DOT/FAA/TC-21/44

Federal Aviation Administration
William J. Hughes Technical Center
Aviation Research Division
Atlantic City International Airport
New Jersey 08405

Application of Machine Learning Techniques to Pavement Performance Modeling

October 2021

Final Report

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/TC-21/44		

4. Title and Subtitle	5. Report Date
APPLICATION OF MACHINE LEARNING TECHNIQUES TO PAVEMENT PERFORMANCE MODELING	October 2021
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Ali Z. Ashtiani	

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
GDIT 600 Aviation Research Blvd. Egg Harbor Township, NJ 08234	
	11. Contract or Grant No. DTFACT-15-D-00007

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
U.S. Department of Transportation Federal Aviation Administration Airport Engineering Division 800 Independence Ave., SW Washington, D.C. 20591	Final Report
	14. Sponsoring Agency Code AAS-100

15. Supplementary Notes

The Federal Aviation Administration Aviation Research and Development Division COR is Jeffrey Gagnon.

16. Abstract

The Federal Aviation Administration (FAA) has initiated research to develop models of long-term airport pavement performance that can be incorporated into a methodology for extended-life pavement design. This report includes a review of machine learning (ML) techniques for prediction of long-term pavement performance. The study examined applicability of three ML techniques, namely artificial neural network, support vector machine, and random forest, based on FAA Extended Airport Pavement Life (EAPL) data. Two modeling approaches, namely static and autoregressive, were implemented for each ML method. Initial models were developed for pavement condition index (PCI) and for a subsidiary index based on PCI but considering only non-load-related components (designated anti-SCI) by considering pavement age as the primary predictor.

The study also implemented various feature selection methods based on supervised learning techniques to determine the key environmental and traffic variables affecting the pavement performance. These methods were implemented to reduce the dimensionality of input space by identifying and removing a subset of irrelevant and redundant environmental and traffic variables. Feature construction methods based on unsupervised learning techniques, including k-means clustering and Principal Component Analysis, were also applied to reduce the dimensionality of input space by constructing new inputs. Findings from the feature selection and construction analyses were used to develop ML predictive models for the performance index anti-SCI, by implementing an autoregressive approach and using random forest as the learning algorithm.

17. Key Words	18. Distribution Statement
Machine learning, Pavement condition index, Extended airport pavement life	This document is available to the U.S. public through the National Technical Information Service (NTIS), Springfield, Virginia 22161. This document is also available from the Federal Aviation Administration William J. Hughes Technical Center at actlibrary.tc.faa.gov.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	128	

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ACRONYMS

C	Penalty parameter
ε	Precision margin
ξ_j	Slack variables
k	User-specified number of clusters
γ	Kernel scale
AADT	Average annual daily traffic
AC	Aircraft categories
AC1	Aircraft category 1
AC2	Aircraft category 2
AC3	Aircraft category 3
ANN	Artificial Neural Network
Anti-SCI	Anti-Structural Condition Index (Non-load-related component of PCI)
ASOS	Automated Surface Observing Systems
AWOS	Automated Weather Observing Systems
BOS	Boston Logan International Airport
BR	Bayesian regularization
BWI	Baltimore/Washington International Thurgood Marshall Airport
CC	Correlation coefficients
CFS	Correlation-based feature selection
CMH	Columbus John Glenn International Airport
CN	Condition number
DOT	State Department of Transportation
DPrec	Days precipitation
EAPL	Extended Airport Pavement Life
ESAL	Equivalent single-axle load
FAA	Federal Aviation Administration
FC	Feature construction
FDD	Freeze degree days
FDOT	Florida Department of Transportation
FOD	Foreign object damage
FPD	Freeze precipitation days
FS	Feature selection
FThC	Freeze-thaw cycles
GSO	Piedmont Triad International Airport
IRI	International Roughness Index
KKT	Karush-Kuhn-Tucker
LGA	LaGuardia Airport
LM	Levenberg-Marquadt
LM-BR	LM with Bayesian regularization
LM-ES	Levenberg-Marquadt with early stopping
LTPP	Long-term pavement performance
MIA	Miami International Airport
MCI	Kansas City International Airport

ML	Machine learning
MLR	Multilinear regression
MSE	Mean squared error
MST	Model selection triple approach
MTOW	Maximum takeoff weights
NOAA	National Oceanic and Atmospheric Administration
PC	Principal components
PCA	Principal component analysis
PCI	Pavement condition index
PCR	Pavement condition rating
R	Pairwise correlation coefficient
R^2	Coefficient of determination
RBF	Radial basis function
RF	Random forest
RMSE	Root mean square error
RNN	Recurrent neural network
RWY	Runway
SCI	Structural Condition Index (Load-related component of PCI)
SFO	San Francisco International Airport
SL	Serviceability level
SN	Structural number
SVM	Support vector machine
Temp90	Days with temperature over 90°F
TempDiff	Average daily temperature difference
TPrec	Total precipitation
TUS	Tucson International Airport
VIF	Variance inflation factor
Weka	Waikato Environment for Knowledge Analysis

EXECUTIVE SUMMARY

Robust performance models that can predict the deterioration of pavement are a key element in developing methodologies for extending the design life of airfield pavements from the current 20 years. These models relate a performance indicator to various interactive features such as age, climate, and traffic. Models based on conventional regression analysis are inadequate for establishing pavement performance prediction models as a function of multitude design features. This study aims to investigate machine learning (ML) techniques for evaluating the most informative features that affect the pavement performance. Machine learning methods are chosen because of their capability to efficiently solve nonlinear problems with high dimensionality and to find predictive functions based on data.

This study included a review of application of ML techniques for prediction of pavement performance. The asphalt-surfaced pavement performance data from the FAA's Extended Airport Pavement Life (EAPL) database were used for this review. The applicability of ML techniques for analyzing the EAPL data was examined by developing pavement performance prediction models for pavement condition index (PCI) per ASTM D5340 requirements and for an anti-structural condition index (anti-SCI), which is similar to PCI but which considers only non-load-related distresses. These models considered pavement age as the primary predictor. Three ML techniques, namely Artificial Neural Network, Support Vector Machine, and Random Forest were employed in this study. Two modeling approaches, namely static and autoregressive, were implemented for each ML method. In the static approach, the pavement functional age was the only predictor. In the autoregressive approach, the previous condition measurement as well as the time lapse between the two measurements were also considered as the predictors. The results indicated that the ML methods based on the static approach can produce pavement deterioration curves as accurately as a polynomial regression model. The results also showed that autoregressive models are capable of predicting the future performance with a high accuracy.

To expand the performance models beyond pavement age, climate and traffic attributes were considered as predictors. A series of climate variables were identified, and various supervised and unsupervised learning techniques were applied to determine the key climate parameters affecting the pavement performance. Feature selection and feature construction methods were implemented to reduce the dimensionality of input space by identifying and removing a subset of irrelevant and redundant variables that can decrease the model accuracy and quality.

This study also examined the correlation between two pavement performance indices (PCI and SCI) and traffic data in the EAPL database. The intention was to determine whether, based on the available data, aircraft traffic is a significant contributing factor to the pavement performance, and if it can be included as an independent variable in the ML prediction model. With limited exceptions, historical runway traffic data are only available for 2014 and later, while the majority of performance data in the database are for years prior to 2014. Therefore, the annual average traffic index was calculated using the data from later years. The regression analysis indicated poor correlation between the average traffic index calculated by this method and the performance indices.

Findings from the climate data analysis were used to develop predictive models for anti-SCI using an autoregressive approach and a random forest as the base learner. Three approaches were implemented to train the prediction model by considering different representations of input data. The results indicated that including climate variables improved the prediction performance.

1. INTRODUCTION

1.1 BACKGROUND

One goal of the Federal Aviation Administration (FAA) Airport Technology Research Program is to develop robust airfield pavement performance models that can be incorporated into a methodology for extended-life pavement design. Pavement life in this context refers to the functional life that is not directly considered in the FAA thickness design models. Pavements designed to meet the current 20-year standard are likely to experience functional failure before reaching their expected design life. Since 2012, the FAA Airport Technology R&D Branch has gathered a variety of airfield pavement data focused on extending airport pavement life and stored them in an Extended Airport Pavement Life (EAPL) database. The FAA EAPL database is designated PA40, as it will be referred to throughout this report. The data collected include performance measures, such as surface friction, profile roughness, and surface distresses that potentially yield pavement functional failure. The PA40 database also contains maintenance work histories, historical pavement condition index (PCI) data as well as historical runway usage and weather data. The FAA has proposed a performance index, called serviceability level (SL), which is a comprehensive index of indexes quantifying the suitability of a pavement for use by aircraft. A preliminary study by the FAA (Ashtiani et al., 2019) proposed a framework for the development of SL as a function of the PCI, roughness, and groove indices. This preliminary study considered pavement age the only independent variable, assuming the age is a surrogate for traffic and weather events. The study developed regression models to predict each component of SL as a function of pavement age using EAPL data. The study also implemented logistic regression models to develop the SL index.

A major challenge in developing the SL solely based on age is that the effects of other factors, such as climate conditions and traffic levels, are missed. Age-based models may predict the same service life for two pavement sections experiencing different levels of traffic and located in different climatic regions. This, in turn, could lead to errors in predicting maintenance timing. The conventional approach in pavement management is to develop separate regression models for related families of pavements. However, models based on conventional regression analysis are inadequate for predicting long-term pavement performance as a function of multiple design features. On the other hand, methods based on computational intelligence (CI) can be promising in contributing multiple features.

1.2 RESEARCH OBJECTIVES

In this study, machine learning (ML) solutions are examined as an analysis tool for the FAA's EAPL database to identify key elements or parameters as part of the development of reliable pavement performance models. The objectives of this study are:

- Investigate available ML technologies and determine which applications are best suited for analysis of the EAPL data.

- Perform preliminary data analysis to identify key parameters or data attributes in the EAPL data.

The report is organized into following five major sections to address the objectives of the study:

1) Investigation of ML technologies to determine the applications that are best suited for analysis of the EAPL data.

2) Validation of the ML concepts and techniques for this data set by developing PCI vs age models.

3) Identification of the most informative attributes to be included in ML predictive models.

4) Development of a preliminary predictive model for pavement performance.

5) Findings and recommendations for future research.

2. MACHINE LEARNING TECHNOLOGIES FOR PAVEMENT PERFORMANCE MODELING

2.1 PAVEMENT PERFORMANCE MODELING

Under the EAPL program, the FAA introduced SL as a means to define pavement life. Serviceability refers to the ability of the pavement to carry aircraft loads while providing: (a) a smooth ride, (b) a tractive surface, and (c) low foreign object damage (FOD) potential. The SL is a combined index of various structural and functional condition indices. Key performance indicators affecting runway serviceability include:

- PCI, or its load-related component (SCI) and non-load-related component (anti-SCI). Anti-SCI is scaled from zero to 100, and for flexible pavements is calculated using all distresses defined in ASTM D5340, except for alligator cracking and rutting

- Roughness indexes. This study used the FAA software ProFAA to compute four key indices: Boeing bump index (BBI), straightedge index (SEI), profilograph index (PI), and international roughness index (IRI).

- Skid resistance. In this study the surface skid resistance is represented by the percent of grooves that have a width or depth less than 0.125 inches, as computed by the FAA's ProGroove software program, which evaluates the grooves of longitudinal profiles.

Pavement performance models attempt to relate the pavement functional condition indices to a set of explanatory variables, such as pavement age, runway usage, and climate conditions, as well as design, construction, and maintenance features. The numerous performance modeling approaches proposed for both highway and airfield pavements fall into three broad categories: deterministic models, probabilistic models, and CI models.

2.1.1 Deterministic Models

Deterministic models predict a condition index over time for a given traffic load, climate, and structural condition. Common models based on linear or nonlinear regression analysis use the deterministic approach where the functional form of the model is explicitly defined. The robustness of these models is in doubt when the performance model considers many explanatory variables.

2.1.2 Probabilistic Models

Probabilistic (or stochastic) models were developed on the premise that the deterministic models cannot address the inherent variability in traffic loads, climate conditions, and pavement condition measurements (Li et al.,1996; Yang et al., 2003). A popular probabilistic model for pavement performance is based on the Markov chain model, in which the probability of a change from a given pavement condition to another condition is determined. This model can handle the uncertainty in pavement prediction performance (Alharbi, 2018).

2.1.3 Computational Intelligence

ML, which is a subcategory of CI, is a sophisticated analysis approach for deriving definitive information out of large sets of data for classification, regression, and prediction. ML uses powerful tools to determine interdependencies among a long list of potentially disparate variables. There is a growing interest in using ML techniques for pavement performance prediction modeling. Popular ML methods for pavement performance prediction include artificial neural networks (ANNs), support vector machines (SVMs), decision trees, random forests (RFs), radial basis functions, and Bayesian networks.

2.2 MACHINE LEARNING MODELS

ML methods operate in two types: supervised and unsupervised. In a supervised learning process, the training input and output data are accurately labeled. Data labeling is the process of parametrizing a group of data features with one or more labels. The labeled inputs and outputs (targets) are mapped to learn the functions that approximate the relationship between them. The goal of an ML model is to increase the learning accuracy while maintaining low variance and avoiding overfitting to noise. In unsupervised learning methods, the outputs are not labeled. The unsupervised methods are used to investigate the clusters, patterns, and correlations among input data. Supervised learning methods are often used for regression, classification, and forecasting analysis:

- Regression is a function approximation method that uses continuous functions to predict the target value with respect to the given set of input variables. Popular ML models for regression are feed-forward ANN, SVM, K-nearest neighbor, Bayesian networks, and RF.

- Classification is a method to predict a data point class as part of a given a set of features. The outputs are discrete-valued classes. Binary classification means that data are classified (grouped) into only two distinct classes. Common ML models for classification include logistic regression, Bayesian network, SVM, RF, decision tree, and ANN.

- Forecasting models are used to predict the future values or behavior of a target given the past observed time-series values (Harvey, 1990).

The input data in these contexts are also expressed as predictors, independent variables, explanatory variables, attributes, or features. The output data can be in the form of target, response, or class.

2.3 MACHINE LEARNING FOR TIME-SERIES MODELS

Time-series models describe the time dependency of the data where the data points are indexed in order by time. Many time-series data exhibit temporal dependency or autocorrelation, meaning that there are relationships between lagged observations.

Models for forecasting time-series data were introduced based on linear statistical methods such as autoregressive integrated moving average (ARIMA) (Box & Jenkins, 1970). It has been observed that the linear methods lack the robustness to deal with many real applications (Bontempi et al., 2013). Most recently, ML models have been proposed as an alternative to linear models. In many applications, one-step forecasting is desired, meaning that the previous time-series data are used to predict one future target. In a method proposed by Bontempi et al (2013), the problem of one-step forecasting can be addressed as a problem of supervised learning. The time-series model is formulated as a nonlinear autoregressive model in which observations from previous time steps (y_{t-1}, y_{t-2}, …, y_{t-n}) are used as input of the supervised learning to predict the value at the next time step (y_t). The autoregressive model can be extended to include exogenous input variables (X) in a time-series manner. Below is an example of how the prediction can be expressed:

$$y(t) = f(y_{t-1}, \dots, y_{t-d}, X_{t-1}, \dots, X_{t-d}) \quad (1)$$

2.4 MACHINE LEARNING IMPLEMENTATION

A key factor for a successful ML process is the quality and representation of data. Noisy and unreliable data as well as irrelevant and redundant information adversely impact the performance of supervised learning (Kotsiantis et al., 2006). A critical step toward the implementation of ML is data preprocessing to ensure data quality, adequacy, and relevancy. Data preprocessing includes, but is not limited to, outlier detection, data normalization, and feature engineering. Data normalization is often applied in ML when the features are in different scales. For example, one climate variable may be in degrees Fahrenheit (average daily temperature) and another may be a count (freeze-thaw cycles). Normalization transforms the features to be on a similar scale. Two common methods for normalization are the min-max normalization where the data are normalized to the range of (-1, +1), and the Z-score that transfers the features to have zero mean and a standard deviation of 1. Normalization may improve the performance and stability of supervised learning. Feature engineering is the process of converting raw data into sets of feature vectors (input variables) that provide the best prediction model performance. Reducing the number of variables to a subset of useful features is desirable for an effective ML prediction. The problem with high-dimensional features is that having more dimensions complicates the ML learning process and increases the difficulty of gauging the influence of each feature on the prediction. In addition, models with high numbers of features relative to the number of data samples are prone to

overfitting. To decrease the dimensionality of the feature domain, feature engineering often implements the two techniques of feature selection (FS) and feature construction (FC). FS is the process in ML that reduces the dimensionality of data by identifying and removing a subset of irrelevant and redundant variables. FC reduces the dimensionality of the data by constructing new features from the original feature set. The product of data pre-processing is the final training set for ML prediction, regression or classification.

The process of building ML models for data-driven applications is an iterative process. Kumar et al. (2016) proposed a model selection triple (MST) approach, which is the process of building a precise prediction model. It has three components: feature engineering, algorithm selection, and parameter tuning. Algorithm selection is the process of picking a ML model suited for the prediction goals. Many ML methods have multiple learning algorithms, which can be used for regression, classification, and prediction. Algorithm selection is often an iterative process requiring technical knowledge and experience. The best candidate ML models are determined depending on the size and type of the data (e.g., time dependent) and sparsity of the data (infrequent, missing, irregular data). Parameter tuning is the process of determining the value of hyper-parameters of ML algorithms. In ML, hyper-parameters refer to parameters that are set prior to training and then tuned to improve the training performance. For each ML model, there are unique hyper-parameter configurations (e.g., loss functions, search strategies, etc.) that impact performance. Hyper-parameters are fine-tuned iteratively to create a trade-off between the performance accuracy (bias) and variance of the ML models. The flowchart in figure 1 depicts the ML process.

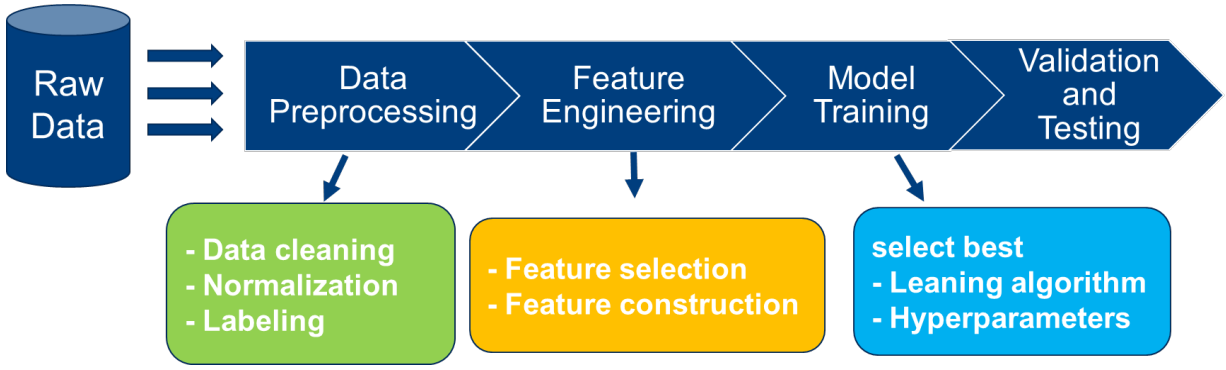


Figure 1. Machine Learning Process

2.5 MACHINE LEARNING MODEL EVALUATION

As part of the iterative MST process, appropriate model performance measures are required. Crucial to estimating these performance measures is the use of cross validation to confirm the stability of the trained models. Common cross validation methods in ML applications include k-fold cross validation, holdout method, and leave-one-out cross validation. These approaches use multiple training and test sets to ensure the model does not overfit, i.e., accurate performance estimates can be obtained when deploying models to new data. The best model is selected based on various criteria including accuracy of prediction (using mean square error [MSE], F1 score, receiver operating characteristic [ROC] curve, etc.), generalization (similar performance on new data), and qualitative criteria. “Qualitative” in this context refers to methods that employ the

judgment of experienced pavement engineering practitioners to verify whether the developed models conform to the conventional engineering logic.

