

Configuration for Predicting Travel-Time Using Wavelet Packets and Support Vector Regression

Adeel Yusuf, Vijay K. Madiseti

School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, USA

Email: adeel@gatech.edu, vkm@gatech.edu

Received May 28, 2013; revised June 28, 2013; accepted July 5, 2013

Copyright © 2013 Adeel Yusuf, Vijay K. Madiseti. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Travel-time prediction has gained significance over the years especially in urban areas due to increasing traffic congestion. In this paper, the basic building blocks of the travel-time prediction models are discussed, with a small review of the previous work. A model for the travel-time prediction on freeways based on wavelet packet decomposition and support vector regression (WDSVR) is proposed, which used the multi-resolution and equivalent frequency distribution ability of the wavelet transform to train the support vector machines. The results are compared against the classical support vector regression (SVR) method. Our results indicated that the wavelet reconstructed coefficient when used as an input to the support vector machine for regression performed better (with selected wavelets only), when compared with the support vector regression model (without wavelet decomposition) with a prediction horizon of 45 minutes and more. The data used in this paper was taken from the California Department of Transportation (Caltrans) of District 12 with a detector density of 2.73, experiencing daily peak hours except most weekends. The data was stored for a period of 214 days accumulated over 5-minute intervals over a distance of 9.13 miles. The results indicated MAPE ranging from 12.35% to 14.75% against the classical SVR method with MAPE ranging from 12.57% to 15.84% with a prediction horizon of 45 minutes to 1 hour. The basic criteria for selection of wavelet basis for preprocessing the inputs of support vector machines are also explored to filter the set of wavelet families for the WDSVR model. Finally, a configuration of travel-time prediction on freeways is presented with interchangeable prediction methods.

Keywords: Travel-Time Prediction; Wavelet Packets; Support Vector Regression; Advanced Traveler Information System

1. Introduction

Accurate travel-time forecast information has become a fundamental component of all ATIS (Advanced Traffic Information Systems). Currently, drivers demand an accurate travel-time calculator that can forecast their commute time in advance. This forecast is even more significant in the morning and evening hours, when the commuters face jammed freeways and they want to avoid the peak-hour congestion. Drivers prefer precise information of the future traffic conditions to manage their route. Presently, most of the State Department traffic websites provide the current traffic conditions, some sites even calculate a forecast of the travel time based on the historical data and/or current data by employing a suitable algorithm [1,2].

The travel-time is dependent on multiple factors that are related through a complex-dependent relationship

with one another. Such factors include weather conditions, driver behavior, and time of the day etc. This complex-dependence makes the traffic data both non-linear and non-stationary. Consequently, accurate prediction of travel time becomes a challenging task.

Travel time prediction method can be classified from different perspectives as shown in **Figure 1**. While, a brief overview of all types is given in Section 2, the focus of this paper is on improving a short-term data driven prediction method.

Table 1 shows a brief overview of the prior art in this area. The prediction horizons in **Table 1** range from 5 minutes to 60 minutes. However, lower forecast horizons are not very useful for commuters in the real-world scenario as there are delays involved in every module of the travel-time prediction process; the process diagram of the prediction process is shown in **Figure 2**.

Artificial Intelligence methods were extensively used

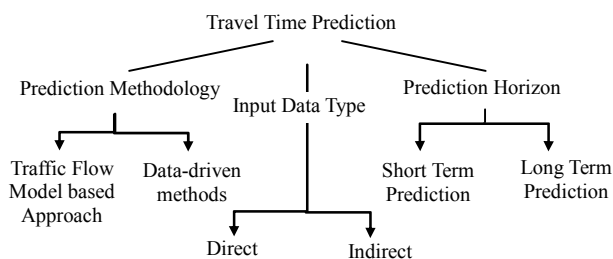


Figure 1. A taxonomy of travel time prediction approaches.

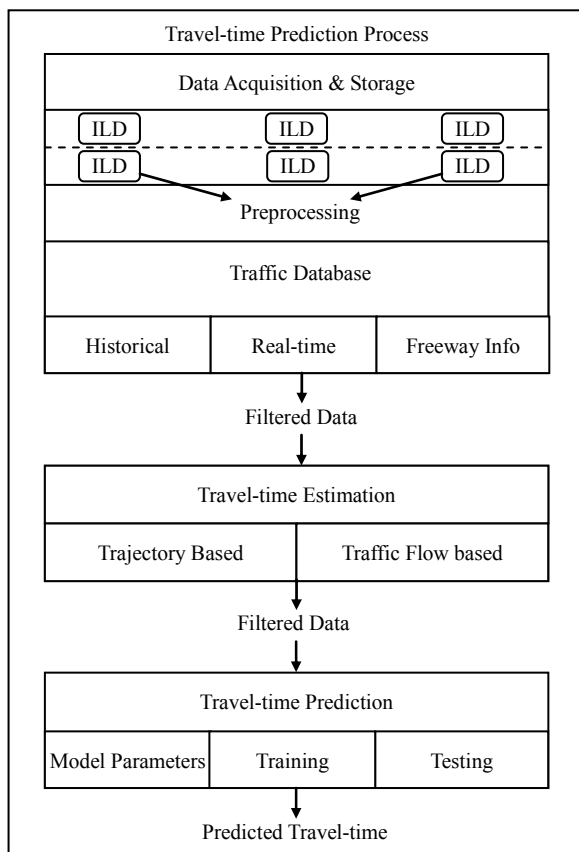


Figure 2. Process diagram of travel-time prediction methods.

in travel-time prediction [7-10]. Most of this work was concentrated on the short-term travel-time prediction, (prediction horizon less than 60 minutes) mainly using the artificial neural network (ANN) technique. On the other hand, machine learning methods, such as support vector regression (SVR), that have shown superior performance when compared with other traditional methods for prediction of non-linear data, have not been applied aggressively in the area of travel-time prediction.

Support vector machines since their inception by Vapnik [11,12] were extensively used in classification and prediction problems. SVM uses a simple geometric interpretation and gives a sparse solution. The solution of SVM is also global and unique as SVM employs the

structural-risk-minimization principle. The support vector regression method [13] approaches the linear regression forecast by addressing it as a convex optimization problem (details in section 4). Its performance in financial time series forecast [14], bioinformatics [15] and various other areas of research also makes it a viable method in intelligent transportation systems (ITS) applications. SVR application as a forecasting tool in ITS was first done by Wu [5], who predicted short-term travel time on the basis of past and current values. Recently, Wang in [16], used wavelet kernel support vector machine for regression to predict traffic flow in ITS applications.

In the recent years many researchers decomposed time series into more informative domains like the wavelets transform [17], S-transform [18] etc., as an input to the SVR that showed more accurate results than the non-decomposed method. This improved performance of SVR along with the ability of SVR to predict non-linear data, formed the motivation of our research to explore the effectiveness of travel-time prediction using wavelet transformed travel-time values as an input to SVR.

The rest of the paper is organized as follows: the problem statement along with some highlights of the past research is given in Section 2. Wavelet theory and Support vector regression are explained in Section 3 and 4, respectively. In Section 5 the proposed model is explained. Then we show the results of our model in Section 6. Finally, the paper is concluded in Section 7, with a brief on the claims made and future research direction.

