Optimization of Dynamic Neural Networks Performance for Short-Term Traffic Prediction

By
Sherif Ishak[1]
Department of Civil and Environmental Engineering
Louisiana State University
Baton Rouge, LA 70803
Phone: (225)-578-4846
Fax: (225)-578-8652
Email: sishak@lsu.edu

Prashanth Kotha
Graduate Student
Department of Civil and Environmental Engineering
Louisiana State University
Baton Rouge, LA 70803

Ciprian Alecsandru
Graduate Student
Department of Civil and Environmental Engineering
Louisiana State University
Baton Rouge, LA 70803

Paper submitted for presentation at the
Transportation Research Board
82nd Annual Meeting
Washington, D.C. 2003
and for publication in the Transportation
Research Record journal

Revised and Re-Submitted on October 25, 2002

[1] Corresponding Author

ABSTRACT

This paper presents an approach to optimize the short-term traffic prediction performance using multiple topologies of dynamic neural networks and various network-related and traffic-related settings. The study emphasizes the potential benefit of optimizing the prediction performance by deploying multi-model approaches under parameters and traffic condition settings. The emphasis of the paper is on the application of temporal-processing topologies in short-term speed predictions in the range of 5 to 20 minute-horizons. Three network topologies are utilized: Jordan/Elman, partially recurrent networks, and time-lagged feedforward networks. The input patterns were constructed from data collected at the target location, as well as the upstream and downstream locations. However, various combinations were also considered. To encourage the networks to associate with historical information on recurrent conditions, a time factor was attached to the input patterns to introduce time recognition capabilities, in addition to information encoded in the recent past data. The optimal prediction settings (type of topology and input settings) were determined such that the performance is maximized under different traffic conditions at the target and adjacent locations. The optimized performance of the dynamic neural networks was compared to that of a statistical non-linear time series approach, which was outperformed in most cases. The study also shows that no single topology has consistently outperformed the others for all prediction horizons considered. However, the results show that the significance of introducing the time factor was more pronounced under longer prediction horizons. A comparative evaluation of performance between optimal and non-optimal settings shows substantial improvement in most of the cases. The procedure applied can also be used to identify the prediction reliability of information dissemination systems.

KEYWORDS

Traffic prediction, speed prediction, freeway operation, artificial neural networks, dynamic neural networks, temporal processing networks, and performance optimization.

INTRODUCTION

Traffic information continues to play a key role in supporting the primary functions of advanced traffic management and information systems (ATMIS). The last few years have witnessed major research and development efforts in the area of intelligent transportation systems (ITS). Hundreds of miles of freeway segments in major urban areas nationwide are now instrumented with traffic surveillance systems that communicate to the traffic management centers real time traffic information. Such information is critical to both transportation system users and providers and is often categorized in three different modes: historical, real-time, and predictive. Transportation system users seek traffic information at the pre-trip planning stage and en-route. With accurate and reliable information, travelers can make appropriate decisions to bypass congested segments of the network, change departure times and/or destination, whenever appropriate. Such decisions can strongly influence the demand at various points in the transportation network and provide opportunities for better utilization of the existing infrastructure capacity. For transportation system providers, traffic information is essential for performance monitoring and decision support systems.

For pre-trip planning and en-route decisions, travelers seek predictive more than real-time information to make trip decisions in advance. Such information can effectively reduce travel costs in terms of travel times and delays. Travelers can essentially avoid congested locations by selecting alternative routes, whenever available. With today's remarkable advancements in technology, traffic information can be disseminated to the traveling public via multiple innovative channels such as the Internet and wireless communications with in-vehicle navigation devices. Such devices have many advantages over conventional devices (radio, TV, etc.) in that they can supplement the travelers, through their interactive capabilities, with information pertinent to their specific travel needs, as opposed to all-purpose information provided by traffic reports. While real-time traffic information is now accessible to the traveling public in several urban areas, predictive information is still unattainable at most locations. Having recognized the importance of predictive information to the users' and providers' decision making process, the research community has addressed this need with a variety of short-term traffic prediction models that attempt to capture the dynamics of traffic conditions. A review of the literature on previously developed models and research activities in this area reveals that there is still need to improve on the prediction performance of existing models. This research study presents a multi-model approach to optimize the performance of dynamic neural networks in short-term traffic prediction under different network settings and traffic conditions. The modeling approach also attempts to incorporate the effect of historical information by inserting an explicit representation of time.

RELATED BACKGROUND

The need to relay reliable predictive information has inspired researchers to seek robust traffic prediction models that are capable of forecasting traffic flow, speed, travel times, and delays in short-term horizons, usually in the order of 5 to 20 minutes. Several research efforts were thus conducted in the last few years to support ITS applications and provide the travelers with travel time information at the pre-trip planning stage and en-route. Kaysi et al. (*1*) and Ben-Akiva et al. (*2*) recommended that traffic routing strategies under recurring and non-recurring congestion be based on forecasting future traffic conditions rather than historical and/or current traffic conditions. This is because travelers' decisions are affected by future traffic conditions rather than current traffic conditions. Several prediction methods have been implemented in research in the past two decades. Ben Akiva et al. (*3*) grouped those methods into three categories: (a) statistical models, (b) macroscopic models, and (c) route choice models based on dynamic traffic assignment. Time series models have been extensively used in traffic forecasting for their simplicity and strong potential for on-line implementation (see for example; *4-18*).

Recently, Chen and Chien (*19*) conducted a study using probe vehicle data to compare the prediction accuracy under direct measuring of path-based travel time versus link-based travel times. The study showed that under recurrent traffic conditions, path-based prediction is more accurate than link-based prediction. Chien and Kuchipudi (*20*) presented the results of using real-time and historical data for travel time prediction. Another study by Kwon et al. (*21*) used an approach to estimate travel time on freeways using flow and occupancy data from single loop detectors and historical travel time information. Forecasting ranged from a few minutes into the future up to an hour ahead. The study showed that current traffic conditions are good predictors for the near future, up to 20 minutes, while long-range predictions need the use of historical data.

Recently, several studies have investigated the use of artificial neural networks to model short-term traffic prediction. For instance, Park and Rilett (*22*) proposed two modular Artificial Neural Networks (ANN) models for forecasting multiple-period freeway link travel times. One model used a Kohonen Self Organizing Feature Map (SOFM) while the other utilized a fuzzy c-means clustering technique for traffic patterns classification. Rilett and Park (*23*) proposed a one-step approach for freeway corridor travel time forecasting rather than link travel time forecasting. They examined the use of a spectral basis neural network with actual travel times from Houston, Texas. Another study by Abdulhai et al. (*24*) used an advanced time delay neural network (TDNN) model, optimized using a Genetic Algorithm, for traffic flow prediction. The results of the study indicated that prediction errors were affected by the variables pertinent to traffic flow prediction such as spatial contribution, the extent of the loop-back interval, resolution of data, and others. Lint et al. (*25*) presented an approach for freeway travel time prediction with state-space neural networks. Using data from simulation models, they showed that prediction accuracy was acceptable and favorable to traditional models. Several other studies applied neural networks for predicting speed, flows, or travel times (see for instance; *26-31*).

OBJECTIVES

The primary objective of this study is to seek optimal settings that maximize the performance of dynamic neural networks in short-term traffic prediction of speed on freeways. The study investigates the performance of three different dynamic neural network topologies under different network and traffic operating condition settings. The three architectures belong to the class of dynamic or temporal-processing networks. For each network, we investigate the effect of inserting time explicitly in the representation of the input patterns to strengthen the networks' ability to learn from historical traffic conditions and their recurrent characteristics, in addition to recent past information that sensitizes the models to non-recurrent conditions.

DESCRIPTION OF THE DYNAMIC NEURAL NETWORK MODELS

The type of neural networks applied in this study belongs to the class of recurrent neural networks that is known to have closed loops (feedback) in their topological structure. Unlike conventional feedforward networks or static mappers such as multilayer perceptron, dynamic networks have temporal processing capabilities in the form of memory or feedback in their topologies. Memories are preserved from past values information in a linear combiner that has a finite transient response. Feedback topologies are referred to as recurrent networks, where the output of a neuron can be fed back to a neuron in the input or intermediate layers. Such networks are popular in applications of time series predictions (*32*). A major distinction between static and dynamic networks lies in the representation of short-term and long-term memory. Static networks have the ability to build long-term memory in their synaptic weights during training. However, they do not have explicit realization of time. On the contrary, dynamic models are sensitive to the sequence of presentation of information as a result of the memory structures or the recurrent connections. In summary, dynamic networks are more capable of codifying enhanced information from the input that leads to better temporal processing capabilities. In order to assess the performance of dynamical neural networks, three topological structures are considered: simple recurrent networks (Jordan/Elman), partial recurrent networks (PRN), and time-lagged feedforward networks (TLFN). The three architectures were applied in the study to compare the prediction performance of each under different network configurations and traffic conditions. A brief description of each model is provided next. More elaborate details can be found in the cited references.