2.6 MACHINE LEARNING FOR PAVEMENT PERFORMANCE MODELING

Pavement performance models can be categorized into two broad approaches: continuous function approximation with regression or discrete time-series prediction. The regression approach is the most commonly adopted method for describing the deterioration of pavement performance. In this approach, the performance at age *t* is estimated for a set of explanatory variables (age, climate, traffic, etc.) at age *t*. This approach is also called *static* modeling (Yang et al., 2003). The majority of existing performance models are based on nonlinear regression. In the time-series approach, the lagged values of a performance index are used in addition to other variables to predict future performance. In this approach, it is required that the explanatory variables be available at lagged intervals. This approach is also called *autoregressive*. The pavement performance data are categorized as *longitudinal* autoregressive when the same samples are repeatedly measured at different time points (Tsagris et al., 2018).

The following paragraphs review applications of ML to the development of pavement performance models. The reviewed studies are mostly from highway pavements, in which data from the Long-Term Pavement Performance (LTPP) study or State Department of Transportation (DOT) databases were used. There are very few studies on the application of ML on performance modeling of airport pavements.

Ozbay and Laub (2001) developed ANN-based models for prediction of highway roughness (IRI) progression using LTPP data and compared the results with linear regression models. They considered three modeling frameworks. The first model considered age, cumulative equivalent single-axle load (ESAL), and structural number (SN) as predictors. The second model considered the first measured IRI as a predictor with a SN and the time and accumulated ESALs between the initial IRI and target IRI. The third model was similar to the second model, except ESALs were not considered. Initially, 16 GPS-2 sites were used with a total of 101 data points for modeling. The backpropagation learning algorithm was used as the training rule. The errors from ANN models were relatively high. The results showed that the first ANN model returned higher errors than the regression model, while the ANN errors were lower in the second and third models with a larger number of input variables. Researchers added 118 more data points from other sites and tested both ANN and regression models. Employing additional data did not increase the accuracy of the linear regression models, but did improve the accuracy of the ANN model. In general, the second model with a higher number of predictors produced higher performance. It was also concluded that one hidden layer with three nodes was the optimal network.

Shekharan (2000) developed a hybrid method of genetic adaptive ANN training models to predict pavement condition ratings (PCR). PCR is a combined index, derived from distresses and roughness, formulated for the Mississippi DOT. The predictors chosen for the training were pavement structure, age, and cumulative ESALs. They also used the same data for creating models based on regression analysis and concluded that the ANN models performed better.

Yan et al. (2003) developed models for prediction of overall PCR using Florida Department of Transportation (FDOT) data. The overall pavement condition was based on three performance

measures including crack, rut, and ride ratings. The minimum of these measures determines the PCR. The FDOT database included time-series performance data from 1976 to 2001. Researchers supplemented missing time-series data by applying moving average methods to smooth the irrational oscillation of time-series data. Individual ANN-based models were developed for prediction of each performance measure. In these models, data from up to the previous 5 years were used to predict the next year's performance. The data types used for modeling the prediction of the future crack included the crack at previous years, the average rut and riding indices of the previous years, and pavement age and maintenance cycles. The ANN architectures with one hidden layer and 11 and 22 neurons were found to be the optimal model for rigid and flexible pavement, respectively. It was concluded that the developed ANN models for each individual condition indices have satisfactory capabilities in predicting the future condition up to a period of five years.

Marcelino et al. (2019) used time-series data to develop 5- and 10-year IRI predictive models. The models were based on an RF algorithm. Data from the LTPP were used in the study. The predictors included lagged values of IRI, climate (annual average precipitation, annual average temperature, annual average freeze index, minimum annual average humidity, and maximum annual average humidity), traffic (cumulative annual average daily truck traffic), and pavement structure (thickness and SN) up to 5 years. It was concluded that the models with five lagged values produced lower errors than the model that used only one lagged value. The 5-year predictive model was more reliable and accurate than the 10-year prediction. A sensitivity analysis concluded that the previous IRI values were the most significant predictor influencing the predictive performance. Other predictors (climate, traffic, and structure) had minimal effect on the prediction of IRI.

Amin et al. (2017) proposed an ANN-based model for PCI prediction using data from collector and arterial roads of Montreal, Canada. They applied backpropagation with the generalized delta rule learning algorithm. The predictors were average annual daily traffic (AADT), ESALs, SN, pavement age, and difference in the PCI between the current and preceding year. The analysis results indicated that the previous PCI significantly influences the estimation of PCI values followed by the AADT and ESALs. Pavement age and structural characteristics of the pavement had insignificant influence in determining the PCI.

Kargah-Ostadi (2013) developed predictive models for IRI using three ML methods. In the first method, feed- and cascade-forward ANN were used, and two backpropagation learning techniques, Levenberg-Marquadt (LM) and Bayesian regularization (BR), were tested. The second method was SVM with polynomial and Gaussian kernel functions. The third method was a radial basis function (RBF) network. The models were built using 14 predictors including previous IRI, the number of ESALs since previous IRI measurement, time lapse since previous IRI, pavement age, three thickness values of pavement layers, climate parameters (precipitation, freezing index, freeze-thaw cycles and number of days above 32°F) since last IRI, and three subgrade material properties. Principal component analysis (PCA) was implemented, which resulted in reduction of the input variables to 10 principal components. It was concluded that the feed-forward ANN with BR learning was the most accurate and had the best generalization capability. However, a sensitivity analysis showed that the predicted IRI had an inverse relationship with the time and traffic, making the model unreliable. While the accuracy of the RBF model was not high, the IRI predicted by the model was more realistic than other methods and was consistent with expected pavement performance.

Alharbi (2018) developed multilinear regression (MLR) ANN models to predict the cracking, riding, and rutting index for asphalt and concrete pavements based on Iowa DOT pavement management system data. The predictors considered for asphalt pavement ride index included pavement age, previous IRI, pavement thickness, subgrade stiffness, average rainfall, and average temperature. The predictors for cracking index were pavement age, previous cracking index, average temperature, and SN. The predictors for rutting index were pavement age, previous rutting index, average temperature, and pavement thickness. The results indicated that ANN produced better accuracy than MLR models in terms of coefficient of determination (R^2) and root mean square error (RMSE). Pavement age followed by average temperature were found to be the most significant predictors for all asphalt pavement performance indices.

Tabatabaee et al. (2013) employed a two-stage model for prediction of pavement serviceability index (PSI) using data from the Minnesota DOT's Office of Materials and Road Research (MnRoad). In the first stage, SVM was used to classify pavement sections into families of sections with similar pavement structure characteristics, such as asphalt material and mix properties. Three pavement structure types were identified and then used as categorical inputs to a recurrent neural network (RNN) model to predict PSI. Other predictors of the RNN model included pavement age, previous PSI, season (summer or winter), yearly ESAL, maintenance type (preventive=1, otherwise=0), and treatment history. The RNN model employed one hidden layer with five hidden neurons and LM as the training algorithm. The results showed that the two-stage model in which the pavement material characteristics were used to group the pavement structures significantly improved the prediction performance, compared to a one-stage model that considers each pavement structure's characteristics as input variables. Sensitivity analysis indicated that treatment history had minimal impact on the model, since the pavement deterioration rate did not change significantly before and after treatment.

Ziari et al. (2016) used SVM for IRI predictions using LTPP database data. Five kernel functions were evaluated in this study. The predictors included pavement age, climate factors (e.g., annual average precipitation, temperature and freezing index), ESAL, asphalt, and total pavement thickness. The results indicated that the Pearson VII universal kernel is capable of predicting pavement performance in the short and long terms of the pavement life cycle. The Gaussian kernel function showed reasonable accuracy for short-term prediction of IRI.

2.7 REVIEW OF MACHINE LEARNING MODELS

A review of the most relevant literature on the application of ML techniques to pavement performance modeling indicated that ANN, SVM, and RF, all yielded acceptable performance. Sections 2.7.1 through 2.7.3 provide a brief introduction to these ML methods.

2.7.1 Artificial Neural Network

ANN is a parallel information processing machine consisting of processing units called neurons. Neurons are organized into three subsets, namely input, hidden (intermediate), and output layers as shown in figure 2. The number of neurons in the input and output layers is often determined by the number of independent and dependent variables, respectively. The number of neurons in the hidden layer is chosen depending on the problem. The connections between the neurons in each

layer to the neurons in the next layer are made by means of weights, biases, and transfer functions as presented in equation 2:

$$y_i = f_i\left(\sum_{j=1}^{n} w_{ij}x_j + b_i\right) \tag{2}$$

where, for neuron i, y_i is the neuron output, x_j is the jth input to neuron i, and w_{ij} is the connection weight between neurons i and j. b_i is the bias of the input neuron and f_i is the transfer function. The neuron output will be the input feature for the connected nodes in the next layer. The most common transfer functions are linear functions and nonlinear sigmoid functions (hyperbolic tangent and logistic). Multiple hidden layers with adequate numbers of neurons and nonlinear transfer functions equip the network to establish nonlinear relationships between input and output vectors.

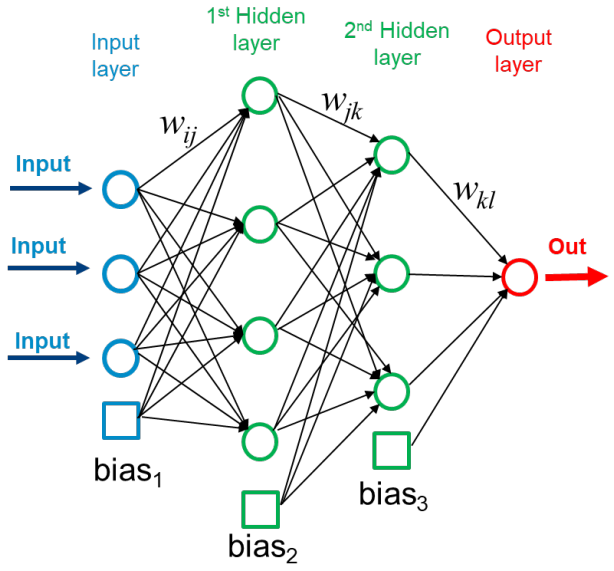


Figure 2. Typical Architecture of ANN Models

The feed-forward multilayer perceptron is the most popular network structure. In this approach, neurons in each layer are only connected to neurons in layers that are closer to the output. In ANN, the relationship between the input and output vectors or the patterns between the data is learned. This learning is also called the training process. The connection weights and biases are adjusted iteratively through the training process with the goal of minimizing the errors between the model output (predicted) and the target output, defined as the objective function. MSE, shown in equation 3, is the most popular performance metric used in training:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{3}$$

where, n is the number of data points, Y is the measured or desired response, and $\hat{Y}$ is the predicted response. The objective of training is to numerically optimize the objective function subject to the model's trainable parameters. The optimization model is based on gradient descent, in which weights and biases are randomly initialized and then adjusted repeatedly such that the performance (e.g., MSE value) moves in the direction of the steepest gradient of the error surface

(Samarasinghe, 2007). This process is called backpropagation. In backpropagation, the error is sent back to the network to adjust the weights and biases to improve the performance (minimize the error).

Backpropagation training methods based on gradient descent are susceptible to finding suboptimal solutions at local minima, thereby reducing the generalization capability of the trained model. A common problem in neural network models is overfitting, which occurs when a network is memorized in the training set but has not yet learned to generalize to new inputs. In this situation, the network produces a relatively small error on the training set but a much larger error when new data is fed into the model. A robust network training should provide a balance between the bias (error) and variance (fitting to the noise).

2.7.2 Support Vector Machine for Classification

SVM is an ML method based on statistical learning theory that builds a model of the data being observed and outputs a function that can be used to predict some features of future data (Cristianini & Schölkopf, 2002). SVM was initially developed for data classification with exactly two classes. The classification operates by finding the best hyperplane that separates all data points of one class from those of the other class. The largest margin between the two classes is indicative of the best hyperplane. The margin represents an area with no data points of either class. Figure 3 illustrates an example of a classification problem in a pavement engineering application where the pavement is classified into serviceable and unserviceable based on the value of two pavement performance measures, PCI and roughness. Consider the training data as a set of points x_j, along with their classes y_j, where $x \in R^d$(d=2, a two-dimensional problem in this example) and $y=\pm 1$. The equation of the hyperplane is $f(x) = w.x + b = 0$. The optimum hyperplane with maximum margin is obtained by minimizing the norm of w as follows:

$$\min \frac{1}{2} \|w\|^2 \tag{4}$$

$$subject\ to \qquad y_j(w.x_j + b) \geq 1 \tag{5}$$

The support vectors are the x_j on the boundary planes where $y_j.(w.x_j+b) = 1$.

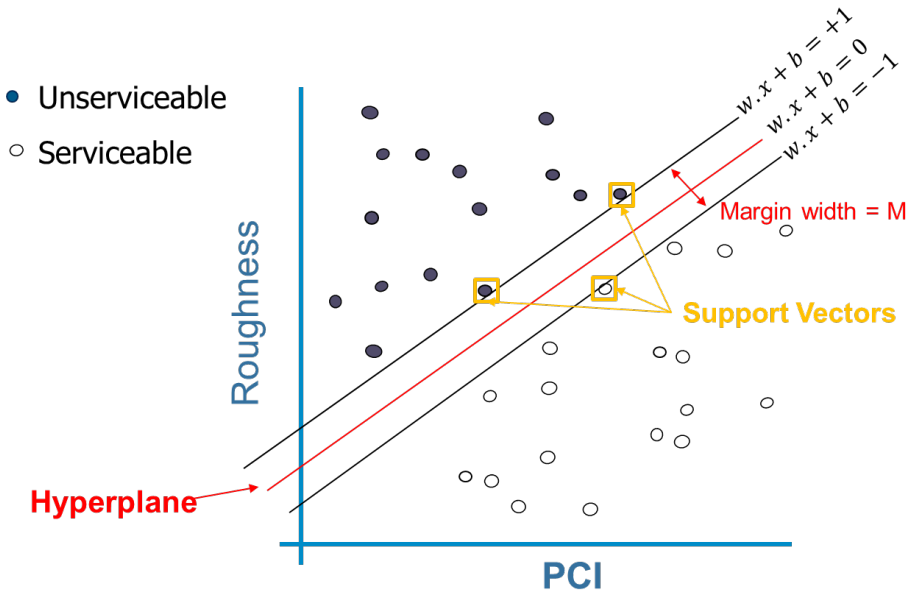


Figure 3. SVM Classification for Pavement Serviceability

The above optimization problem is only feasible for data that can be separated by a linear hyperplane. For inseparable data, SVM applies the soft margin concept, meaning that the hyperplane separates most but not all of the data (figure 4). In this case, the soft margin formulations allow for some errors in classifications to avoid overfitting to noise. This is done by introducing slack variables ξ_j and a penalty parameter C to the optimization function:

$$\min \frac{1}{2}\|w\|^2 + C\sum_j \xi_j \tag{6}$$

$$subject\ to \qquad y_j\left(w.x_j + b\right) \geq 1 - \xi_j \tag{7}$$

The first term in equation 6 refers to margin error, and the second term refers to classification error. The optimization problem can be expressed as a dual quadratic optimization by constructing the Lagrange function from the objective function. The final form of the optimization function is shown in equation 8:

$$\begin{aligned} &\max \sum_j \alpha_i - \frac{1}{2}\sum_j \sum_k \alpha_j \alpha_k y_j y_k x_j{}' x_k \\ &\alpha_j = C - \mu_j \quad \left(\mu_j \text{ is the Lagrange multiplier}\right) \\ &\text{subject to} \qquad \sum_j y_j \alpha_j \\ &0 \leq \alpha_j \leq C \end{aligned} \tag{8}$$

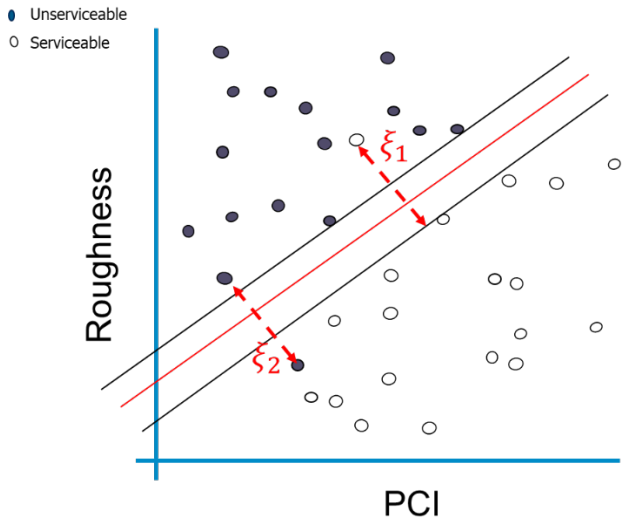


Figure 4. Soft Margin SVM Classification

Classification of data is not always feasible with a simple hyperplane. Nonlinear SVM methods have been proposed to solve these problems. In these methods, instead of using nonlinear hyperplanes, the training data are transferred into a higher dimensional feature space R^D by means of a mapping function ϕ, where the training data can be separated by a linear hyperplane.

Therefore, the dot product ($x_j.x_k$) in R^d for the linear case is equivalent to the dot product ($\varphi(x_j).\ \varphi(x_k)$) in R^D for the nonlinear case. The dot product in R^D space is expressed by the kernel functions $K(X_j,X_k) = (\varphi(x_j).\ \varphi(x_k))$. Three popular kernel functions are shown in equations 9 to 11:

$$\text{Linear:} \quad K(X_j, X_k) = X_j{}^T X_k \tag{9}$$

$$\text{Polynomial:} \quad K(X_j, X_k) = (1 + X_j{}^T X_k)^p \tag{10}$$

$$\text{Gaussian or Radial Basis Function:} \quad K(X_j, X_k) = \exp\left(-\gamma \left\| X_j - X_k \right\|^2\right) \tag{11}$$

The $\gamma > 0$ in equation 11 is the kernel scale, which controls the width of the Gaussian. A small value of γ results in a smooth margin boundary. As γ is increased, the curvature of the margin boundary increases, and the value of the hyperplane function becomes constant outside the close proximity of the region where the data are concentrated. In this situation, the classifier is clearly overfitting the data (Ben-Hur &Weston, 2010).

