the American Statistical Association*, 82:231-238.
- Nair, A., Liu, J., Rilett, L. & Gupta, S. (2000?) "Non-Linear Analysis of Traffic Flow." TransLink® Research Reports,
<http://translink.tamu.edu/docs/Research/LinearAnalysisTrafficFlow/chaos1.pdf>
- Okutani, I. & Stephanedes, J. (1984). "Dynamic Prediction of Traffic Volume Through Kalman Filter Theory." *Transportation Research*, Part B, Vol. 18B, 1-11, Feb 1984.
- Park, D. & Rilett, L. (1998). "Forecasting Multiple-Period Freeway Link Travel Times Using Modular Neural Networks." TRB 1998.
- Ran, B. (2000). "Using Traffic Prediction Models for Providing Predictive Traveller Information." *Int. J. Technology Management*, Vol. 20. Nos. 3/4, 2000.
- Rathi, A. & Santiago, A. (1989). "The New NETSIM: TRAF-NETSIM 2.00 Simulation Program." Paper Presented at the Transportation Research Board 68th Annual Meeting, Washington, D.C.

Rice, J. & Van Swet, E. (2001). "A Simple and Effective Method for Predicting Travel Times on Freeways." *PATH Working Paper*, Institute of Transportation Studies, University of California, Berkeley, June 2001.

Robinson, P. (1983). "Nonparametric Estimators for Time Series." *Journal of Time Series Analysis*, Vol. 4, No.3, pp 185-207.

Seifert, B. & Gasser, T. (1996). "Finite Sample Variance of Local Polynomials: Analysis and Solutions." *Journal of the American Statistical Association*, Vol. 91, No. 433. (Mar., 1996), pp. 267-275.

Smith, B., Williams, B. & Oswald, K. (1999). "Parametric and Nonparametric Traffic Volume Forecasting." TRB Preprints CDROM 1999.

Van Arem, B, Kirby, H., Van Der Vlist, M. & Whittaker, J. (1997). "Recent Advances and Applications in the Field of Short-Term Traffic Forecasting." *International Journal of Forecasting* 13 (1997) 1-12.

Van Vliet, D. (1980). "SATURN: A Simulation-Assignment Model for the Evaluation of Traffic Management Schemes." *Traffic Engineering and Control*, Vol.21 (4).