Short-Term Traffic Flow Prediction: Neural Network Approach

BRIAN L. SMITH AND MICHAEL J. DEMETSKY

Much of the current activity in the area of intelligent vehicle-highway systems (IVHS) focuses on one simple objective: to collect more data. Clearly, improvements in sensor technology and communication systems will allow transportation agencies to more closely monitor the condition of the surface transportation system. However, monitoring alone cannot improve the safety or efficiency of the system. It is imperative that surveillance data be used to manage the system in a proactive rather than a reactive manner. Proactive traffic management will require the ability to predict traffic conditions. Previous predictive modeling approaches can be grouped into three categories: (*a*) historical, data-based algorithms; (*b*) time-series models; and (*c*) simulations. A relatively new mathematical model, the neural network, offers an attractive alternative because neural networks can model undefined, complex nonlinear surfaces. In a comparison of a backpropagation neural network model with the more traditional approaches of an historical, data-based algorithm and a time-series model, the backpropagation model was clearly superior, although all three models did an adequate job of predicting future traffic volumes. The backpropagation model was more responsive to dynamic conditions than the historical, data-based algorithm, and it did not experience the lag and overprediction characteristics of the time-series model. Given these advantages and the backpropagation model's ability to run in a parallel computing environment, it appears that such neural network prediction models hold considerable potential for use in real-time IVHS applications.

An emerging group of technologies and systems known as intelligent vehicle-highway systems (IVHS) have the potential to serve as powerful tools in combating transportation safety and congestion problems by improving the manner in which the nation's extensive existing surface transportation system operates. The backbone of IVHS is the "smart highway"—advanced traffic management systems (ATMS). ATMS collect, utilize, and disseminate real-time data on the status of the surface transportation system. ATMS rely on extensive traffic surveillance systems, thereby providing all other IVHS components with accurate, real-time information. Furthermore, ATMS provide traffic control in both time and space through techniques such as the optimization of traffic signal timing and ramp metering. These techniques have proven benefits, such as freer traffic flows, shorter journey times, and fuel savings (*1*).

The challenge of effectively using real-time data extends to the area of advanced traveler information systems (ATIS), the basic premise of which is to provide travelers with accurate and timely information to allow them to make sound decisions. Clearly, there is a well-defined link between ATMS and ATIS in that both rely on accurate real-time data that describe the status of the transportation network. In addition, it is possible that ATIS will serve as an additional control measure for ATMS by encouraging individual route selection, which would spread demand across all available capacity.

B. L. Smith, Virginia Transportation Research Council, Box 3817 University Station, Charlottesville, Va. 22901. M.J. Demetsky, Department of Civil Engineering, University of Virginia, Charlottesville, Va. 22901.

Although many of the physical components of ATMS and ATIS are still some years away from wide-scale deployment, the preliminary development of software support systems is feasible and should receive immediate attention. At this time, most research has focused on specific applications, such as incident detection and ramp-metering algorithms; very little consideration has been given to developing more general support systems, such as real-time, short-term prediction of traffic conditions. The development of such software support systems will enhance the performance of current systems and serve as a critical step in developing ATMS and ATIS.

REAL-TIME INFORMATION

Real-time data primarily will consist of vehicle counts, vehicle locations, and vehicle speeds. Clearly, vehicle counts alone cannot help a traveler make a routing decision or a traffic manager set a series of signal timings. It is critical that the raw data be processed to derive true information that will support intelligent decision making.

A particularly important function in transforming raw data into information is the prediction of traffic conditions. The current focus on real-time applications is likely to result in reactive control of the transportation system. There is certain to be some lag between the collection of real-time data and the implementation of a control strategy. Therefore, the system will operate under control strategies that are based on past conditions. To control the system in a proactive manner, ATMS must have some sort of predictive capability: "The ability to make and continuously update predictions of traffic flows and link times for several minutes into the future using real-time data is a major requirement for providing dynamic traffic control" (*2*, p. x). Traffic prediction is also an important function for ATIS: "the rationale behind using predictive information (for route guidance) is that travelers' decisions are affected by future traffic conditions expected to be in effect when they reach downstream sections of the network" (*3*). In fact, traveler information services are hampered by the lack of a capability to predict future traffic conditions. For example, changeable message signs are rarely used to provide travel time information because they are inaccurate. Clearly, the success of IVHS is dependent on the development of a traffic prediction capability. Consequently, "special attention should be given to the ability to make short-term traffic predictions with real-time sensor data" (*2*, p. viii).

