The Study on Establishing the Pavement Performance Prediction by Artificial Neural Networks

S. H. Lin
Department of Civil Engineering, De-Lin Institute of Technology, Taipei, Taiwan, ROC.

H. Fang
Department of Information Management, Shu-Zen College of Medicine and Management, Kaoshong, Taiwan, ROC.

ABSTRACT: Measuring the condition of current pavement is accomplished by collecting field distress data and synthesizing data to identify appropriate alternatives for rehabilitation or reconstruction. Many agencies have pavement management systems (PMS) to assist with data collection, evaluation, and decision-making during this process. The present serviceability index (PSI) is a common tool for quantifying information concerning the serviceability of the pavement. A primary factor used in establishing the PSI is the roughness of the surface profile. The PSI can also include standard distress criteria such as rutting, fatigue cracking, and thermal cracking. However, the actual causes and conditions of pavement distress are very complex. The statistical modeling can only consider no more than a few of the parameters, in a simplified manner, and in some cases various transformations of the original data. Because of the statistical nature of models this does not mean that cracking and rutting are not important, since they will react on the roughness of the surface profile. The artificial neural networks (ANNs) offer a number of advantages over the traditional statistical methods, caused by their generalization, massive parallelism and ability to offer real time solutions. In this paper, real pavement condition and the subjective present serviceability rating (PSR) in Taiwan are used to develop a generic intelligent pavement performance prediction using ANNs. In contrast to statistical analysis, it is concluded that the good predictive results can be obtained from the pavement performance model established by neural network.

KEY WORDS: Pavement management systems, present serviceability index, pavement distress, artificial neural networks.

1 INTRODUCTION

Pavement Management System (PMS) has been developed and applied all over the world for many years. Many pavement engineers and planners believed that a systems approach could provide more cost-effective utilization of limited resources. In Taiwan, several PMSs based on different road categories, pavement types, and applied levels were also developed by governments and engineering agencies cooperating with the academia since 1983. Pavement performance models play a pivotal role in the PMS decision-making process, especially with PMS that optimizes among several maintenance options for each subsection of the roadway.

Therefore, without performance models, deferring maintenance would have no technical or economic consequence (Jansen and Schmidt 1994).

The most common approaches for performance prediction are regression analyses and Markov chains. However, the pavement deterioration process is so complex that it is tedious to find an appropriate functional form to model it as used in traditional statistical modeling. The statistical modeling can only consider no more than a few of the parameters, in a simplified manner, and in some cases various transformations of the original data (Attoh-Okine 1999). Hence a new approach, which can be categorized as “soft computing techniques”, is taking the territory from its traditional counterpart in terms of modeling complex processes. Soft computing includes three principal constituents (Flintsch 2003): artificial neural networks (ANNs), fuzzy mathematical programming, and evolutionary computing (including genetic algorithms), and it has been used in infrastructure management with various degrees of success.

Artificial neural networks offer a number of advantages over the traditional statistical methods, caused by their generalization, massive parallelism and ability to offer real time solutions. Literature review shows that ANNs and other soft computing techniques are increasingly used instead of the traditional methods. Neural models have been used for predicting roughness (Attoh-Okine 1994, Huang and Moore 1997, La Torre et al. 1998.), cracking development (Lou et al. 2001, Owusu-Ababia 1998), pavement condition (Attoh-Okine 1999, Yang et al. 2003) and the Present Serviceability Rating (Shekharan 1998). In this study, the ANNs methodology was utilized to establish an efficient, rational and practical pavement performance model. All data types for 25 constructed pavements in Taiwan were collected from the database of “Research and Development (R&D) of Monitoring Pavement Performance Project”. The Institute of Transportation (IOT) has been conducting this project for four years. For the ANNs model and the Regressive statistical model, this paper compares measured and predicted values of the subjective Present Serviceability Rating (PSR). The purpose of these comparisons is to evaluate the performance of the ANNs model, and identify the model that more accurately predict pavement quality.

2 METHODOLOGY

Neural network technology mimics the brain's own problem solving process. Just as humans apply knowledge gained from past experience to new problems or situations, a neural network takes previously solved examples to build a system of ‘neurons’ that produce valid answers from noisy data. The architecture of a neural network is characterized by a large number of simple neuron-like processing units interconnected by a large number of connections. The pattern of connectivity among the processing units and the strength of the connections encode the knowledge of a network. The main advantages of neural networks are their learning capabilities and their distributed architecture that allows for highly parallel implementation. In order to construct a neural network for solving a particular problem, three key components need to be determined first. They are (1) Architecture, (2) Neuron activation function, and (3) Learning method.