2. Problem Description

The travel-time prediction problem can be viewed from the perspective of the input data type, prediction methodology and prediction horizon as shown in **Figure 1**. Irrespective of the class of travel-time prediction, the fundamental components of the process are similar as shown in **Figure 2**. Below we explain each component with a review of the main published work done in each area.

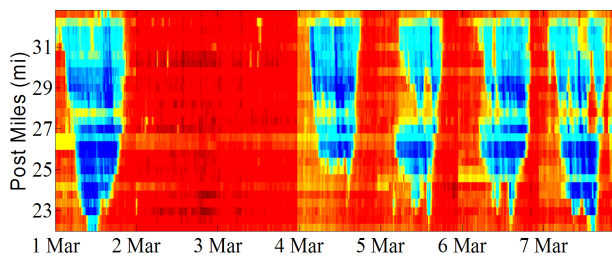
2.1. Data Acquisition and Storage (ILD)

Formulation of an accurate predictive inference relies significantly on the quality of the traffic data. A typical speed plot constructed using a portion of the dataset we used is shown in **Figure 3**. The blue area represents congestion, while the red part shows the free flow speeds.

Inductive Loop Detector (ILD) data based on its abundance and known quality issues has been used as input data in most travel-time prediction papers [6,19-25]. The scalability of the model also biased the choice of the researcher towards choosing ILD as a data source. Other forms of datasets include probe vehicle data, traffic cam-

Table 1. Comparison of related work.

Prior Art Related to Short-Term Travel-Time Prediction			
Prediction Methods	Author/Year of Publication	Length of Roadway	Accuracy/Prediction Horizon
Neural Networks	J.W.C. Van Lint (2004) [3]	5.28 Mi (8.5 Km)	RMSEP: 7.7% MRE: 0.49% SRE 6% Horizon: 15 min
Kaman Filter	Chen and Steven Chien (2001) [4]	8 Mi (12.88 Km)	MARE: 0.0173 - 0.0208 Horizon: 5 min
Support Vector Regression	Wu, Ho and Lee (2004) [5]	28 - 217.5 Mi (45 - 350 Km)	RME:0.96% - 4.42%, RMSE 1.33% - 7.35% Horizon: 3 min
PCA/Nearest Neighbor	Rice and Zwet (2004) [1]	48 Mi (77.25 Km)	RMSE: 2.6 - 11 (Approx) Horizon: 60 min
Regression	Kwon, Coifman and Bickel (2000), [6]	6.2 Mi (10 Km), 20 Mi (32.19 Km)	MAPE: (Tree Method) 6.9% - 28.7%, (Regression) 7.7% - 23.3% Horizon 10 - 60 min

**Figure 3. Speed plot of a portion of the dataset.**

era feeds, and satellite data, data obtained from microwave radar, license plate matching, and automated vehicle tag matching.

Before using ILD data as our data source, certain known issues required attention in context of the site selection and data pre-processing phases. Spacing between consecutive loop detectors directly affects the quality of the data captured. The standard spacing requirement between consecutive loop detectors is not defined in literature. However, [26] concluded that the detector spacing of 1 to 1.5 km is optimum for the use of short-term forecasting of traffic parameters. In [27], it was shown that a detector spacing of 0.33 to 1 mile does not destabilize the travel-time estimation errors, while [28] concluded that a detector spacing of 0.5 miles is sufficient to represent traffic congestion with acceptable accuracy.

After data acquisition preprocessing steps are performed on this data to ensure its validity. ILDs are prone to a number of errors [29]. These data errors are usually detected and removed using imputation methods [29,30]. [29] gave a linear model based on historical data using neighboring detectors to detect faulty values and through linear regression imputed the missing or bad values. The method proposed in [29] was adopted by CALTRANS for data processing of the loop detector data in California roadways.

2.2. Travel-Time Estimation

Like any prediction problem, the ground truth (estimated

travel time) is essential to calculate and evaluate the results (predicted travel time). The travel-time estimation methods are divided into two broad categories: *trajectory-based* and *flow-based*.

2.2.1. Trajectory-Based Methods

The trajectory-based methods convert the time-mean speeds collected from detectors to space-mean speed. Different methods are proposed to calculate link travel-time from this speed. The two common methods are the mid-point method and the average speed method. Both of these methods assume a constant speed between links, which in reality is never the case especially when traffic is in transition from free flow to congestion or vice versa. Hence, the algorithms proposing a constant speed lose their accuracy with the increase in congestion [31]. Van Lint and Van der Zijpp proposed an alternate approach, the “Piecewise Linear Speed” method [32], which solved the function of the travel-time based on the time mean speed using an ordinary differential equation to calculate the trajectory of the vehicle in the section based on space mean speed.

2.2.2. Flow-Based Methods

An alternate way of estimating travel-time is through flow-based models which focus on capturing the dynamics of traffic using traffic-flow theory concepts, and through traffic data simulation, draw the travel-time of the segment. Accurate flow information is also required for a precise estimation; however, in most cases it is difficult to collect data from all on-ramps and off-ramps using the existing infrastructure, which becomes a bottleneck for flow-based estimation methods. These models are, however, more popular in research involving traffic flow simulation.

2.3. Travel-Time Prediction

The travel-time prediction approach is mainly classified w.r.t. the prediction horizon, modeling approach and type of input data as shown in **Figure 1**. Further classification

is also possible w.r.t. the road type (freeways, arterials); but, since the scope of this proposal is confined to freeways; we would not discuss the arterial travel-time prediction problem.

The historical data of traffic parameters can represent a traffic profile, which could be implemented to predict future values, in similar traffic conditions. This approach demands *offline* processing. The data is classified into different subtypes based on their characteristics. In [33] the data was sub-classified into the “type of day”, for prediction of travel-time. This forecast method does not take into account the dynamics of traffic for travel-time prediction, which makes this method less robust for short-term prediction. Consequently, it produces low accuracy results, when the current traffic is not representative of its historical profile. Historical predictor is normally used for long-term prediction.

A hybrid approach of combining historical data with current data was used in [34] where real-time data was captured directly from the road side terminals, and using it with aggregated historical data showed improved results. [1] used principal component analysis and windowed nearest neighbor, while combining historical and instantaneous data.

Traffic data shares similarities when compared with historical data of the same day and time as the current data. Regression methods with coefficients varying with the time of the day were used by [1], [35] and [36] to predict travel-time. [6] also used linear regression with step wise variable selection method. Regression models involve the examination of historical data, thereby, extracting parameters, which represent traffic characteristics, and projecting them into the future to predict travel-time. Autoregressive integrated moving average (ARIMA) was introduced by [37] and [38] as an alternate to model the stochastic nature of traffic. [39] used autoregression model to predict travel time. Non-linear time series with multifractal analysis was implemented in [40] and [41] for travel time prediction.

Kalman and Extended Kalman Filters used in [2,42] provide good performance in predicting travel-time for one time-step ahead horizon, which is normally not more than 5 minutes, as the state model needs real observations to calculate each error term.