The objective function in the case of nonlinear SVM takes the following form:

$$max \sum_j \alpha_i - \frac{1}{2} \sum_j \sum_k \alpha_j \alpha_k y_j y_k K(x_j, x_k) \tag{12}$$

2.7.3 Support Vector Machine for Regression

SVM classification can be extended for regression applications. SVM regression is considered a single-classification problem. SVM regression relies on kernel functions and is a nonparametric technique, i.e., the model structure is not specified a priori but is determined from data. Suppose that the training data includes a set of N observations of multivariate predictors x_n and target values y_n. The SVM aims to find a prediction function $f(x)$ that deviates from target y_n by a value less than or equal to ε for each training point x_n, and at the same time is as flat as possible. Consider a linear function $f(x)$ as:

$$f(x) = w.x + b \tag{13}$$

To satisfy the flatness of the function, the norm of $(w.w)$ should be minimized by treating this as a convex optimization problem:

$$\min \frac{1}{2}\|w\|^2 \tag{14}$$

subject to the constraint that the error for each training data point is less than the precision error ε:

$$y_n - (w.x_n + b) \leq \varepsilon \tag{15}$$

Convex optimization is not always possible, meaning that no such function $f(x)$ exists that approximates all data points under the constraints. Analogous to the soft margin concept for the SVM classification, the optimization can become feasible by allowing regression errors to exist up to the value of slack variables ξj and ξ*j. The optimization function is expressed as:

$$\text{Min} \quad \Phi(\text{w}) = \frac{1}{2}\|w\|^2 + C\sum_j(\xi_j + \xi_j^*) \tag{16}$$

subject to:

$$\begin{gathered} y_j - \left(w.x_j + b\right) \leq \epsilon + \xi_j \\ \left(w.x_j + b\right) - y_j \leq \epsilon + \xi_j^* \\ \xi_j \geq 0, \xi_j^* \geq 0 \end{gathered} \tag{17}$$

The constant C controls the overfitting (regularization) by imposing a penalty on observations that have an error greater than the precision margin (ε). This value designates the trade-off between the flatness of $f(x)$ and the deviations exceeding the limiting threshold ε. SVM employs a linear ε-insensitive loss function to penalize estimation error. This loss function is based on the observed values y beyond the ε boundary (figure 5). The errors that are within a distance ε of the observed value are set to zero. This can be described as shown in equation 18:

$$L_\varepsilon = \begin{cases} 0 & if\ |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon & otherwise \end{cases} \tag{18}$$

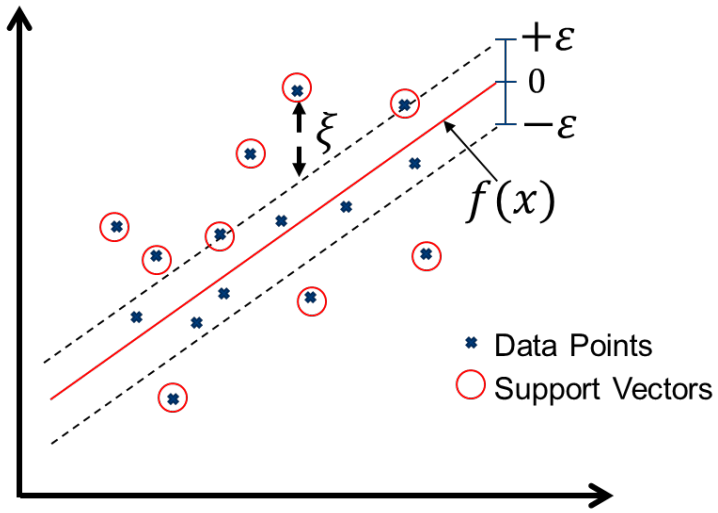


Figure 5. The Soft Margin ε-Insensitive Loss in SVM Regression

The optimization problem in equation 16 can be solved in its Lagrange dual formulation. This is done by introducing nonnegative Lagrange multipliers α_n and α^*_n for each observation x_n:

$$L(\alpha) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\, x_i' x_j + \varepsilon \sum_{i=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} y_i\,(\alpha_i^* - \alpha_i) \qquad (19)$$

Subject to,

$$\sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0 \qquad (20)$$

$$0 \le \alpha_i, \alpha_i^* \le C \qquad (21)$$

The Lagrange multipliers α_i and α^*_i are obtained by solving the optimization function in equation 19. The w and b in the regression function in equation 16 and 17 will be determined as:

$$w = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)(x_i' x) \qquad (22)$$

$$b = \frac{1}{N}\sum_{i=1}^{N}(y_i - \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)(x_i' x)) \qquad (23)$$

The optimal solution is obtained by implementing the complementarity conditions of Karush-Kuhn-Tucker (KKT) (Karush, 1939):

$$\begin{gathered} \alpha_i(\varepsilon + \xi_i - y_i + x_i' w + b) = 0 \\ \alpha_i^*(\varepsilon + \xi_i^* + y_i - x_i' w - b) = 0 \\ \xi_i(C - \alpha_i) = 0 \\ \xi_i^*(C - \alpha_i^*) = 0 \end{gathered} \qquad (24)$$

Under the KKT conditions all observations inside the ε tube have Lagrange multipliers of zero. If either α_i or $\alpha^* i$ is not zero, the corresponding observation is called a support vector.

A linear model is not always adequate in describing the variations of target values. The nonlinear SVM regression model can be obtained by replacing the dot product $x_i'.x$ with a nonlinear kernel function $K(x_i,x_j) = (\varphi(x_i).\varphi(x_j))$, where $\varphi(x)$ is a transformation function that maps x to a high dimensional feature space (R^D). The optimal function *f(x)* for the nonlinear SVM is governed by a transformed feature space in lieu of the input space:

$$L(\alpha) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)\left(\alpha_j - \alpha_j^*\right)K\left(x_i, x_j\right) + \varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) - \sum_{i=1}^{N} y_i\,(\alpha_i^* - \alpha_i) \quad (25)$$

subject to:

$$\sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0 \quad (26)$$

$$0 \le \alpha_i, \alpha_i^* \le C \quad (27)$$

The prediction function is determined as:

$$f(x) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)K\left(x_i, x_j\right) + b \quad (28)$$

The KKT complementarity conditions are:

$$\begin{gathered}\alpha_i(\varepsilon + \xi_i - y_i + f(x)) = 0 \\ \alpha_i^*(\varepsilon + \xi_i^* + y_i - f(x)) = 0 \\ \xi_i(C - \alpha_i) = 0 \\ \xi_i^*(C - \alpha_i^*) = 0\end{gathered} \quad (29)$$

2.7.4 Random Forest

RF is an ensemble of regression and classification techniques that makes predictions by combining the results from many individual decision trees. RF uses bagging, also known as bootstrap aggression, for combining the outputs of multiple decision trees into an RF. Decision trees are trained on randomly sampled subsets of the data. While an individual decision tree is often prone to overfitting, the bagging technique decreases the variance in the model. One significant advantage of RF is that it can handle the collinearity of the features in the input domain by randomly sampling subsets of data and then training the different decision trees on these subsets instead of on the entire dataset. The training subsets are then replaced as many times as the ensemble makes trees. The final result of a prediction model is determined by averaging over all predictions from the sampled trees (Breiman, 2001; Dietterich, 2000; Svetnik et al., 2003). Some of the RF hyperparameters are: number of trees in the forest, minimum leaf size, minimum number of samples required to split an internal node, and number of learning cycles.

3. VALIDATION OF ML TECHNIQUES BY DEVELOPING PCI VS AGE MODELS

In this section, three ML methods, ANN, SVM, and RF, are used to develop deterioration models for two pavement condition indices, PCI and anti-SCI. The indices will be modeled against

pavement age. The deterioration models based on ML methods will be compared to the regression models. Since the purpose of this study is to validate the application of ML models, only the performance data from asphalt surfaced pavements of the EAPL database were used for the analysis. A similar study can be done using rigid pavement data. The following runways (RWY) were considered in this study:

- Boston Logan International Airport (BOS) RWY 4L-22R
- John Glenn Columbus International (CMH) RWY 10L-28R
- CMH RWY 10R-28L
- Piedmont Triad International Airport (GSO) RWY 5L-23R
- La Guardia Airport (LGA) RWY 4-22
- Kansas City International Airport (MCI) RWY 9-27
- Miami International Airport (MIA) RWY 12-30
- Tucson International Airport (TUS) RWY 3-21
- TUS RWY 11L-29R

Two modeling approaches are used to predict each condition index. In the first model, the pavement functional age is the only predictor. This model, which is similar to regression, is referred to as a static model and is used for function approximation. The second model is an autoregressive model in which the previous condition measurement as well as the time lapse between the two measurements are also considered as the predictors. Table 1 illustrates the input variables and the output in each model.

Table 1. Input and Output Variables in ML Models

Model	Input Variable	Variable Description	Output
Static	Age (t)	Time since last major rehabilitation or reconstruction	PCI (t)
Autoregressive	Age(t)	Time since last major rehabilitation or reconstruction	PCI (t)
	PCI (t-1)	Previous PCI measurement	
	delatTime	Time since previous PCI measurement	

The average and standard deviation (SD) of the time lapse in the training dataset are 3.9 years and 2.5 years, respectively. The histogram of the time lapse is shown in figure 6. The autoregressive model is expected to predict the future pavement condition in increments from 1 to 6 years (average ±SD).

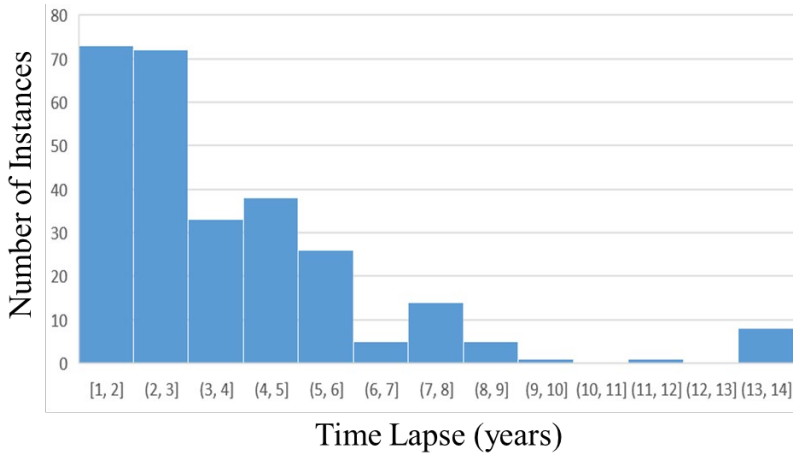


Figure 6. Time Lapse Between Two Anti-SCI Measurements

3.1 DATA PREPARATION AND PREPROCESSING

Data from the PA40 database for flexible pavement keel sections were used for developing the predictive models. Before ML training, various data preprocessing methods were applied to both the input and output data to make sure that the database was comprehensive, complete, and free of outliers. The performance data have an inherent spatial variability and errors that need to be identified, and possibly removed, prior to performing ML modeling. For example, records with improved PCI between two consecutive inspections need to be examined before using for ML prediction modeling.

The functional age at the time of each performance measurement was calculated for every pavement section. For each modeling approach, the required input and output vectors were created. In the autoregressive model, every pavement record for which the PCI was greater than the previous measurement, without the pavement having undergone rehabilitation, was considered an outlier and removed from the database.

3.2 MODEL PERFORMANCE EVALUATION

The performance of the trained ML models is evaluated based on three criteria:

1) Accuracy of prediction. The accuracy is defined as the ability of the model to achieve maximum likelihood or minimum errors. Two accuracy measures considered in this study are root mean square error (RMSE) and relative root mean square error (RRMSE) as defined in equations 30 and 31:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_{m,i} - \hat{Y}_{p,i})^2} \tag{30}$$

$$RRMSE(\%) = 100 \times \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_{m,i} - Y_{p,i}}{Y_{m,i}}\right)^2} \quad (31)$$

where, Y_m is the measured target and Y_p is the predicted target. RMSE is the standard deviation of prediction error. It indicates how concentrated the data are around the prediction fitted curve. Compared to mean average error, RMSE gives higher weights to larger errors. It is not a bounded metric, meaning that its value could be from zero to infinity.

The other accuracy measure is the coefficient of determination (R^2) of the linear regression curve of the measured versus predicted targets.

2) Generalization. Generalization is achieved when the error in training data is not significantly different than that on testing or new data.

3) Qualitative. The predicted performance should be consistent with pavement engineering principles. In the context of this study, qualitative evaluation means that the performance must exhibit a monotonically decreasing trend.

3.3 ANTI-SCI PREDICTION MODEL

This section illustrates the results of static and autoregressive anti-SCI prediction models using the three ML methods (ANN, SVM, and RF).

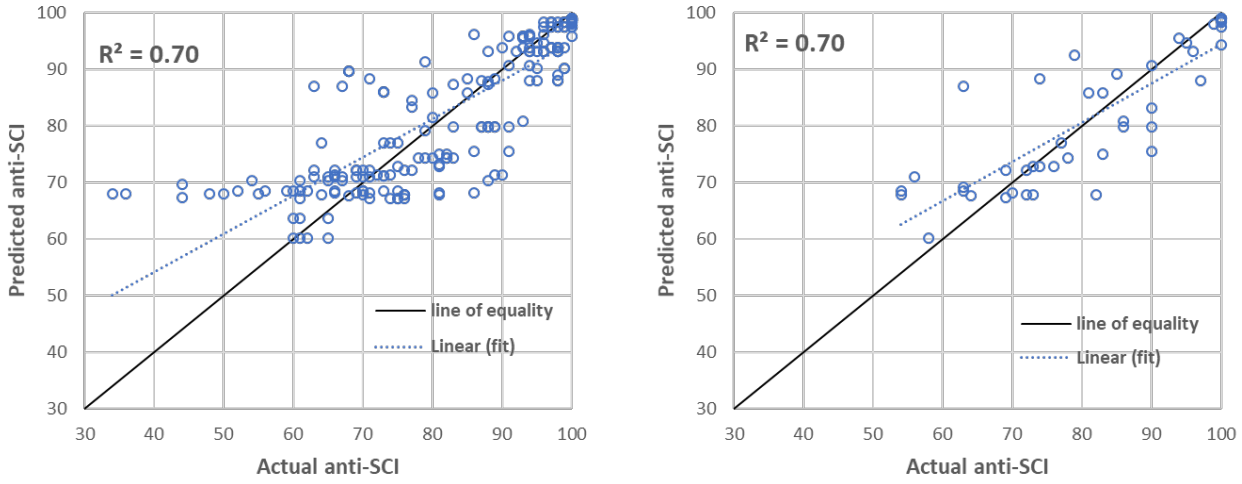
3.3.1 Anti-SCI Prediction Using ANN—Static Model

A feed-forward multilayer perceptron is used as the network structure. Two backpropagation learning algorithms were used in this study: Levenberg-Marquadt with early stopping (LM-ES) and LM with Bayesian regularization (LM-BR). In the LM-ES method, the training patterns are divided randomly into training and validation subsets. Error (MSE) on the training subset is minimized until the MSE on the validation subset starts to increase. The training process is then stopped at the minimum MSE on the validation data. Bayesian regularization minimizes a linear combination of squared errors and weights and then produces a network that generalizes well (MacKay, 1992). In this algorithm, a regularizing term is added to be minimized along with the error (MSE). This additional term typically penalizes larger weights.

To obtain the highest predictive efficiency, alternative architectures consisting of multiple hidden layers, neuron numbers, and transfer functions (sigmoid logarithm, sigmoid tangent, or pure linear) were considered. For developing the static model, a network architecture with one input layer (single neuron), one hidden layer (three neurons), and one output layer (single neuron) was identified as the optimal architecture for the prediction. This model can be described as 1-3-1 architecture. The sigmoid logistic transfer function was used for the hidden layer and a linear transfer function was used for the output layers.

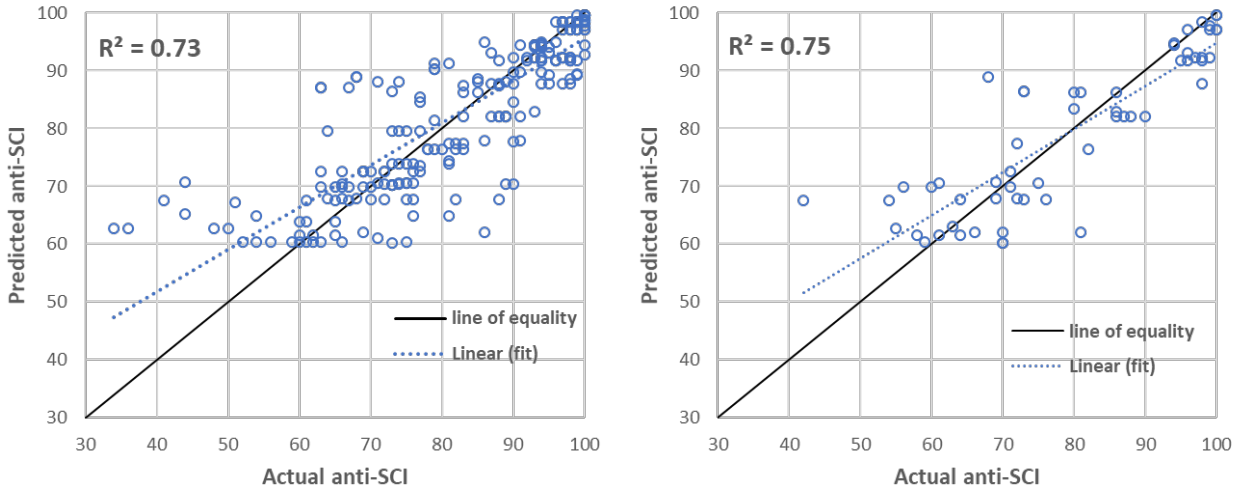
To implement the LM-ES model, from the 293 data points, 205 (70%) were selected randomly for training, 44 (15%) for testing, and 44 (15%) for validation. To implement the ANN model with Bayesian Regularization, 235 (80%) were selected randomly for training and 58 (20%) were selected randomly for testing. One section from each runway was set aside for independent testing (36 data points). The MATLAB® for Deep Learning Toolbox™ software application was used for model development. All the input and output variables were normalized to the range of [-1, +1] before starting the ANN training in order to have uniformly distributed variables. The "mapminmax" preprocessing function in the MATLAB program was used for data normalization. Following the established optimum network architecture, the model was trained multiple times using both training algorithms. The testing dataset was not used for training the ANN models.

The RMSE of the testing dataset in both LM-ES and LM-BR models is 7.7. This number indicates that the average anti-SCI prediction error is less than 7.7 units (on the 0-100 scale) from the prediction curves. Figures 7 and 8 show the scatter plots of the measured (target) versus predicted anti-SCI developed by the ANN LM-ES and LM-BR approaches for the training and testing datasets. The training and testing datasets demonstrated comparable R^2 in both models. This indicates that both training methods performed well in terms of generalization, meaning that the probability of obtaining an overfitted prediction is low. Figures 9 and 10 show the fitted curves of the predicted anti-SCI values in both LM-ES and LM-BR models, respectively. It is of concern that the LM-BR model does not exhibit a monotonically decreasing anti-SCI deterioration.



(a) Training Subset (b) Testing Subset

Figure 7. Measured vs Predicted Anti-SCI of Training (a) and Testing (b) Subsets Using the LM-ES Algorithm



(a) Training Data (b) Testing Data

Figure 8. Measured vs Predicted Anti-SCI of Training (a) and Testing (b) Data Using the LM-BR Algorithm