Jordan/Elman Network

The Jordan/Elman network is referred to as the simple recurrent network (SRN) (*33*). It is a single hidden-layer feedforward network with feedback connections from the outputs of the hidden-layer neuron to the network input layer. It was originally developed to learn temporal sequences or time-varying patterns. As shown in FIGURE 1(a), the network contains context units located in the upper portion and used to replicate the hidden-layer output signals at the previous time step. The context units are introduced to resolve conflicts arising from patterns that are similar, yet result in similar outputs. The feedback provides a mechanism to discriminate between identical patterns occurring at different times. The context unit is also referred to as a low-pass filter that creates a weighted average output of some of the more recent past inputs. Therefore, the context units are also called “memory units” since they tend to remember information from the past events. The training phase of this network is achieved by adapting all the weights using standard backpropagation procedures. More details on this topology can be found in (*33, 36*).

Partially Recurrent Networks (PRN)

This network is considered a simplified version of the Jordan/Elman network without hidden neurons. It is composed of an input layer of source and feedback nodes, and an output layer, which is composed of two types of computation nodes: output neurons and context neurons. The output neurons produce the overall output, while the context neurons provide feedback to the input layer after a time delay. The topological structure of the network is illustrated in FIGURE 1(b). More details can be found in (*34, 36*).

Time-Lagged Feedforward Networks (TLFN)

In dynamic neural networks time is explicitly included in mapping input-output relationships. As a special type, TLFN extends nonlinear mapping capabilities with time representation by integrating linear filter structures in a feedforward network. The type of topology used in this study is also called focused TLFN and has memory only at the input layer. The TLFN is composed of feedforward arrangement of memory and nonlinear processing elements. It has some of the advantages of feedforward networks such as stability, and can also capture information in input time signals. FIGURE 1(c) shows a simplified topological structure of the focused TLFN. The figure shows that memory PE (processing elements) are attached in the input layer only. The input-output mapping is performed in two stages: a linear time-representation stage at the memory PE layer and a nonlinear static stage between the representation layer and the output layer. Further details the underlying mathematical operations of TLFN can be found in (*33*, *35*, *36*).

STUDY AREA AND DATA COLLECTION

The study was conducted on a freeway segment of I-4 in Orlando, Florida, using 30-second loop detector speed data. The traffic surveillance system compiles data from a 40-mile six-lane corridor instrumented with 70 inductive dual loop detector stations spaced at nearly 0.5 miles in both directions. The real time and archived data is accessible on the web at http://trafficinfo.engr.ucf.edu. The loop detector data is collected in real time via a T1 link between the Regional Traffic Management Center (RTMC) in Orlando and the intelligent transportation system lab at the University of Central Florida. Speed, volume counts, and lane occupancies are downloaded and compiled into an SQL server that supports multiple publicly accessible web applications such as real time and short-term travel time predictions between user-selected on- and off-ramps. The web-based short-term traffic prediction system was implemented using a nonlinear time series model that was tested extensively in a previous study (*18*). In this study the data was collected from a one-mile test section that is covered by three adjacent stations. A total of 28 weekdays were randomly selected in the year 2001, where data was available during the entire morning peak period from 6:00 AM to 10:00 AM in the westbound direction. To suppress the noise and random fluctuations in speed, the data was aggregated over 5-minute moving time windows. The 28 days were randomly split into 14 days for training, 4 for cross-validation, and 10 for testing.

TRAINING

The model development process was completed using NeuroSolutions (*36*). In order to achieve optimal performance, different settings were attempted by varying the number and type of inputs to each network topology. The input patterns were composed of two portions: the recent past information (mandatory) and the time factor (optional). The recent past information component was represented by speed data observed in the past 10 minutes at the three adjacent stations, individually and collectively to capture temporal and spatial variations of traffic conditions. For optimization purposes, the input patterns were constructed from four spatial settings: current station (y), current and upstream (y, z), current and downstream (y, x), or the three stations combined (x, y, z). The time factor was appended to the input patterns and toggled on and off to test the network's ability to learn from similar historical traffic conditions observed at the same time on other weekdays. The essence of inserting the time factor, as explained earlier, is to make the network time-cognizant, and thus, improve its predictive performance at relatively longer prediction horizons under recurrent conditions. Each network was trained to predict the average 5-minute speeds at 5, 10, 15, or 20-minute horizons. During the training phase the performance of the networks was monitored via the cross-validation set to avoid overtraining. Training was terminated when the mean square error (MSE) for the cross-validation set did not decrease for 50 consecutive training cycles, a common procedure to prevent overtraining.

For each network considered, the input patterns were generated from one of the following four input scenarios: {$y(t)$}, {$y(t), z(t)$}, {$y(t), x(t)$}, or {$y(t), x(t), z(t)$}, where:
$y(t)$ = average 5-minute speed at target station at time t
$x(t)$ = average 5-minute speed at downstream station at time t
$z(t)$ = average 5-minute speed at upstream station at time t

The time component was optionally attached to each of the four input scenarios. Each network produces one single output that represents the average speed at the target station $S(t+\delta)$, where δ, the prediction horizon, was set to 5, 10, 15, or 20 minutes.

TESTING

In order to test the performance of each network after training was completed, a set of 10 peak periods collected from 10 different weekdays was presented to each network. For each input pattern in the testing set, predictions were evaluated for 5, 10, 15, and 20 minute horizons. Each predicted value was compared against the actual observed value to calculate two measures of performance: average absolute relative error (*AARE*) and root mean square error (*RMSE*). Each measure is defined as follows:

$$AARE = \frac{\sum_{i=1}^{N} \left| \frac{S_i - V_i}{V_i} \right|}{N} \tag{1}$$

And

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (S_i - V_i)^2}{N}} \tag{2}$$

Where
AARE = Average absolute relative error of speed
RMSE = Root mean square error (mph)
S_i = Predicted speed (mph) for observation i
V_i = Actual speed (mph) for observation i
N = Number of observations

The two previous measures were used to compare the performance of the three network topologies under the following settings: desired prediction horizon (5, 10, 15, and 20 minutes), input type (*y*, *xy*, *yz,* and *xyz*), and presence of time factor (Yes/No). In the meantime, the performance was also tested under various combinations of traffic conditions at each of the three stations. At each station, traffic conditions were broken down into four levels of congestion: level 1 (speed >60 mph), level 2 (40-60 mph), level 3 (20-40 mph), and level 4 (<20 mph). For each combination of the four levels of congestion at each station, the network performance was evaluated to determine the optimal settings. This procedure was primarily used to answer two questions. First, which network topology and what input settings are optimal for predictions at each of the four prediction horizons considered? Second, what combinations of traffic conditions are associated with the largest errors or worst prediction performance in order to identify the level of confidence in our predictions? Both questions are critical to the successful online implementation of a traffic prediction system. The answer to the first question will identify which network and what settings are best for variable traffic conditions. This allows us to deploy a multi-model traffic prediction system that is optimized under different prediction horizons, input settings, and traffic conditions. The answer to the second question underlines the prediction accuracy and level of uncertainty associated with the application of each model and each traffic condition. This can be directly used as a reliability measure of the traffic information dissemination system.

OPTIMAL SETTINGS

The optimal settings were selected for each of the four levels of prediction horizon by minimizing *AARE* and *RMSE* independently. Despite the fact that both measures are used to quantify the performance of prediction models, consistency between the two measures is not, by their mathematical definition, guaranteed, and thus, the optimal settings obtained from each may not necessarily be identical. While some optimal settings were different, others were similar, indicating that both measures are in agreement. For each prediction horizon, a total of 25 cases resulted from different combinations of congestion levels at the three stations. The 25 cases, however, were determined by the testing set, and therefore, do not include all possible combinations of congestion levels. Theoretically, with four levels of congestion at each station, the total number of combinations would amount to 64 cases (4x4x4). However, some combinations were not observed in the testing set due either to its limited size or to their infrequent occurrence in general. Therefore, the optimization was based on the 25 cases only. Clearly, additional testing with a larger data set can be similarly applied to optimize the performance for the remaining cases.

