Framework for Development and Comprehensive Comparison of Empirical Pavement Performance Models

Nima Kargah-Ostadi, Ph.D.[1]; and Shelley M. Stoffels, P.E.[2]

Abstract: Empirical performance-prediction models are a central part of every network-level pavement management system. In this regard, a variety of novel techniques including computational intelligence have been applied, mainly without a systematic approach to ensure compliance with principles of pavement engineering. In this study, a framework is provided for development and comprehensive comparison of alternative techniques for pavement performance modeling. As an example, several machine-learning techniques are compared in developing flexible pavement-roughness prediction models using Federal Highway Administration (FHWA's) long-term pavement performance (LTPP) data. Three important principles of model development—maximum likelihood, consistency, and parsimony—are considered in providing a robust parameterization guideline. Variant architectures of artificial neural networks (ANN), radial basis function (RBF) networks, and support vector machines (SVM) are tested to determine the optimum parameters. Final developed models are compared through quantitative and qualitative evaluations by means of a testing database that has not been used for model development. The example comparison gives the generalized RBF network model an edge over other machine-learning techniques in predicting pavement performance. This framework can be implemented by roadway agencies to develop a robust and representative performance-prediction model for pavement management systems. Moreover, the provided framework can be used to benchmark and compare alternative modeling paradigms for specific prediction problems. **DOI: 10.1061/(ASCE)TE.1943-5436.0000779.** *© 2015 American Society of Civil Engineers.*

Introduction

Empirical performance-prediction models are a central part of network-level pavement management systems (PMS) to predict future performance of the pavement network, identify maintenance and rehabilitation (M&R) needs, and estimate the network conditions after the application of various M&R alternatives (Haas et al. 1994). Performance models are implemented to guide M&R project selection, together with life-cycle cost analysis (LCCA) and engineering judgment.

The pavement performance model's development can be conceptualized through two distinct approaches; discrete time series prediction, or nonlinear function approximation (regression). In order to construct a discrete time series prediction scheme out of the performance progression trends, the performance measurements need to have consistent time intervals, which involves interpolation of missing data. Also, the factors affecting performance need to be summarized into a time-independent site factor through a nonlinear equation. While this simplifying approach might be an attractive option in very large pavement networks, it has numerous inherent errors. Hence, the majority of previous efforts have been focused on nonlinear multivariate regression where pavement performance is considered as the dependent variable and factors affecting performance are employed as independent variables (Perera et al. 1997).

Various studies have been conducted regarding pavement deterioration trends and factors affecting performance. Most of these studies have limitations such as the large number of variables used, correlation of input variables, difficulty of variables' data collection, experimental shape of performance curves, and intricacy of the equations, among others (Perera and Kohn 2001). Consequently, the need for a simpler and more efficient empirical model to use in network-level PMS is still an ongoing pursuit of the pavement engineering community.

Due to the large number of variables and the complex ways in which they affect one another, use of simple statistical approaches such as linear regression does not seem to be an appropriate means to develop performance-prediction models (Hunt and Bunker 2003). In addition, the shape of pavement performance curves are not known beforehand and multiple arrangements have to be tested in order to develop a model using nonlinear regression (Perera et al. 1997; Von Quintus et al. 2001). Hence, several studies have attempted to use computational intelligence techniques such as artificial neural networks (ANN) to develop more accurate models (Attoh-Okine 1994; Yang et al. 2003).

Past studies using ANN to develop pavement performance models have largely implemented feed-forward architectures and back-propagation learning algorithms. Other promising machine-learning techniques exist that could potentially be suitable for pavement performance modeling. Most of the past applications of machine learning lack a systematic approach to robust parameterization of these algorithms. In addition, most of the previously developed models have been evaluated based on prediction accuracy alone, and only a few have been evaluated regarding generalization capability. Among others, one major limitation of these studies is the lack of qualitative evaluations of model behavior according to principles of pavement engineering.

[1]Graduate Engineer, Fugro Roadware, Inc., 8613 Cross Park Dr., Austin, TX 78754 (corresponding author). E-mail: nimako82@gmail.com

[2]Associate Professor, Dept. of Civil and Environmental Engineering, Pennsylvania State Univ., 208 Sackett Building, University Park, PA 16802.

Note. This manuscript was submitted on June 29, 2014; approved on February 20, 2015; published online on April 21, 2015. Discussion period open until September 21, 2015; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Transportation Engineering***, © ASCE, ISSN 0733-947X/04015012(7)/$25.00.

Following a systematic approach, a comprehensive framework is devised in this study for comparison of performance models developed using various machine-learning techniques. Along with quantitative evaluations of performance models, several qualitative assessments are utilized in this comparison. To demonstrate application of this framework, a deterministic modeling approach is adopted for prediction of the flexible pavement international roughness index (IRI). In this context, a range of machine-learning techniques including ANN, radial basis function (RBF) networks, and support vector machines (SVM) are compared against each other. Data from Federal Highway Administration (FHWA's) long-term pavement performance (LTPP) program are implemented for model development.