PREDICTION OF TRAFFIC CONDITIONS: PREVIOUS EFFORTS

"The short-term forecasting of traffic conditions has had an active but somewhat unsatisfying research history" (*1*). Most efforts have focused on traffic prediction for surface signal control systems, such as the Urban Traffic Control System (UTCS). There have been a limited number of freeway traffic prediction applications. The approaches used for traffic prediction are largely dictated by the fact that traffic conditions are time dependent and often follow fairly well-defined patterns. Previous traffic prediction efforts can be classified as historical, data-based algorithms; time-series models; or simulations.

Historical, Data-Based Algorithms

The basic premise behind historical, data-based algorithms is that traffic patterns are cyclical. In other words, a knowledge of typical traffic conditions on Tuesday at 5:30 p.m.will allow one to predict the conditions on any particular Tuesday at 5:30 p.m. AUTOGUIDE, an ATIS demonstration project in London, utilizes the simplest historical, data-based algorithm possible. AUTOGUIDE simply uses a historical traffic data base to predict travel times on the basis of time of day (*4*). Such an algorithm is attractive in that it requires no real-time data.

The UTCS traffic control system utilizes predictions of traffic conditions in an attempt to control signals in a proactive manner. In general, UTCS relies on historical data as support for predictions. A weakness of this method is that UTCS requires an extensive set of historical data; consequently, it is difficult to install the system in a new setting (*6*). An enhancement to the prediction capabilities of the second-generation UTCS (UTCS-2) is that the system uses "current traffic measures to correct for the traffic deviation from the average historical pattern" (*5*, p. 28). Finally, it is interesting to note that the third-generation UTCS (UTCS-3) does not utilize historical data; it predicts conditions on the basis of current traffic measurements only. Although the predictive models of both UTCS-2 and UTCS-3 have serious problems with time lag, UTCS-3 is incapable of performing at a level comparable to UTCS-2 (*5*).

LISB, which is a European traveler information experiment, uses a simple methodology to predict future traffic conditions. LISB uses both historical data and real-time data. A projection ratio of the "historical travel time on a specific link to the current travel time as reported by equipped vehicles" is used to predict travel times on the link for future intervals. A major weakness of this methodology is that it implicitly assumes that the projection ratio will remain constant (*3*, p. 4.)

Time-Series Analysis Techniques

In a traffic management system, detectors are used to measure the system's condition at time t, $x(t)$. These measurements can easily be stored for use in predicting the system's condition at time $t + D$, where D is the prediction interval. As such, the prediction problem boils down to forecasting $x(t + D)$, given $x(t)$, $x(t - D)$, $x(t - 2D)$, and so on. This representation of the prediction problem describes a time series. There have been a number of techniques developed in the field of statistics to model time series. Transportation researchers have applied many of these time series analysis techniques to traffic prediction.

The Box and Jenkins technique is a widely used approach to specifying a variety of time-series models (*7*). It has been shown to yield accurate forecasting results in a number of application areas. The most developed Box and Jenkins technique is the autoregressive integrated moving average (ARIMA) method. ARIMA models require very little computational time for execution, which makes them useful for applications in real-time traffic management. However, ARIMA models have not shown great promise in traffic applications. For example, in attempts to apply ARIMA models to UTCS, it has been found that they "resulted in unsatisfactory goodness of fit and high errors; in certain cases they have not been more accurate than a simple moving average" (*6*, p. 1).

Simulation Models

Simulation models provide predictive capability because they demonstrate how the system is likely to react to varying conditions and control strategies. Given the importance of predictive capabilities in ATMS, it is natural to consider the application of simulation in a real-time environment: "An effective on-line simulation model would enable the ATMS control center to project promptly future traffic patterns considering any previously implemented strategies in a real-time operating environment" (*8*, p. 13). Unfortunately, at this time, the real-time application of traffic simulation is not feasible because existing model/algorithm constructs cannot support real-time applications (*9*). A need exists for new approaches to the simulation of transportation systems.