The optimal settings for each of the four prediction horizons based on *AARE* are shown in TABLE 1 and TABLE 2. Each table shows the optimal settings for each of the 25 cases in an ascending order of the *AARE*. Each record in the tables shows the optimal network topology, the type of input, and the relevance of time factor corresponding to each case. The tables also compare the performance of the NN approach to that of a non-linear time series approach that was developed and tested in a previous study (*16*, *18*). The time series model was tested under different traffic conditions using prediction horizons of 5, 10, and 15 minutes. Comparisons are provided in each table in terms of *AARE*, except for 20-minute prediction horizons where the time series performance was not available. The tables show that, except for a few cases that are shown in grey, the optimized NN approach has outperformed the time series approach in terms of *AARE*. It should be noted also that the time series approach was derived from the past recent information at the target station only, and therefore, the effect of traffic conditions at adjacent upstream and downstream stations was not accounted for.

Considering the optimized NN approach, both tables show that no particular network topology seemed to have outperformed the others for all cases. The same applies to the type of inputs and the relevance of time factor. For 5-minute predictions of speed, the *AARE* did not exceed 7.3% for all 25 cases. Six cases produced errors in the range of 5% to 7% while the rest were consistently below 3%. For 10-minute predictions, 19 cases exhibited errors less than 10%. Four cases showed errors greater than 20% and were mostly characterized with severe congestion at one or more of the three stations (levels 3 or 4). Similar conclusions can be made for 15- and 20-minute predictions, which show consistently larger errors for most of the cases. Intuitively, larger prediction errors are expected with longer prediction horizons. Even when such errors are uncontainable, it is imperative that we identify cases associated with such large errors for reliability purposes. Similarly, the optimal settings were also selected based on the *RMSE* for each prediction horizon. Comparisons show some discrepancies between optimal settings for each case based on each measure. Cases with identical optimal settings are marked with an asterisk in both tables.

PERFORMANCE ENVELOPES

The minimum values of *AARE* and *RMSE* corresponding to the optimal settings for all prediction horizons are plotted in FIGURE 2(a) and FIGURE 2(b), respectively. The figures show the lower and upper bounds or performance envelopes for each case, which are associated with the minimum (5-minute) and maximum (20-minute) prediction horizons, respectively. The performance envelopes can be used to identify cases where prediction errors do not meet a maximum acceptable threshold value that is appropriate for online implementation. For example, if we set the *AARE* threshold value to 10%, then cases with prediction errors larger than 10% should be avoided. Consequently, when such cases are encountered in real world, the high levels of uncertainty in predictions will then be realized and treated appropriately by traffic information dissemination systems.

It should be emphasized here that the large errors associated with such cases may be attributed to under-representation of those cases in the training data set. This often leads to the network's inability to generalize under such conditions. Additional training with data collected from such infrequent conditions could essentially lead to improvement in the overall prediction performance. Despite the potential improvement of performance that could result from additional training, it is still reasonable to expect that there will be certain conditions where the optimal settings cannot yield the minimum acceptable prediction performance. As such, traffic prediction systems must have the capability to identify large-error prediction cases to measure the reliability of predictive information.

EFFECT OF NETWORK TOPOLOGY AND TIME COMPONENT

The optimal settings defined in the study included the network topology and the relevance of attaching the time component to the network input. In this section we investigate the optimal network topology that is prevalent in the optimal settings for each prediction horizon. FIGURE 3(a) shows that no specific topology appears to be consistently prevalent in the optimal settings. While TLFN produced better performance in 5- and 10-minute predictions, the performance of Jordan/Elman and PRN was slightly better for 15- and 20-minute predictions, respectively. For all cases combined, the performance of the three topologies was very comparable.

Another critical factor that was examined in this study is the influence of time component on the prediction performance. This component was introduced to strengthen the time realization feature of the network and its ability to learn from historical information. This feature can be extremely useful in predicting the onset of congestion and in improving the accuracy of longer-horizon predictions, when predicted conditions are more likely to be less dependent on the information relayed by the recent past values. This can essentially lead to a model that is more capable of predicting traffic under both recurrent and non-recurrent conditions. FIGURE 3(b) shows the percentage of cases whose optimal settings included the time component. The figure clearly shows that the relevance of the time realization feature is more pronounced in longer-horizon predictions. For instance, nearly 96% of the cases favored the time component in their optimal settings for 20-minute predictions, as opposed to nearly 63% for 5-minute predictions. This trend suggests the increasingly critical role of time when making longer-term predictions.

COMPARISON BETWEEN OPTIMAL AND NON-OPTIMAL PERFORMANCE

In order to quantify the performance improvements achieved by optimization with traffic conditions and network settings versus optimization with network settings only, we compare the optimal with the non-optimal performance for each case. In our comparison, non-optimal performance refers to optimization with network settings only and without consideration of traffic conditions. Optimal performance, on the other hand, refers to optimization with both network settings and traffic conditions. To facilitate the comparison, the reduction in errors of both scenarios was utilized. FIGURE 4(a) shows the percentage reduction in *AARE* for each prediction horizon. The figure shows significant improvements in prediction performance for most of the cases as a result of optimization with traffic conditions. Improvements as high as 30% in case 11 and 25% in cases 11, 13, and 17 were observed. In most cases, the reduction in *AARE* was in the range of 5% to 10% as a result of optimization with traffic conditions. Similar results were attained in the comparative evaluation of performance based on *RMSE*. FIGURE 4(b) shows the improvements in terms of the reduction in *RMSE* (mph) for each case and each prediction horizon. The reduction of *RMSE* was as high as 10 mph for case 9 and nearly 8 mph for cases 13, 19, 20, and 23. A better illustration of the performance improvements in all cases combined can be seen in FIGURE 5(a) and FIGURE 5(b). Each figure shows the cumulative percentage of cases with a corresponding reduction in *AARE* and *RMSE*. For instance, the figures show that 90% of the cases showed improvements of 13% or less in terms of *AARE* reduction and nearly 4.5 mph in terms of *RMSE* reduction. Such improvements were exclusively attributed to performance optimization with traffic conditions.

CONCLUSIONS

This study presented an approach to optimize the performance of dynamic neural networks in short-term traffic prediction. Three neural network architectures, Jordan/Elman, partially recurrent, and focused time-lagged feedforward networks, were trained and tested under various network configurations and traffic conditions settings. The input to the networks was divided into two main components: recent past information and a time factor. The recent past information was represented by spatiotemporal observations in the past 10 minutes. The time factor was arbitrarily attached to the input patterns in order to strengthen the time recognition capability of the networks. This technique was primarily introduced to allow the networks to learn from historical information on traffic conditions that are likely to be recurrent. This was necessary to improve the prediction performance during recurring conditions when future predictions are less dependent on information relayed by recent past data. The predictive performance was measured in terms of two types of errors: average absolute relative error and root mean square error.

The optimal settings were selected to minimize the prediction errors under different network and traffic condition settings for variable prediction horizons. The optimal settings were based on the testing results obtained from the three network topologies trained with the same data set. The network settings were varied by changing the input source from three adjacent detector stations and by toggling the time component on and off. Traffic conditions were broken down into four levels of congestion at each of the three stations. This resulted in a total of 25 combinations of different traffic conditions in the testing data set. Each of the 25 cases was optimized independently to identify the optimal network topology and the optimal network settings. The results of the study showed that no particular network topology has consistently outperformed the others for all prediction horizons and all cases. However, the performance optimization for different traffic conditions has the advantage of identifying cases where none of the trained networks were able to produce acceptable performance. While additional training with more data may improve the performance for some of those cases, it is still unequivocally critical to identify cases where the expected prediction performance may fall below the minimum acceptable by traffic management centers. This is a critical factor in addressing information reliability issues of information dissemination systems developed for the traveling public. The optimized prediction performance was also compared to that of a statistical non-linear time-series approach, which was outperformed by the NN approach in most of the cases and prediction horizons.

Another important finding is the effect of the time component on the optimal performance. The results showed that the time factor appeared more frequently in the optimal settings of longer prediction horizons. Nearly 96% of the cases included the time component in their optimal settings for 20-minute predictions, as opposed to 63% for 5-minute predictions. This finding emphasizes the critical role of inserting time explicitly in the models to improve longer horizon predictions. The study also presented a comparative evaluation of the prediction performance under optimal and non-optimal traffic condition settings. Using the reduction in *AARE* and *RMSE* the performance improvement in each case and for each prediction horizon was evaluated. The comparative evaluation clearly demonstrates the benefit of optimizing the performance under different traffic conditions at the same station, and both upstream and downstream stations.