Machine Learning

Machine-learning (ML) techniques are a subcategory of computational intelligence techniques mainly employed for deriving definitive information out of large sets of data for pattern recognition, classification, function approximation, prediction, and so forth (Jain et al. 1996). For function approximation applications, neural networks, radial basis function networks, and support vector machines are the most frequently used ML algorithms. They have very similar designs, which consist of a large number of parallel but connected simple processors. In the ANN terminology, these computing nodes are called neurons, which are organized in a number of layers and connected to each other via model parameters. These similar networks of processors are trained using different learning paradigms to estimate model parameters based on observed input-output data records. While ANNs are inspired by biological neurons, RBF networks and SVM are based on statistical learning theory.

The learning paradigm for ANN is a recursive stochastic approximation used in supervised training of multilayer perceptrons (MLP). Network parameters are randomly initialized and then adjusted recursively in a manner such that the mean squared error (MSE) between predicted versus measured outputs moves in the direction of steepest slope. The RBF learning paradigm is termed hypersurface reconstruction, which is essentially smooth curve fitting using regularized interpolation. The SVM learning paradigm is an approximate implementation of the principle of structural risk minimization adopted from the statistical learning theory. The maximum degradation (distance) of an approximating hyperplane from the observed data is minimized. To learn more about the details of these paradigms, the reader is referred to Haykin (1999) and Smola and Scholkopf (2004).

In two independent studies, Funahashi (1989) and Hornik et al. (1989), proved the Universal Approximation Theorem for ANN. They proved that an MLP having a single hidden layer is enough to approximate any continuous nonlinear input-output relationship to any degree of accuracy. However, this does not mean that the MLP would be the most efficient one or that it would have a good generalization capability. Similar studies have proved that RBF networks and SVM are also universal approximators (Haykin 1999). Any superior performance of each technique compared to others has proved to be problem-specific and a systematic benchmarking and comparison approach is required to determine the most appropriate.

The four implemented learning machines are an RBF network, a Gaussian SVM, and two ANN models. For improving generalization in neural networks, two approaches were compared: early stopping using the validation data set (ANN-ES), and regularization using Bayesian inference (ANN-BR) technique (MacKay 1992). If the classical training process of an ANN involving the minimization of prediction error on the training data alone is carried out, often the resultant approximating function fits the inherent noise within the data as well as the real patterns, and thus it will have poor generalization capability if tested on other data. This phenomenon is called overfitting and efforts to address it range from early stopping to regularization (Haykin 1999).

In early stopping, the training patterns are divided randomly into training and validation subsets. The error on the training subset is minimized until the error on validation subset starts to increase. The training process is then stopped at the minimum error on validation data. In regularization training, a regularizing term is added to be minimized along with the error. This additional term typically penalizes larger weights. Therefore, this type of regularization is called weight decay and advocates smaller weights for less important network connections. Instead of trial and error, some studies suggest the application of Bayesian inference to balances bias (lack of fit) and variance (overfitting due to noise) (MacKay 1992).

Data Extraction and Preprocessing

Data from LTPP-specific pavement studies (SPS) numbers one and five were used in developing the IRI models (LTPP 2013). The SPS-1 and SPS-5 experiments study long-term performance of flexible pavements with varying structural factors and alternative rehabilitation treatments, respectively. Tables 1 and 2 show the design factorials for LTPP SPS-1 and SPS-5 studies, respectively. These pavement sections have been constructed in a variety of climatic regions and on various subgrade soil types.

Factors affecting pavement surface roughness were used as input variables to develop a model that predicts the output variable, roughness. In most of the previous studies regarding factors affecting flexible pavement roughness, several factors have consistently been deemed as having considerable influence. Those factors are initial roughness, pavement age, traffic, climatic conditions (average annual precipitation and average annual freezing index), pavement structural properties, subgrade properties (Tighe 2002) (moisture content and percent passing No. 200 sieve), drainage type

Table 1. Design Factorial and Nomenclature of Test Sections in the LTPP SPS-1 Study

Base thickness (mm)	AC thickness (mm)	Without drainage			With drainage	
		DGAB	ATB	DGAB/ATB	PATB/DGAB	ATB/PATB
203	102	0113	0103	0105	0107	0122
	178	0101	0115	0117	0119	0110
305	102	0102	0116	0118	0120	0111
	178	0114	0104	0106	0108	0123
406	102	—	—	—	0121	0112
	178	—	—	—	0109	0124

Note: ATB = asphalt treated base; DGAB = dense graded aggregate base; PATB = permeable asphalt treated base.

