Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 8 (2012) 473 – 478

Procedia
Computer Science

New Challenges in Systems Engineering and Architecting
Conference on Systems Engineering Research (CSER)
2012 – St. Louis, MO
Cihan H. Dagli, Editor in Chief
Organized by Missouri University of Science and Technology

Development of Void Prediction Models for Kansas Concrete Mixes Used in PCC Pavement

Hakan Yasarer*[a], Yacoub Najjar[b], a*

[a]PhD Student, Department of Civil Engineering, Kansas State University, Manhattan, KS 66506, USA
[b]Professor, Department of Civil Engineering,, Kansas State University, Manhattan, KS 66506, USA

Abstract

Permeability of the concrete material used in Portland Cement Concrete (PCC) pavement structures is a major factor for long-term durability assessment. To properly characterize the permeability response of a PCC pavement structure, the Kansas Department of Transportation (KDOT) generally runs the Boil Test (BT) to determine the % void response. The BT typically measures the volume of permeable pore space within the concrete samples over a period of five hours at a concrete age of 7, 28, and 56 days. In this study, backpropagation Artificial Neural Network-(ANN) and Regression-based % void response prediction models for the BT are developed by using the database provided by KDOT in order to reduce the duration of the testing period or ultimately eliminating the need to conduct the BT. The noted excellent prediction accuracy of the developed models proved that the ANN and the Regression models have efficiently characterized the BT response. Therefore, they can be considered as effective and applicable models to predict the permeability (% void response) response of concrete mixes used in PCC pavements.

© 2012 Published by Elsevier Ltd. Selection Open access under CC BY-NC-ND license.

Keywords: Boil Test; Artificial Neural Networks; Civil Engineering; Concrete Pavement; Modeling

1. Background

Permeability of the concrete in a PCC is a major factor for long-term durability. The permeability of concrete depends on its pore network, which comes primarily from the excess water used during mixing in the initial hardening process. The porosity of concrete consists of closed or logged pores in addition to a network of interconnected pores [6]. Pore size ranges from a few angstroms to about 100 A° for the so called 'gel pores', from 100 to 100000 A° in 'capillary pores', and a few millimeters in 'air or large

* Corresponding author. Tel.: +17853179767; fax: +1-785-532-7717.
E-mail address: hiy@ksu.edu.

1877-0509 © 2012 Published by Elsevier B.V. Open access under CC BY-NC-ND license.
doi:10.1016/j.procs.2012.01.088

pores'. Inter connected pores endow the concrete permeability. All the hydrated cement products are subjected to attack by sulphates, chlorides, acids, and water. It is a common practice to evaluate the water permeability characteristics when assessing the concrete durability characteristics. Permeability can be measured by conducting standard test methods. In this study, % of water absorption, % of permeable voids and % of total voids have been determined as per ASTM C 642-97 [1]. This test was performed as per procedure given in ASTM C 642-97 [1] by the oven-drying method. The measurements as part of ASTM C 642-97 [1] such as Oven-dry mass (A), Saturated surface-dry weight (B) and Curing time (CT) were used to develop prediction models by ANN and Regression to predict Saturated surface-dry weight after boiling (C), and Weight in water after boiling (D). Therefore, two models are developed to predict C and D individually using the same database. A, B, C, D and CT are the only values used for model development. However, % volume of permeable pore space is the final value calculated out of A, C and D variables and thus was used for accuracy measure comparisons. In this study, both regression and ANN approaches were used to characterize the % volume of permeable voids of concrete.

2. ANN Model Development

Typically, the desired ANN model is developed by following four sequential development stages. In the first stage, the ANN architecture is determined based on problem characteristics and ANN knowledge. Therefore, input and output variables are chosen accordingly. This step also includes classifying the datasets as training, testing or validation sets. In the second stage, the network is trained and tested on the experimental data to obtain the optimum number of hidden nodes and iterations for the desired ANN architecture determined in stage one. In the third stage, the best performing network obtained from the second stage is validated on the validation database. If accuracy measures for training, testing and validation databases are very comparable, then the developed model may not need to be retrained on all data. In the fourth stage, if needed, the best performing network obtained in the second stage is retrained on all experimental data in order to allow the ANN model to better characterized the desired behavior by exposing it to the testing and validation datasets. Generally, retraining the network on all datasets is expected to provide more reliable model and better accuracy measures if the dataset classification task is done in an appropriate manner. However, it has been shown through several research studies by Najjar and Co-workers [[2], [3], [4] and [5]] that the stage four is recommended to arrive at a better performing network. In this chapter, the four sequential stages have been conducted twice to arrive at two desired prediction models for C and D. In order to develop boil test permeability prediction model, two models for predicting C and D have been proposed. The three best performing models for predicting C and D have been evaluated in order to select the most appropriate model to predict C and D. The networks developed for C and D have fully connected internal structure (i.e. any node in one layer connects to all the nodes in the next layer) with one hidden layer. ANN model development issues for C and D networks are explained in this section.