2.1 Architecture

The ANNs architecture design process includes determining input and output variables, number of hidden layers, and number of hidden neurons in each hidden layer. After finishing the ANN architecture design, the ANN architecture needs to be trained, tested and validated.

In this study, inputs for the developed model take into considerations various pavement distress and surface conditions. The eight inputs used for the analysis are: roughness, expressed in IRI (International Roughness Index, m/km) and rutting in wheel path (RD, mm). Other rated surface distresses are fatigue cracking (FC), longitudinal cracking (LC), raveling (R), shoving (S), patching (P) and block cracking (BC). Roughness and rutting measurements were collected using the Automatic Road Analyzer (ARAN). For other rated surface distresses, the severity extent was estimated subjectively by a panel of 12, mostly engineers and experts. The output of the subjective rating is called the PSR. It is consistent with the well-known, dimensionless, 0-to-5 scale used for AASHTO Present Serviceability Rating (PSR). The subjective estimates were obtained by a panel of 23, riding over selected pavements that were judged to represent a wide range of conditions.

In ANNs, the number of hidden neurons determines how well a problem can be learned. If you use too many, the network will tend to try to memorize the problem, and thus not generalize well later. If you use too few, the network will be unable to learn sufficiently from the training data set. However, a rule of thumb is suggested that the default number of hidden neurons (N) is computed with the following formula (Ward 2000):

$$\mathrm{N} = \frac{\text{Inputs} + \text{Outputs}}{2} + \sqrt{\text{the number of patterns in the training data set}} \tag{1}$$

Through experience and literature reviews (Funahashi 1989, Hornik et al. 1989, Lin 2001), we have found that any continuous function can be approximated with an arbitrary accuracy using the three-layered network if we use enough hidden neurons. When we use more than one hidden layer (the layers between the input and output layers) training time may be increased by as much as an order of magnitude.

2.2 Neuron Activation Function

Each neuron in an ANN is an independent processing element (PE). It combines the inputs and produces an output in accordance with an activation function. The output of one neuron is connected to the input paths of other processing elements through connecting weights. The hidden layers produce outputs based upon the sum of weighted values passed to them. The activation function, also called the squashing function, maps this sum into the output value, which is then fired on to the next layer. The output is calculated by the equation:

$$\mathrm{O}_j = f\left(\sum_{i=1}^{n} a_i w_i\right) \tag{2}$$

Where, a_i is the activity level of the ith PE or input; w_{ij} represents the connection weight associated with ith input; Oj is output of jth neuron, and f is the transfer function.

Although the sigmoid function is the most popular, there are other functions which may be used. Five typical transfer functions are generally used as neuron activation functions: sigmoid, linear, hyperbolic tangent, sine, and Gaussian.

The sigmoid function is a bound, monotonic, non-decreasing function that provides graded, nonlinear response within a specified range, 0 to 1. Usually we use this function when the outputs are categories. The linear function produces a linearly modulated output from the input. It is useful for problems where the output is a continuous variable, as opposed to several outputs which represent categories. The linear activation function is often ineffective if there are a large number of connections coming to the output layer because the total weight sum generated will be high. For the hyperbolic tangent and sine function, it is sometimes better for continuous valued outputs, however, especially if the linear function is used on the output layer. The Gaussian activation function is useful in a small set of problems, and it tends to bring out meaningful characteristics in the middle-range of the data.

2.3 Learning Method

An important characteristic of neural networks is its ability to 'learn'. The network 'learns' by adjusting the interconnection weights between layers. The answers that the network is producing are repeatedly compared with the correct answers, and each time the connecting weights are adjusted slightly in the direction of the correct answers. The Back-propagation (BP) supervised learning method, which is used in this research. It is the most widely used learning method and it presents a clear mathematical concept and ease of programming. "NeuroShell 2" is a software program employed in this research.

In the BP, each presentation of data sets and the input values are compared with desired output values and adaptive weights within the network and are incrementally adjusted to minimize the output error. The error function is expressed as:

$$E_t = \frac{1}{2}\sum\sum[Y_j(c) - D_j(c)]^2 \tag{3}$$

Where, E_t = square of the output error for all the patterns in the data sample; Y = actual output; j = output neuron index; c = input sample case; and D = Desired output.