Table 2. Design Factorial and Nomenclature of Test Sections in the LTPP SPS-5 Study

Section number	Surface preparation	Type of AC	Overlay thickness (mm)
0501	Routine maintenance		0
0502	Minimum surface	Recycled	50
0503	preparation		125
0504		Virgin	125
0505			50
0506	Intensive surface	Virgin	50
0507	preparation (milling)		125
0508		Recycled	125
0509			50

and conditions, and maintenance and rehabilitation treatments (Perera and Kohn 2001; Hunt and Bunker 2003; Haider et al. 2007; Kutay 2007). It is difficult and not cost effective to collect all these data for all pavements making up a network. Therefore, considering all factors affecting pavement roughness in its modeling is typically not practical. Of course, statistical significance testing and engineering judgment are necessary means in order to influence the inclusion of various factors as input variables (Von Quintus et al. 2001; Kargah-Ostadi et al. 2010).

For each IRI reading on each pavement section, a data record was composed of measures of the input variables at that specific time for the pavement section. As a result, the number of data records was the same as the number of IRI readings between every two rehabilitation efforts. Since in a PMS, condition data are usually available on a regular basis and predictions are only desired for the next few time periods, an incremental form is preferred for a model that is intended to be used at the network level (Prozzi and Madanat 2003). Consequently, instead of initial roughness, which may not be readily available for all pavement sections, the previous IRI measurement was considered as an input variable for each IRI reading. Accordingly, incremental and cumulative time and incremental traffic and climatic factors were also considered as other potential input variables.

Based on LTPP data availability, Table 3 shows the 14 input variables (and statistics) that were considered to model roughness progression in the right wheel path. There are a total of 1,192; 1,118; 483; and 609 data records in the wet-freeze (WF), wet-no freeze (WNF), dry-freeze (DF), and dry-no freeze (DNF) climatic regions, respectively. Out of the total data records, about 25%are on fine-grained soils and 75% are on coarse-grained subgrade.

Several preprocessing steps were taken for smoothing, outlier detection, normalization (mean removal), and decorrelation. In order to alleviate the irrational fluctuations in the time-dependent performance data, a three-point moving average scheme was used for smoothing the IRI versus age curves. This scheme was time normalized according to the following equation:

$$\mathrm{IRI}_{\mathrm{TNMA}} = \frac{1}{2}\left[\mathrm{IRI}_C + \frac{(\mathrm{IRI}_P)\times(D_N - D_C) + (\mathrm{IRI}_N)\times(D_C - D_P)}{(D_C - D_P) + (D_N - D_C)}\right] \quad (1)$$

where $\mathrm{IRI}_{\mathrm{TNMA}}$ = time normalized moving average IRI. IRI_C, IRI_P, and IRI_N correspond to the current, previous, and next IRI measurements, respectively. D_C, D_P, and D_N are the current, previous, and next IRI measurement dates, respectively. This is the first step in separating actual performance trends from the inherent noise within the data.

Data records where the IRI measurement was less than its previous IRI reading (while no maintenance or rehabilitation treatment had happened) were considered outliers and were eliminated from the database. It should be noted that a variety of reasons such as compaction by traffic and differences in testing seasons could cause the IRI to decrease with time, and corresponding investigations should be considered in this regard.

Next, the input variables were normalized to have zero mean and standard deviation of one across all data records. The output variable was mapped to a range of [−1, +1] for consistency. After normalization and outlier detection, principal component analysis (PCA) was used to simplify the developed models through reduced dimensionality and eliminate input variable correlations, thereby reducing overfitting probability.

Principal components (PC) are new uncorrelated variables created through a linear combination of correlated input variables (Jolliffe 1986). Through decomposition of the covariance matrix of the original correlated inputs, PCA arranges the resultant components in an order according to the portion of the total input data variance that they account for. The considered 14 input variables in Table 3 were combined using the PCA coefficients to form the principal components. In this study, the first 10 components explained more than 95% of the total variance within input variables. These 10 independent principal components were selected as uncorrelated input parameters for model development. Unlike