Artificial neural networks (ANN) were extensively used for marking non-linear boundaries. To address the problem of a time series forecast, a subtype of ANN called the recurrent neural network (RNN) was considered suitable [19,24,43]. RNN has an internal state, which keeps track of the temporal behavior between classes. Different architectures of the Multilayer perceptron have been used to predict travel-time with an improved accuracy [7,8,10,19,20,23,24,43-45]. The support vector regression method was also investigated in [5,46].

On the other hand, traffic flow models work on the concept of correlating the theory of fluid dynamics with vehicular flow. From the perspective of traffic flow models, travel-time prediction is more of a boundary condition prediction problem, because the flow model is designed offline, and it would predict the time based on the values of demand and supply at on-ramps and off-ramps respectively. The model is run using a simulation scheme, which is based on the assumptions of the car-following, gap acceptance, and risk avoidance parameters. The simulation model predicts the aggregated parameters of simulated vehicles to display the predicted travel-time [47,48]. This makes traffic flow models very complex and requires a high degree of expertise and long man-hours for design and maintenance.

Traffic flow models give us a better understanding of the traffic flow dynamics, but as far as their accuracy for travel time prediction is concerned, they demand a precise infrastructure of input detectors, whose location would be defined by the flow model. To manage the supply and demand parameters, the flow models require additional detectors on each off and on-ramp. Traffic flow based models are a good method to evaluate the cause and effect of traffic phenomenon, but applying them for travel-time prediction would entail a huge design and maintenance cost for every freeway section. Due to their modular design, precision of traffic flow models, for travel-time prediction, would be as accurate, as the precision of the predicted inputs and boundary conditions.

3. An Overview of Wavelets

Wavelets are functions, which present a multi-resolution decomposition of a signal x using a mother function ψ and a linear combination of its dilated and/or shifted versions (1).

$$\psi_{u,s}(x) = \frac{1}{\sqrt{s}} \psi\left(\frac{x-u}{s}\right), \quad (1)$$

where s defines the dilation and u defines the shift. To ensure orthonormality of basis functions [49] the time-scale parameters are sampled on a dyadic grid on the time-scale plane. Thus Equation (1) becomes

$$\psi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t-n}{2^j}\right).$$

The orthonormal wavelet transform is then given by

$$\langle x(t), \psi_{j,n}(x) \rangle = \frac{1}{\sqrt{2^j}} \int_{-\alpha}^{\alpha} x(t) \psi_{j,n}(2^{-j}t - n) dt$$

To make the transform computationally effective the concept of sub-band coding [50] was used to filter the signal with a series of high pass and low pass filters to

analyze its high frequency and low frequency components respectively. The input signal $x(t)$ can now be represented in discrete domain as

$$x(t) = \sum_{n \in \mathbb{Z}} c_{j,n} \phi_{j,n}(t) + \sum_{j=J}^{\alpha} \sum_{n \in \mathbb{Z}} d_{j,n} \psi_{j,n}(t).$$

The sampled scaling $c_{j,n}$ and wavelet coefficients $d_{j,n}$ can now be defined using high pass h_l and low pass filter g_l .

$$c_{j,n} = \sum_{l \in \mathbb{Z}} g_l c_{j-1, 2n-1}.$$

$$d_{j,n} = \sum_{l \in \mathbb{Z}} h_l c_{j-1, 2n-1}.$$

To add translation-invariance in discrete wavelet transform (DWT), maximum overlap discrete wavelet transform (MODWT) was introduced, which instead of down sampling and up sampling the signal introduces high and low pass filters up sampled by a factor of 2^{j-1} . The up sampling filters also introduce redundancy in the output, since the number of samples at output in every level is equal to the number of samples in the input signal. This makes multi-resolution analysis much more effective especially from the perspective of using this transform as an input to another system.

$$d_{j,n}^{(M)} = \sum_{l=0}^{L_j-1} \tilde{h}_l c_{j-1, n-2^{j-1}l \bmod N}^{(M)}.$$

$$c_{j,n}^{(M)} = \sum_{l=0}^{L_j-1} \tilde{g}_l c_{j-1, n-2^{j-1}l \bmod N}^{(M)}.$$

The filters can now be represented as a circular filter of the original time series.

$$d_{j,n}^{(M)} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} x_{n-l \bmod N}^{(M)}.$$

$$c_{j,n}^{(M)} = \sum_{l=0}^{L_j-1} \tilde{g}_{j,l} x_{n-l \bmod N}^{(M)}.$$

To generate the wavelet packet tree, both the approximation and detail coefficients are decomposed instead of just the approximation coefficients as in the case of the DWT. Hence the wavelet packet distributes the frequency of the original signal evenly between all coefficients as opposed to the wavelet transform where 50% of the signal frequency is in the first detail as shown in **Figure 4**. In the WDSVR model, we chose the wavelet packet transform to evenly distribute the signal frequency in each support vector module.

4. Support Vector Regression

Support vector machines (SVM) work on the concept of Structural Risk Minimization [12] by transforming a low dimensional input x into a high dimensional feature space

through a mapping function ϕ and then approximating the function $f(x)$ using linear regression

$$f(x) = \sum_{i=1}^D w_i \phi_i(x) + b,$$

where b is the threshold. w is the normal vector to the hyperplane. The coefficients can be determined from the data by minimizing the regression risk function.

$$\text{Reg}[w] = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N |y - f(x)|_{\epsilon}, \tag{2}$$

where C is the cost function, which defines the tradeoff between training error and model complexity. The ϵ -SVR algorithm discards the training points that lie beyond the threshold ϵ defined by the user. Mathematically

$$|f(x) - y|_{\epsilon} = \begin{cases} |y - f(x)| & \text{for } |y - f(x)| \geq \epsilon \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Equation (3) is also known as the Vapnik's ϵ -insensitive loss function. Both Equation (3) and the regression risk function Equation (2) can be minimized by introducing Langrangian multipliers α and α_i^* to this quadratic problem, yielding the solution

$$f(x, \alpha, \alpha^*) = \sum_{i=1}^N (\alpha - \alpha_i^*) k(x_i, x) + b$$

with $\alpha, \alpha_i^* = 0, \alpha_i, \alpha_i^* \geq 0$ for $i = 1, \dots, N$. $k(x_i, x)$ is the kernel function $k(x_i, x)$, which is computed by calculating the dot product of some feature space.

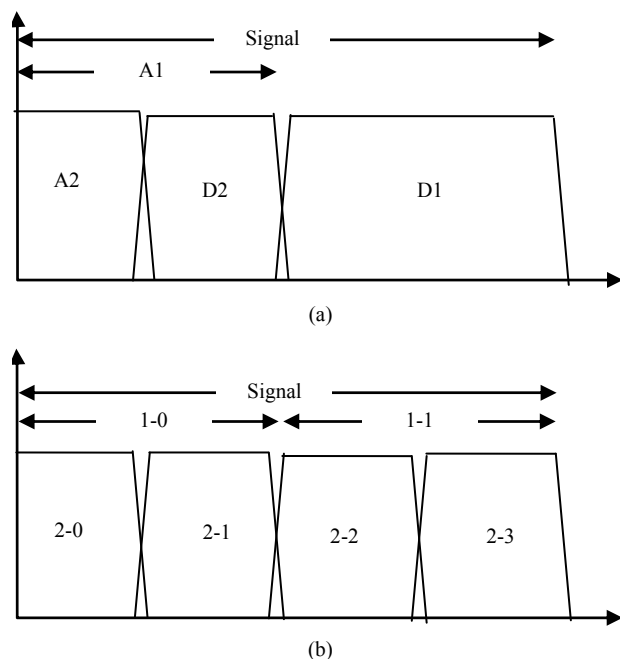


Figure 4. Frequency allocation of 2 level DWT. Frequency allocation of 2 level wavelet packet transform.

$$k(x, y) = \sum_{j=1}^D \phi_j(x) \phi_j(y).$$

It is important to note that the kernel $k(x, y)$ has a known analytical form and must obey the Mercer's condition.