Finally, the study presented an approach towards the development of a more efficient traffic prediction system with multiple neural network topologies and multiple network and traffic condition settings. The approach was extensively examined at one location and the results were encouraging for the approach to be applied to other locations as well. For locations that exhibit similar traffic conditions during the peak periods, the settings obtained in this study may be transferable directly without retraining. However, testing is still recommended with data collected from the other locations first. If the testing results are not satisfactory, then the current settings may not be applicable and the performance must be optimized at the new locations following the steps described in this study.

REFERENCES

1. Kaysi, I., M. Ben-Akiva, and H. Koutsopoulos. "An Integrated Approach to Vehicle Routing and Congestion Prediction for Real-Time Driver Guidance." Paper presented at the 72nd TRB Annual Meeting, Washington, D.C., 1993
2. Ben-Akiva M., A. de Palma, and I. Kaysi. "Dynamic Network Models and Driver Information Systems". *Transportation Research*, Vol. 25A, No. 5, 1991.
3. Ben-Akiva, M., E. Kroes, J. de Ruiter, S. Smulders, J. Whittaker, and K. Lindveld. "Real-Time Prediction of Traffic Congestion." Vehicle navigation & information systems: conference record of papers. New York, N.Y. Institute of Electrical and Electronics Engineers, 1992. p. 557-562, 1992.

4. Stamatiadis, C. and W. Taylor. "Travel Time Predictions for Dynamic Route Guidance with a Recursive Adaptive Algorithm." Paper presented at the 73rd TRB Annual Meeting, Washington, D.C., 1994.
5. Ahmed, M., and A. Cook. "Analysis of Freeway Traffic Time Series Data by Using Box-Jenkins Techniques." *Transportation Research Record* 722, TRB, National Research Council, Washington, D.C., pp 1-9, 1982.
6. Gazis, D., and C. Knapp. "On Line Estimation of Traffic Densities from Time Series of Traffic and Speed Data." *Transportation Science*, Vol. 5, No. 3, 1971.
7. Kyte, M., J. Mark, and D. Frith. "Using Multivariate Time Series Analysis to Model Freeway Traffic Flow." Paper presented at the 68th TRB Annual Meeting, Washington, D.C., 1989.
8. Lu Jian. "Prediction of Traffic Flow by an Adaptive Prediction System." *Transportation Research Record* 1287, TRB, National Research Council, Washington, D.C., pp 54-61, 1990.
9. Nihan, N., and G. Davis. "Application of Prediction-Error Minimization and Maximum Likelihood to Estimate Intersection O-D Matrices from Traffic Counts." *Transportation Science*, Vol. 23, No. 2, pp 77-90, 1989.
10. Okutani, I. and Y. Stephanedes. "Dynamic Prediction of Traffic Volume through Kalman Filtering Theory." *Transportation Research B*, Vol. 18B, No. 1, pp 1-11, 1984.
11. Hamed, M. M., H.R. Al-Masaeid, and Z.M.B. Said. "Short-Term Prediction of Traffic Volume in Urban Arterials." *ASCE Journal of Transportation Engineering*, Vol. 121, No. 3, pp 249-254, 1995.
12. Lee, S. and D. Fambro. "Application of the Subset ARIMA Model for Short-Term Freeway Traffic Volume Forecasting." *Transportation Research Record. No. 1678* (1999) p. 179-188.
13. Chang, J. L. and S.P. Miaou. "Real-Time Prediction of Traffic Flows using Dynamic Generalized Linear Models." *Transportation Research Record, No. 1678*, pp 168-178, 1999.
14. Jiang, Y. "Dynamic Prediction of Traffic Flow and Congestion at Freeway Work Zones." Transportation Research Board Annual Meeting in Washington D.C., 1999.
15. Iwasaki, M. and K. Shirao. "A Short Term Prediction of Traffic Fluctuations Using Pseudo-Traffic Patterns." ITS world congress conference proceedings, 1996.
16. D'Angelo, M., H. Al-Deek, and M. Wang. "Travel Time Prediction for Freeway Corridors**."** *Transportation Research Record. No. 1676* p. 184-191, 1999.
17. Williams, B. M. "Multivariate Vehicular Traffic Flow Prediction: An Evaluation of ARIMAX Modeling" Transportation Research Board Annual Meeting in Washington D.C., 2001.
18. Ishak, S. and H. Al-Deek. "Performance Evaluation of a Short-Term Time-Series Traffic Prediction Model" *ASCE Journal of Transportation Engineering*, Vol. 128, No. 6, 2002.
19. Chen, M. and S. Chien. "Dynamic Freeway Travel Time Prediction Using Probe Vehicle Data: Link-based vs. Path-based." Transportation Research Board Annual Meeting in Washington D.C., 2001.
20. Chien S. and C. M. Kuchipudi. "Dynamic Travel Time Prediction with Real-Time and Historical Data". Transportation Research Board Annual Meeting in Washington D.C., 2002.
21. Kwon, J., B. Coifman, and P. Bickel. "Day-to-Day Travel Time Trends and Travel Time Prediction from Loop Detector Data." Transportation Research Board Annual Meeting in Washington D.C., 2000.

22. Park, D. and L. Rilett. "Forecasting Multiple-Period Freeway Link Travel Times Using Modular Neural Networks." *Transportation Research Record. No. 1617* p. 163-170, 1998.
23. Rilett, L., D. Park. "Direct Forecasting of Freeway Corridor Travel Times Using Spectral Basis neural Networks." *ASCE Journal of Transportation Engineering*, Vol. 125, No. 6 1999.
24. Abdulhai, B., H. Porwal, and W. Recker. "Short-Term Freeway Traffic Flow Prediction Using Genetically Optimized Time-Delay-Based Neural Networks." Transportation Research Board Annual Meeting in Washington D.C., 1999.
25. Lint, V., S. P. Hoogendoorn, and H.J. Zuylen. "Freeway Travel Time Prediction with State-Space Neural Networks." Transportation Research Board 81st Annual Meeting, 2002.
26. Fu, L. and L.R. Rilett. "Estimation of Time-Dependent, Stochastic Route Travel Times using Artificial Neural Networks." *Transportation Planning and Technology*, 24(1), pp. 25-36, 2000.
27. Park, D., L.R. Rilett, and G. Han. "Spectral Basis Neural Networks for Real-Time Travel Time Forecasting." *ASCE Journal of Transportation Engineering*, Vol. 125, No. 6, pp. 515-523, 1999.
28. Van Der Voort, M., M. Dougherty, and S. Watson. "Combining Kohonen Maps with ARIMA Time Series Models to Forecast Traffic Flow." *Transportation Research part C*, 4(5), pp. 307-318, 1996.
29. Yasdi, R. "Prediction of Road Traffic using a Neural Network Approach" *Neural Computing and Applications*, 8, pp. 135-142, 1999.
30. Zhang, H. M. "Recursive Prediction of Traffic Conditions with Neural Network Models" *ASCE Journal of Transportation Engineering,* Vol. 126, No. 6, pp. 472-481, 2000.
31. Zhang, H., S. G. Ritchie, and Z.P. Lo. "Macroscopic Modeling of Freeway Traffic using an Artificial Neural Network" *Transportation Research Record, No. 1588,* pp 110-119, 2000.
32. Duda, R., P. Hart, and D. Stork. Pattern Classification. Second Edition, John Wiley & Sons, Inc., 2001.
33. Ham, F. M. and I. Kostanic. *Principles of Neurocomputing for Science and Engineering*. McGraw Hill, NY, 2001.
34. Haykin, S. *Neural Networks, A Comprehensive Foundation*. Macmillan College Publishing Company, Inc., 1994.
35. Principe, J. C., N. R. Euliano, and W.C. Lefebvre. *Neural and Adaptive Systems*. John Wiley and Sons, Inc., NY, 2000.
36. Lefebvre, Curt, "Neuro Solutions, Version 4.10", NeuroDimension, Inc., Gainesville, FL, 2001.