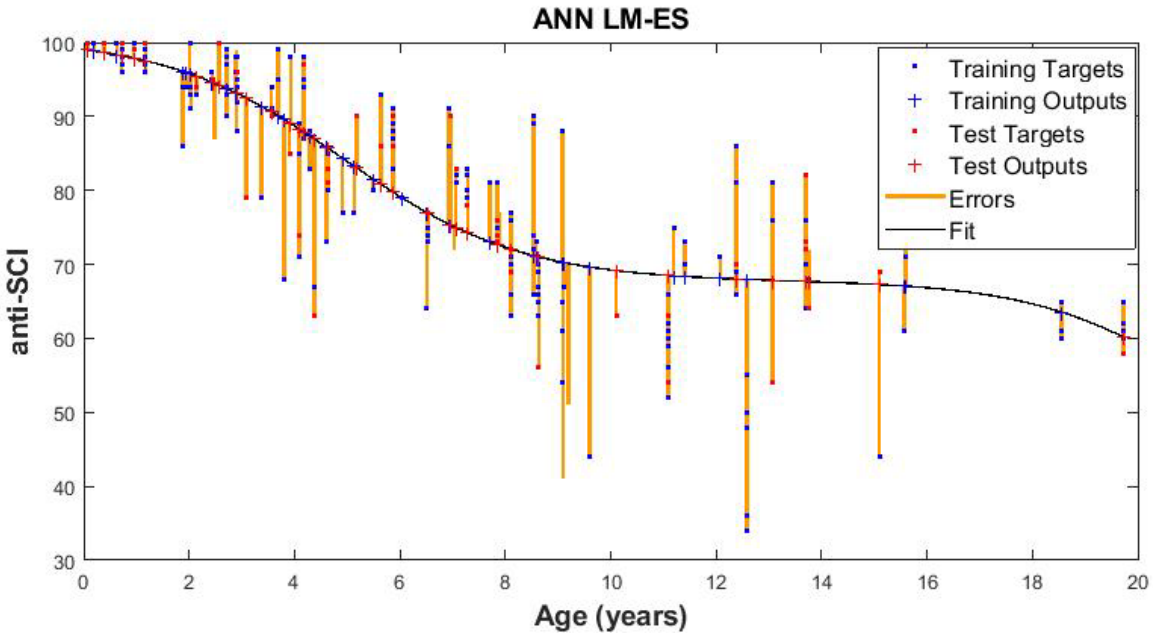


Figure 9. Anti-SCI Prediction Curve in the ANN LM-ES Model

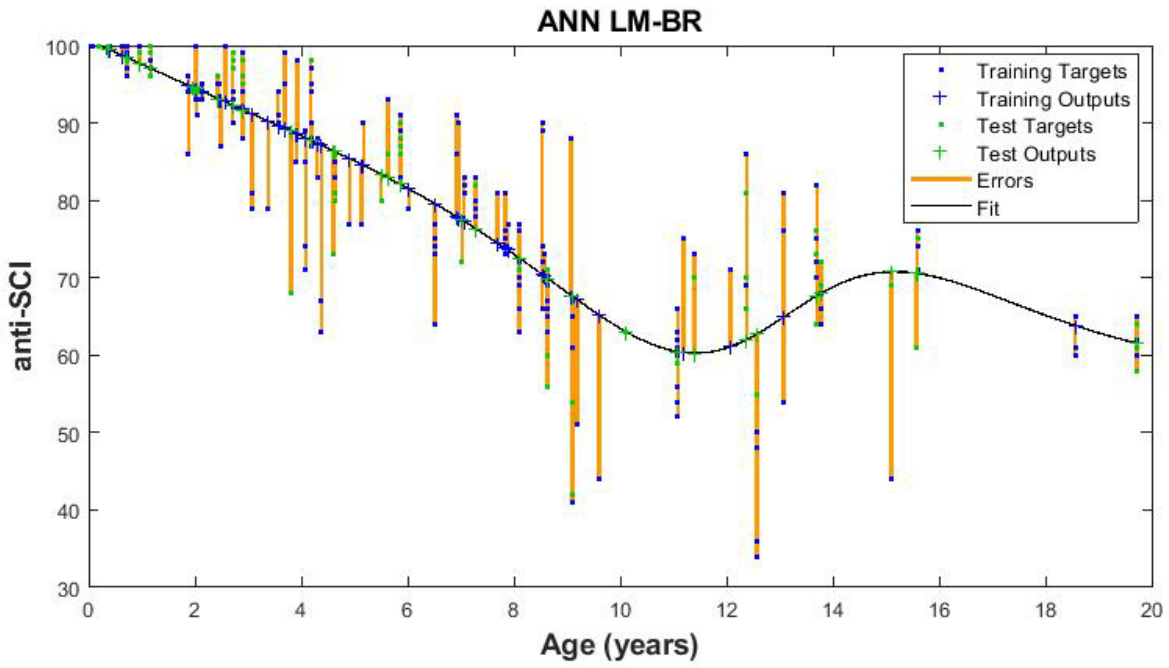


Figure 10. Anti-SCI Prediction Curve in the ANN LM-BR Model

3.3.2 Anti-SCI Prediction Using SVM—Static Model

The polynomials and Gaussian kernel functions were tested in this study. For these kernel functions, the hyperparameters that need to be tuned are kernel scale, penalty parameter C, and precision margin ε. The optimum SVM hyperparameters were determined by performing a three-dimensional grid search. The grid points are chosen from a predefined range for kernel scale γ (0.1, 0.5, 1, …, 10), penalty parameter C (0, 0.001, …, 20), and precision margin ε (0.001, 0.01, …., 10). The Gaussian kernel function (equation 11) was used in this study since it showed better performance than the polynomials. The grid points that produced the highest accuracy and at the same time resulted in a monotonically decreasing deterioration trend were selected as the SVM model hyperparameters. A 10-fold cross validation was used for evaluation of model performance. In each round of cross-validation, data were partitioned into 10 randomly chosen subsets (folds) of roughly equal size. One subset was used to validate the model trained using the remaining nine subsets. This process is repeated 10 times such that each subset is used exactly once for validation. The optimum hyperparameters resulting from the grid search and the cross-validation analysis are shown in table 2.

Table 2. SVM Static Model Hyperparameters

SVM Model Hyperparameter	Value
Kernel function	Gaussian
Kernel scale, γ	1.5
Penalty parameter, C	12
Precision margin, ε	0.001

Figure 11 shows the predicted anti-SCI deterioration curve, and figure 12 shows the scatter plot of predicted versus measured anti-SCI with R^2 equal to 0.69. It is observed from the figures that the error of the SVM model is relatively high in predicting anti-SCI less than 60.

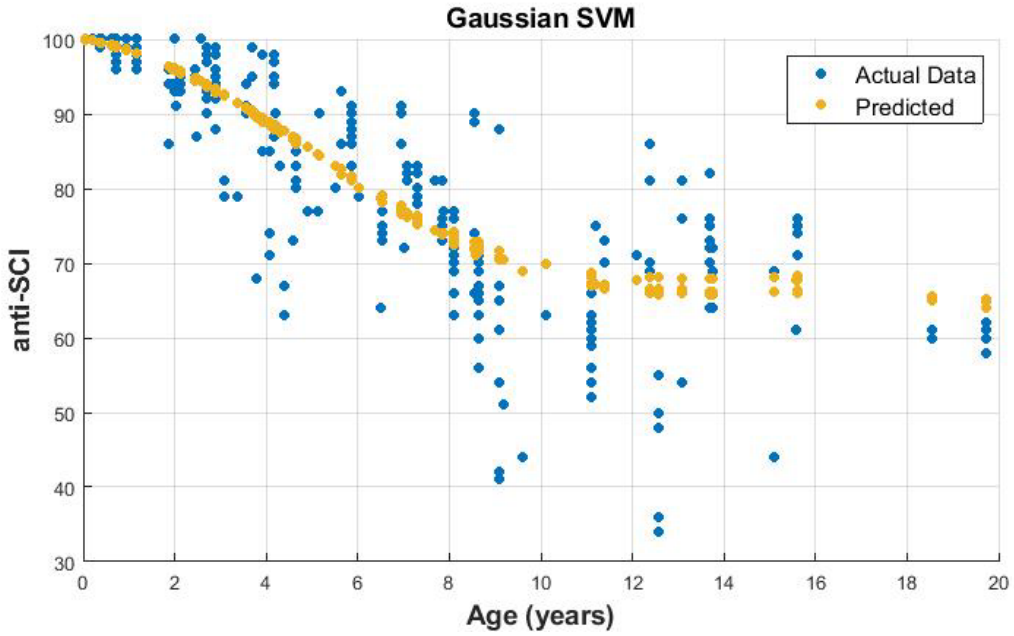


Figure 11. Anti-SCI Prediction Curve in the SVM Static Model

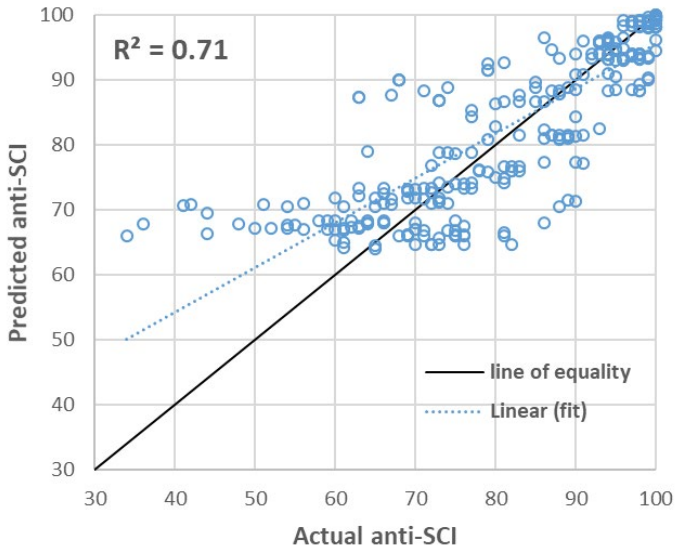


Figure 12. Measured vs Predicted Anti-SCI in the SVM Static Model

3.3.3 Anti-SCI Prediction Using RF—Static Model

The RF model employed a bagged ensemble of regression trees to predict the performance index. The out-of-bag technique was used for cross-validation to tune the regression tree hyperparameters and improve the predictive performance. The RF hyperparameters were optimized using random search. The following hyperparameters yielded the best performance:

- Number of trees in the forest = 300;
- Minimum leaf size = 8;
- Minimum number of samples required to split an internal node = 5;
- Number of learning cycles = 30.

Figure 13 shows the predicted anti-SCI deterioration curve, and figure 14 shows the scatter plot of predicted versus measured anti-SCI with R^2 equal to 0.79. The averages of RMSE and RRSE from 10 cross-validation analyses are 7.01 and 11.95. Notwithstanding that the errors are lower than the ANN and SVM models, the deterioration trend is not monotonically decreasing. This means that the developed static RF model is not suitable for representing the pavement performance over age.

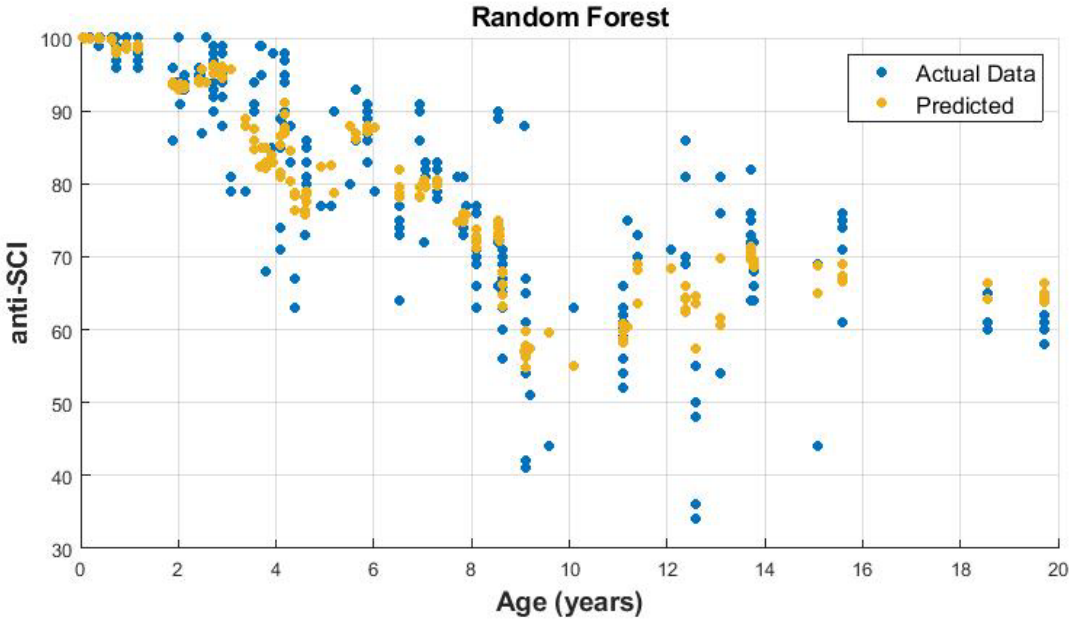


Figure 13. Anti-SCI Prediction Curve in the RF Model

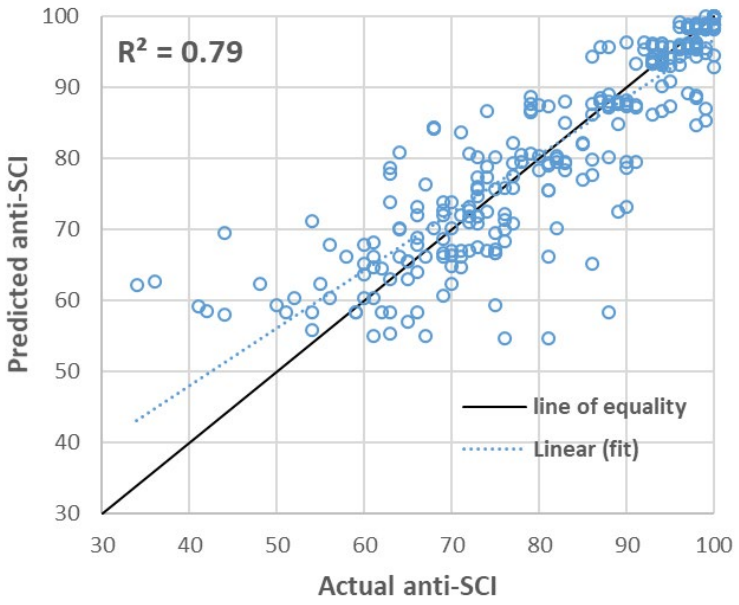


Figure 14. Measured vs Predicted Anti-SCI in the RF Model

3.3.4 Comparison of Static Models for Anti-SCI Prediction

An anti-SCI model based on regression analysis was developed by fitting a fourth-degree polynomial equation. The developed ML models were then compared to the regression model. Figure 15 shows the comparison between the anti-SCI deterioration curves from ANN LM-ES and SVM models to the regression model. Both ANN and SVM models predict a slightly higher value than the regression model, especially for the anti-SCI values beyond 6 years. Table 3 summarizes the architecture and performance of the various models used to predict anti-SCI in this study. Results indicate that while the RF model produced a lower error, the predicted deterioration curve does not exhibit a monotonically descending trend. The ANN LM-ES and SVM models outperformed the fourth-degree polynomial regression model by producing a lower prediction error and exhibiting a meaningful deterioration trend. Overall, the prediction error was high in all the models for anti-SCI lower than 60. The high error is attributed to the fact that the pavement age is incapable of explaining all the variability in anti-SCI. Other variables such as climate and weather condition can affect the deterioration of pavement. Figure 16 shows the actual versus predicted anti-SCI values on the independent testing data. All models exhibited R^2 values on the testing data set that were reasonable and comparable to the R^2 values from the training data set. In other words, they yielded good generalization performance.

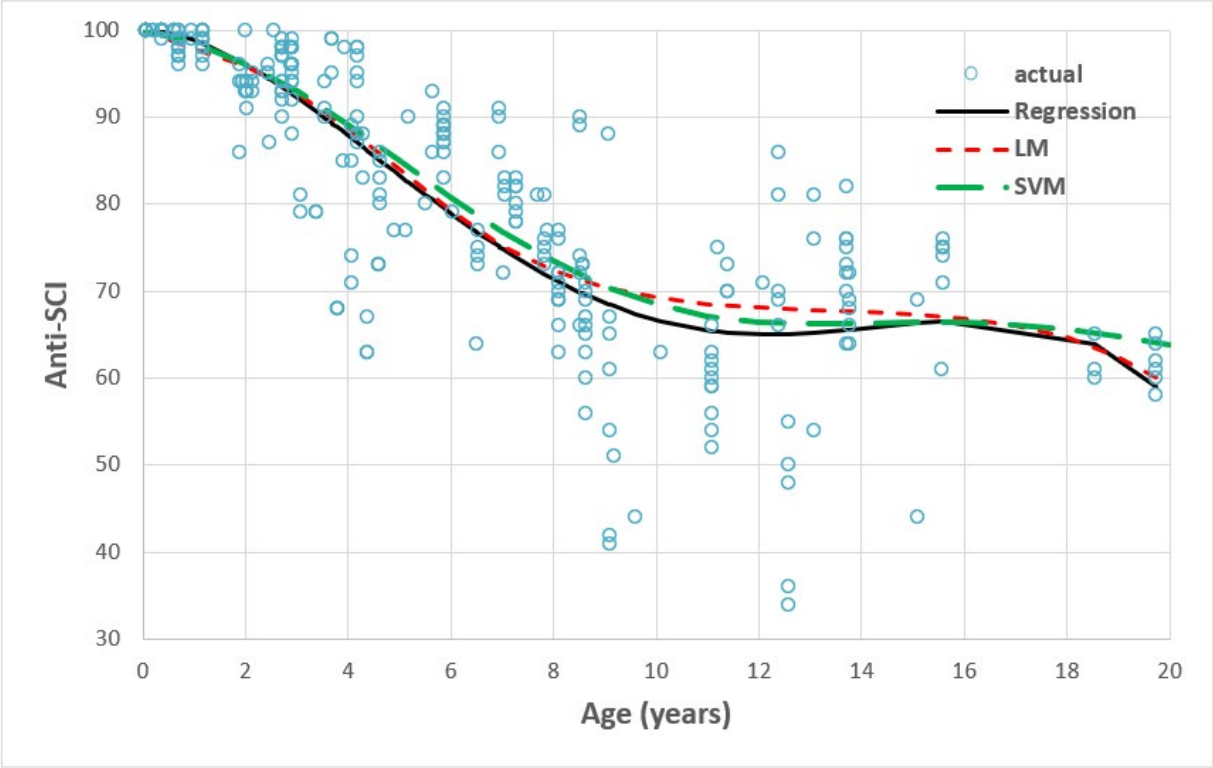


Figure 15. Deterioration Curves Predicted by Three Models

Table 3. Summary of Model Performance in Predicting Anti-SCI for the Static Model

Model	Network Architecture	Model Testing Subset			Independent Testing Subset			Decreasing Trend
		RMSE	RRSE	R^2	RMSE	RRSE	R^2	
ANN LM	1-3-1	7.77	11.43	0.69	7.74	10.90	0.63	Yes
ANN BR	1-3-1	7.75	12.81	0.74	8.33	11.78	0.59	No
SVM	Gaussian kernel	8.07	14.84	0.71	8.02	11.38	0.62	Yes
RF	Bagged ensemble of trees	7.01	11.95	0.79	7.50	9.91	0.66	No
Regression	Fourth-degree polynomial	8.09	14.04	0.72	7.98	11.1	0.64	No

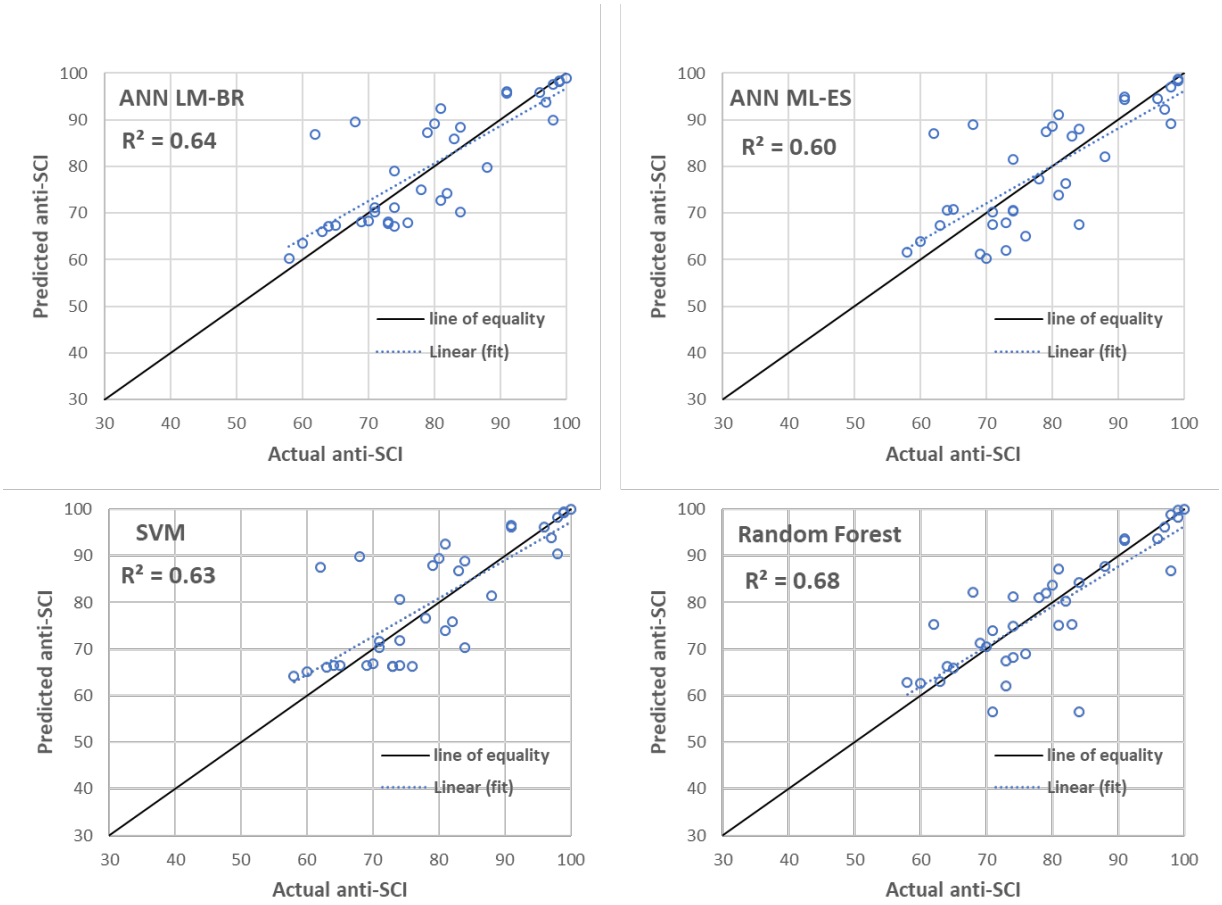
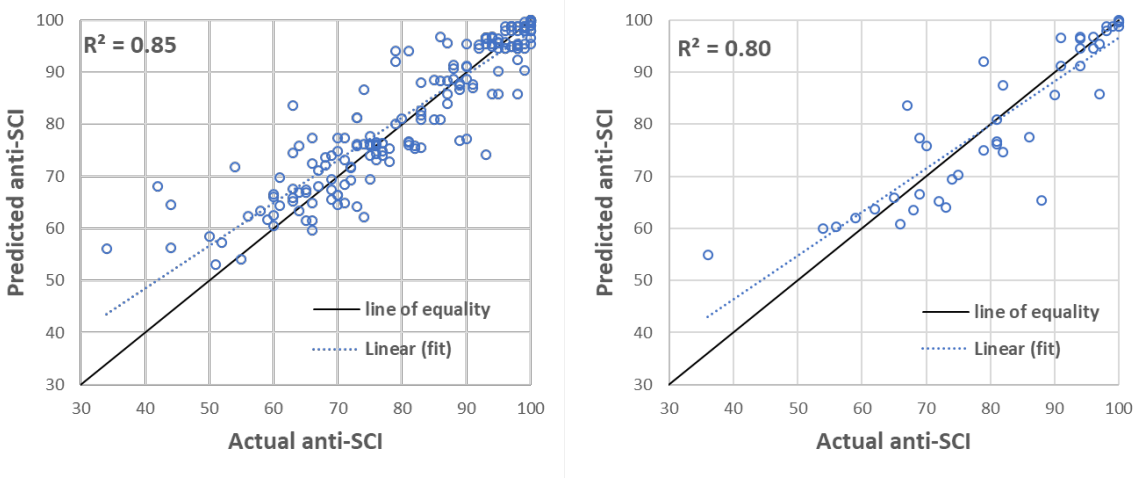