5. Wavelet Packet Support Vector Regression

The structure of wavelet packet support vector regression is schematically outlined in **Figure 5**. The model works by evenly distributing the original signal's frequency using the wavelet packet transform into the SVR modules. The time series signal, which represented the travel-time of the freeway was sampled from the database, based on the prediction horizon selected. The time signal was then transformed using the wavelet packet decomposed signals, such as $\sum_{n=0}^{2^j-1} W_{j,n}$, where j is the level of the decomposition. The wavelet decomposition was done using a sliding window as shown in **Figure 6**. The window size determines the number of input features given to the support vector machine. In our case the window size of 8 was selected and the decomposition was done at level 2. These wavelet coefficients were stored for the support vector regression module. The four frequency components were processed through their respective support vector machines leading to compute one time-step ahead output, where the step was equal to the time interval between the consecutive input values. The support vector regression output was finally aggregated to calculate the travel-time forecast. **Table 2** gives the step by step implementation of the wavelet packet support vector regression algorithm.

6. Experiments and Results

6.1. Selection of Mother Wavelet

The major computational load of the proposed travel-time prediction model was divided into two parts: computation of the wavelet packet reconstructed time-series data, and training of the support vector regression machines using the optimal cost and epsilon values.

The grid search method was used for searching for epsilon and cost values.

A definite procedure for selection of mother wavelets is yet to be established for wavelet decomposed support vector regression models. However, analyzing the wavelet reconstructed signal in context of the characteristics of the support vector machines helped us in filtering the relevant wavelets basis.

The accuracy of the proposed model is superior to the classical SVR model, if the condition in Equation (4) is met.

$$\varepsilon_{SVR} > \varepsilon_{SVR(2,0)} + \varepsilon_{SVR(2,1)} + \varepsilon_{SVR(2,2)} + \varepsilon_{SVR(2,3)}, \quad (4)$$

where, ε_{SVR} is the error of the classical support vector method. It is clear from Equation (4) that WDSVR would not produce more accurate results than SVR for shorter time horizons, knowing that prediction error is proportional to the prediction horizon. In our datasets, the WDSVR gave more accurate results than the SVR method for prediction horizons of 45 minutes or more.

We conducted two basic tests for the admissibility of all wavelets for the support vector machine module.

6.1.1. Cross-Correlation of Wavelet Decomposed Data

For accurate predictions of a non-linear and non-stationary dataset the reconstructed wavelet coefficients of successive windows should not be correlated with one another. A positive linear correlation of +1.0 would indicate a similar pattern to the SVR module for every input and would adversely affect its prediction accuracy. To test our hypothesis we computed the cross-correlation of each window with the other.

6.1.2. Recurrence Relationship

The second test was to detect if the reconstructed wavelet coefficients windows were following a certain pattern at a particular location. We know that the input data of the successive windows is non-linear. The existence of a unique pattern at a similar location in the input signal would indicate a similar pattern to the support vector machine in every iteration, which in reality is not the case. Consequently, it would adversely affect the performance of the SVR module. To detect such events we calculated the first difference of each successive window.

Table 2. Algorithm for wavelet decomposed support vector regression.

1) Sample travel-time array into subsets for their respective prediction horizons using

$$y[t\Delta] = \sum_{k=0}^N x \left[\frac{hk}{5} + 1 \right],$$

where h is the prediction horizon in minutes.

2) Initialize $p = 0$ and decompose the sampled signal using wavelet packet decomposition at level $j = 2$

$$W_{j,n} \left[\sum_{k=p}^{p+7} y[t\Delta - k] \right].$$

3) Store $W_{j,n}$ computed in step 2 for the SVR module and increment $p = p + 1$.

4) Repeat steps 2 and 3 until the end of the input array $y[t\Delta]$.

5) Increment $n = n + 1$ and repeat steps 2 - 4 until $n = 2^j$.

6) Divide $W_{j,n}$ into training and testing sets and compute one step ahead prediction value using their respective SVR modules.

7) Aggregate the predictions of all 4 SVR modules to calculate the predicted travel time.

To identify the above characteristics in the wavelet signal we used a subset of the data chosen at random ranging four days. In **Figure 7(a)** the wavelet reconstructed difference signal converged to zero at a similar point in every iteration. **Figure 7(b)** is the first difference of the of the Biorthogonal 1.1 filter output at level 2,3, which indicates a linear correlation among the successive windows. On the other hand, the best performing wavelet at one hour prediction horizon, the Reverse Biorthogonal 6.8 wavelet, showed no cross-correlation or recurrence relationship as shown in **Figures 7(b)** and **(d)**. Based on our admissibility tests, 9 wavelets were filtered out of a total of 42, hence reducing the computational load of our project by 21.43%. While a detailed study on wavelet selection for WDSVR is needed,our results on the selection of wavelets for the support vector machines have shown encouraging results to motivate further work in this area.

6.2. An Alternate Configuration for Interchangeable

The WDSVR and SVR have both proven suitable for travel-time prediction depending on the selected forecast

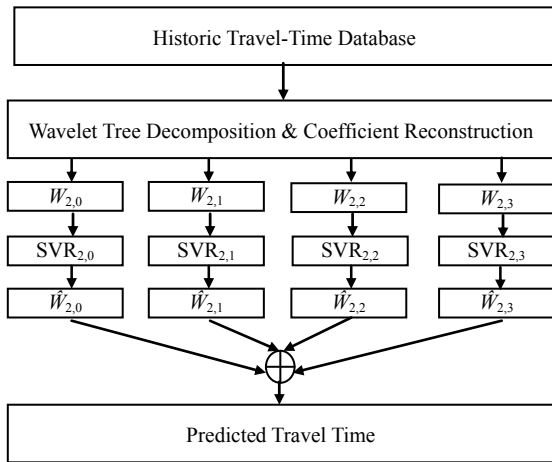


Figure 5. Schematic diagram of the wavelet decomposed support vector regression model.

horizon. In our dataset, we observed that SVR is more accurate for prediction horizons of less than 45 minutes. From 45 minutes onwards, WDSVR gives more accurate results. Considering the effectiveness of both models in different horizons, we have proposed an interchangeable configuration in **Figure 8**, where travel-times using both models were computed in parallel and then *switch* to the configuration for active use depending on the selected prediction horizon. The cloud component, which houses both the prediction models is flexible and can be either scaled horizontally or vertically to accommodate for the computation overhead.

6.3. Experimental Setup

The data for our model validation and testing was collected from the Caltrans Performance Measurement System (PeMS) website [2].