Table 3. Considered Input Variables and Database Statistics

Number	Variable name	Variable description	Units	Minimum	Maximum	Average	Standard
1	pIRIR	Previous IRI measurement in the right wheel path	m/km	0.42	4.04	1.08	0.44
2	dKESAL	Number of equivalent single-axle loads since previous IRI measurement	thousands	2.76	10,021.92	496.05	833.54
3	dTime	Time since previous IRI measurement	years	0.03	6.92	1.16	0.75
4	AGE	Time since last hot mix asphalt concrete placement	years	0.50	35.96	8.40	5.34
5	tAC	Total thickness of all asphalt concrete layers	mm	76.20	393.70	182.27	63.08
6	tTL	Total thickness of all treated base and subbase layers	mm	0.00	891.54	177.19	173.57
7	tUL	Total thickness of all untreated base and subbase layers	mm	0.00	1,930.40	352.51	329.87
8	P200	Percent fines (passing number 200 sieve) in subgrade soil	%	0.30	97.50	37.86	24.68
9	PCL	Percent clay (smaller than 0.002 mm) in subgrade soil	%	0.00	75.60	12.90	10.90
10	PI	Plasticity index of subgrade soil	NA	0.00	69.00	8.71	11.29
11	dPRCP	Amount of precipitation since previous IRI measurement	mm	0.05	10,501.40	937.01	984.00
12	dFZI	Freezing index since previous IRI measurement	degC-days	0.00	6,130.00	316.59	658.69
13	dFTC	Number of freeze-thaw cycles since previous IRI measurement	count	0.00	438.00	76.96	66.22
14	dN32°C	Number of days above 32°C since previous IRI measurement	count	0.00	688.00	69.69	89.17

stepwise regression, PCA allows for implementation of all input variables in model development while eliminating the correlations among them.

Data records for one section from each climatic region were selected randomly for testing of the developed models regarding roughness progression trends. The remaining 3,361 data records were divided into training (70%), validation (15%), and testing (15%) databases. The validation database was used in early stopping learning paradigms in order to avoid overfitting due to noise. In other generalization learning algorithms, the validation and training databases were jointly used for model development. For quantitative evaluation of model prediction accuracy, the testing database was used because it had not been utilized by the model at the development stage.

Algorithm Benchmarking

For each learning machine, various architectures were compared according to the benchmarking flowchart depicted in Fig. 1 so that the architecture with highest accuracy and generalization would be chosen. Small changes in the architectures (number of processing elements and their arrangement) and control parameters of machine-learning algorithms might result in significant changes in the final developed model. The final architecture selected according to this flowchart will have an optimum and robust parameterization. For each specific learning problem, this benchmarking step is essential before any objective comparison of the algorithms can be made.

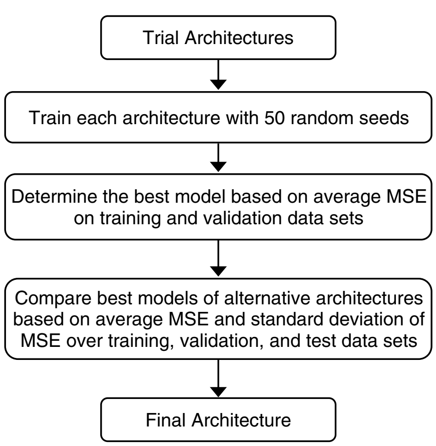

Fig. 1. Flowchart for benchmarking each learning machine on the specific problem

Comparison Framework

In order to provide a comprehensive comparison of various machine-learning techniques, both the learning process and the developed models need to be evaluated (Fig. 2). The learning processes are compared in terms of effectiveness (average MSE of the final model over all data sets), efficiency (total amount of time required to evaluate alternative architectures and determine the optimum), and reliability (subtracting from 100, the COV of MSE over 50 random trials of the final machines).

While having lower efficiency and reliability, Bayesian regularization of ANN (ANN-BR) was the most effective learning algorithm in reducing the error between measured and predicted pavement performance. Gaussian SVM (GSVM) and generalized RBF networks (GRBF) follow regularized ANN in terms of effectiveness. Except regularized ANN, the other learning methods have very high reliability.

In this study, three important principles of model development (Gupta et al. 2008) were considered: (1) maximum likelihood (minimizing MSE on training data); (2) consistency (generalization); and (3) parsimony (selecting the simplest learning network, with equal error). The same principles are incorporated in quantitative and qualitative evaluation of the final developed models (Fig. 2).

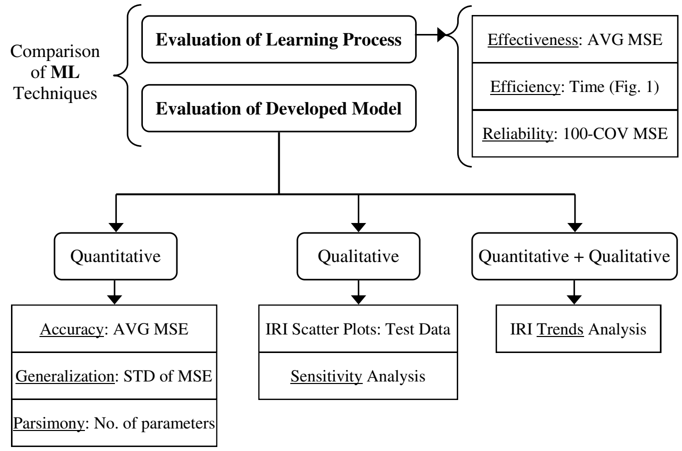

Fig. 2. Framework for comprehensive comparison of various machine-learning techniques