The overall objective is to minimize the error function by adjusting the interconnection weights. The initial weights are chosen randomly in a range of values. Using a gradient descend method, the adjusted weights (w_{jk}) for the connections between hidden layer and output layer can be expressed as equation (4):

$$w_{jk} = w_{jk}^{current} - \eta\frac{\partial E_t}{\partial w_{jk}} + \alpha(w_{jk}^{current} - w_{jk}^{previous}) \tag{4}$$

Where, η = network learning rate and α = the momentum term. Selection of η and α can sometimes be difficult (Sorsa and Koivo 1993). Similarly, weight adjustment (w_{ij}) for the connections between input layer and hidden layer can be written as equation (5):

$$w_{ij} = w_{ij}^{current} - \eta\frac{\partial E_t}{\partial w_{ij}} + \alpha(w_{ij}^{current} - w_{ij}^{previous}) \tag{5}$$

Given the architecture of ANNs, the weights of links among the neurons are resolved through the training process as defined by Equations (4) and (5). The total error, as defined in Equation (3), for the training data set will continue to get smaller forever, or at least gets to the constant. To avoid over-training the model, the trained network is exposed to the testing dataset at intervals, and the testing error is calculated. The training process should be stopped in this study when the number of events since the minimum error for the test data set reaches 20000 events. The last and also the most critical step is to verify the model using a validation data set. If the validation error is still acceptably small, the ANN model is considered as a reasonable model. In this study, three data sets were randomly chosen from 215 inputs. About 129 patterns (60%) were used for training, 43 patterns (20%) for testing and about 43 patterns (20%) for validation.

3 PERFORMANCE EVALUATION OF THE ANN MODELS

Practical applications of ANNs require some way of choosing the number of neuron, the number of layers, activation function, and other specifications such as the learning rate and momentum term. The correct selection of these parameters separate the signal from the noise and avoid over-fitting of the signal. As mentioned above, the three layer BP network forms the basis of the analysis. The linear transfer function was applied on the input layer and output layer.

3.1 Effect of Activation Function and the Number of Hidden Neurons

Various neurons with different activation functions were investigated to identify most 'optimum architecture' for the analysis. Depending on the characteristics of the pavement performance model in this study, we selected the sigmoid, tanh, and Gaussian activation functions to evaluate. As defined in Equation (1), the number of hidden neurons is 16 ($0.5\times(8+1)+\sqrt{129}=16$). In addition, 8 and 32 hidden neurons were also selected for investigation. A learning rate of 0.1 and a momentum term of 0.1 were used as the default values for these architectures (8-8-1, 8-16-1, and 8-32-1).

The coefficient of multiple determination (R^2) is a statistical indicator. It compares the accuracy of the model to the accuracy of a trivial benchmark model wherein the prediction is just the mean of all of the samples. A perfect prediction model would result in the R^2 near 1. The corresponding R^2 were calculated by the validation data set (N=43). Figure 1 shows there is higher R^2 for hidden neurons with the Gaussian function. However, unlike other activation functions, the performance of models using the Gaussian function is more unstable for different neurons. A 'stable' performance presents that the model is more difficult to occur over-learning.

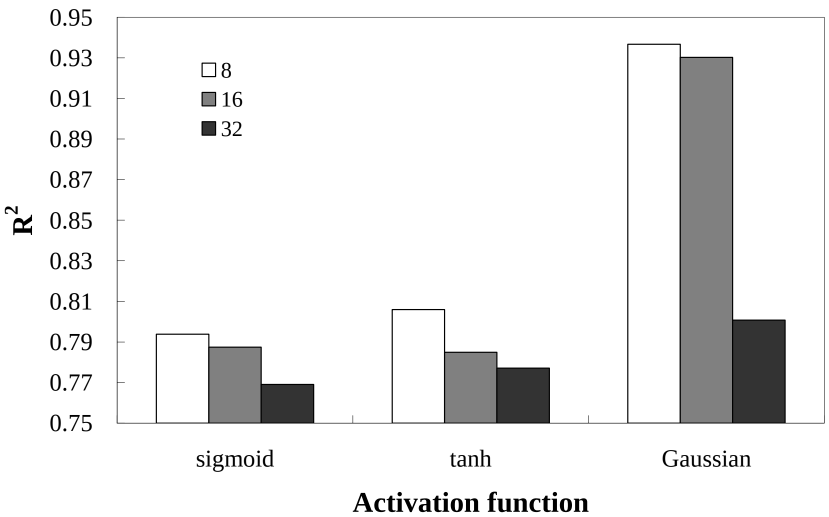

Figure 1: Effect of various neurons with different activation functions.