The route of 9.13 miles on I-5N was selected with a detector density of 2.73. The data was observed for 214 consecutive days commencing from March 01, 2011 to September 30, 2011 from 1 pm to 8 pm. The time slot was selected after observing the daily pattern of congestion during this period. The data revealed daily congestion in the evening hours except holidays and most weekends. This loop detector data was collected over a 5 minutes interval. The speed data was converted to travel-time series using the PLSB travel-time estimation method [32]. We decomposed the time series using the wavelet packet decomposition at level 2. The data was then reshaped into a $u \times v$ matrix with $u = N - 7$ and $v = 8$. The decomposed and reshaped wavelet transform of travel-time matrix gave us 2^j matrices at level j represented as

$$W_{j,n} = \begin{bmatrix} W_{j,n,t-1} & \dots & W_{j,n,t-8} \\ \vdots & \ddots & \vdots \\ W_{j,n,N-7} & \dots & W_{j,n,N} \end{bmatrix}$$

The four matrices were given as input to their respective support vector machines with $(N - 7) \times 0.7$ rows for training while the remaining 30% for evaluation. The

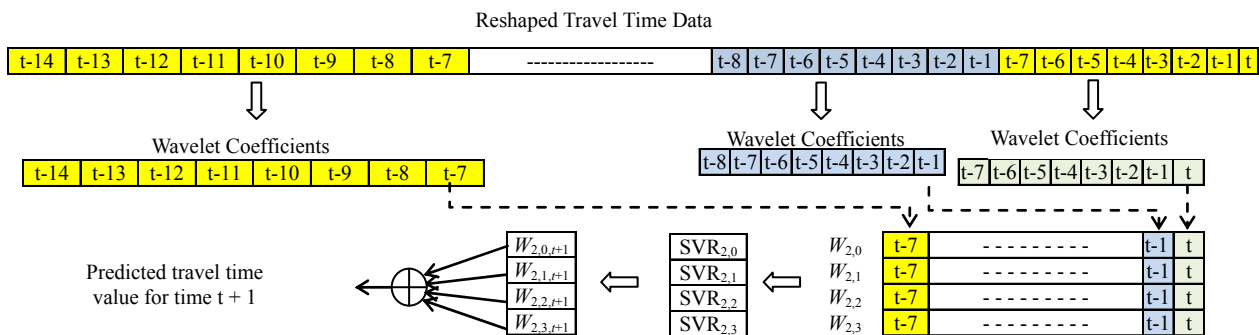


Figure 6. Flow diagram of the algorithm for wavelet decomposed support vector regression.

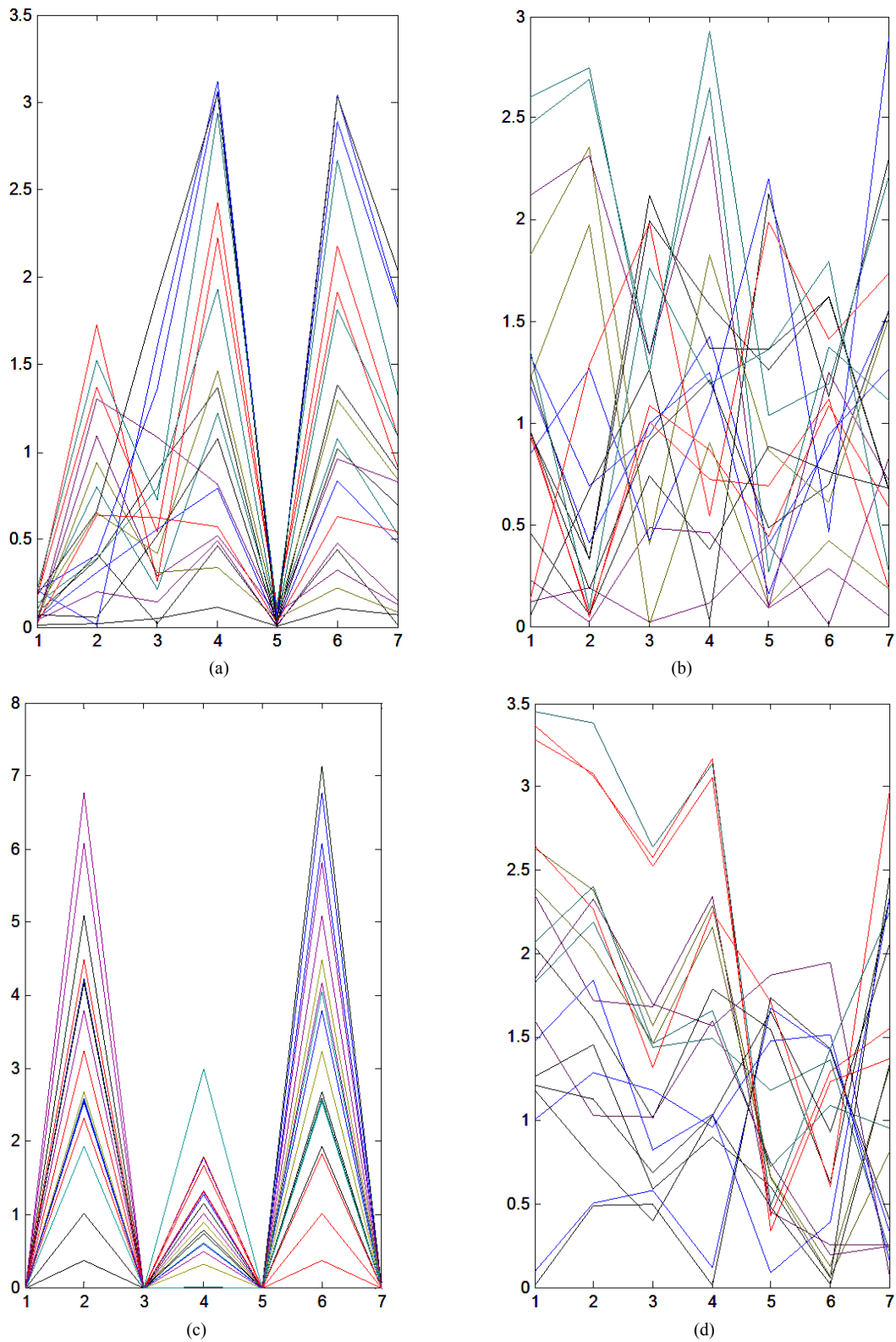


Figure 7. A comparison of wavelet recurrence relationship and cross correlation of better and worse performing wavelets: (a) First difference signal of wavelet Packet Reconstructed time series at level 2,3 using Biorthogonal 3.3; (b) First difference signal of wavelet Packet Reconstructed time series at level 2,3 using Reverse Biorthogonal 6.8; (c) First difference signal of wavelet Packet Reconstructed time series at level 2,3 using Biorthogonal 1.1; (d) First difference signal of wavelet Packet Reconstructed time series at level 2,3 using Reverse Biorthogonal 6.8.

evaluation matrix for each $W_{j,n}$ above was represented as

$$\text{label}_{j,n} = \begin{bmatrix} W_{j,n,t} \\ W_{j,n,t-1} \\ \vdots \\ W_{j,n,N+1} \end{bmatrix}$$

The predicted labels of each support vector machine were aggregated to compute the forecast time value. Finally the values generated by SVR were evaluated for errors.