Figure 16. Performance of ML Models on Independent Testing Data

3.3.5 Anti-SCI Prediction Using ANN—Autoregressive Model

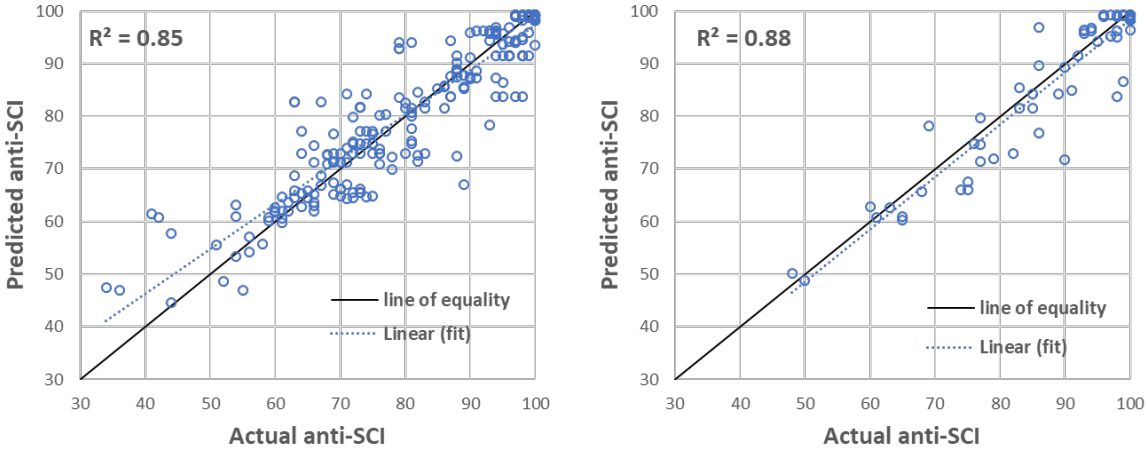
The input variables in the autoregressive models are pavement age, the previous anti-SCI measurement, and the time lapse since the previous measurement. Similar to the static model, a feed-forward multilayer perceptron network with two backpropagation learning algorithms, LM-ES and LM-BR, was used for the autoregressive modeling. The optimum model architecture consists of an input layer with three neurons (consistent with the three predictors), one hidden layer with five neurons, and an output layer with one neuron. The sigmoid logistic and linear transfer functions were used for the hidden and output layers, respectively.

The RMSE of the testing datasets are 6.0 in the LM-ES model and 5.5 in the LM-BR model. Figures 17 and 18 show the scatter plots of the measured (target) versus predicted anti-SCI developed by the ANN LM-ES and LM-BR approaches for the training and testing datasets. The R^2 on both testing datasets are greater than those obtained from the ANN static models. However, the accuracy of the trained LM-ES model on the testing dataset is less than that of the training dataset, indicating that the model generalization is questionable. In the LM-BR model, the accuracy of the testing dataset is comparable to that of the training dataset, indicating suitable generalizability. Overall, the autoregressive predictive models seem to have greater power than the corresponding static models, in that they produce lower errors and greater R^2.



(a) Training Subset (b) Testing Subset

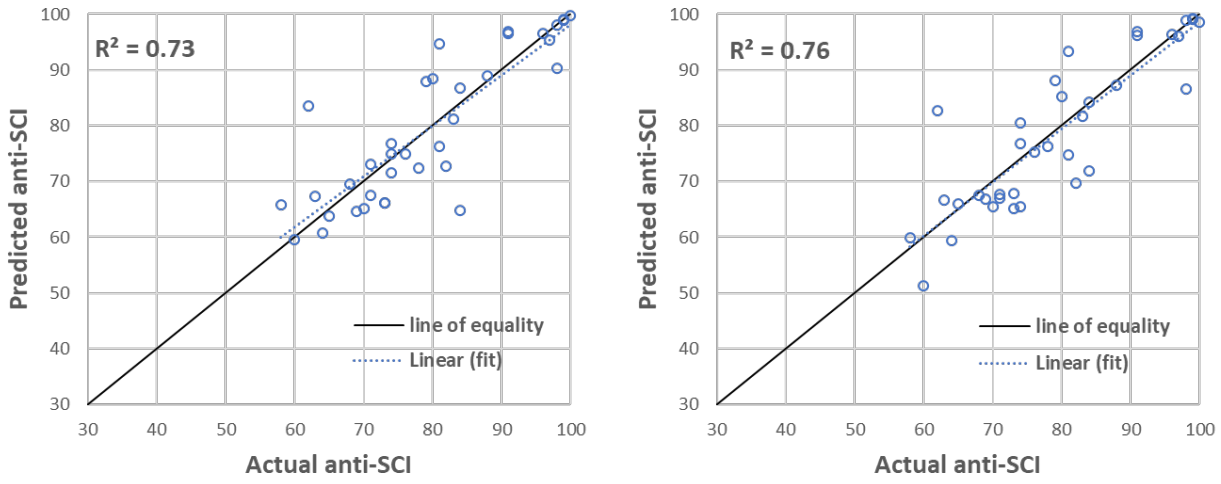
Figure 17. Measured vs Predicted Anti-SCI of (a) Training and (b) Testing Subsets Using the ANN LM-ES Autoregressive Algorithm



(a) Training Subset (b) Testing Subset

Figure 18. Measured vs Predicted Anti-SCI of (a) Training and (b) Testing Subsets Using the ANN LM-BR Autoregressive Algorithm

The predictive performance of the trained ANN models was evaluated using the independent testing dataset as shown in figure 19.



(a) ANN LM-ES Autoregressive Model (b) ANN LM-BR Autoregressive Model

Figure 19. Measured vs Predicted Anti-SCI on the Independent Testing Dataset

3.3.6 Anti-SCI Prediction Using SVM—Autoregressive Model

Similar to the static model, a 10-fold cross-validation was used for the evaluation of model performance. The optimum hyperparameters for the autoregressive model resulting from the cross-validation analysis are shown in table 4. The scatter plot of the actual and predicted anti-SCI values for the testing dataset from all 10 folds is shown in figure 20. Results indicate that the accuracy of the autoregressive SVM model is greater than the static model. The trained SVM model was tested with an independent testing dataset as shown in figure 21. The performance of the SVM model is acceptable as determined by the RMSE and R^2 for the testing dataset.

Table 4. SVM Autoregressive Model Hyperparameters

SVM Model Hyperparameter	Value
Kernel function	Gaussian
Kernel scale, γ	5
Penalty parameter, C	4
Precision margin, ε	0.001

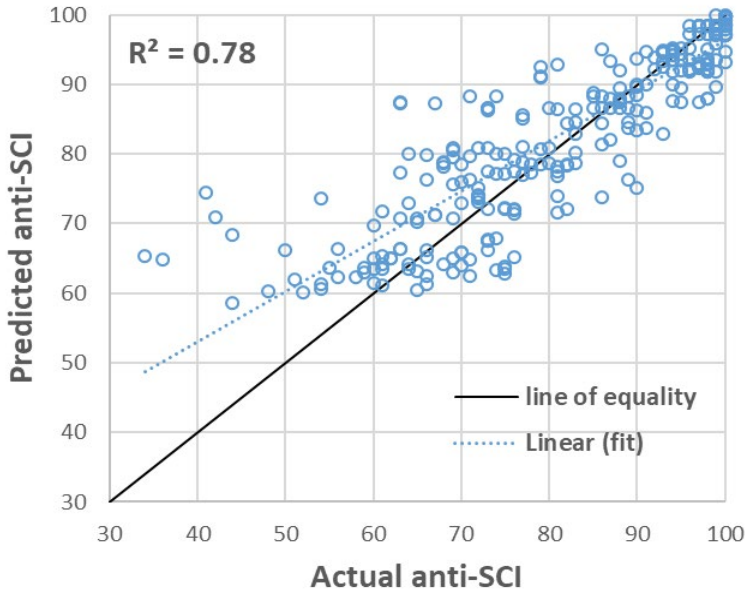


Figure 20. Measured vs Predicted Anti-SCI in the SVM Autoregressive Model

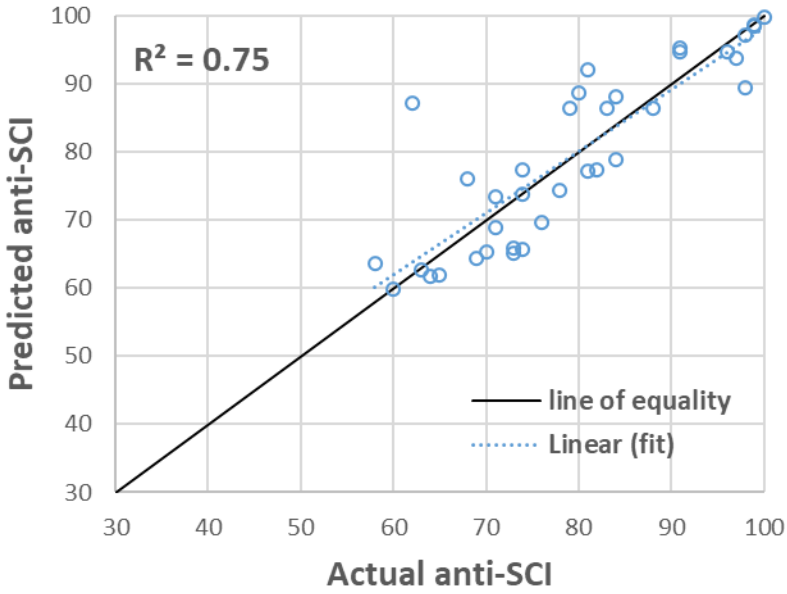


Figure 21. Performance of the SVM Autoregressive Model on Independent Testing Dataset

3.3.7 Anti-SCI Prediction by RF—Autoregressive Model

Similar to the static model, a bagged ensemble of regression trees was used for the prediction of performance index using the autoregressive approach. The following hyperparameters resulting from multiple iterations of cross-validation analysis yielded the best performance:

- Number of trees in the forest = 100
- Minimum leaf size = 3
- Minimum number of samples required to split an internal node = 5
- Number of learning cycles = 30

Figure 22 shows the scatter plot of predicted versus measured anti-SCI with $R^2 = 0.80$. The average values of RMSE and RRSE from the 10 cross-validation analyses are 6.8 and 13.4, respectively. The trained RF model was tested with the independent testing dataset as shown in figure 23. The performance of the trained RF model on the independent testing dataset is better than for the training data, as determined by the higher R^2.

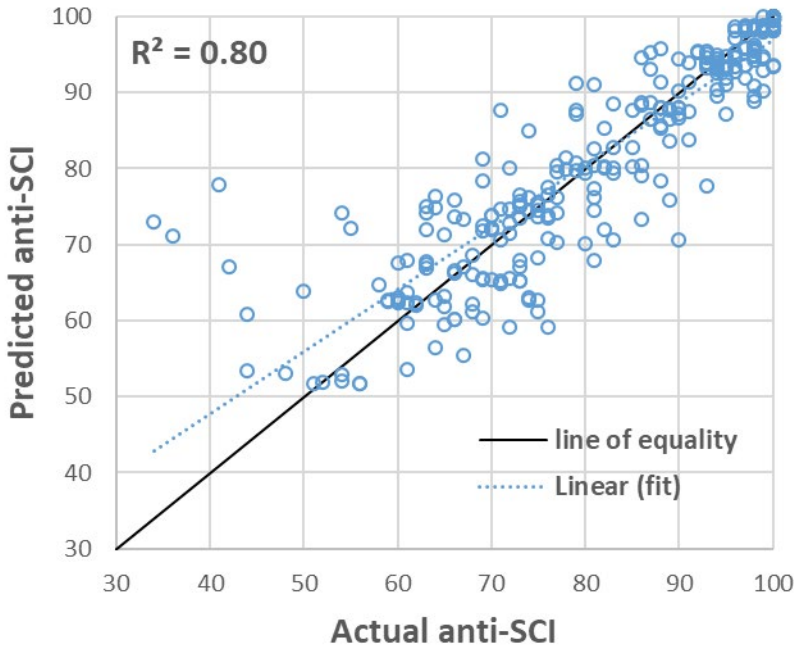


Figure 22. Measured vs Predicted Anti-SCI in the RF Autoregressive Model

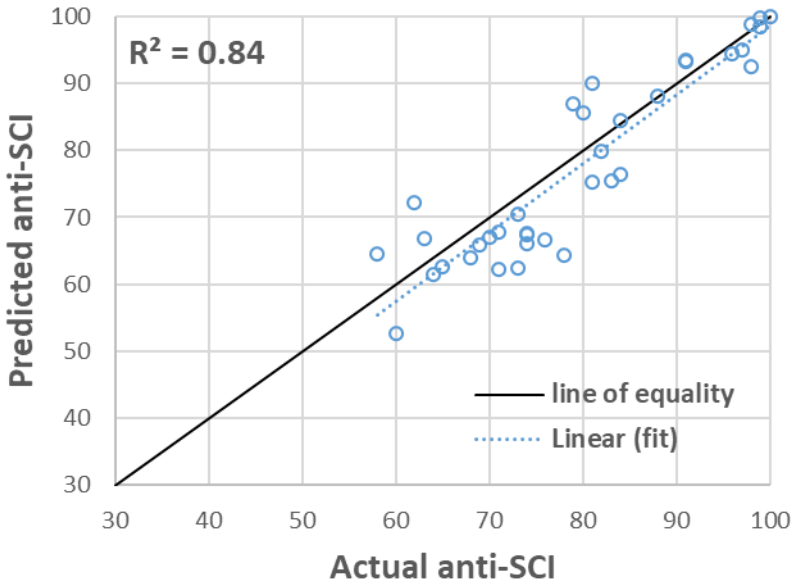


Figure 23. Performance of the RF Autoregressive Model on Independent Testing Dataset

3.3.8 Comparison of Autoregressive Models for Anti-SCI Prediction

Table 5 summarizes the performance of the autoregressive models in predicting anti-SCI. The performance of the developed autoregressive models for incremental prediction of the pavement condition was evaluated using data from three runway sections that were set aside for independent testing. The first example uses performance data from section 20C of CMH RWY 10L-28R. The anti-SCI was measured at ages 2, 4, 6, 9, 12, and 15 years. The measured anti-SCI of 80 at the age of 4 years was used as the base value. Then the subsequent anti-SCI values were calculated incrementally at the age intervals for which measured data were available. The predicted anti-SCI values using the four ML models were compared to the measured values, as shown in figure 24. Of the four models, the RF model predictions were the most consistent with actual measured values. The ANN LM-ES and SVM models underpredicted the anti-SCI at higher ages. Since the predicted anti-SCI at each increment was used as the input for the next increment, the errors accumulated.

Table 5. Summary of Model Performance in Predicting Anti-SCI for the Autoregressive Model

Model	Network Architecture	Model Testing Subset			Independent Testing Subset		
		RMSE	RRSE	R^2	RMSE	RRSE	R^2
ANN LM	1-3-1	6.9	11.4	0.80	6.9	9.4	0.73
ANN BR	1-3-1	5.3	6.2	0.88	6.7	9.2	0.76
SVM	Gaussian kernel	7.4	13.6	0.78	6.5	9.3	0.75
RF	Bagged ensemble of trees	6.8	13.4	0.80	5.9	8.1	0.84

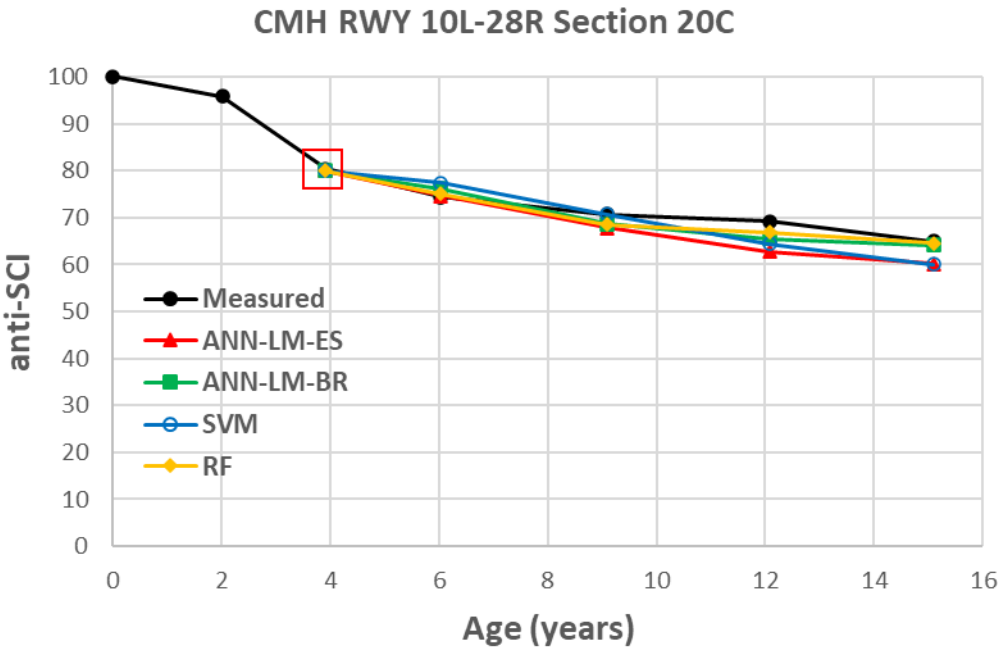


Figure 24. Performance Comparisons of Autoregressive Models to the Measured Anti-SCI, CMH RWY 10L-28R

In the next example, performance data from section 4B of MCI RWY 9-27 were used. The anti-SCI was measured at the approximate ages of 4, 8, 12, 14, 17, and 18.5 years. The measured anti-SCI of 74 at the age of 8 years was used as the base value. Then the subsequent anti-SCI values were calculated incrementally at the age intervals in agreement with the measured data. The predicted anti-SCI values using the four ML models were compared to the measured values in figure 25. All the models underpredicted the anti-SCI, especially at the ages of 12 and 14 years. The gap in prediction could be due to relatively good performance of this pavement section compared to other runway pavements at the same ages. Figure 25 shows the anti-SCI remained above 70 after 14 years. It is also due to the fact that the primary predictor in these models is pavement age, which may not be able to explain all the variability of the pavement performance.