As shown in Figure 1, the ANN models using 8 hidden neurons have higher R^2, and there is little difference in R^2 between 8 and 16 hidden neurons. The R^2 of models with 32 hidden neurons is lowest for different activation function. To consider goodness of fit and stability of the ANN models, we used two hidden groups with different activation functions in a hidden layer, and the hidden neurons were divided evenly into two groups. Again, 8 and 16 hidden neurons were selected for investigation. As an example, the schematic architecture of pavement performance model (8-4/4-1) is shown in Figure 2, where hidden neurons with different color represent different activation functions. Similarly, a learning rate of 0.1 and a momentum term of 0.1 were used for these architectures (8-4/4-1 and 8-8/8-1).

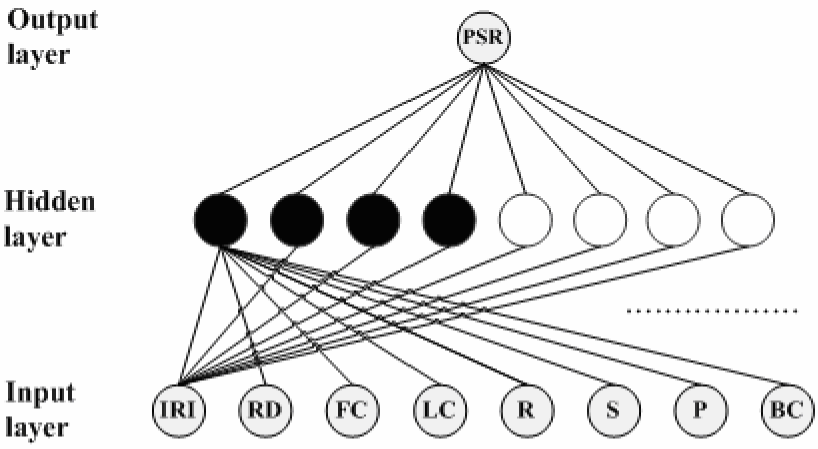

Figure 2: The schematic architecture of pavement performance model (8-4/4-1).

It was found that the ANN models using 16 hidden neurons (8/8) have more stable and better performance in Figure 3. In contrast to the results of Figure 1, the R^2 of models are raised by using two hidden groups with different activation functions in a hidden layer, especially as sigmoid and tanh function.

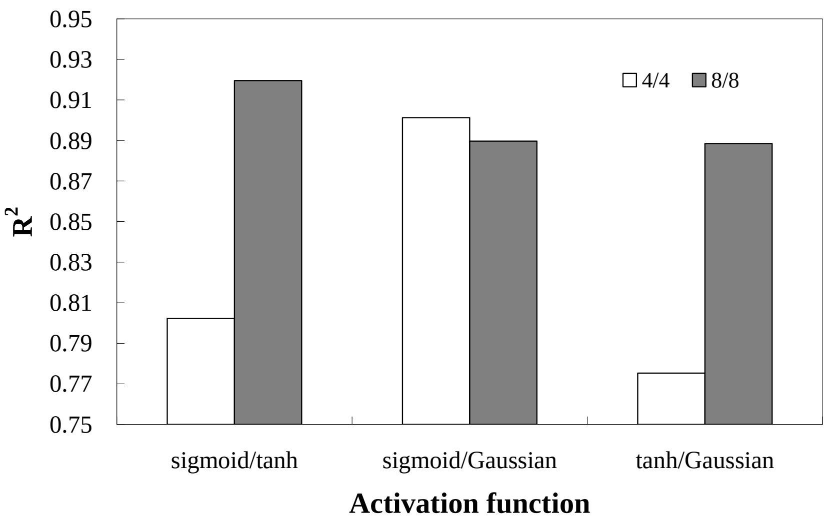

Figure 3: Effect of two hidden groups with different activation functions.

3.2 Effect of Learning Rate and Momentum

The learning rate is a parameter that determines the size of the weights adjustment each time the weights are changed during training. The larger the learning rate, the larger the weight changes, and the faster the learning will proceed. Large learning rates often lead to oscillation of weight changes and learning never completes, or the model converges to a solution that is not optimum. One way to allow faster learning without oscillation is to make the weight change as a function of the previous weight change to provide a smoothing effect. The momentum factor determines the proportion of the last weight change that is added into the new weight change.