We tested our model using Debauchies, Coiflets, Symlets, Reverse Biorthogonal and Biorthogonal wavelets in 42 different configurations, with different values of cost and epsilon. It was observed that not all wavelets gave better results than the benchmark SVR predicted values. However, some of the worse performing wavelets were filtered out using our wavelet selection process to save computational cost. The best outputs in each time horizon sub-category were shown in **Tables 1-3**.

Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE) and Pearson Product-Moment

Correlation were the three indicators chosen for evaluation of our model and for comparison with the classical Support Vector Regression model. **Table 4** shows the comparison of MAPE between SVR and SVR with wavelet decomposed inputs. **Table 5** shows comparison of Pearson product-moment correlation between SVR and SVR with wavelet decomposed inputs.

Our results indicated that the wavelet decomposed support vector regression model consistently showed better performance for prediction horizon of 45 minutes and above but below 45 minutes the classical SVR method was more accurate. **Figure 9** showed the better tracking ability of the proposed model in comparison with the SVR model.

7. Summary of Results

The proposed wavelet packet decomposed SVR method showed improved results for travel-time data prediction over the conventional SVR method for prediction horizons of 45 minutes and above. For accurate state estimation through machine learning methods large datasets are

Table 3. Comparison of RMSE between SVR and SVR with wavelet decomposed inputs (our approach).

Prediction Methods	Prediction Horizon							
	45-min		50-min		55-min		60-min	
Wavelet Packet SVR	bior2.6	$\epsilon = 0.1, C = 100$	bior6.8	$\epsilon = 0.01, C = 100$	coif5	$\epsilon = 0.1, C = 100$	db6	$\epsilon = 0.001, C = 100$
		2.2		2.31		2.41		2.46
SVR Predictor		$\epsilon = 0.01, C = 100$		$\epsilon = 0.1, C = 1$		$\epsilon = 0.001, C = 100$		$\epsilon = 0.1, C = 10$
		2.26		2.4		2.48		2.88

Table 4. Comparison of MAPE (%) between SVR and SVR with wavelet decomposed inputs.

Prediction Methods	Prediction Horizon							
	45-min		50-min		55-min		60-min	
Wavelet Packet SVR	bior2.6	$\epsilon = 0.1, C = 1$	rbio2.8	$\epsilon = 0.1, C = 100$	rbio2.8	$\epsilon = 0.001, C = 100$	rbio6.8	$\epsilon = 0.01, C = 100$
		12.35		13.1		13.66		14.74
SVR Predictor		$\epsilon = 0.01, C = 10$		$\epsilon = 0.01, C = 100$		$\epsilon = 0.1, C = 1$		$\epsilon = 0.1, C = 100$
		12.57		13.5		13.96		15.06

Table 5. Comparison of Pearson product-moment correlation between SVR and SVR with wavelet decomposed inputs.

Prediction Methods	Prediction Horizon							
	45-min		50-min		55-min		60-min	
Wavelet Packet SVR	bior2.6	$\epsilon = 0.1, C = 1$	bior6.8	$\epsilon = 0.01, C = 100$	coif5	$\epsilon = 0.1, C = 100$	db6	$\epsilon = 0.001, C = 100$
		0.8767		0.8623		0.8486		0.8441
SVR Predictor		$\epsilon = 0.01, C = 100$		$\epsilon = 0.1, C = 100$		$\epsilon = 0.1, C = 10$		$\epsilon = 0.1, C = 10$
		0.8702		0.8498		0.8381		0.8406

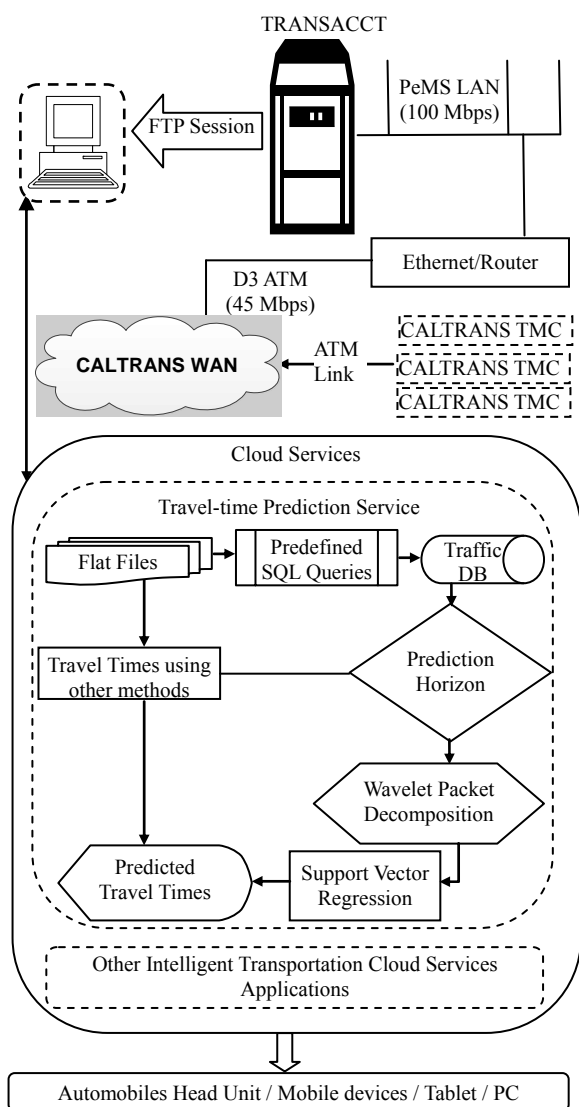


Figure 8. Proposed configuration for travel-time prediction for ATIS.

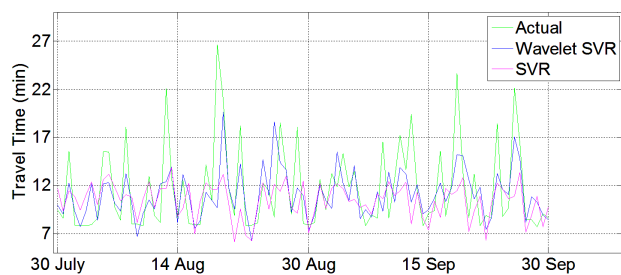


Figure 9. Comparison of actual travel time, and predicted travel-time by Support Vector Regression and Wavelet decomposed Support Vector Regression methods.

needed, many of which, are now available online. Their training with multiple methods would require significant computation cost and storage, for which, we have proposed an alternate framework with a cloud component,

which could be scaled horizontally or vertically to cater for the memory and computation requirements. We proposed a modular prediction method, where multiple prediction algorithms are stored in the cloud and the best performing algorithm is selected based on the prediction horizon. We also investigated wavelet properties in conjunction with their effectiveness for support vector machines. We observed that wavelet basis, whose cross-correlation between the wavelet reconstructed coefficients of successive windows resulted in a linear correlation of +1.0 or the ones with recurrent relationships are not useful for WDSVR model and should be discarded to reduce the computation cost. In our dataset it reduced computational cost by 21.43%. Further improvements to our model might be made possible by subdividing the dataset based on its patterns, some examples are by congested and free flow parts or by day of the week or both. The scalability of the model also makes it a viable option for its application to calculate arterial travel times.