Results and Discussion

The comparison framework depicted in Fig. 2 was applied to the roughness prediction models developed with the four different ML techniques. For quantitative evaluation of alternative models, accuracy is calculated as the average MSE over training and test data sets. Generalization capability is measured as the standard deviation of MSE between training and test data sets. In addition, model complexity is calculated in terms of number of parameters involved. Fig. 3 represents the quantitative comparison of the four IRI models, where the number of parameters is indicated in front of the models names. In order to provide a baseline for comparison, a multivariate nonlinear regression model with an exponential form (NL-REG) was developed based on the same principal components as inputs

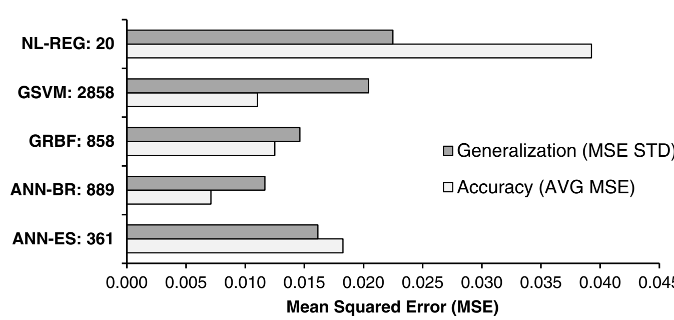

Fig. 3. Quantitative comparison of accuracy, generalization, and parsimony of the final developed models; number of model parameters is indicated in front of each model name

$$IRI = a_1 \exp(a_2 PC_1) + a_3 \exp(a_4 PC_2) + \cdots + a_{19} \exp(a_{20} PC_{10}) \quad (2)$$

Fig. 3 indicates that all ML models result in lower prediction MSE compared to the nonlinear regression model. While the exponential form was only provided as a baseline for comparison, there are numerous possible forms, and the optimum equation for nonlinear regression is not known. Meanwhile, the number of possible variants of each ML algorithm is limited and the best can be determined through the systematic approach provided in Fig. 1.

The ML models provide better accuracy and generalization compared to the nonlinear regression, at the cost of higher complexity of the final model. ANN-BR is the most accurate and has the best generalization capability. However, if model parsimony is considered, GRBF might be selected as the best model. For some models, such as the GSVM, there is a significant difference between the accuracy and generalization capability.

For qualitative evaluation, scatter-plots of measured versus predicted IRI over the testing database (504 records not used in model development) are examined. Fig. 4 shows IRI scatter plots for the four selected models. The ANN-BR, GRBF, GSVM, and ANN-ES in this order provide the best scatter plots, and this does not agree with the quantitative evaluation (Fig. 3).

A more important qualitative evaluation tool is sensitivity analysis of the model output to variations in input factors. This step is essential to evaluate consistency in model function according to principles of pavement engineering. Regarding machine-learning techniques, evaluation of model form is not viable. However, for nonlinear regression techniques, dimensional analysis and other mechanistic principles can be employed to evaluate model form. Nevertheless, the developed empirical models for pavement management at the network level are intended to support prediction rather than scientific reasoning.

All but one input variable are kept constant at database average values (Table 3) and one variable is increased or decreased by 25% to examine sensitivity of model output to each input factor. In agreement with previous studies (Perera and Kohn 2001), all the

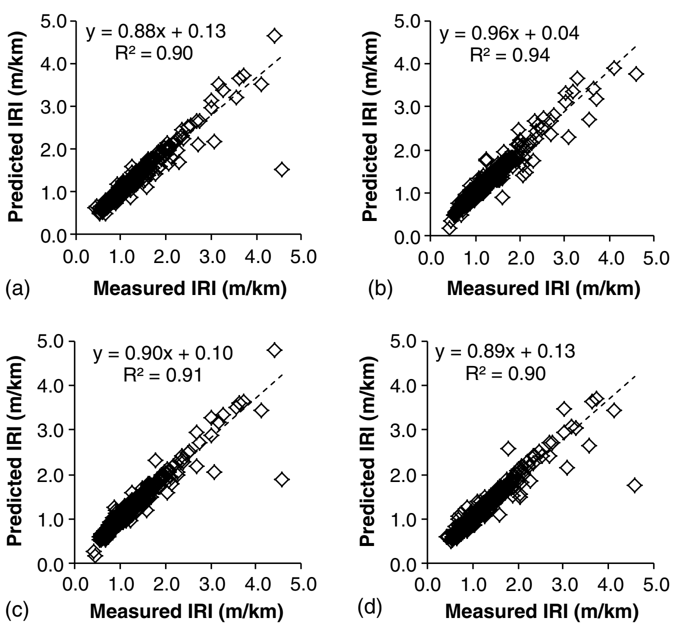

Fig. 4. IRI scatter plots on test database (504 records not used in model development) using (a) ANN-ES; (b) ANN-BR; (c) GRBF; (d) GSVM