ANN Model Architecture

Based on the knowledge gained from experimental data analysis, ANN model architecture for C has been built by considering 3 inputs and 1 output, which respectively are:

Inputs:

1- (A) Mass of oven-dried sample in air (grams)
2- (B) Mass of surface-dry sample in air after immersion (grams)
3- (CT) Curing Time (days)

Output:

1- (C) Mass of surface-dry sample in air after immersion and boiling (grams)

ANN model architecture for D has been built by considering 3 inputs (the same ones used for Model C) and 1 output (D). In this network, output D represents the apparent mass of sample in water after immersion and boiling (in grams).

In this study, 3 models, from among many considered, giving appropriate accuracy statistical measures have been selected based on optimum hidden nodes, minimum values of Mean Absolute Relative Error (MARE) and Averaged-Squared-Error (ASE) and maximum values of Coefficient of Determination (R^2). A total of 414 datasets are used to build the desired database; 211, 112 and 91 sub-databases are used, respectively, for training, testing and validation purposes. Datasets that include minimum and maximum values of each input and output variables are included in the training sub-database in order for the network to represent the wide characteristics of the response (i.e., to assure that the developed network will always act in an interpolation mode).

Model Training and Testing

Based on statistical accuracy measures such as ASE, R^2 and MARE, the optimal network structures for three models to predict C were obtained from among many networks. The corresponding accuracy measures for the selected models are listed in Table 1 with the best performing network is identified in bold. This network is labeled as 3-3-1 to indicate that this network uses 3 inputs variables and 3 hidden nodes to produce 1 output representing C.

Table 1 Statistical Accuracy Measures of ANN-Models of C

Model		Model C1	**Model C2**	Model C3
Training Architecture		3-(1-3-18-20000)-1	**3-(2-3-18-19500)-1**	3-(3-4-18-19900)-1
Training	MARE(%)	0.336%	**0.164%**	0.184%
	R^2	0.989	**0.997**	0.996
	ASE	0.000103	**0.000027**	0.000035
Testing	MARE(%)	0.310%	**0.149%**	0.177%
	R^2	0.984	**0.996**	0.994
	ASE	0.00009	**0.000023**	0.000033
Validation	MARE(%)	0.319%	**0.174%**	0.189%
	R^2	0.989	**0.996**	0.996
	ASE	0.000087	**0.000033**	0.000037
All Data	MARE(%)	0.33%	**0.164%**	0.33%
	R^2	0.989	**0.997**	0.989
	ASE	0.000092	**0.000025**	0.000092
Final Architecture		3 - 3 - 1	**3 - 3 - 1**	3 - 4 - 1

In a similar manner, the optimal network structures for three models to predict D were obtained. Corresponding statistical accuracy measures for these networks are given in Table 2. The best performing network is identified in bold.

Model Validation

After Subdividing the database into training, testing, and validation sub-databases, the desired network (C or D) was trained on 211 datasets and tested online on 112 datasets in order to obtain the optimum number of hidden nodes and iterations for the ANN architecture determined in stage one. Model validation (i.e., the third modeling stage) is then performed by utilizing the best performing network identified in stage two to predict the outputs of the 91 validation datasets. Accordingly, statistical accuracy measures such as ASE, R^2 and MARE can be calculated for the validation datasets. The resulting validation statistical accuracy measures for network models C and D are, respectively, listed in Tables 1 and 2.

Model Selection

Statistical accuracy measures for training, testing and validation databases at optimal ANN structure with 3 hidden nodes and 19,500 iterations showed better overall prediction accuracy when compared with those for models C1 and C3. Even though Model C1 has same amount of hidden nodes as Model C2, Model C2 attains better accuracy measures than those of Model C1. It is to be noted that all of the three models, listed in Table 1, can be used as prediction models since they all have considerably excellent accuracy statistics. In this case, the best-performing model (identified in bold) is considered in the final selection. For this reason, Model C2 has been chosen to be used as the best network structure. Thus, all of the 414 datasets from the Boil test were used to retrain (i.e., the stage four modeling) the network at this optimal structure to obtain the generalized response throughout the entire database. Statistical accuracy measures of the selected model trained with all data are: ASE_{all}=0.000025, R^2_{all} =0.997 and $MARE_{all}$=0.164%. More information regarding ANN model development issues can be found in Yasarer (2010) [7].