LIST OF TABLES

TABLE 1 OPTIMAL SETTINGS FOR 5- AND 10-MINUTE PREDICTIONS BASED ON *AARE*
TABLE 2 OPTIMAL SETTINGS FOR 15- AND 20-MINUTE PREDICTIONS BASED ON *AARE*

LIST OF FIGURES

FIGURE 1 SIMPLIFIED TOPOLOGIES OF THE APPLIED DYNAMIC NEURAL NETWORKS.
FIGURE 2 PERFORMANCE ENVELOPES ASSOCIATED WITH OPTIMAL SETTINGS FOR EACH CASE.
FIGURE 3 EFFECT OF NETWORK TOPOLOGY AND TIME FACTOR ON OPTIMAL SETTINGS.
FIGURE 4 REDUCTIONS IN *AARE* AND *RMSE* ACHIEVED BY OPTIMIZATION WITH TRAFFIC CONDITIONS.
FIGURE 5 PERFORMANCE IMPROVEMENTS ACHIEVED BY OPTIMIZATION WITH TRAFFIC CONDITIONS.

TABLE 1 Optimal Settings for 5- and 10-Minute Predictions Based on *AARE*

(a) 5-Minute Prediction Horizon

Case	Congestion Indicator			Optimal Settings for NN Approach			Average Absolute Relative Error	
	Downstream Station (X)	Current Station (Y)	Upstream Station (Z)	Network	Inputs	Time Factor	NN Approach	Time Series Approach
21	3	4	4	JORDAN-ELMAN	Y	Yes	<0.001	0.018
8	2	2	4	TLFN	XYZ	Yes	0.001	0.169
18	3	3	3	TLFN	Y	No	0.001	0.031
25	4	4	4	TLFN	ZY	Yes	0.001	0.018
4*	2	1	4	PRN	XY	Yes	0.002	0.362
6	2	2	2	PRN	XYZ	No	0.002	0.169
19	3	3	4	JORDAN-ELMAN	ZY	Yes	0.002	0.031
9	2	3	3	TLFN	XY	No	0.003	0.031
10	2	3	4	TLFN	ZY	No	0.003	0.031
13*	3	1	3	JORDAN-ELMAN	ZY	No	0.003	0.362
20	3	4	3	TLFN	XYZ	No	0.003	0.018
22	4	3	3	JORDAN-ELMAN	XYZ	No	0.003	0.031
24	4	4	3	TLFN	ZY	No	0.003	0.018
15	3	2	2	JORDAN-ELMAN	ZY	Yes	0.005	0.169
14	3	2	1	TLFN	XY	Yes	0.008	0.169
23*	4	3	4	TLFN	Y	No	0.011	0.031
3	2	1	3	PRN	Y	Yes	0.013	0.362
11*	3	1	1	PRN	Y	Yes	0.027	0.362
5*	2	2	1	TLFN	XY	Yes	0.029	0.169
1	2	1	1	PRN	Y	Yes	0.05	0.362
2*	2	1	2	PRN	Y	Yes	0.051	0.362
17*	3	2	4	JORDAN-ELMAN	XY	Yes	0.054	0.169
12	3	1	2	PRN	Y	Yes	0.058	0.362
16*	3	2	3	JORDAN-ELMAN	XY	Yes	0.072	0.169
7	2	2	3	PRN	Y	Yes	0.073	0.169

(b) 10-Minute Prediction Horizon

Case	Congestion Indicator			Optimal Settings for NN Approach			Average Absolute Relative Error	
	Downstream Station (X)	Current Station (Y)	Upstream Station (Z)	Network	Inputs	Time Factor	NN Approach	Time Series Approach
9	2	3	3	TLFN	ZY	Yes	<0.001	0.030
21	3	4	4	Jordan-Elman	XYZ	No	0.002	0.015
25	4	4	4	PRN	XYZ	Yes	0.002	0.015
5	2	2	1	PRN	Y	Yes	0.003	0.199
6	2	2	2	PRN	XYZ	No	0.003	0.199
8	2	2	4	TLFN	Y	Yes	0.005	0.199
15	3	2	2	TLFN	Y	Yes	0.005	0.199
24	4	4	3	TLFN	ZY	No	0.006	0.015
20	3	4	3	TLFN	XYZ	Yes	0.007	0.015
23	4	3	4	PRN	Y	Yes	0.01	0.030
18*	3	3	3	TLFN	XY	Yes	0.01	0.030
14	3	2	1	PRN	ZY	Yes	0.015	0.199
4	2	1	4	TLFN	ZY	No	0.02	0.465
22*	4	3	3	PRN	XY	No	0.08	0.030
10	2	3	4	TLFN	XYZ	No	0.088	0.030
19*	3	3	4	Jordan-Elman	XYZ	Yes	0.109	0.030
12	3	1	2	TLFN	Y	Yes	0.111	0.465
2	2	1	2	TLFN	Y	Yes	0.113	0.465
1*	2	1	1	TLFN	Y	Yes	0.117	0.465
3	2	1	3	PRN	ZY	Yes	0.137	0.465
7	2	2	3	Jordan-Elman	Y	No	0.144	0.199
16*	3	2	3	Jordan-Elman	XY	Yes	0.196	0.199
11*	3	1	1	TLFN	Y	Yes	0.356	0.465
13*	3	1	3	Jordan-Elman	XY	Yes	0.394	0.465
17*	3	2	4	Jordan-Elman	XY	Yes	0.409	0.199

Cases marked with * have the same optimal settings for both *AARE* and *RMSE*.

TABLE 2 Optimal Settings for 15- and 20-Minute Predictions Based on *AARE*

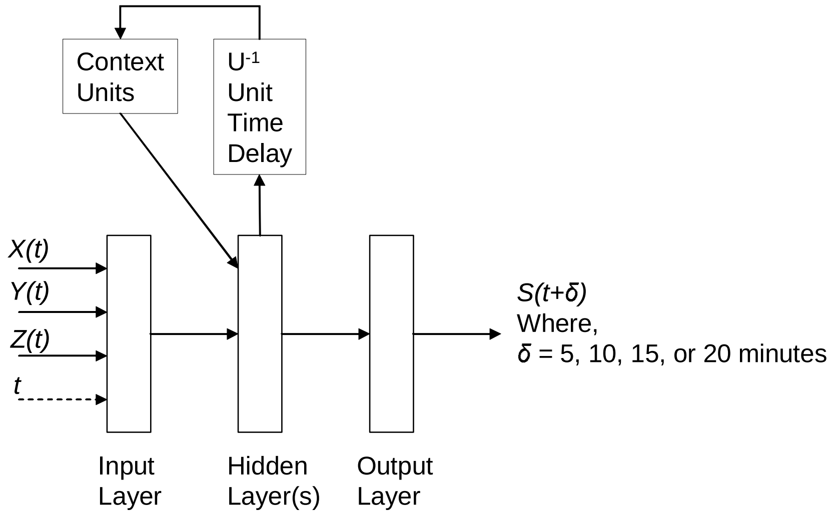
(c) 15-Minute Prediction Horizon

Case	Congestion Indicator			Optimal Settings for NN Approach			Average Absolute Relative Error	
	Downstream Station (X)	Current Station (Y)	Upstream Station (Z)	Network	Inputs	Time Factor	NN Approach	Time Series Approach
14	3	2	1	Jordan-Elman	ZY	Yes	<0.001	0.199
23	4	3	4	Jordan-Elman	Y	Yes	0.002	0.030
24	4	4	3	Jordan-Elman	ZY	Yes	0.002	0.015
5	2	2	1	TLFN	Y	Yes	0.003	0.199
6	2	2	2	TLFN	XYZ	Yes	0.003	0.199
9	2	3	3	Jordan-Elman	ZY	Yes	0.005	0.030
22	4	3	3	TLFN	XYZ	No	0.005	0.030
25*	4	4	4	Jordan-Elman	XYZ	Yes	0.006	0.015
21	3	4	4	Jordan-Elman	XYZ	Yes	0.007	0.015
4*	2	1	4	Jordan-Elman	XYZ	Yes	0.01	0.465
8	2	2	4	TLFN	Y	No	0.01	0.199
18*	3	3	3	Jordan-Elman	ZY	Yes	0.015	0.030
20*	3	4	3	Jordan-Elman	XY	Yes	0.02	0.015
19*	3	3	4	Jordan-Elman	ZY	Yes	0.08	0.030
15*	3	2	2	PRN	Y	Yes	0.088	0.199
2*	2	1	2	PRN	Y	Yes	0.109	0.465
1*	2	1	1	Jordan-Elman	XY	Yes	0.111	0.465
10	2	3	4	PRN	ZY	No	0.113	0.030
12*	3	1	2	PRN	Y	Yes	0.117	0.465
11*	3	1	1	PRN	Y	Yes	0.137	0.465
3	2	1	3	Jordan-Elman	XY	Yes	0.144	0.465
7*	2	2	3	PRN	Y	Yes	0.196	0.199
17	3	2	4	PRN	Y	Yes	0.356	0.199
13*	3	1	3	PRN	Y	Yes	0.394	0.465
16	3	2	3	PRN	Y	Yes	0.409	0.199