developed IRI models are most sensitive to previous IRI measurements. With a 25% change in previous IRI values, all models predict about 20% of the change in the predicted IRI. Fig. 5 demonstrates change in IRI output (percent) with varying input variables. In this figure, the bounds for the vertical axis values are restricted to 5% in order to better examine sensitivity of the models to parameters other than previous IRI measurement. Generally, it seems that the GRBF model provides a sensitivity pattern that agrees better with previous studies of factors affecting flexible pavement roughness.

The ANN-BR model shows an inverse relationship of IRI changes with respect to traffic and time between IRI measurements. However, as all the other models show, IRI should increase with an increase in these variables. An increase in any of the thickness variables is expected to decrease roughness, and this is the case in all models except for the ANN-ES and GSVM models, which show an increase in IRI with increasing thickness of the AC layer. Finer grained subgrade soils typically contribute to faster roughness progression in asphalt pavements (Haider et al. 2007). While most of the developed models show limited sensitivity to subgrade parameters, the ANN-BR model shows a decrease in IRI with increasing percent fines and clays in the subgrade, contrary to expectations.

Except for the ANN-BR, the developed models show lower IRI values with an increase in precipitation, which is a questionable behavior. All of the models show a direct relationship between roughness and the freezing index, which agrees with previous studies (Kutay 2007). In contrast, there is a questionable inverse relationship between predicted IRI and number of freeze-thaw cycles between measurements, except for the GSVM model. ANN-ES and GRBF models correctly indicate an increase in IRI with an increase in the number of warmer days between measurements (dN32°C).

A combination of quantitative and qualitative evaluation is also carried out by examining variations of model output with time (Gupta et al. 2008). Previously unimplemented data records from one section in each of the four climatic regions are used to compare average absolute error (AAE) of observed versus predicted IRI change rates (Fig. 6). While the generalized RBF network (GRBF) model does not have the best quantitative capacities (Fig. 3), it performs better than the other models in predicting roughness progression rates.

Overall, quantitative and qualitative evaluations of the developed IRI models give an edge to the generalized RBF network. While it has moderate accuracy and generalization capability compared to other models, it performs best in terms of sensitivity behavior and prediction of roughness progression trends. Fig. 7 shows the measured IRI trends along with trends predicted with the GRBF model in four climatic regions.

The predicted roughness progression trends via the GRBF model matches the observed pavement performance in all climatic regions. For flexible pavement networks that have similar attributes to the LTPP SPS-1 and SPS-5 studies, it is suggested that generalized RBF networks be implemented to develop empirical roughness prediction models.

This comparison framework demonstrates that a model having better accuracy might not necessarily generalize well. At the same time, quantitative evaluation alone is not sufficient to determine the best model, and other qualitative criteria should also be considered. Furthermore, in some of the above evaluations, performance differences between models are not very significant. Above all, these comparisons are problem specific and should be carried out for every learning problem. Roadway agencies interested in application of novel techniques such as computational intelligence