By examining statistical accuracy measures for training, testing and validation databases listed in Table 2, it becomes clear that Model D2, at optimal ANN structure with 3 hidden nodes and 20,000 iterations shows better prediction accuracy when compared with those for Models D1 and D3. Even though all of the three models have performed considerably well, Model D2 yields the least ASE and MARE values and gives the highest R^2 value. For this reason, Model D2 has been chosen to be used as the best network model for D. Thus, all of the 414 datasets from the Boil test were used to retrain the network at this optimal structure to obtain the generalized response throughout the entire database. The corresponding accuracy measures of selected model trained with all data are: ASE_{all}=0.000643, R^2_{all} =0.934 and $MARE_{all}$=1.110%.

Table 2 Statistical Accuracy Measures of ANN-Models of D

Model		Model D1	**Model D2**	Model D3
Training Architecture		3-(1-2-12-20000)-1	**3-(1-3-12-20000)-1**	3-(2-4-12-20000)-1
Training	MARE(%)	1.203%	**1.144%**	1.144%
	R^2	0.926	**0.929**	0.929
	ASE	0.000776	**0.00073**	0.000729
Testing	MARE(%)	1.132%	**1.077%**	1.076%
	R^2	0.943	**0.948**	0.948
	ASE	0.0006	**0.000536**	0.000536
Validation	MARE(%)	1.21%	**1.14%**	1.15%
	R^2	0.918	**0.926**	0.925
	ASE	0.00072	**0.000631**	0.000633
All Data	MARE(%)	1.112%	**1.110%**	1.111%
	R^2	0.933	**0.934**	0.933
	ASE	0.000644	**0.000643**	0.000644
Final Architecture		3 - 2 - 1	**3 - 3 - 1**	3 - 4 - 1

3. Regression Models

Regression model development issues for Model C and D have been accomplished using the Excel Data Analysis Toolkit. The 414 datasets used for ANN-Model development were used herein to obtain the prediction models. The input and output variables used herein are similar to the ones used in the ANN model development tasks. Using linear regression approach, the following equation for Model C was developed.

$$C = 7.555 - 0.0727A + 1.065B - 0.010CT \quad (1)$$

Statistical accuracy measures of the linear regression C model obtained using Excel Data Analysis Toolkit are: MARE (%) = 0.171%, R^2_{all} = 0.996 and Standard Deviation of Error, SDE, (%) = 0.255%. The accuracy comparison of ANN Model and Regression Model is listed in Table 3. It is very clear from Table 3 that the ANN model is slightly outperforming the regression-based model. This indicates that the modeled behavior is fully linear. For this reason, as expected, the ANN-based model will not show significant improvements over the linear regression type model.

Similarly, using linear regression approach, the following equation for Model D was developed.

$$D = -129.1371 + 0.1679A + 0.5463B + 0.0175CT \quad (2)$$

Corresponding statistical accuracy measures for this model are: MARE (%) = 1.30%, R^2_{all} = 0.909 and Standard Deviation of Error, SDE, (%) = 1.762%. The statistical comparison of ANN Model and Regression Model are depicted in Table 4. It can be concluded from Table 4, that the ANN model is outperforming the regression-based model. This indicates that the modeled behavior is fairly linear. Accordingly, ANN-based model will generally show notable improvements over its linear regression counterpart, as was noted in various similar previous research studies.

Table 3 Statistical Accuracy Measures for Models of C

Statistical Measures	ANN (3 – 3 – 1)	REGRESSION
MARE (%)	0.164%	0.171%
SDE (%)	0.245%	0.255%
R^2	0.997	0.996

Table 4 Statistical Accuracy Measures for Models of D

Statistical Measures	ANN (3 – 3 – 1)	REGRESSION
MARE (%)	1.110%	1.300%
SDE (%)	1.449%	1.762%
R^2	0.934	0.909

4. Excel Application for the Void Model

By using the connection weights of the optimal ANN models C and D, an excel-based application is developed. In this application, the developed models by ANN and Regression to predict C and D are combined into one Excel sheet where the connection weights of Models C and D as well as regression coefficients of linear regression equations are utilized. In other words, operations of one function for Model C and one function for Model D are merged in one user-friendly application. By entering the compatible input variables for A, B and Curing time in the Excel interface, ANN- and Regression-based models utilize all 3 input values (user-provided) to predict the C and D values. Percentage of volume permeable pore space (voids) is then calculated utilizing the predicted C and D values.