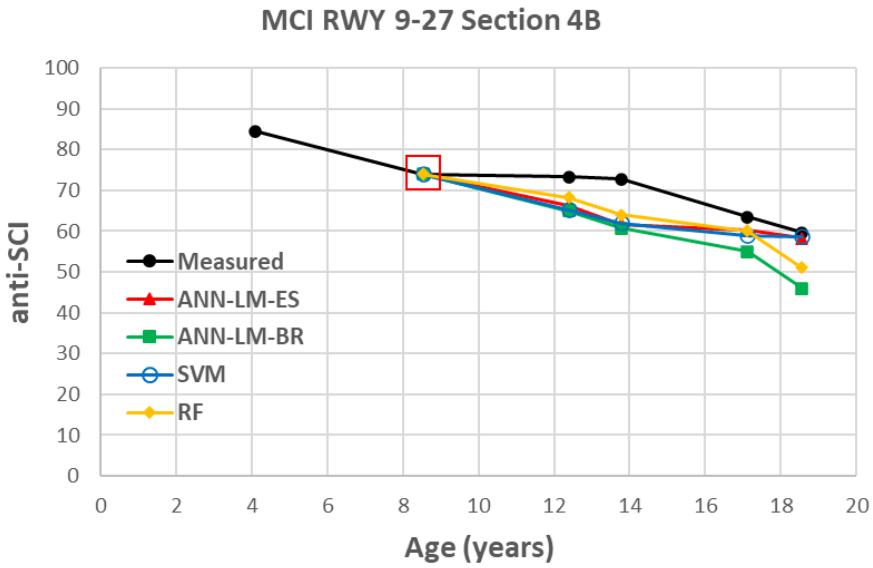


Figure 25. Performance Comparisons of Autoregressive Models to the Measured Anti-SCI, MCI RWY 9-27

A third example illustrates the capability of the autoregressive models in predicting future performance. In section 200 of San Francisco International Airport (SFO) RWY 10R-28L, the last available anti-SCI measurement was 68, at approximate age 11 years. The developed models were used to predict the future anti-SCI in increments of 2 years. In figure 26, there is a 15-unit difference in the predicted anti-SCI at age 19 between ANN LM-BR and RF models and ANN LM-ES and SVM.

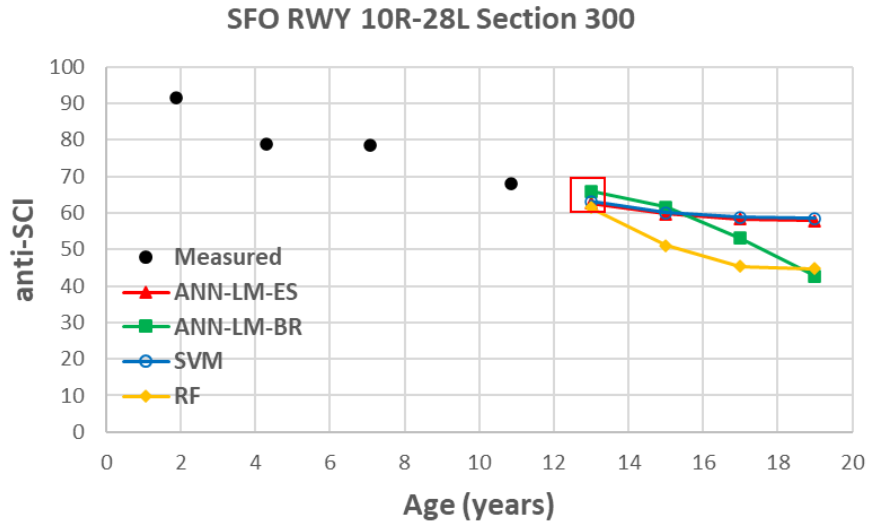


Figure 26. Performance Comparisons of Autoregressive Models to the Measured Anti-SCI, SFO RWY 10R-28L

3.4 PCI PREDICTION MODEL

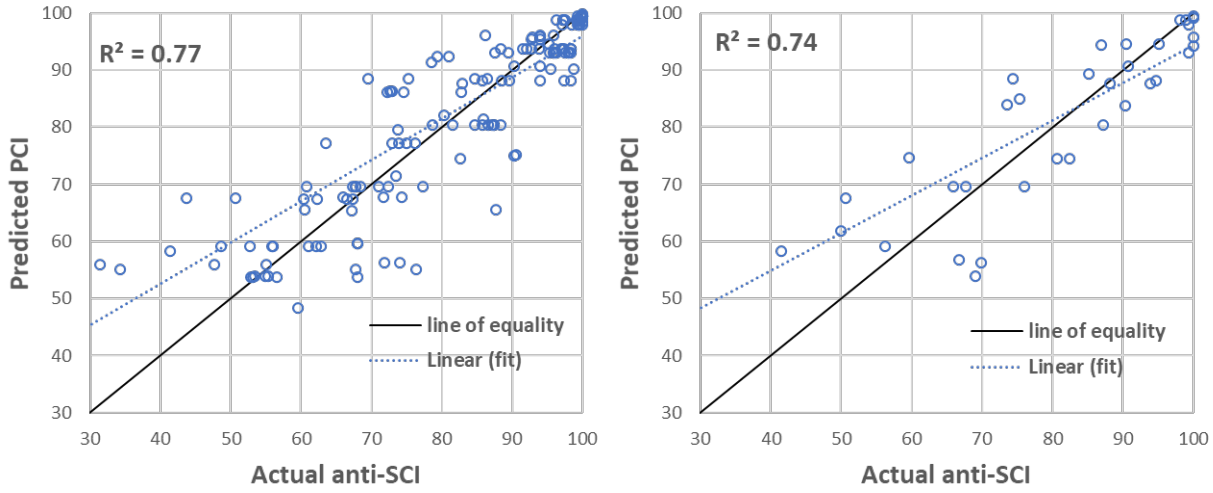
This section illustrates the results of static and autoregressive PCI predictive models using the same three ML methods that were used in the previous section (ANN, SVM, RF).

3.4.1 PCI Prediction Using ANN—Static Model

Similar to the anti-SCI modeling, feed-forward multilayer perceptron with LM-ES and LM-BR backpropagation learning algorithms were implemented for developing PCI predictive modeling. For developing the static model, a network architecture with one input layer (single neuron), one hidden layer (three neurons), and one output layer (single neuron) was identified as the optimal architecture for the prediction. The sigmoid logistic transfer function was used for the hidden layer, and linear transfer function was used for the output layer.

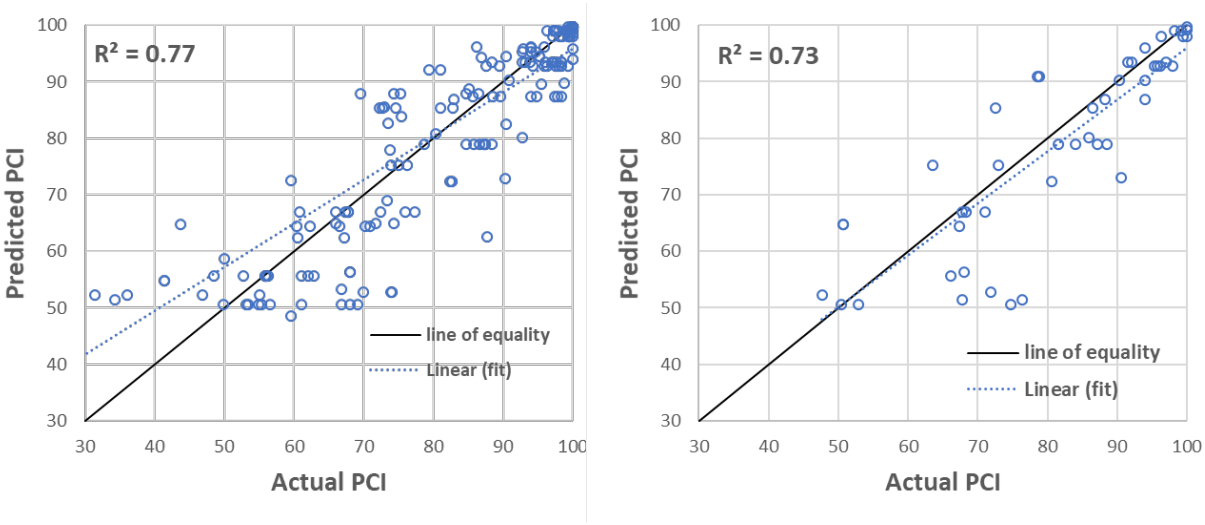
To implement the LM-ES model, 70% of the data were selected randomly for training, 15% for testing, and 15% for validation. To implement the ANN model with Bayesian regularization, 80% were selected randomly for training and 20% for testing. One section from each runway was set aside for independent testing (30 data points). All input and output variables were normalized to the range of [-1, +1] before starting the ANN training to have uniformly distributed variables.

Figures 27 and 28 show the scatter plots of the measured (target) versus predicted PCI developed by ANN LM-ES and LM-BR approaches for the training and testing datasets. The RMSE error of the testing datasets in the LM-ES and LM-BR models are 10.8 and 9.0, respectively. While the testing datasets demonstrated slightly lower R^2 than the trained models, both models can be characterized as having acceptable generalizability. Figures 29 and 30 show the fitted curves of the predicted PCI values in the LM-ES and LM-BR models, respectively. Both models exhibit monotonically decreasing PCI deterioration.



(a) Training Subset (b) Testing Subset

Figure 27. Measured vs Predicted PCI of (a) Training and (b) Testing Subsets Using the LM-ES Algorithm



(a) Training Data (b) Testing Data

Figure 28. Measured vs Predicted PCI of (a) Training and (b) Testing Data Using the LM-BR Algorithm

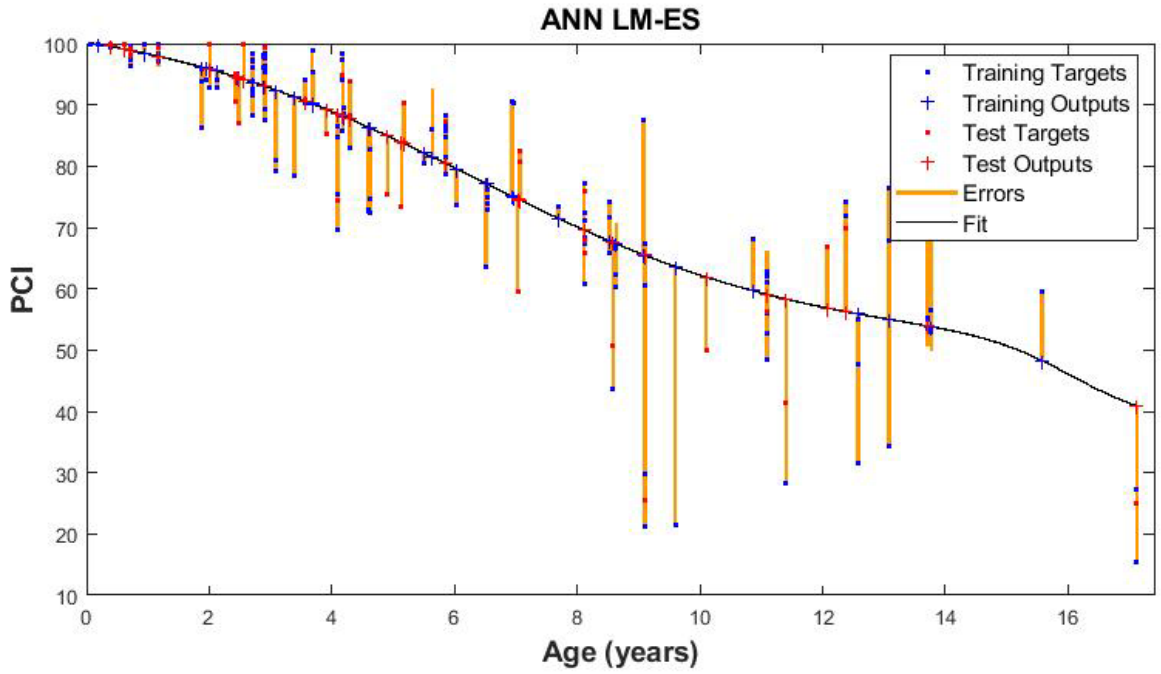


Figure 29. The PCI Prediction Curve in the ANN LM-ES Model

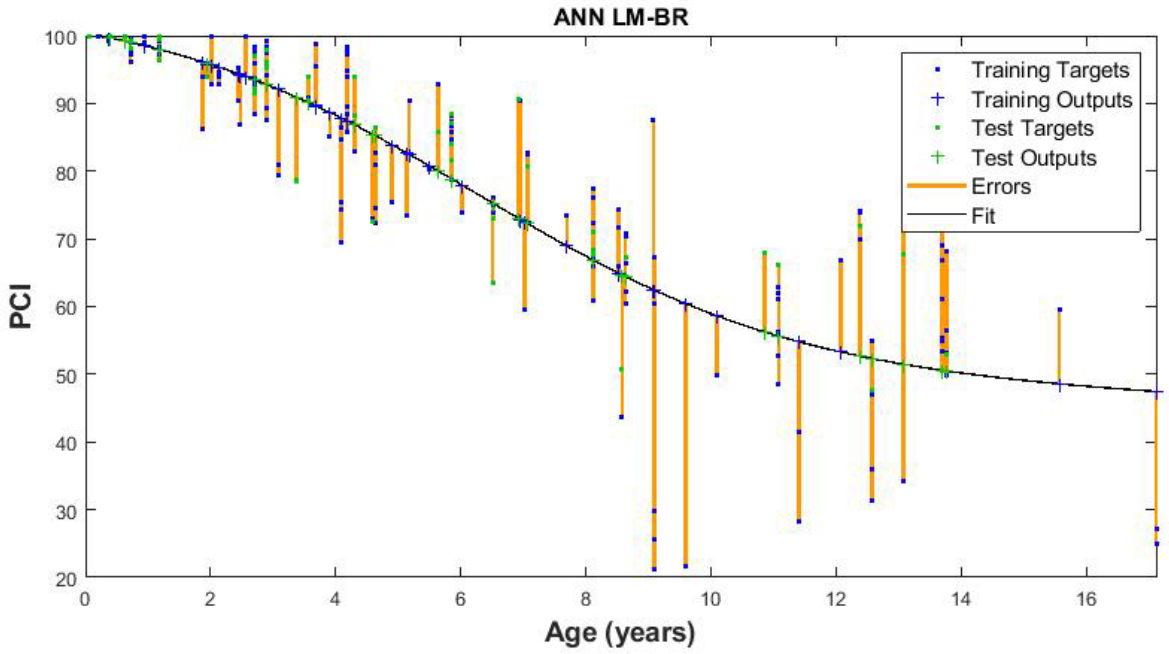


Figure 30. The PCI Prediction Curve in the ANN LM-BR Model

Figure 31 shows the performance of the ANN LM-ES and ANN LM-BR models using the independent testing dataset. Both models returned lower errors on the independent data than the training data.

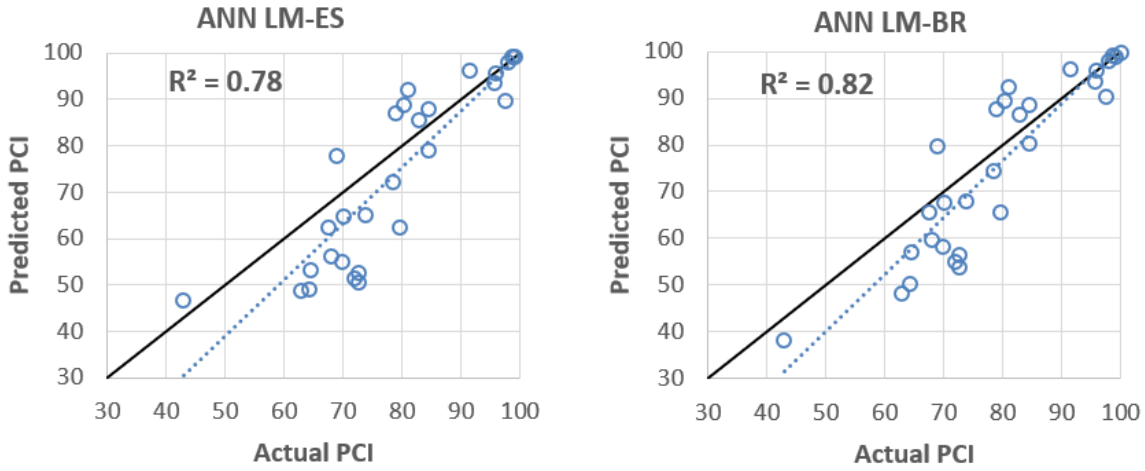


Figure 31. Performance of the ANN Models on Independent Testing Data

3.4.2 PCI Prediction Using SVM—Static Model

The optimum hyperparameters that yielded the best model performance were determined by performing a 10-fold cross validation as shown in table 6.

Table 6. SVM Hyperparameters for PCI Static Model

SVM Model Hyperparameter	Value
Kernel function	Gaussian
Kernel scale, γ	4
Penalty parameter, C	12
Precision margin, ε	0.001

Figure 32 shows the predicted PCI deterioration curve, which is relatively linear and follows a descending trend. Figure 33 shows the scatter plot of predicted versus measured PCI with R^2 and RMSE equal to 0.75 and 9.87, respectively. Figure 34 shows the performance of the trained SVM model on the independent testing dataset. The generalization performance of the trained model is acceptable, as it produced lower RMSE and higher R^2 on the independent test dataset.