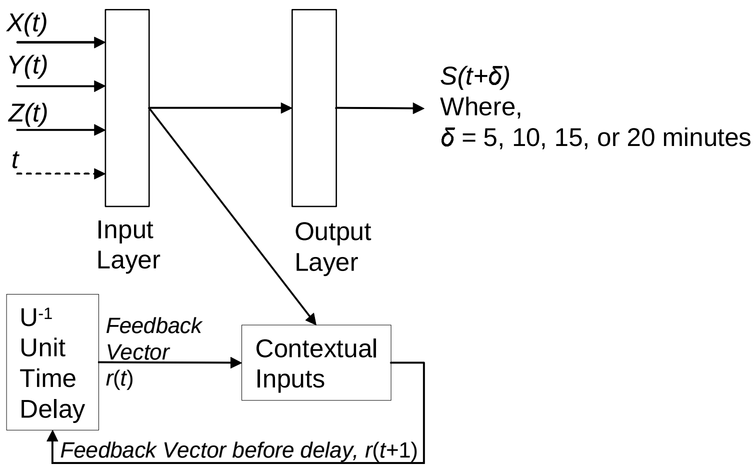
(d) 20-Minute Prediction Horizon

Case	Congestion Indicator			Optimal Settings for NN Approach			Average Absolute Relative Error **
	Downstream Station (X)	Current Station (Y)	Upstream Station (Z)	Network	Inputs	Time Factor	NN Approach
8	2	2	4	PRN	ZY	Yes	<0.001
9	2	3	3	TLFN	XY	Yes	<0.001
18	3	3	3	TLFN	XY	Yes	0.003
23	4	3	4	Jordan-Elman	XYZ	Yes	0.003
22*	4	3	3	PRN	Y	Yes	0.008
14	3	2	1	TLFN	ZY	Yes	0.010
5	2	2	1	PRN	XYZ	Yes	0.015
20	3	4	3	TLFN	XY	Yes	0.015
4*	2	1	4	TLFN	ZY	Yes	0.017
25*	4	4	4	Jordan-Elman	XYZ	Yes	0.019
21	3	4	4	TLFN	XY	Yes	0.030
6	2	2	2	PRN	XY	Yes	0.064
24	4	4	3	TLFN	XYZ	Yes	0.073
1	2	1	1	PRN	XYZ	Yes	0.076
19	3	3	4	TLFN	XY	Yes	0.113
12*	3	1	2	Jordan-Elman	ZY	Yes	0.134
10	2	3	4	Jordan-Elman	XYZ	No	0.137
2*	2	1	2	Jordan-Elman	XYZ	Yes	0.170
3*	2	1	3	Jordan-Elman	XYZ	Yes	0.232
15	3	2	2	PRN	XY	Yes	0.270
11*	3	1	1	PRN	XY	Yes	0.316
7*	2	2	3	PRN	XY	Yes	0.418
17*	3	2	4	PRN	ZY	Yes	0.436
13	3	1	3	PRN	XY	Yes	0.527
16*	3	2	3	Jordan-Elman	XYZ	Yes	0.600

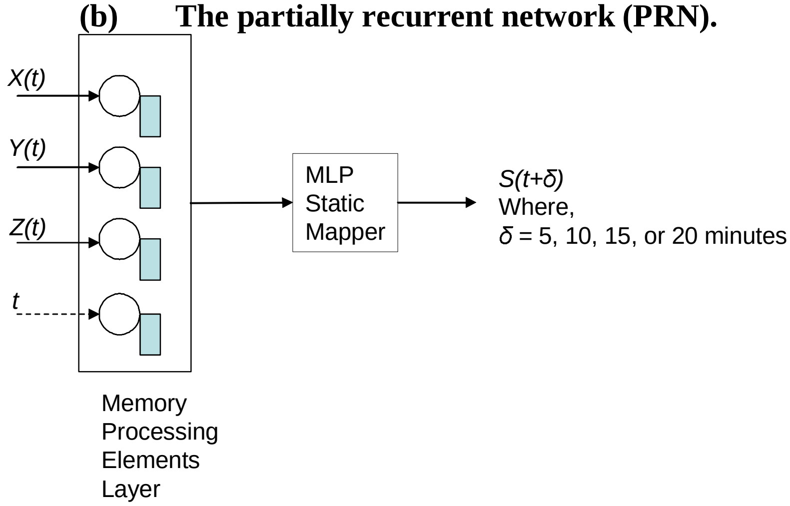
Cases marked with * have the same optimal settings for both *AARE* and *RMSE*.
** Time series predictions were not available for 20-minute prediction horizons



(a) The Jordan/Elman network.

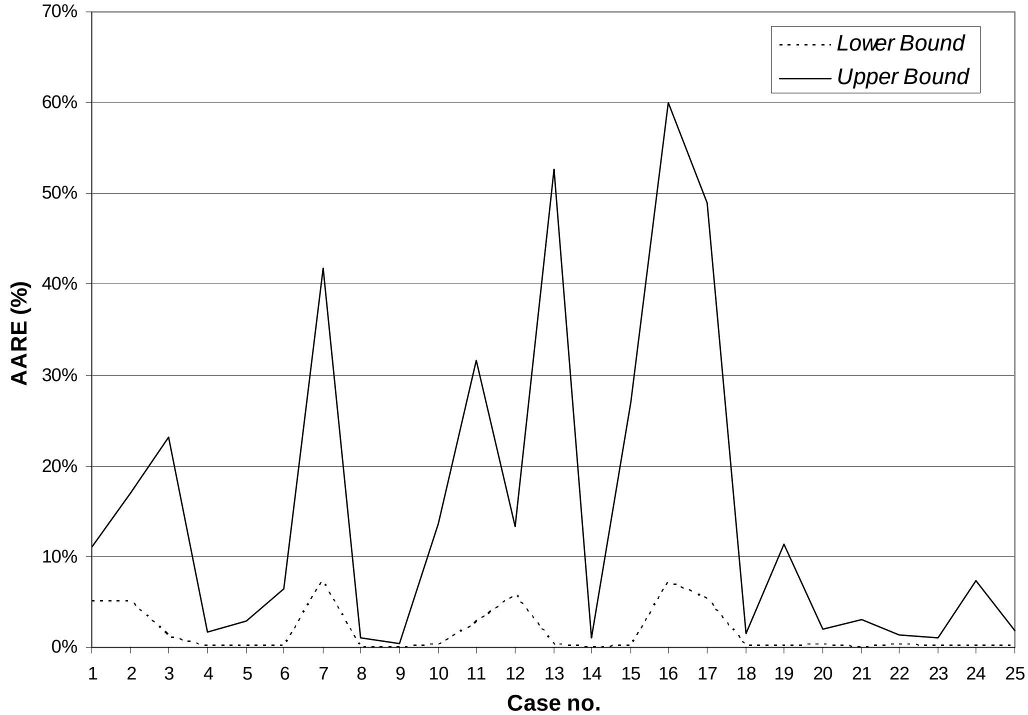


(b) The partially recurrent network (PRN).