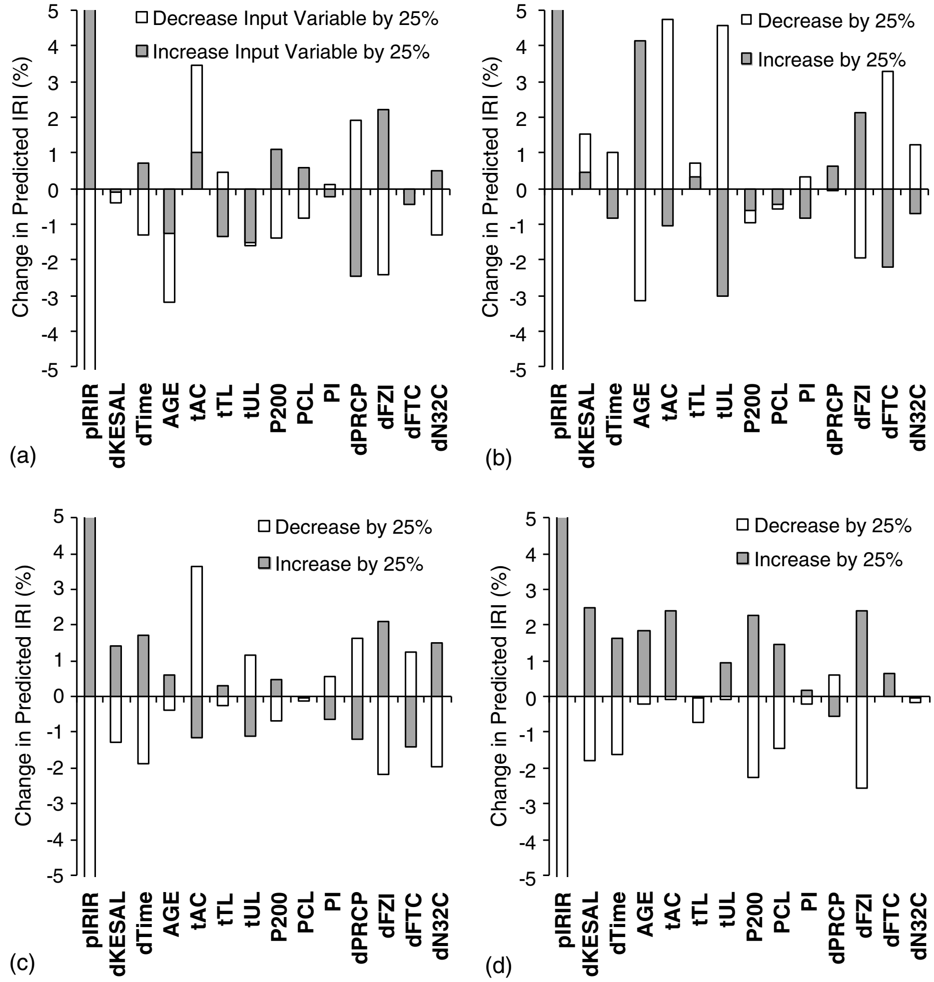

Fig. 5. Sensitivity analysis of (a) ANN-ES; (b) ANN-BR; (c) GRBF; (d) GSVM models to changes in each input variable

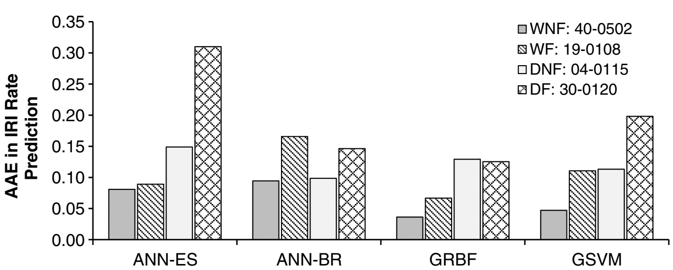

Fig. 6. Comparison of AAE of the models in predicting IRI change rate in each of the four climatic regions

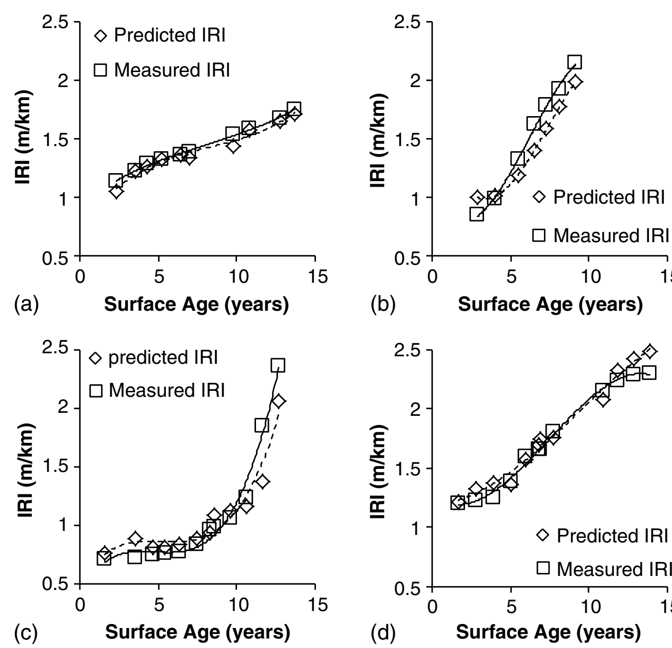

Fig. 7. IRI progression trends predicted using the generalized RBF network (GRBF) for pavement sections in (a) WNF; (b) WF; (c) DNF; (d) DF climates

algorithms need to implement the developed benchmarking approach (Fig. 1) and comparison framework (Fig. 2) in order to determine the most appropriate pavement performance model.

Conclusions and Recommendations

In this study, a comprehensive framework was provided to benchmark and compare alternative machine-learning techniques for development of network-level pavement performance models. The comparison was based on an assessment of the learning processes, followed by quantitative and qualitative evaluations of the final developed models. Preprocessing steps included smoothing, outlier detection, normalization, and decorrelation via principal component analysis. For algorithm benchmarking, variant architectures of neural networks, radial basis function networks, and support vector machines were tested to determine the optimum parameters.