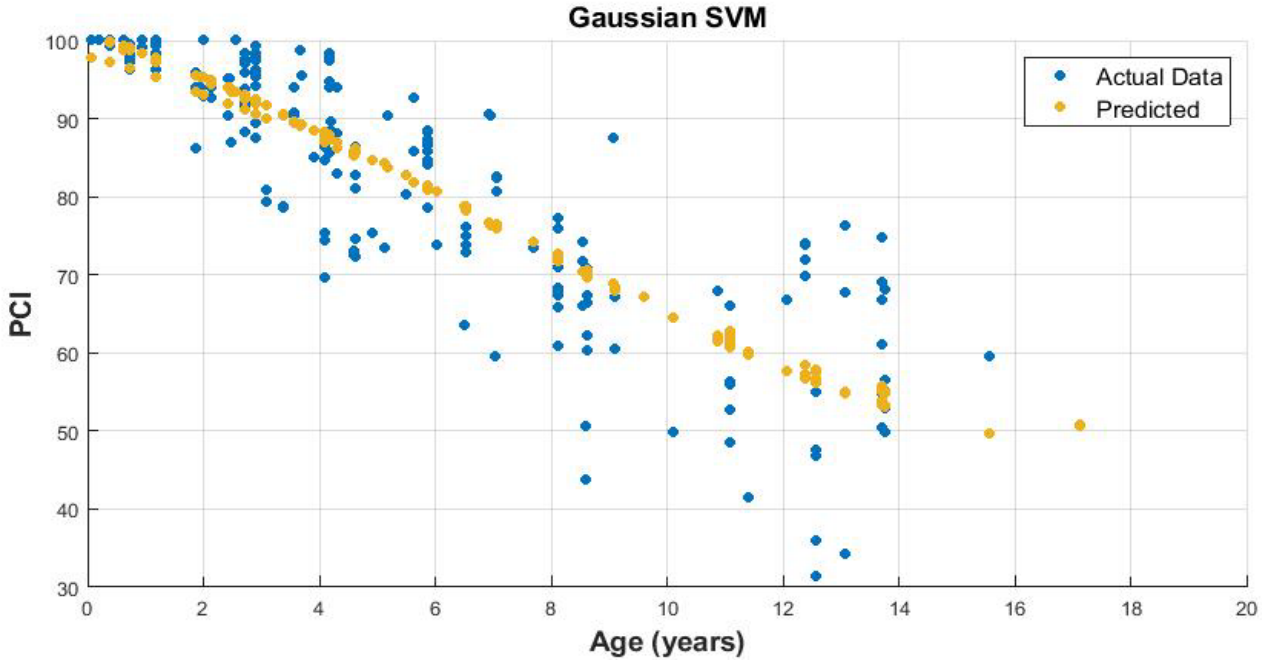


Figure 32. The PCI Prediction Curve in the SVM Model

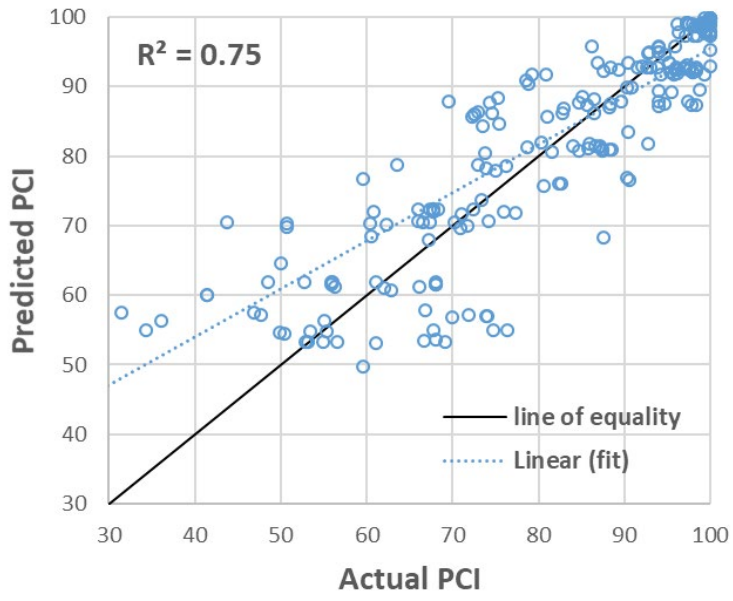


Figure 33. Measured vs Predicted PCI in the SVM Model

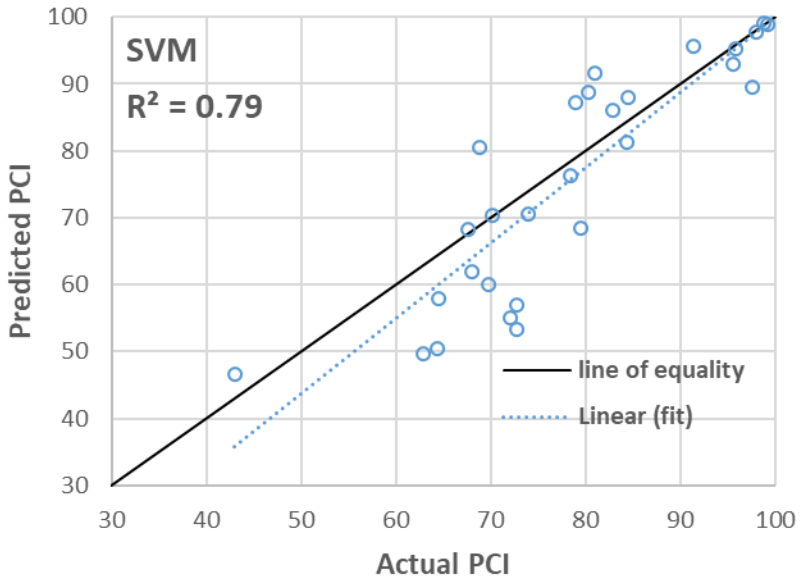


Figure 34. Performance of the SVM Model on Independent Testing Data

3.4.3 PCI Prediction Using RF—Static Model

The RF hyperparameters were optimized using out-of-bag cross-validation analysis. The following hyperparameters yielded the best performance:

- Number of trees in the forest = 300
- Minimum leaf size = 8
- Minimum number of samples required to split an internal node = 5
- Number of learning cycles = 30

Figure 35 shows the predicted PCI deterioration curve, and figure 36 shows the scatter plot of the predicted versus measured PCI with R^2 and RMSE equal to 0.8 and 8.6, respectively. Although the errors are lower than the ANN and SVM models, the deterioration curve is not smooth and does not follow a monotonically decreasing trend. This means that the RF model is not suitable for representing the pavement performance over time. The independent test dataset shows comparable R^2 (figure 37), but greater RMSE than the training dataset, suggesting unreliable generalization performance of the trained model.

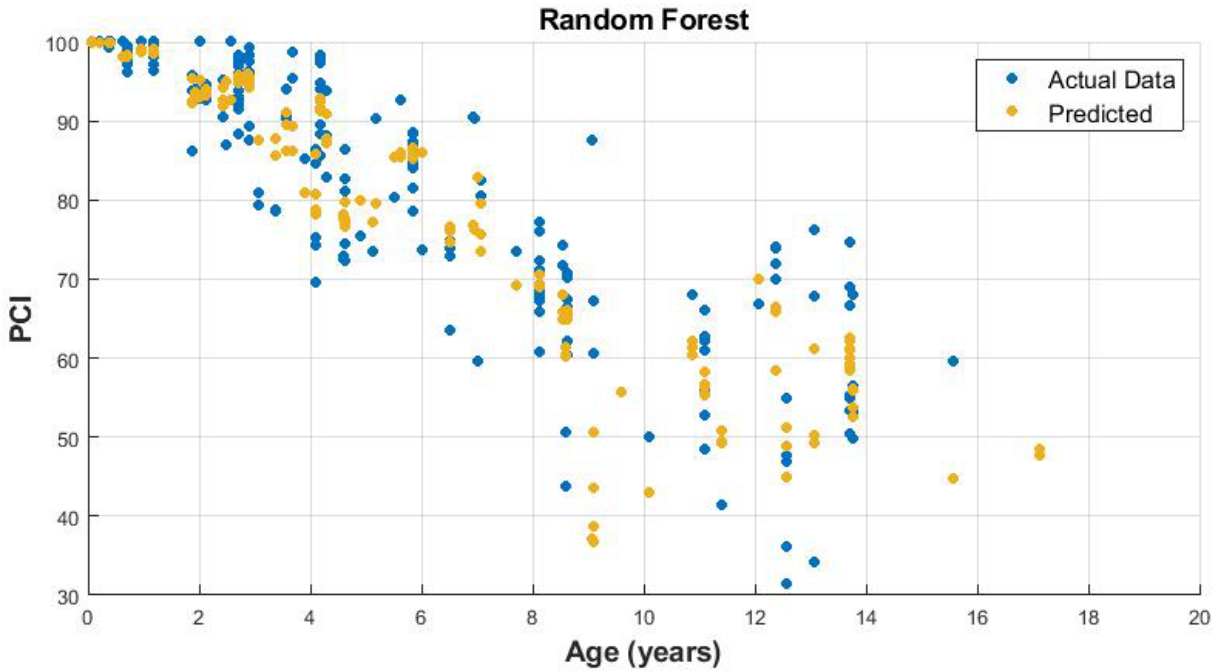


Figure 35. The PCI Prediction Curve in the RF Model

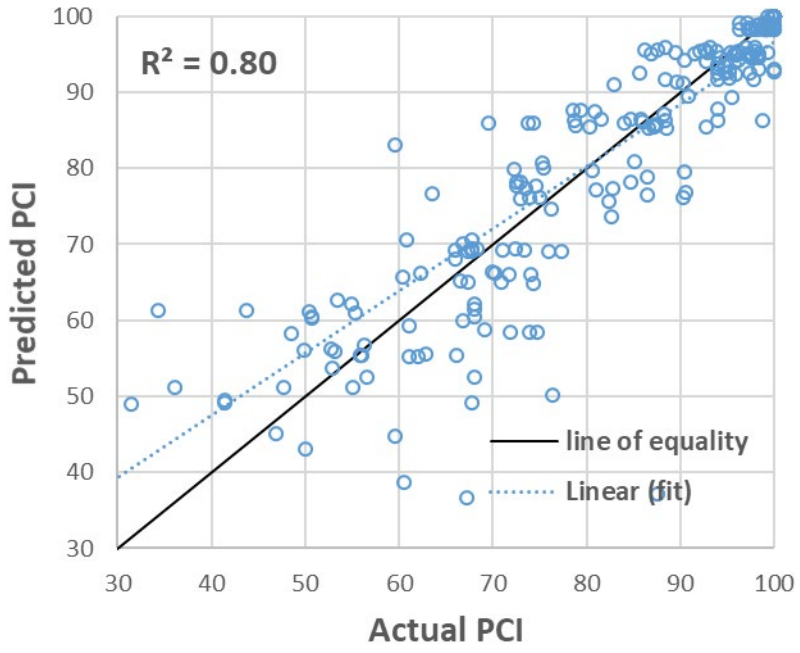


Figure 36. Measured vs Predicted PCI in the RF Model

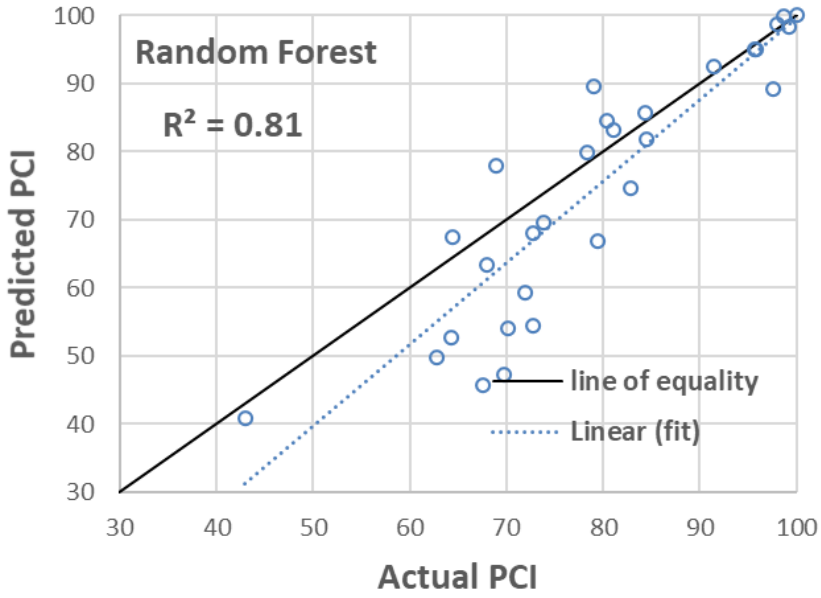


Figure 37. Performance of the RF Model on Independent Testing Data

3.4.4 Comparison of Static Models for PCI Prediction

A fourth-degree polynomial equation was fitted to predict the PCI and then was compared to the ML models. The PCI deterioration curves obtained from the two ANN and SVM models were compared to the regression model, as shown in figure 38. The deterioration curves from the ANN

and SVM models were not consistent with the regression curve. Table 7 summarizes the architecture and performance of different models used in this study to predict PCI. Results indicate that while the RF model produced a lower error, the predicted deterioration curve is not smooth and does not exhibit a monotonically descending trend. The ANN LM-BR model performed better than the fourth-degree polynomial regression model by producing lower prediction error. However, its performance was less promising when tested on the independent dataset. Overall, the prediction errors were relatively high in all the models, especially for PCI lower than 50. The high error is attributed to the fact that the pavement age is incapable of explaining all the variations in PCI. Some of the models did not exhibit comparable performance when applied to the independent test dataset. However, given the limited data available for independent testing, the overall generalization performance of the models was acceptable.

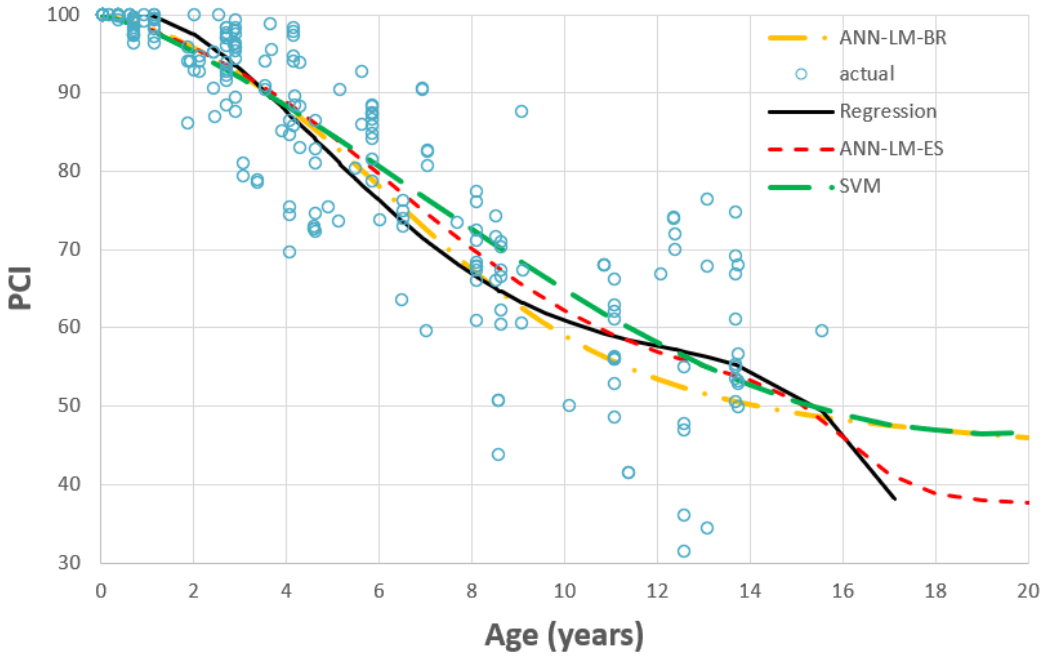


Figure 38. The PCI Deterioration Curves Predicted by Four Models

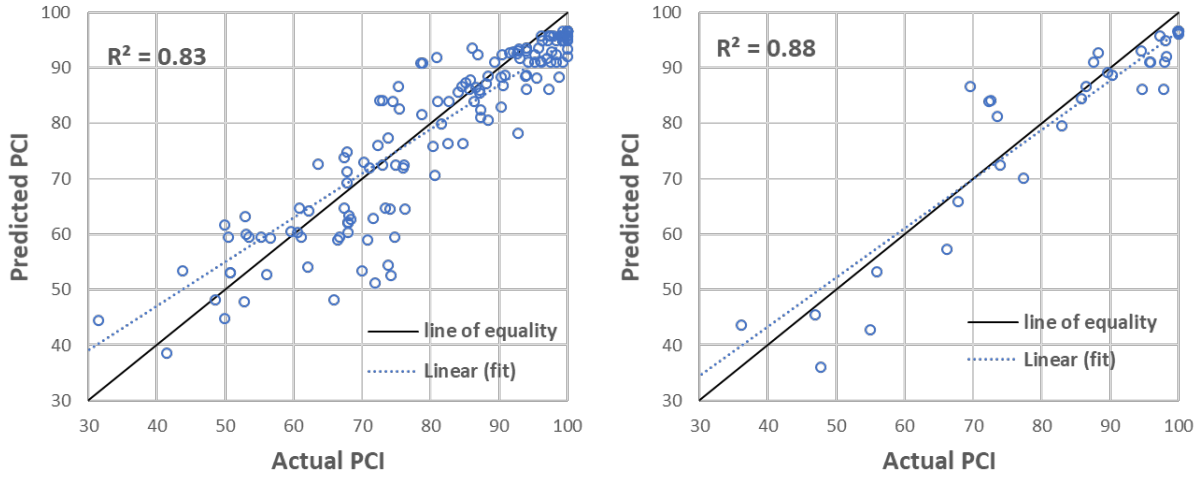
Table 7. Summary of Model Performance in Predicting PCI—Static Models

Model	Network Architecture	Model Testing Subset			Independent Testing Subset			Decreasing Trend
		RMSE	RRSE	R^2	RMSE	RRSE	R^2	
ANN LM	1-3-1	10.83	31.14	0.73	9.07	12.72	0.81	Yes
ANN BR	1-3-1	9.01	12.70	0.73	10.48	14.65	0.77	Yes
SVM	Gaussiankernel	9.87	29.44	0.75	8.55	11.96	0.78	Yes
RF	Bagged of ensemble trees	8.63	21.49	0.80	9.56	13.54	0.80	No
Regression	Fourth-degree Polynomial	9.31	26.09	0.76	9.98	16.02	0.87	Yes

3.4.5 PCI Prediction Using ANN—Autoregressive Model

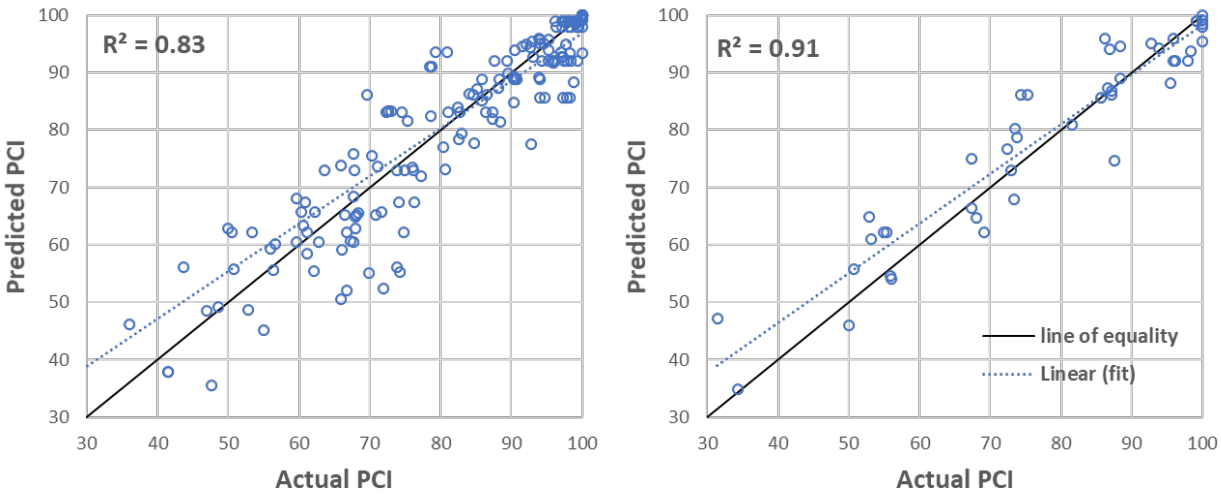
Similar to the static model, a feed-forward multilayer perceptron network with two backpropagation learning algorithms, LM-ES and LM-BR, was used for the autoregressive modeling. The optimum model architecture was an input layer with three neurons (consistent with the three predictors), one hidden layer with five neurons, and an output layer with one neuron. The sigmoid logistic and linear transfer functions were used for the hidden and output layers, respectively.

The RMSE for the testing dataset are 6.9 in the LM-ES model, and 5.8 in the LM-BR model. Figures 39 and 40 show the scatter plots of the measured (target) versus predicted PCI developed by the ANN LM-ES and LM-BR approaches for the training and testing datasets. The RMSE and R^2 for the testing dataset were reasonable in both models and were even greater than the error on the training dataset, indicating suitable generalizability. Overall, the autoregressive predictive models seemed to be more powerful than the static modes by producing lower errors and greater R^2.



(a) Training Subset (b) Testing Subset

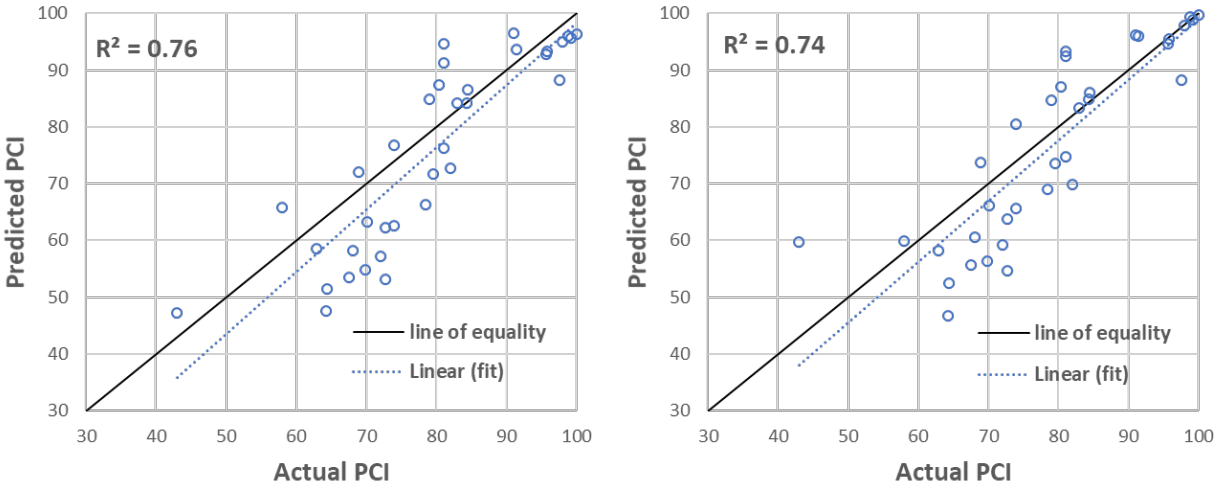
Figure 39. Measured vs Predicted PCI of Training (a) and Testing (b) Subsets Using the ANN LM-ES Autoregressive Algorithm



(a) Training Subset (b) Testing Subset

Figure 40. Measured vs Predicted PCI of Training (a) and Testing (b) Subsets Using the ANN LM-BR Autoregressive Algorithm

The predictive performance of the trained ANN models was evaluated using the independent testing dataset as shown in figure 41. The models produced higher errors on the independent testing dataset.