The network 8–8/8–1 with the sigmoid/Gaussian combination is selected as a basis for following investigation. This is based on the stability and performance of the various networks investigated, as previously discussed. Analyses were performed on nine sets of learning rate and momentum terms, as shown in Figure 4.

Figure 4 shows there is little difference in R^2 when learning rate and momentum are less than or equal to 0.5. In other words, a relatively small learning rate and momentum seem to provide the appropriate stability of the model to avoid potential over-learning.

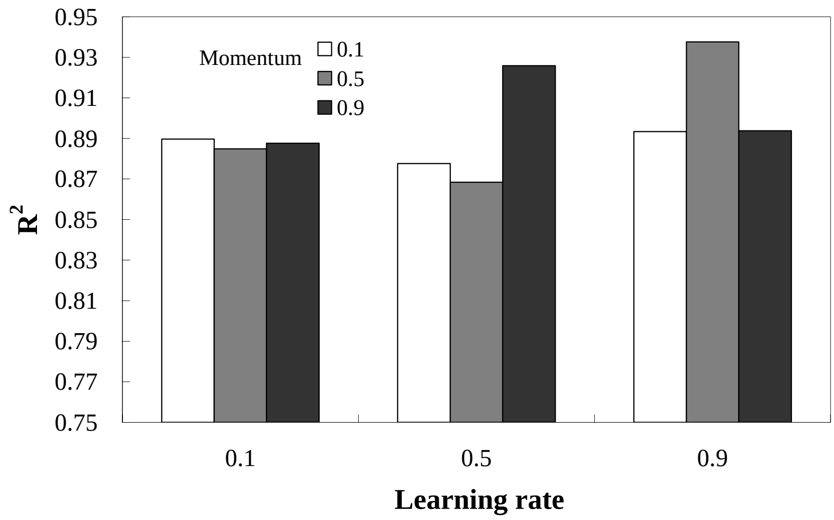

Figure 4: Effect of learning rate and momentum.

3.3 Comparison of the ANN Model with the Regression Model

Using training and testing data set, the multiple regression equation of PSR is similar to the AASHTO present serviceability index (PSI) can be express as:

$$PSR = 4.194 - 0.025\sqrt{FC} - 0.328 \times RD^2 - 0.076\sqrt{BC} - 0.029 \times IRI^2 \tag{6}$$

Where, $R^2 = 0.743$.

As an illustration, correlation graphs for computed and actual PSR are shown in Figure 5. It can be seen from Figure 5 that data used for validation are more closely clustered around the equality line for ANN model than for regression model, and ANN model exhibit a much higher R^2 value than multiple regression model.

4 CONCLUSIONS

As a key component of PMS, pavement performance models play a crucial role. This study was conducted to develop appropriate pavement performance models based on artificial neural networks. At this stage of the research it is very difficult to generalize the effect of the number of neuron, activation function, learning rate and momentum terms on pavement performance prediction using ANNs. Several conclusions can be drawn from this study:

- It is an appropriate method to choose the number of hidden neurons by Equation (1).

- The performance of ANN model can be raised by using two hidden groups with different activation functions in a hidden layer.
- A relatively small learning rate and momentum seem to provide the appropriate stability of the model to avoid potential over-learning.
- The ANN model provided an effective alternative to the current pavement performance models. Compared with the multiple regression model, better predictive results can be obtained from the ANN model.