The different learning processes were compared in terms of effectiveness, efficiency, and reliability. Quantitative evaluation revealed that having higher accuracy (lower average error on training data) does not always result in models that generalize well on a testing database that had not been implemented in model development. Qualitative evaluations included scatter plots of predicted versus measured performance and sensitivity analysis of the models with respect to variations in factors affecting performance. Prediction models that demonstrate low quantitative errors might not necessarily succeed in qualitative evaluations of model behavior according to engineering principles. Combining quantitative and qualitative aspects, the predicted variation of model output with time was compared to measured performance progression trends.

A flexible pavement-roughness prediction model was used as an example to demonstrate application of the framework that was devised based on the principles of model development. This systematic approach to benchmarking and comparison, while comprehensive, is problem specific and should be repeated for other performance modeling efforts.

References

Attoh-Okine, N. O. (1994). "Predicting roughness progression in flexible pavements using artificial neural networks." *Proc., 3rd Int. Conf. on Managing Pavements*, Vol. 1, National Research Council, Transportation Research Board, Washington, DC, 55–62.

Funahashi, K. (1989). "On the approximate realization of continuous mappings by neural networks." *Neural Networks J.*, 2(3), 183–192.

Gupta, H. V., Wagener, T., and Yuqiong, L. (2008). "Reconciling theory with observations: Elements of a diagnostic approach to model evaluation." *Hydrol. Processes*, 22(18), 3802–3813.

Haas, R., Hudson, W. R., and Zaniewski, J. (1994). *Modern pavement management*, Krieger Publishing, Malabar, FL.

Haider, S., Chatti, K., Buch, N., Lyles, R., Pulipaka, A., and Gilliland, D. (2007). "Effect of design and site factors on the long-term performance of flexible pavements." *J. Perform. Constr. Facil.*, 10.1061/(ASCE)0887-3828(2007)21:4(283), 283–292.

Haykin, S. (1999). *Neural networks: A comprehensive foundation*, 2nd Ed., Prentice Hall, NJ.

Hornik, K., Stinchcombe, M., and White, H. (1989). "Multilayer feedforward networks are universal approximators." *Neural Networks J.*, 2(5), 359–366.

Hunt, P. D., and Bunker, J. M. (2003). "Study of site specific roughness progression for a bitumen-sealed unbound granular pavement network." *Transportation Research Record 1819*, Transportation Research Board, Washington, DC, 273–281.

Jain, A. K., Mao, J., and Mohiuddin, K. M. (1996). "Artificial neural networks: A tutorial." *Computer*, 29(3), 31–44.

Jolliffe, I. T. (1986). *Principal component analysis*, Springer, New York.

Kargah-Ostadi, N., Stoffels, S. M., and Tabatabaee, N. (2010). "Network-level pavement roughness model for rehabilitation recommendations." *Transportation Research Record 2155*, Transportation Research Board, Washington, DC, 124–133.

Kutay, M. E. (2007). "Spectral analysis of factors affecting roughness in flexible pavements." *23rd World Road Congress*, World Road Association (PIARC), Paris.

LTPP (Long-Term Pavement Performance). (2013). "Long-term pavement performance (LTPP) standard data release 27.0." *USB Version*, Federal Highway Administration, U.S. Dept. of Transportation, McLean, VA.

MacKay, D. J. C. (1992). "Bayesian interpolation." *Neural Comput.*, 4(3), 415–447.

Perera, R. W., Byrum, C., and Kohn, S. D. (1997). "Investigation of development of pavement roughness." *FHWA-RD-97-147*, Office of Infrastructure Research and Development, FHWA, McLean, VA.

Perera, R. W., and Kohn, S. D. (2001). "LTPP data analysis: Factors affecting pavement smoothness." *NCHRP Web Document 40*, National Cooperative Highway Research Program, Transportation Research Board, Washington, DC.

Prozzi, J. A., and Madanat, S. M. (2003). "Empirical-mechanistic model for estimating pavement roughness." *82nd Annual Meeting of Transportation Research Board*, Transportation Research Board, Washington, DC.

Smola, A. J., and Scholkopf, B. (2004). "A tutorial on support vector regression." *Stat. Comput.*, 14(3), 199–222.

Tighe, S. (2002). "Evaluation of subgrade and climatic zone influences on pavement performance in the Canadian strategic highway program's (C-SHRP) long-term pavement performance (LTPP) study." *Can. Geotech. J.*, 39(2), 377–387.

Von Quintus, H. L., Eltahan, A., and Yau, A. (2001). "Smoothness models for hot-mix asphalt-surfaced pavements; developed from long-term pavement performance program data." *Transportation Research Record 1764*, Washington, DC, 139–156.

Yang, J., Lu, J. J., Gunaratne, M., and Xiang, Q. (2003). "Forecasting overall pavement condition with neural networks: Application on Florida highway network." *Transportation Research Record 1853*, Washington, DC, 3–12.