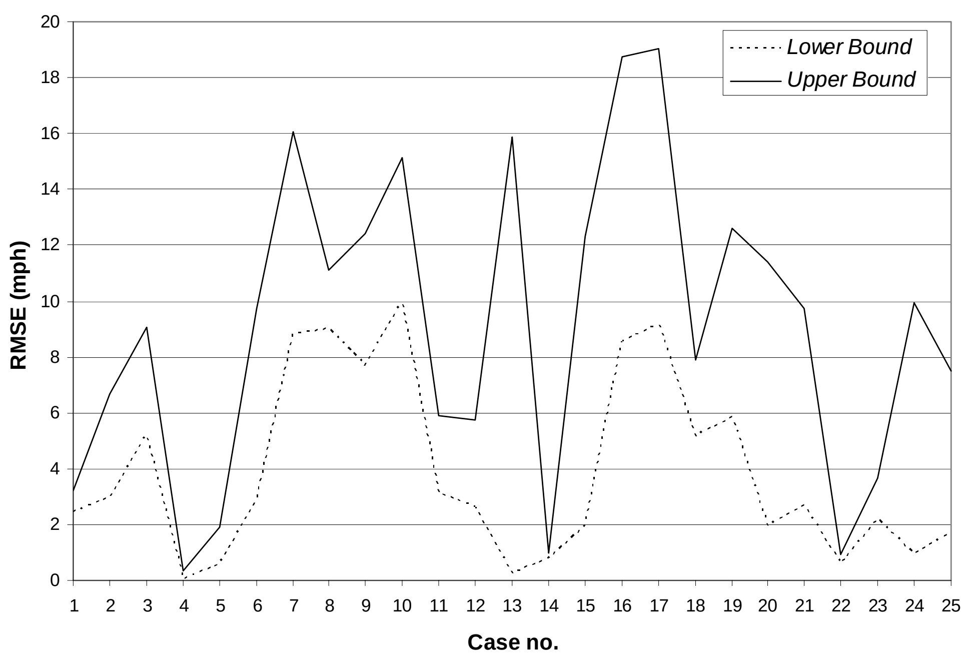


(c) The time-lagged feedforward network (TLFN).

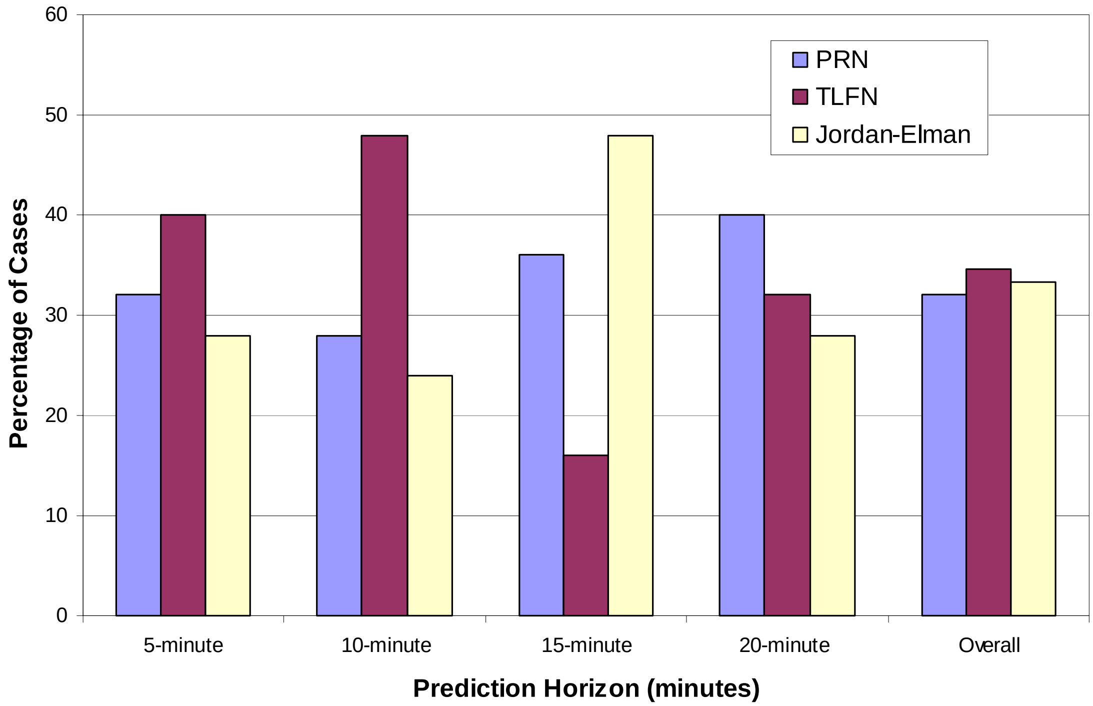
FIGURE 1 Simplified topologies of the applied dynamic neural networks.



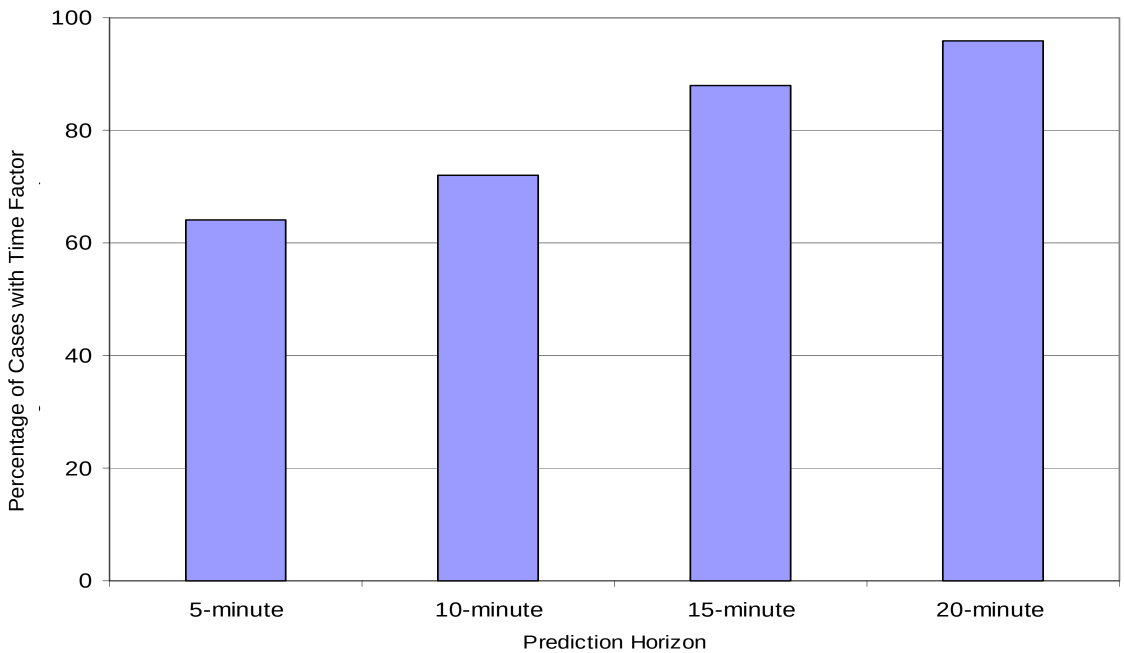
(a) Performance envelopes based on *AARE*.



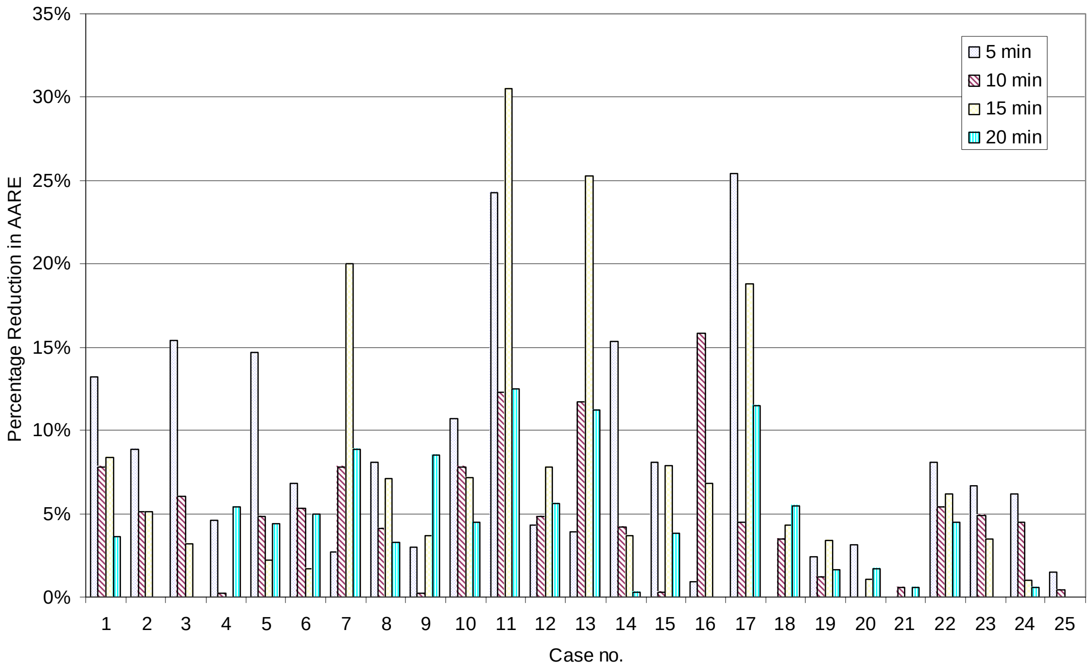
(b) Performance envelopes based on *RMSE*.
FIGURE 2 Performance envelopes associated with optimal settings for each case.