An exciting development that may support real-time simulation is parallel computing. Parallel computing, or processing, is defined as "an efficient form of information processing which emphasizes the exploitation of concurrent events" (*8*, p. 14). In other words, a parallel computer has multiple processors that work simultaneously (in parallel). Of course, this parallelization allows for tremendous increases in the speed of execution. However, the programming of a parallel computer is extremely challenging because of the need to synchronize certain procedures. A recent research effort attempted to develop an architecture for a parallel traffic simulation application. Although it shows promising results, the effort is still in preliminary stages (*8*). The wide-scale deployment of parallel traffic simulation appears to be far from realization.

Assessment

Although a number of approaches to the prediction problem have been described in this section, the fact remains that very few traffic control systems include any proven forecasting capability. There is, thus, a need to develop efficient and accurate real-time traffic prediction models. To be effective, such a model must be able to recognize patterns, use historical or time-series data or both, and represent complex, nonlinear relationships. The next section will introduce neural networks, which have shown considerable promise in these areas.

NEURAL NETWORKS

Over the past several years, both in research and in practical applications, neural networks have proven to be a very powerful method of mathematical modeling. In particular, neural networks are well suited for pattern recognition, offer efficient execution, and model nonlinear relationships effectively. Clearly, neural networks are well worth exploring as a tool for the short-term prediction of traffic.

Neural networks may be defined as "an information processing technology inspired by studies of the brain and nervous system" (*10*, p. 30). This inspiration obviously led to the use of the word neural. However, neural networks in no way attempt to produce biological clones; rather, they are simply models with a rigorous mathematical basis (*11*). Although neural networks are typically associated with the field of artificial intelligence, they function as a sophisticated form of regression. The use of neural networks has been proven successful in a number of applications for the following reasons (*12*):

1. Neural networks can perform highly nonlinear mappings between input and output spaces;
2. The parallel structure of neural networks lends them to implementation on parallel computers, which offers the potential for extremely fast processing; and
3. The neural networks approach is nonparametric; therefore, one need not make any assumptions about the functional form of the underlying distribution of the data.

These characteristics have attracted the attention of researchers from a number of disciplines to problems such as classification, forecasting, process control, and signal processing (*10*).

Neural Networks Basics

To gain a fuller understanding of the underlying mechanics of neural networks, it is instructive to consider the following definition: "a neural network is a computing system made up of a number of simple, highly interconnected processing elements" (*13*, p. 71). The basic structure of a neural network is illustrated in Figure 1. A description of the elements follows.

- *Processing element:* The processing element is the basic building block of a neural network. Processing elements on the input layer simply pass the input value to the adjoining connection weights. Processing elements on the hidden and output layers sum their inputs and compute an output according to a transfer function.
- *Connection weight:* Connection weight serves to join processing elements within the neural network. The connections are of varying strength, which *weight* the value that the connection "transports" between processing elements. In effect, the connection weights may be compared with coefficients in a regression model.
- *Layers*: Layers are sets of processing elements in which all processing elements in adjacent sets are connected. A neural network generally has an input layer, a hidden layer (in which all connection weights are internal to the network), and an output layer.
- *Bias:* The bias is a constant input to each processing element. The input is defined solely by the connection weight between the bias input (which outputs a constant value of 1.0) and the processing element.
- *Transfer function:* The transfer function is an operator, usually nonlinear, that is applied to the summed inputs of a processing element to produce the output value.

In a basic feed-forward neural network, raw input data are presented to processing elements in the input layer. The input values are then weighted and passed to the hidden layer through the connections. Processing elements in the hidden layer sum and process their inputs and then pass the output to the output layer. Processing elements in the output layer sum and process their weighted inputs to produce the network output. The following equation represents this process in a functional form:

$$Y = \Phi[W_2\Phi(W_1 X + \Theta_1) + \Theta_2]$$

where

Φ = transfer function,
W_1 = array of connection weights for layer *I*,
X = input values, and
Θ_1 = array of bias values for layer *I*.