5. Predicting % of Voids

By using the developed Excel sheet described in Section 4, % volume of permeable pore space (voids) are calculated for all 414 datasets. Actual and predicted values are then compared. The statistical accuracy measures using ANN-based models are: MARE (%) = 3.431%, R^2_{all} = 0.894 and Standard Deviation of Error, SDE (%) = 4.822%. Similarly, the statistical accuracy measures utilizing the linear regression models are: MARE (%) = 3.698%, R^2_{all} = 0.883 and SDE (%) = 4.928%. As can be seen from the comparison, the ANN-based prediction is slightly outperforming the regression-based prediction. Therefore, in this case, both ANN- and Regression-based predictions can be used efficiently to predict % voids typically obtained from the boil test. These models can also be used to verify experimentally-based boil test results regarding the %voids in concrete samples.

6. Concluding Remarks

In this study, a static artificial neural network with a backpropagation learning algorithm was developed to predict the Boil Test-based % voids in concrete mixes. The comparisons of the predicted responses by ANN and Regression shown in paper indicate that ANN-based models attain better prediction accuracy than the Regression models. It is apparent that both ANN and Regression models have efficiently characterized the Boil test response. Therefore, ANN- and Regression-based model can reliably be used for % void prediction tasks to reduce the duration of the 5 hours testing period as long as the input variables fall within the applicable ranges. Moreover, developed ANN and Regression models can be used to verify measured responses for planned-to-be conducted Boil tests without the need for any additional experimentally-based information. Even though the development of the ANN model requires good fundamental understanding of the Boil Test procedure and ANN knowledge, Excel-based application described in section 4, which is the utilization tool of the developed ANN models, is simple to use and does not require the user to acquire specific knowledge about ANN model development. ANN and Regression models overcome the drawback of the 5 hours testing time making it a powerful, rapid and low cost alternative to obtain the % void of concrete mixes with a reliable level of accuracy. Due to fact that the database for model development was provided by KDOT, the developed Boil Test % void prediction tool described herein may be applicable only for KDOT applications. The research procedure discussed in this study can be performed to develop similar reliable prediction models for other national or international transportation entities.

Acknowledgement

Financial support received from Kansas State University Transportation Center (K-State UTC), and cooperation from the Kansas Department of Transportation (KDOT) in collecting the experimental data, to conduct this research study are gratefully appreciated.

References

[1] ASTM C 642 (1997). “Standard Test Method for Density, Absorption, and Voids in Hardened Concrete,” ASTM International, West Conshohocken, PA, pp. 3.

[2] Najjar, Y. M. and Zhang, X. (2000). “Characterizing the 3D Stress-Strain Behavior of Sandy Soils: A neuro-Mechanistic Approach”, *ASCE Geotechnical Special Publications*, No. 96, pp. 43-57.

[3] Najjar, Y. M. and McRyenold, R. (2003) "Modeling the Joint Deterioration Behavior of PCC Kansas Pavements via Dynamic ANN Approach," ASCE Geotechnical Special Publication # 123, pp. 209-222.

[4] Najjar, Y. M. and Huang, C. (2007) “Simulating the Stress-Strain Behavior of Georgia Kaolin via Recurrent Neuronet Approach," *Computers and Geotechnics*, Vol. 34, pp. 364-361.

[5] Najjar, Y. M. and Mryyan, S. (2009). “ANN-Based Profiling: Data Importance,” *Intelligent Engineering Systems through Artificial Neural Networks*, Vol. 19, pp. 155-162.

[6] Saraswathy, V. and Song, Ha-Won (2008) “Evaluation of Cementitious Repair Mortars for Corrosion Resistance”, *Portugaliae Electrochimica Acta* 26(5), pp. 417-432.

[7] Yasarer, H. I. (2010). “Characterizing the Permeability of Concrete Mixes Used in Transportation Applications: A Neuronet Approach”, *Master’s Thesis*, Kansas State University, Manhattan, KS, 105 pp.