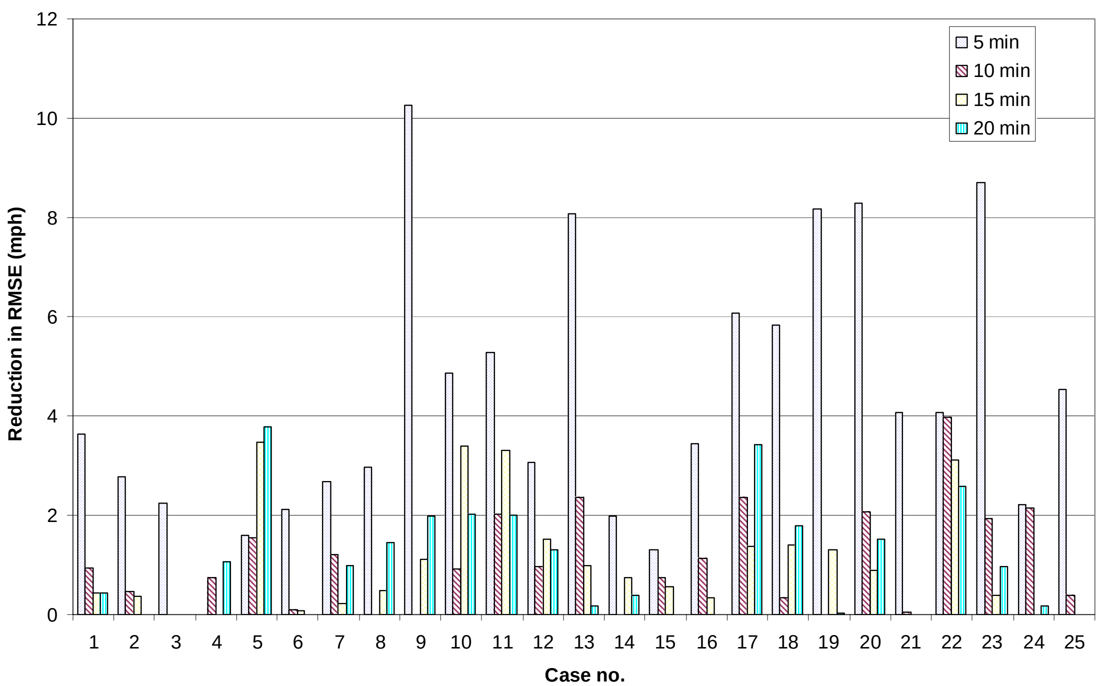
(a) Frequency of optimal networks with prediction horizons based on *AARE*.



(b) Percentage of cases including time factor in optimal settings.
FIGURE 3 Effect of network topology and time factor on optimal settings.

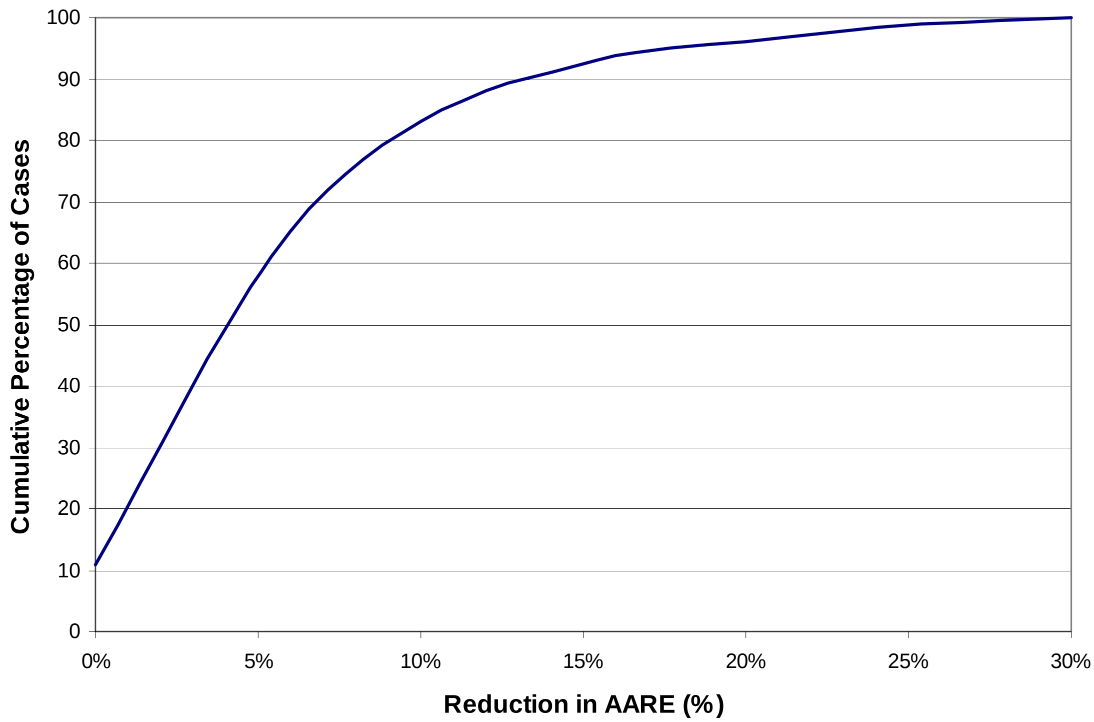


(a) Percentage reductions in *AARE*.

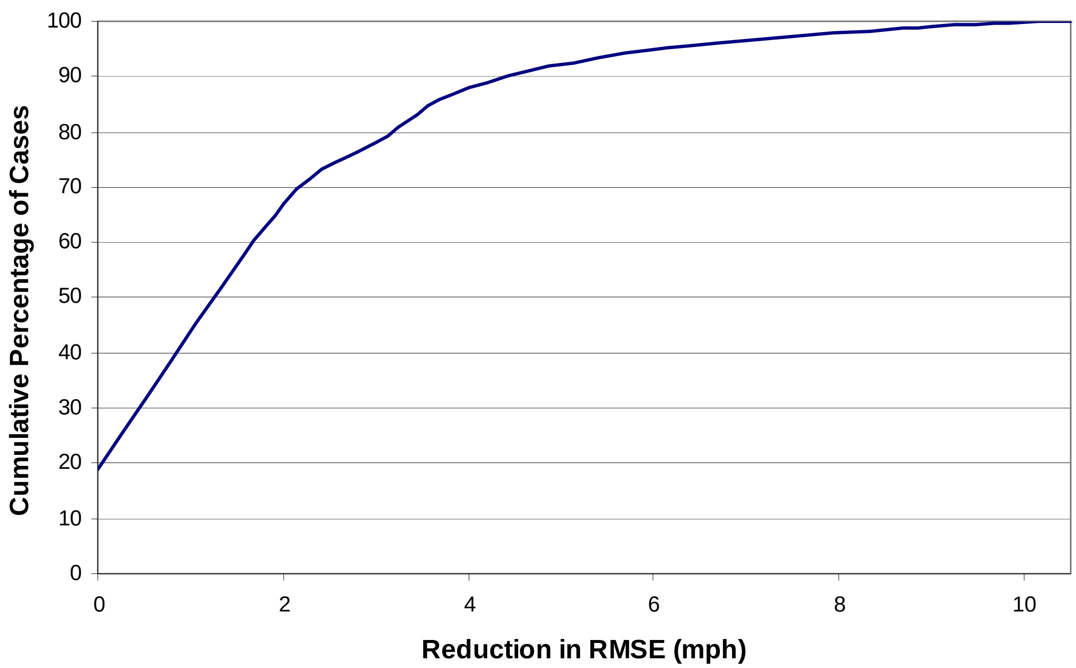


(b) Reduction in *RMSE*.

FIGURE 4 Reductions in *AARE* and *RMSE* achieved by optimization with traffic conditions.



(a) Percentage reductions in *AARE*.



(b) Reduction in *RMSE*.

FIGURE 5 Performance improvements achieved by optimization with traffic conditions.