The description presents a neural network as a graphical mathematical modeling technique. In a neural network, the fundamental variables are the set of connection weights. The definition of the connection weights, much like the definition of coefficients in a re-

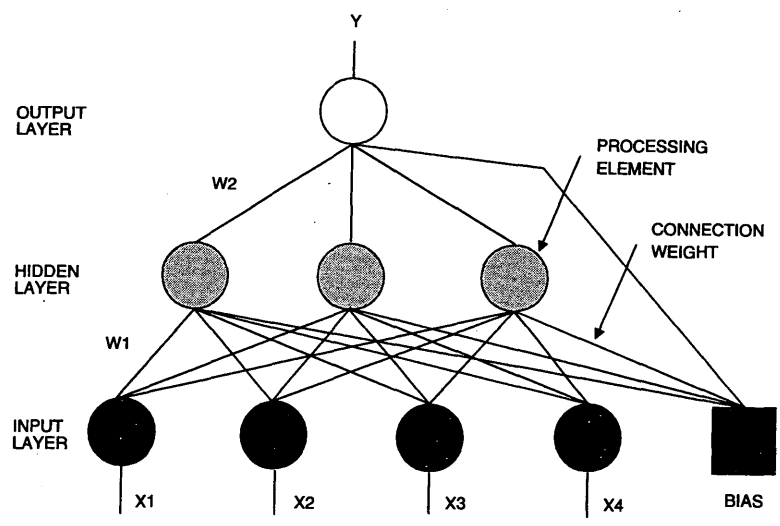

FIGURE 1 Neural network.

gression model, allows the model to be "fitted" to the data. In regression, the method of least squares is utilized to define coefficient values. In the field of neural networks, the process of defining connection weights is generally referred to as "learning." The learning method utilized by a network defines the neural network paradigm. The backpropagation paradigm was chosen for this application. The strengths of this paradigm have led to the conclusion that "almost universally backpropagation has become the standard network paradigm for modeling, forecasting, and classification" (*10*, p. 29). A complete description of this paradigm, can be found in a previous publication (*14*).

APPLICATION OF NEURAL NETWORKS TO TRAFFIC FLOW PREDICTION

The characteristics of neural networks make them excellent candidates for application to the traffic flow prediction problem. In this section, recent studies examining neural network traffic flow prediction models are described.

Gilmore et al. (*15*) applied a backpropagation neural network to predict congestion on surface streets. On the basis of current and past volumes on the surface system, the network predicts traffic flow over the next half hour, in 5-min intervals. The development of this network is based strictly on data obtained from a simulation model (*15*). Although the effort illustrates the potential of neural networks, it is difficult to generalize the results. In effect, a mathematical model (the neural network) was developed to predict the behavior of another mathematical model (the simulation model).

A similar study illustrates the potential of neural networks for the prediction of freeway traffic volume. Zhang et al. describe a backpropagation network to emulate a macroscopic traffic flow model. They chose such an approach on the basis of the fact that "traffic flow on freeways is a complex process that is often described by a set of highly non-linear dynamic equations" (*12*, p. 2). After training and testing the network on data developed by the macroscopic model, it was concluded that the neural network captured the traffic dynamics of the macroscopic model. Clearly, this is an encouraging conclusion. However, again one will note that a mathematical model (the neural network) was developed to predict the behavior of another mathematical model (the macroscopic model).

Another important research effort exploring the applicability of neural networks to the traffic flow prediction problem was conducted at the University of Leeds. A short-term traffic forecasting model was developed for a surface system using a backpropagation neural network. The model simply relies on current network flow levels as well as flow levels 5 and 10 min in the past. Data from a SCOOT traffic control system in England were utilized. Although the neural network model performed well, it was outperformed by a traditional Box and Jenkins time-series model (*16*). Although this result may seem disappointing, this effort is encouraging because it describes a viable neural network application in a real-world situation.

Clearly, these efforts illustrate the potential of neural networks. However, the need remains to demonstrate the effectiveness of a neural network prediction model using data from an actual freeway facility. Data available from a traffic management system often leave much to be desired, particularly when compared with simulation data. The challenge of maintaining loop detectors, noise in communication systems, and other system problems results in data streams that often look much different from those available from a simulation model. Clearly, it is important to examine the effectiveness of a neural network prediction model using data collected in an operational traffic management system.

CASE STUDY: FREEWAY VOLUME PREDICTION

The purpose of the case study was to develop short-term volume prediction capability at a site on the Capital Beltway. The site selected for this study is on the inner loop of the Beltway near the Telegraph Road interchange in Alexandria, Virginia. At this location, the Beltway is a four-lane freeway, carrying a high volume of local and interstate traffic. In addition, the section is affected by one the region's most notorious bottlenecks, the Woodrow Wilson Bridge.