REFERENCES

- [1] J. Rice and E. Van Zwet, "A Simple and Effective Method for Predicting Travel Times on Freeways," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 5, No. 3, 2004, pp. 200-207. [doi:10.1109/TITS.2004.833765](https://doi.org/10.1109/TITS.2004.833765)
- [2] C. M. Kuchipudi and S. I. J. Chien, "Development of a Hybrid Model for Dynamic Travel-Time Prediction," *Transportation Data Research: Planning and Administration*, Washington DC, 2003, pp. 22-31.
- [3] H. van Lint, "Reliable Travel Time Prediction for Freeways," Ph.D. Thesis, TU Delft, Delft, 2004.
- [4] M. Chen and S. I. Chien, "Dynamic Freeway Travel-Time Prediction with Probe Vehicle Data: Link Based versus Path Based," *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1768, 2001, pp. 157-161. [doi:10.3141/1768-19](https://doi.org/10.3141/1768-19)
- [5] C. H. Wu, J. M. Ho and D. Lee, "Travel-Time Prediction with Support Vector Regression," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 5, No. 4, 2004, pp. 276-281. [doi:10.1109/TITS.2004.837813](https://doi.org/10.1109/TITS.2004.837813)
- [6] J. Kwon, B. Coifman and P. Bickel, "Day-to-Day Travel-Time Trends and Travel-Time Prediction from Loop-Detector Data," *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1717, 2000, pp. 120-129. [doi:10.3141/1717-15](https://doi.org/10.3141/1717-15)
- [7] A. Dharia and H. Adeli, "Neural Network Model for Rapid Forecasting of Freeway Link Travel Time," *Engineering Applications of Artificial Intelligence*, Vol. 16, No. 7-8, 2003, pp. 607-613. [doi:10.1016/j.engappai.2003.09.011](https://doi.org/10.1016/j.engappai.2003.09.011)
- [8] D. Park, L. R. Rilett and G. Han, "Spectral Basis Neural Networks for Real-Time Travel Time Forecasting," *Journal of Transportation Engineering*, Vol. 125, No. 6, 1999, pp. 515-523. [doi:10.1061/\(ASCE\)0733-947X\(1999\)125:6\(515\)](https://doi.org/10.1061/(ASCE)0733-947X(1999)125:6(515))

- [9] D. J. Park and L. R. Rilett, "Forecasting Multiple-Period Freeway Link Travel Times Using Modular Neural Networks," In: *Land Use and Transportation Planning and Programming Applications*, 1998, pp. 163-170.
- [10] L. R. Rilett and D. Park, "Direct Forecasting of Freeway Corridor Travel Times Using Spectral Basis Neural Networks," *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1752, 2001, pp. 140-147.
- [11] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, Vol. 20, No. 3, 1995, pp. 273-297. [doi:10.1007/BF00994018](https://doi.org/10.1007/BF00994018)
- [12] V. N. Vapnik, "The Nature of Statistical Learning Theory," Springer Verlag, Berlin, 2000. [doi:10.1007/978-1-4757-3264-1](https://doi.org/10.1007/978-1-4757-3264-1)
- [13] V. Vapnik, S. E. Golowich and A. Smola, "Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing," *Advances in Neural Information Processing Systems*, Vol. 9, 1997, pp. 281-287.
- [14] T. B. Trafalis and H. Ince, "Support Vector Machine for Regression and Applications to Financial Forecasting," *IJCNN 2000, Proceedings of the IEEE International Joint Conference on Neural Networks*, Como, 27 July 2000, pp. 348-353.
- [15] M. Song, C. M. Breneman, J. Bi, N. Sukumar, K. P. Bennett, S. Cramer and N. Tugcu, "Prediction of Protein Retention Times in Anion-Exchange Chromatography Systems Using Support Vector Regression," *Journal of Chemical Information and Computer Sciences*, Vol. 42, No. 6, 2002, pp. 1347-1357. [doi:10.1021/ci025580t](https://doi.org/10.1021/ci025580t)
- [16] F. Wang, G. Tan and Y. Fang, "Multiscale Wavelet Support Vector Regression for Traffic Flow Prediction," *3rd International Symposium on Intelligent Information Technology Application (IITA 2009)*, Nanchang, 21-22 November 2009, pp. 319-322.
- [17] S. Yao, C. Hu and W. Peng, "Server Load Prediction Based on Wavelet Packet and Support Vector Regression," *2006 International Conference on Computational Intelligence and Security*, Guangzhou, 3-6 November 2006, pp. 1016-1019.
- [18] M. Faisal and A. Mohamed, "A New Technique to Predict the Sources of Voltage Sags using Support Vector Regression based S-Transform," *IASTED, International Conference on Modelling, Simulation, and Identification/658: Power and Energy Systems/660, 661, 662*, Beijing, 2009.
- [19] H. van Lint, S. P. Hoogendoorn and H. J. van Zuylen, "State Space Neural Networks for Freeway Travel Time Prediction," In: J. R. Dorronsoro, Ed., *Artificial Neural Networks—ICANN 2002*, Madrid, 28-30 August 2002, pp. 1043-1048.
- [20] S. Innamaa, "Short-Term Prediction of Travel Time Using Neural Networks on an Interurban Highway," *Transportation*, Vol. 32, No. 6, 2005, pp. 649-669. [doi:10.1007/s11116-005-0219-y](https://doi.org/10.1007/s11116-005-0219-y)
- [21] J. W. C. van Lint, "Reliable Real-Time Framework for Short-Term Freeway Travel Time Prediction," *Journal of Transportation Engineering*, Vol. 132, No. 12, 2006, pp. 921-932. [doi:10.1061/\(ASCE\)0733-947X\(2006\)132:12\(921\)](https://doi.org/10.1061/(ASCE)0733-947X(2006)132:12(921))
- [22] N. Zou, J. W. Wang and G. L. Chang, "A Reliable Hybrid Prediction Model for Real-Time Travel Time Prediction with Widely Spaced Detectors," *11th International IEEE Conference on Intelligent Transportation Systems*, Beijing, 12-15 October 2008, pp. 91-96.
- [23] N. Zou, J. W. Wang, G. L. Chang and J. Paracha, "Application of Advanced Traffic Information Systems Field Test of a Travel-Time Prediction System with Widely Spaced Detectors," *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2129, 2009, pp. 62-72. [doi:10.3141/2129-08](https://doi.org/10.3141/2129-08)
- [24] J. W. C. van Lint, S. P. Hoogendoorn and H. J. van Zuylen, "Accurate Freeway Travel Time Prediction with State-Space Neural Networks under Missing Data," *Transportation Research Part C: Emerging Technologies*, Vol. 13, No. 5-6, 2005, pp. 347-369. [doi:10.1016/j.trc.2005.03.001](https://doi.org/10.1016/j.trc.2005.03.001)
- [25] H. J. M. Van Grol, M. Danech-Pajouh, S. Manfredi and J. Whittaker, "DACCORD: On-Line Travel Time Prediction," *World Transport Research: Selected Proceedings of the 8th World Conference on Transport Research*, Pergamon, Oxford, 1999.
- [26] H. Chen, M. S. Dougherty and H. R. Kirby, "The Effects of Detector Spacing on Traffic Forecasting Performance Using Neural Networks," *Computer-Aided Civil and Infrastructure Engineering*, Vol. 16, No. 6, 2001, pp. 422-430. [doi:10.1111/0885-9507.00244](https://doi.org/10.1111/0885-9507.00244)
- [27] P. Yi, S. Ding, H. Wei and G. W. Saylor, "Investigating the Effect of Detector Spacing on Midpoint-Based Travel Time Estimation," *Journal of Intelligent Transportation Systems*, Vol. 13, No. 3, 2009, pp. 149-159.
- [28] J. Kwon, K. Petty and P. Varaiya, "Probe Vehicle Runs or Loop Detectors?: Effect of Detector Spacing and Sample Size on Accuracy of Freeway Congestion Monitoring," *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2012, 2007, pp. 57-63. [doi:10.3141/2012-07](https://doi.org/10.3141/2012-07)
- [29] C. Chen, J. Kwon, J. Rice, A. Skabardonis and P. Varaiya, "Detecting Errors and Imputing Missing Data for Single-Loop Surveillance Systems," *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1855, 2003, pp. 160-167. [doi:10.3141/1855-20](https://doi.org/10.3141/1855-20)
- [30] L. N. Jacobson, N. L. Nihan and J. D. Bender, "Detecting Erroneous Loop Detector Data in a Freeway Traffic Management System," *Transportation Research Record*, Washington DC, 1990.
- [31] C. D. R. Lindveld, R. Thijs, P. H. L. Bovy and N. J. Van der Zijpp, "Evaluation of Online Travel Time Estimators and Predictors," *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1719, 2000, pp. 45-53. [doi:10.3141/1719-06](https://doi.org/10.3141/1719-06)
- [32] J. W. C. van Lint and N. Van der Zijpp, "Improving a Travel-Time Estimation Algorithm by Using Dual Loop Detectors," *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1855, 2003, pp. 41-48. [doi:10.3141/1855-05](https://doi.org/10.3141/1855-05)
- [33] M. Saito and T. Watanabe, "Prediction and Dissemination