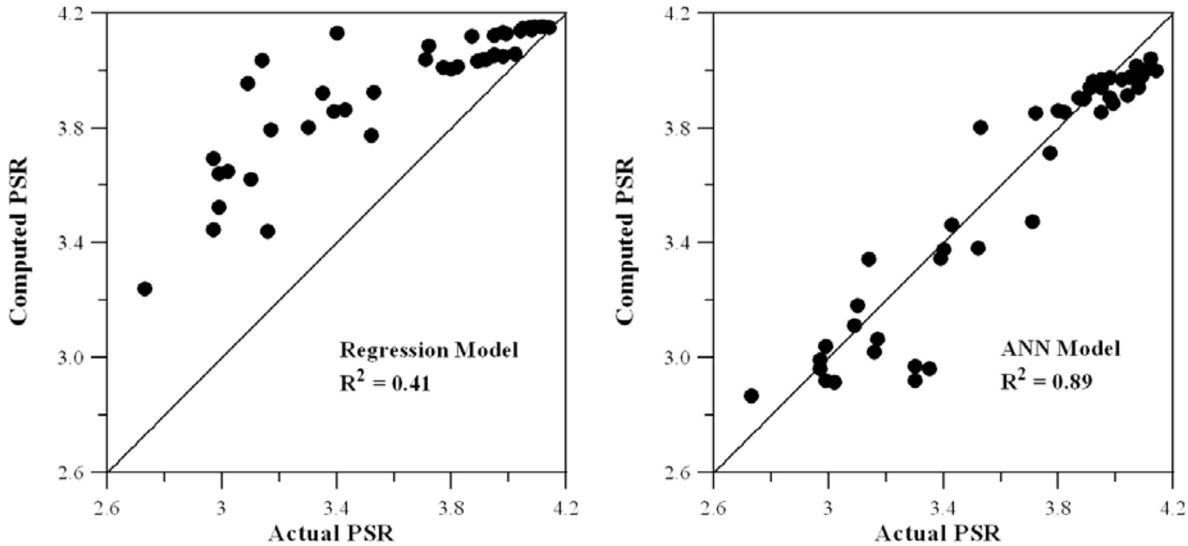

Figure 5: Plot of PSR.

REFERENCES

Attoh-Okine, N. O., 1994. *Predicting Roughness Progression in Flexible Pavements Using Artificial Neural Networks.* Proceedings of the Third International Conference on Managing Pavements, Vol. 1, San Antonio, TX, USA.

Attoh-Okine, N. O., 1999. *Analysis of Learning Rate and Momentum Term in Backpropagation Neural Network Algorithm Trained to Predict Pavement Performance.* Advances in Engineering Software.

Flintsch, G. W., 2003. *Soft Computing Applications in Pavement and Infrastructure Management: State-of-the-Art.* TRB 2003 Annual Meeting CD-ROM (001767.pdf), National Research Council, Washington DC, USA.

Funahashi, K., 1989. *On the Approximate Realization of Continuous Mappings by Neural Networks.* Neural Networks, Vol. 2.

Hornik, K., Stinchcombe, M. and White, H., 1989. *Multilayer Feedforward Networks Are Universal Approximators.* Neural Networks, Vol. 2.

Huang, Y. and Moore, R. K., 1997. *Roughness Level Probability Prediction Using Artificial Neural Networks.* Transportation Research Record 1592, TRB, National Research Council, Washington DC, USA.

Jansen, J. M. and Schmidt, B., 1994. *Performance Models and Prediction of Increase Overlay Need in Danish State Highway Pavement Management System.* Proceedings of the Third International Conference on Managing Pavements, Vol. 1, San Antonio, TX, USA.

La Torre, F., Domenichini, L. and Darter, M. I., 1998. *Roughness Prediction Based on the Artificial Neural Network Approach.* Proceedings of the Fourth International Conference on Managing Pavements, Vol. 2.

Lin, S. H., 2001. *Artificial Neural Network Modeling of Layer Modulus from Falling Weight Deflectometer*. Proceedings of the 17th International Conference on Advanced Science and Technology, Chicago, USA.

Lou, Z., Gunaratne, M., Lu, J. J. and Dietrich, B., 2001. *Application of a Neural Network Model to Forecast Short-Term Pavement Crack Condition: Florida Case Study.* Journal of Infrastructure Systems, Vol. 7, ASCE.

Owusu-Ababia, S., 1998. *Effect of Neural Network Topology on Flexible Pavement Cracking Prediction.* Computer-Aided Civil and Infrastructure Engineering, Vol. 13.

Shekharan, A. R., 1998. *Effect of Noisy Data on Pavement Performance Prediction by Artificial Neural Networks.* Transportation Research Record 1643, TRB, National Research Council, Washington, DC, USA.

Sorsa, T. and Koivo, H.N., 1993. *Application of Artificial Neural Networks in Process Fault Diagnosis.* Automatica, Vol. 29.

Ward Systems Group, 2000. *NeuroShell 2 Help.* Ward Systems Group, Inc, USA.

Yang, J., Lu, J. J., Gunaratne, M. and Xiang, Q., 2003. *Overall Pavement Condition Forecasting Using Neural Networks – An Application to Florida Highway Network.* TRB 2003 Annual Meeting CD-ROM (001441.pdf), National Research Council, Washington DC, USA.