The Northern Virginia Traffic Management System (TMS) monitors this site with a video camera and full loop detector stations in each of the four lanes. The stations provide the following data continuously to the TMS:

- Volume (vehicles/hour),
- Average speed (miles/hour), and
- Average occupancy (percent).

In addition, Virginia Department of Transportation operates an automatic weather monitoring system (SCAN) to collect pertinent weather data. The SCAN station in Rosslyn, Virginia, roughly 8 mi from the freeway site, is utilized to access the following data:

- Air temperature and
- Pavement condition (wet/dry).

To develop predictive models, a traffic and weather data base was created. The data are stored in 15-min intervals from June 3, 1993, through August 11, 1993, resulting in 3,000 records. This set of data was divided into training and test sets for model development. The training set consisted of 2,550 records, and the test set consisted of 450. Each set of data consisted of roughly a uniform distribution of volume levels. Finally, a third set of data was collected after August 11 to serve as a validation data base.

Models Developed

Three models were developed to predict the link volume at the Telegraph Road site on the Beltway. A 15-min prediction interval was utilized. These models were used to compare traditional approaches to short-term predictions of traffic conditions with a neural network model. A brief description of each model follows.

Historical Average

This model is a simple historical, data-based algorithm. The model developed in this case study simply used the historical average volume, which was calculated using the training data set, as the basis for predicting future volume. In other words, to predict volume on Monday, September 10, at 3:00 p.m., the historical average volume on Mondays at 3:00 p.m. was used.

ARIMA

ARIMA models are among the most powerful and advanced statistical time-series techniques. On the basis of an analysis of autocorrelations and partial autocorrelations of the volume time-series, an ARIMA (2,1,0) model was selected for this application. Such a model describes a second-order autoregressive process that is integrated with no moving average. Model coefficients were based on an analysis of the training data set.

Backpropagation Model

The backpropagation neural network was developed using the following variables as inputs: volume (t), volume (t − 15 min), historical volume (t), historical volume (t + 15 min), average speed (t), and wet pavement (t) (a binary variable). It was trained using a learning rate of 0.3 and a momentum of 0.4. The network architecture consisted of one hidden layer of 10 processing elements.

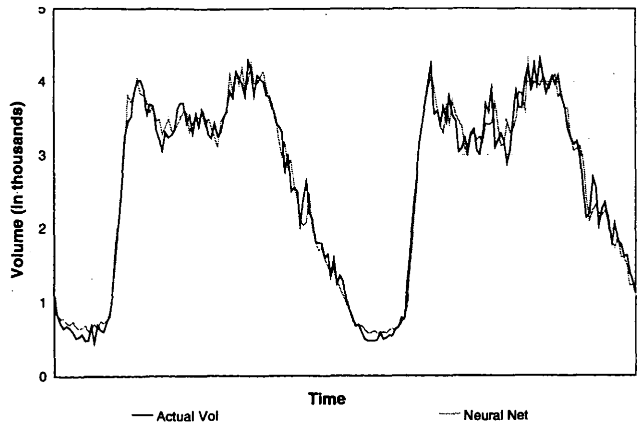

FIGURE 2 Backpropagation mode performance: validation data set.

Performance Analysis

To compare the models, the third, independent validation data set was used. This data set was gathered on two consecutive days (a Monday and Tuesday) in September 1993. In general, all three of the models did an excellent job of predicting future volumes in the short term. In fact, a comparison of error measures in Table 1 reveals that the historical average and backpropagation models displayed comparable error measures, whereas the ARIMA model was less accurate. Figure 2 illustrates graphically the performance of the neural network model on the validation data.

Table 2 displays the average estimate percentage of error for all cases and for cases in peak conditions (defined as any period in which the volume exceeds 3,000 vehicles per hour). Interestingly, the historical data model outperforms the neural network when considering all periods, whereas the neural network model demonstrates better accuracy during peak periods. This indicates that the historical data model can be expected to consistently produce estimates within 5 to 6 percent error levels. On the other hand, one would expect better performance from a neural network during peak periods. This expectation is likely because of the neural network's capability to accurately model the complex characteristics of traffic flow during peak conditions. Clearly, performance during peak periods is of the utmost importance to traffic management and traveler information applications. Therefore, a peak period of the validation data in detail will be examined. The period considered is the p.m. peak, from noon to 7:00 p.m. on Monday.