- System for Travel Time Utilizing Vehicle Detectors,” *Steps Forward: Intelligent Transport Systems World Congress*, Yokohama, 9-11 November 1995, p. 106.
- [34] I. Steven, J. Chien and C. M. Kuchipudi, “Dynamic Travel Time Prediction with Real-Time and Historic Data,” *Journal of Transportation Engineering*, Vol. 129, No. 6, 2003, p. 608.
[doi:10.1061/\(ASCE\)0733-947X\(2003\)129:6\(608\)](https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(608))
- [35] X. Zhang and J. A. Rice, “Short-Term Travel Time Prediction,” *Transportation Research Part C: Emerging Technologies*, Vol. 11, No. 3-4, 2003, pp. 187-210.
[doi:10.1016/S0968-090X\(03\)00026-3](https://doi.org/10.1016/S0968-090X(03)00026-3)
- [36] H. Sun, H. X. Liu, H. Xiao, R. R. He and B. Ran, “Use of Local Linear Regression Model for Short-Term Traffic Forecasting,” *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1836, 2003, pp. 143-150. [doi:10.3141/1836-18](https://doi.org/10.3141/1836-18)
- [37] M. S. Ahmed and A. R. Cook, “Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques,” *Transportation Research Record*, No. 722, 1979, pp. 1-9.
- [38] M. Levin and Y. D. Tsao, “On Forecasting Freeway Occupancies and Volumes (Abridgment),” *Transportation Research Record*, No. 722, 1980, pp. 47-49.
- [39] T. Oda, “An Algorithm for Prediction of Travel Time Using Vehicle Sensor Data,” *Third International Conference on Road Traffic Control*, London, 1-3 May 1990, pp. 40-44.
- [40] M. P. D’Angelo, H. M. Al-Deek and M. C. Wang, “Travel-Time Prediction for Freeway Corridors,” *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1676, 1999, pp. 184-191.
[doi:10.3141/1676-23](https://doi.org/10.3141/1676-23)
- [41] S. Ishak and H. Al-Deek, “Performance Evaluation of Short-Term Time-Series Traffic Prediction Model,” *Journal of Transportation Engineering*, Vol. 128, No. 6, 2002, pp. 490-498.
[doi:10.1061/\(ASCE\)0733-947X\(2002\)128:6\(490\)](https://doi.org/10.1061/(ASCE)0733-947X(2002)128:6(490))
- [42] H. F. Ji, A. G. Xu, X. Sui and L. Y. Li, “The Applied Research of Kalman in the Dynamic Travel Time Prediction,” *Geoinformatics*, 2010 18th International Conference on, Beijing, 18-20 June 2010, pp. 1-5.
- [43] J. W. C. van Lint, S. P. Hoogendoorn and H. J. van Zuylen, “Freeway Travel Time Prediction with State-Space Neural Networks-Modeling State-Space Dynamics with Recurrent Neural Networks,” *Advanced Traffic Management Systems for Freeways and Traffic Signal Systems 2002: Highway Operations, Capacity, and Traffic Control*, No. 1811, 2002, pp. 30-39.
- [44] C. P. I. van Hinstiergen, J. W. C. van Lint and H. J. van Zuylen, “Bayesian Training and Committees of State-Space Neural Networks for Online Travel Time Prediction,” *Transportation Research Record*, Vol. 2105, 2009, pp. 118-126. [doi:10.3141/2105-15](https://doi.org/10.3141/2105-15)
- [45] D. Park, L. R. Rilett and G. H. Han, “Forecasting Multiple-Period Freeway Link Travel Times Using Neural Networks with Expanded Input Nodes,” *Applications of Advanced Technologies in Transportation*, No. 1617, 1998, pp. 325-332.
- [46] L. Vanajakshi and L. R. Rilett, “Support Vector Machine Technique for the Short Term Prediction of Travel Time,” *2007 IEEE Intelligent Vehicles Symposium*, Istanbul, 13-15 June 2007, pp. 600-605.
- [47] M. Ben-Akiva, M. Bierlaire, H. Koutsopoulos and R. Mishalani, “DynaMIT: A Simulation-Based System for Traffic Prediction,” *DACCORD Short Term Forecasting Workshop*, Delft, 1998.
- [48] M. Ben-Akiva, M. Bierlaire, D. Burton, H. N. Koutsopoulos and R. Mishalani, “Network State Estimation and Prediction for Real-Time Traffic Management,” *Networks and Spatial Economics*, Vol. 1, No. 3-4, 2001, pp. 293-318. [doi:10.1023/A:1012883811652](https://doi.org/10.1023/A:1012883811652)
- [49] S. G. Mallat, “A Wavelet Tour of Signal Processing,” Academic Press, Edinburgh, 1999.
- [50] M. Vetterli and J. Kovačević, “Wavelets and Subband Coding,” Prentice Hall-PTR, Upper Saddle River, 1995.