Peak Period Analysis

Figure 3 illustrates the performance of the historical data model. The model predicts consistently low values for this period. For whatever reason, higher-than-"normal" volumes occurred on this Monday, volumes that the historical model had no capability to predict. This illustrates the significant weakness of such a model; it cannot react to external or abnormal factors that may affect the volume level.

TABLE 1 Error of Prediction Models

Model	Root Mean Square Error	Average Absolute Error
Historical Average	2730	146
ARIMA	3490	195
Backpropagation	2620	144

TABLE 2 Average Percent Forecast Error

Model	Average % Forecast Error All Cases	Average % Forecast Error Volume > 3,000 veh/hr
Historical Average	6.4%	5.0%
ARIMA	9.0%	10.8%
Backpropagation	7.5%	4.3%

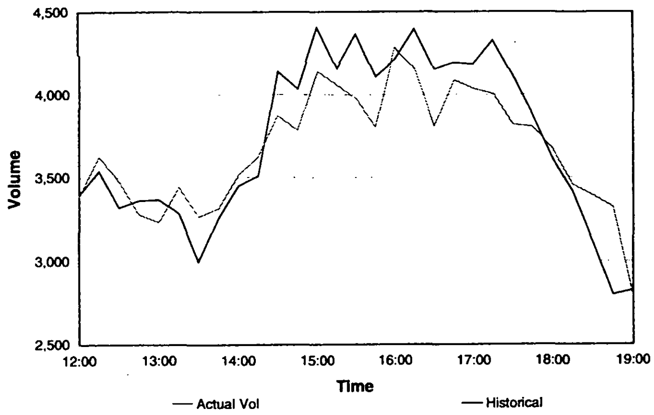

FIGURE 3 Historical data model: p.m. peak.

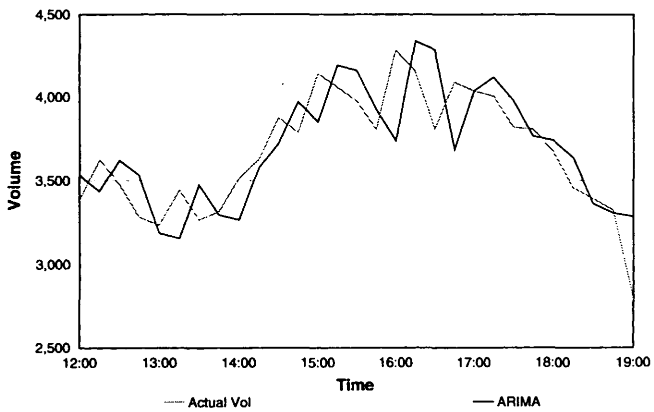

FIGURE 4 ARIMA model: p.m. peak.

Figure 4 illustrates the performance of the ARIMA model. It is clear that the predictions of the ARIMA model tend to lag roughly one interval (15 min) In addition, the ARIMA model tended to overpredict values. This is evident in that volume peaks for the ARIMA model are consistently more extreme than those of the actual volume. The lagging and overpredicting are not surprising given the fact that the ARIMA model uses only time-series data.

Finally, Figure 5 displays the performance of the backpropagation model during the p.m. peak period. This model does an excellent job of predicting volume levels without the lag or overprediction problems of the ARIMA model. This example shows that although all three models have roughly comparable overall error, the backpropagation model clearly does the best job of modeling the underlying relationship between the state of the system and future traffic volume during peak conditions.

CONCLUSION

IVHS technology allows for vastly improved data collection and data communication capability. However, a very real risk is that the world will become data rich and information poor. Thus, a critical effort in the development of IVHS is to create real-time decision support software that will rely on advanced technology, such as expert systems and models. A critical element of such support software that has been identified in this paper is a short-term traffic condition prediction model.

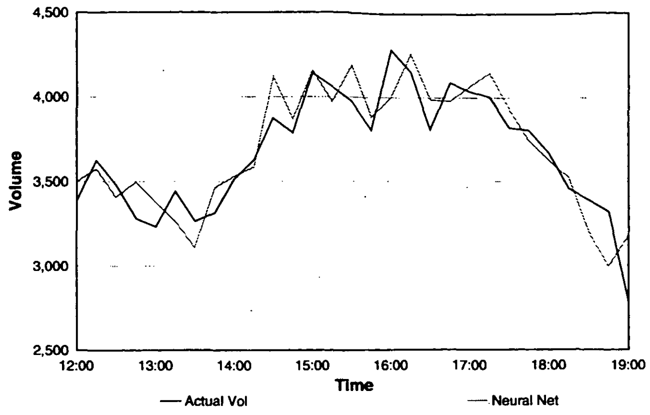

FIGURE 5 Backpropagation model: p.m. peak.

This paper has demonstrated the potential of neural networks to accurately predict short-term traffic conditions in real time. A neural network developed with data from an operational traffic management system performed comparably to traditional prediction approaches when tested with an independent set of validation data. The neural network model, however, outperformed other models during peak conditions, demonstrating its ability to model complex traffic characteristics. On the basis of these promising results, research is continuing to further refine neural network models for ultimate implementation in traffic management systems.

REFERENCES

1. Davis, G. A., and N. L. Nihan. Nonparametric Regression and Short-Term Freeway Traffic Forecasting. *Journal of Transportation Engineering,* 1991, pp. 178–188.
2. Cheslow, M., S. G. Hatcher, and V. M. Patel. *An Initial Evaluation of Alternative Intelligent Vehicle Highway Systems Architectures.* MITRE Report 92W0000063. MITRE Corp., McLean, Va., 1992.
3. Kaysi, I., M. Ben-Akiva, and H. Koutsopoulos. An Integrated Approach to Vehicle Routing and Congestion Prediction for Real-Time Driver Guidance. In *Transportation Research Record* 1408, TRB, National Research Council, Washington, D.C., 1993.
4. Jeffery, D. J., K. Russam, and D. I. Robertson. Electronic Route Guidance by AUTOGUIDE: The Research Background. *Traffic Engineering and Control,* 1987, pp. 525–529.
5. Stephanedes, Y. J., P. G. Michalopoulos, and R. A. Plum. Improved Estimation of Traffic Flow for Real-Time Control. In *Transportation Research Record 795,* TRB, National Research Council, Washington, D.C., 1981, pp. 28–39.
6. Okutani, I., and Y. J. Stephanedes. Dynamic Prediction of Traffic Volume Through Kalman Filtering Theory. *Transportation Research Part B,* 1984, pp. 1–11.
7. Nihan, N. L., and K. O. Holmesland. Use of the Box and Jenkins Time Series Technique in Traffic Forecasting. *Transportation,* 1980, pp. 125–143.
8. Junchaya, T., G. Chang, and A. Santiago. Advanced Traffic Management System: Real-Time Network Traffic Simulation Methodology with a Massive Parallel Computing Architecture. In *Transportation Research Record 1358;* TRB, National Research Council, Washington, D.C., 1992. pp. 13–21.
9. Rathi, A. K., and A. J. Santiageo, *Simulation and Intelligent Vehicle Highway Systems.* FHWA Paper CONF-9209176-1, 1992.
10. Klimasauskas, C. C. Applying Neural Networks—Parts 1–6. *PC AL,* 1991.

11. Nelson and Illingworth. *A Practical Guide to Neural Nets.* Addison-Wesley, Inc., Reading, Mass., 1991.
12. Zhang, H., S. G. Ritchie, and Z. Lo, Macroscopic Modeling of Freeway Traffic Using an Artificial Neural Network. Presented at 72nd Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 1993.
13. Faghri, A., and J. Hua. Evaluation of Artificial Neural Network Applications in Transportation Engineering. In *Transportation Research Record 1358;* TRB, National Research Council, Washington. D.C., 1992, pp. 71–80.
14. Rumelhart, D. E., and J. L. McClelland. *Parallel Distributed Processing,* Vol. 1. The MIT Press, Cambridge, Mass, 1986.
15. Gilmore, J. F., K. J. Elibiary, and M. Abe. Traffic Management Applications of Neuro Networks Systems. *Working Notes,* AAAI-93 Workshop on AI in Intelligent Vehicle Highway Systems, 1993, pp. 85–95.
16. Dougherty, M. S., H. R. Kirby, and R. D. Boyle. The Use of Neural Networks to Recognize and Predict Traffic Congestion. *Traffic Engineering and Control,* 1993, pp. 311–314.

The opinions, findings, and conclusions expressed in this report are those of the authors and not necessarily those of the sponsoring agencies.

Publication of this paper sponsored by Committee on Artificial Intelligence.