(a) ANN LM-ES Autoregressive Model (b) ANN LM-BR Autoregressive Model

Figure 41. Measured vs Predicted PCI on Independent Testing Dataset

3.4.6 PCI Prediction Using SVM—Autoregressive Model

Table 8 lists the optimal hyperparameters for the autoregressive model resulting from the 10-fold cross-validation analysis. The scatter plot of actual versus predicted PCI values for the testing dataset from all 10 folds is in figure 42. The autoregressive SVM model was more accurate than the static model. The trained SVM model was tested with the independent testing dataset as shown in figure 43. The SVM model performance was inferior for new sets of data, as determined by higher R^2 for the testing dataset.

Table 8. SVM Autoregressive Model Hyperparameters

SVM Model Hyperparameter	Value
Kernel function	Gaussian
Kernel scale, γ	5
Penalty parameter, C	4
Precision margin, ε	0.001

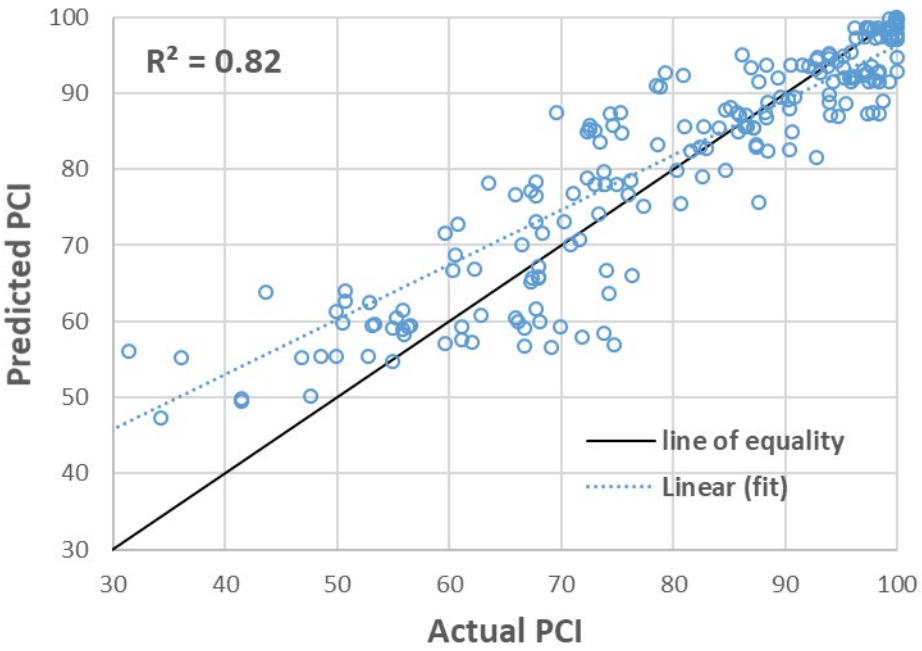


Figure 42. Measured vs Predicted PCI in the SVM Autoregressive Model

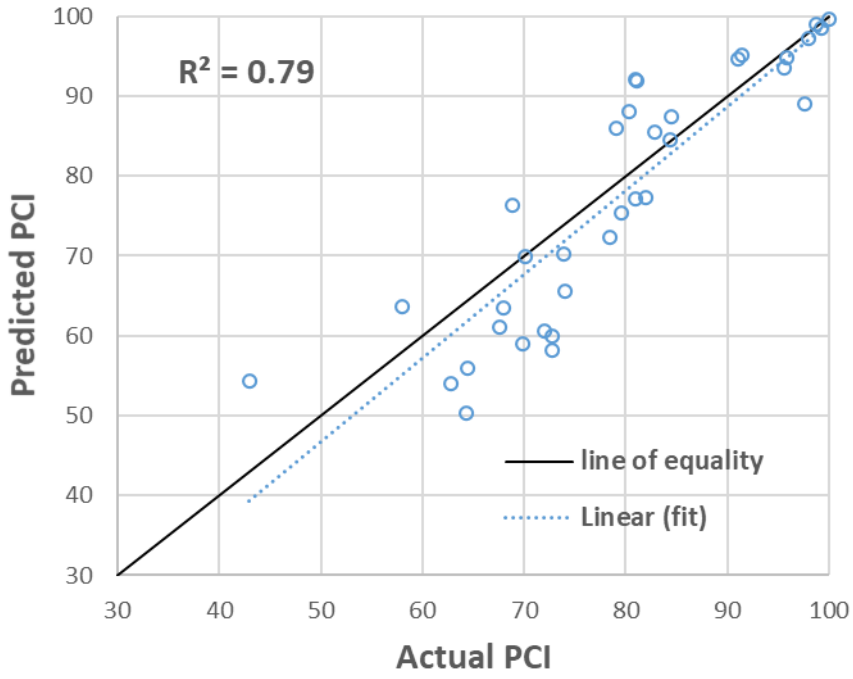


Figure 43. Performance of the SVM Autoregressive Model on Independent Testing Dataset

3.4.7 PCI Prediction Using RF—Autoregressive Model

The following hyperparameters yielded the best performance:

- Number of trees in the forest = 100
- Minimum leaf size = 3
- Minimum number of samples required to split an internal node = 5
- Number of learning cycles = 30

Figures 44 and 45 show the scatter plots of predicted versus measured PCI on the cross-validation test data and the independent testing data. The trained RF model yielded acceptable performance on the independent test data as determined by consistent R^2 and RMSE.

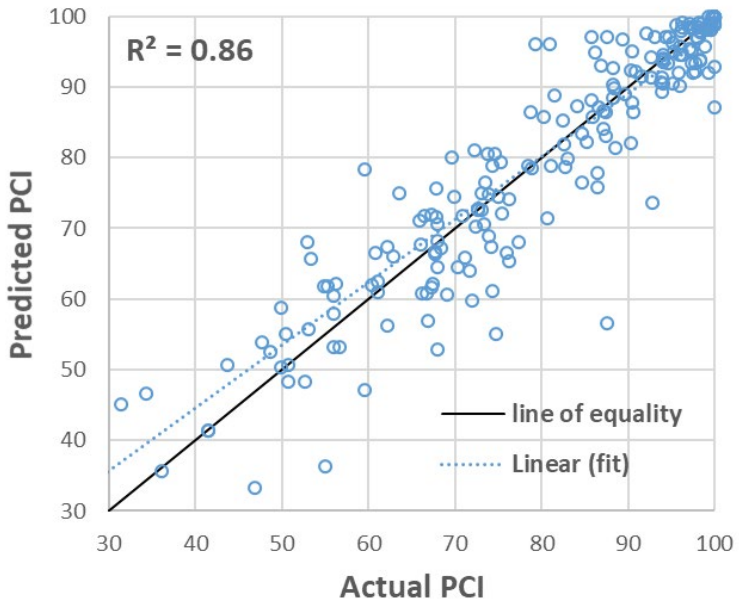


Figure 44. Measured vs Predicted PCI in the RF Autoregressive Model

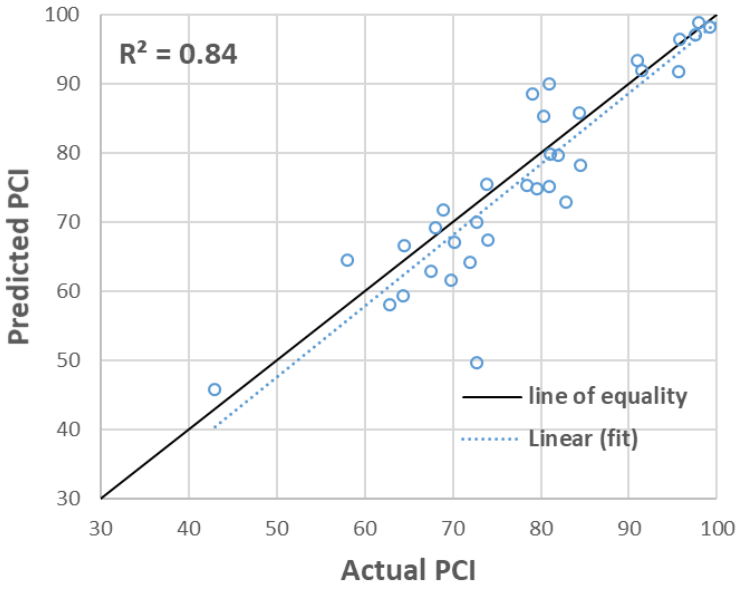


Figure 45. Performance of the RF Autoregressive Model on Independent Testing Dataset

3.4.8 Comparison of Autoregressive Models for PCI Prediction

Table 9 summarizes the autoregressive ML model performance in predicting PCI. The performance of four autoregressive models was evaluated in this section using data from two runway sections that were set aside for independent testing. The PCI of section 20C of CMH RWY 10L-28R was measured at ages 2, 4, 6, 9, 12, and 15 years. The measured PCI of 80 at the age of 4 years was used as the base value, and the subsequent PCI values were calculated incrementally

at similar age intervals as the measured data. The predicted PCI values using the four ML models were compared to the measured values as shown in figure 46. The figure shows that the ANN and RF models underpredicted PCI by a large margin for the age beyond 8 years. SVM model prediction was closer to the measured PCI, but the error increases as the pavement age increases. One reason for the high discrepancy could be an unusual deterioration trend for this section, where the PCI remained almost unchanged after age 6.

Table 9. Summary of Model Performance in Predicting PCI—Autoregressive Models

Model	Network Architecture	Model Testing Subset			Independent Testing Subset		
		RMSE	RRSE	R^2	RMSE	RRSE	R^2
ANN LM	1-3-1	6.9	12.8	0.88	9.2	12.9	0.8
ANN BR	1-3-1	5.8	10.5	0.91	8.8	13.6	0.8
SVM	Gaussian kernel	8.6	26.8	0.8	7.5	11.1	0.8
RF	Bagged ensemble of trees	7.3	19.7	0.9	6.1	8.3	0.9

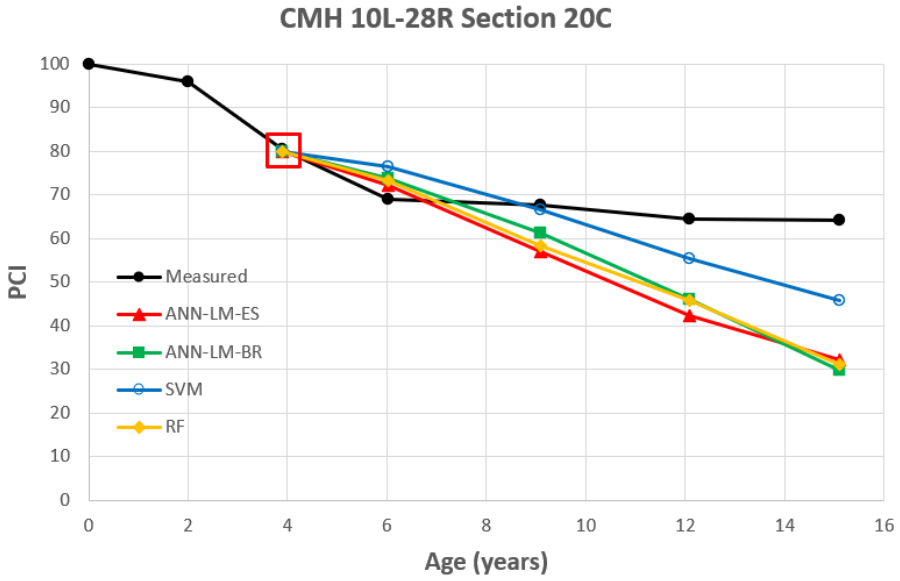


Figure 46. Comparisons of Performance of Autoregressive Models to the Measured PCI, CMH RWY 10L-28R

PCI data for section 4B of MCI RWY 9-27 were measured at the approximate ages of 4, 8, 12, 14, 17, and 18.5 years. The measured PCI of 74 at the age of 8 years was used as the base value, and the subsequent PCI values were calculated incrementally at the age intervals concurrent with the measured data. The predicted PCI values using the four ML models were compared to the measured values, as shown in figure 47. All the models underpredicted the PCI. The gap in prediction could be due to unusually good performance of this pavement section, where the PCI remained above 70 after 14 years. It is also due to the fact that the primary predictor in these

models is pavement age, which may not be able to explain all the variability of the pavement performance.

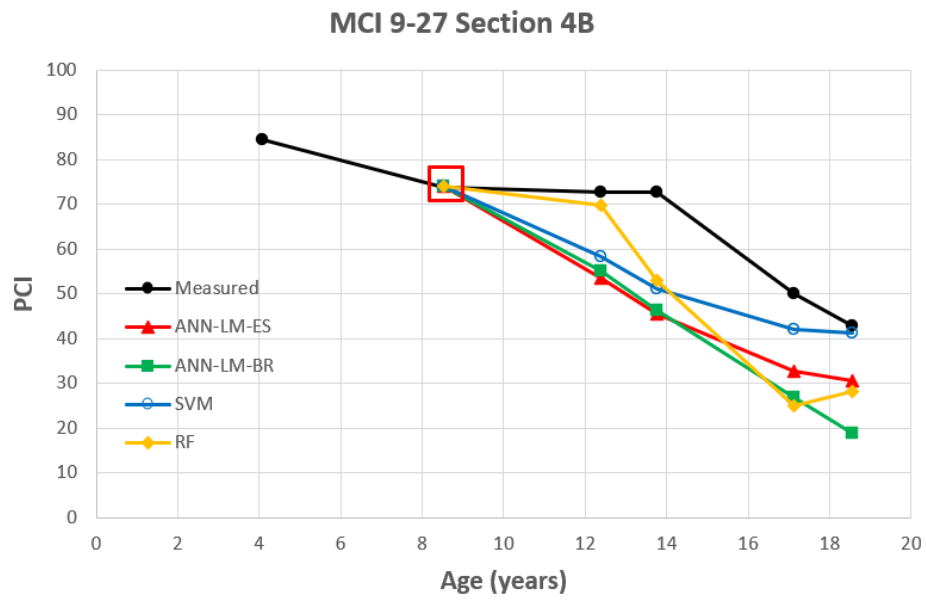


Figure 47. Comparisons of Performance of Autoregressive Models to the Measured PCI, MCI RWY 9-27

4. IDENTIFICATION OF THE MOST INFORMATIVE FEATURES FOR PAVEMENT PERFORMANCE

4.1 INTRODUCTION

The fundamental goal of a data-driven approach to pavement performance modeling is to identify and characterize relationships between a series of features (sometimes referred to as predictors or independent variables) in the data and a target (which is the performance metric). In the deterioration models presented in previous sections, the pavement age was the only feature considered. The resulting models had relatively high errors, suggesting that the pavement age alone cannot explain all the variation in pavement performance. Pavement performance is influenced by various interacting features such as: climate and drainage conditions, traffic, pavement structure and material properties, and quality of construction. Maintenance history may also play a major role. Pavement performance models are typically underspecified, meaning that the predictors do not fully represent the underlying factors inducing pavement deterioration. One reason is that the complete set of predictors is not often available, because acquiring it is difficult and expensive. The other reason is the inherent uncertainty in the quality of predictors, e.g., variability in construction materials and variation in environmental and traffic loads over time. Prediction models can also be over-specified, typically by including a large set of predictors in the hope that they will capture some significant part of the variation in pavement performance. Underspecified models tend to have low predictive accuracy and robustness. In over-specified models, by contrast, the ML learning process is more complicated and the chance of overfitting is high, especially when the set of incorporated features is large in comparison with the number of samples. The presence of a large number of insignificant predictors tends to increase the prediction variance, making the

predictive model unreliable. In addition, an increased number of features increases the difficulty in exploring the effect of each feature on the prediction.

The historical climate data are the most accessible feature in the PA40 database for predicting pavement performance. Seventeen climate variables that may influence the pavement performance were initially identified in this study. Given that the performance data are only available for 11 airports with flexible pavements, there are more independent variables than the number of climate locations. This may pose challenges in developing reasonable predictive models. Therefore, it is desirable to select or construct a subset of features that is useful to build a good prediction model.

The objective of this section is to identify the key environmental (i.e., climate or weather) variables that affect the pavement performance. The candidate environmental variables will then be included as input variables in the ML models to improve their predictive performance. This study uses data-driven approaches based on supervised and unsupervised learning algorithms.

4.2 IDENTIFYING THE ENVIRONMENTAL VARIABLES

This study used the PA40 database to examine the influence of weather on the performance of airfield pavements. Weather records for each airport were obtained from the National Oceanic and Atmospheric Administration (NOAA) and added to the PA40 database. PA40 incorporates a method to summarize the weather events experienced by a given pavement section between the date of construction and the date of a subsequent inspection. Details of how weather events were summarized and the implementation in PA40 can be found in the document entitled, "Development of Historical Climate and Weather Data Links Between PA40 and Existing FAA Databases," which is included in this report as appendix A.

Environmental variables (table 10) in the PA40 database are typically represented as the summation or average of 6 principal variables: (1) temperature, (2) precipitation, (3) dew point, (4) solar radiation, (5) sky cover, and (6) wind speed—each evaluated between the last major construction date and the date of a given inspection. Another variable, average daily temperature difference, was defined for this study as the average of high temperature minus low temperature for each day of a given year. Daily temperature fluctuations induce a temperature gradient through the depth of the asphalt. An extreme temperature gradient may be associated with certain types of distresses, such as block cracking in flexible pavements.

Weather variables can be characterized by both cumulative and average values. Some of the variables that were originally presented as cumulative values in PA40 were calculated as average values between the last rehabilitation/construction date and inspection in this study. Table 10 lists the weather variables both ways. As shown in the table, seven of the weather variables can be expressed as either cumulative or average values, but six others can only be expressed as average values.

Table 10. Environmental Variables Used in This Study

Average Environmental Variables	Cumulative Environmental Variables	Units
Freeze degree days (FDD)	Freeze degree days (FDD)	°F-days
Freeze-thaw cycles (FThC)	Freeze-thaw cycles (FThC)	cycles
Days temperature over 90°F (Temp90)	Days temperature over 90°F (Temp90)	days
Days precipitation (DPrec)	Days precipitation (DPrec)	days
Total precipitation (TPrec)	Total precipitation (TPrec)	inches
Freeze precipitation days (FPD)	Freeze precipitation days (FPD)	days
Hydration days (HydD)	Hydration days (HydD)	days
Average daily temperature (Avg Temp)		°F
Average daily temperature difference (Temp Diff)		°F
Relative humidity (RH)		%
Average wind speed (Wind)		mph
Thornthwaite index		%
Sky cover		oktas

An initial visual examination was performed on each weather variable to develop an understanding of the range and variability of the data in each airport's database. To this end, the average annual values of each climate variable were calculated from 1990 to the end of 2016. Box plots were drawn to provide a visual summary of the distribution of each climate data based on five numbers: minimum, first quartile (25%), median, third quartile (75%), and maximum. The box plots are presented in figure 48 for 10 airports with flexible pavements. Box plots can also be useful in detecting right and left skewness. Days Precipitation (DPrec) and Total Precipitation (TPrec) roughly follow the same pattern. Expect for SFO, the precipitation patterns are also very similar to the patterns for hydration days (HydD) and relative humidity (RH). Two sites with similar winter weather conditions (e.g., similar FDD), such as BOS and Salt Lake City International Airport (SLC), can have different summer conditions (as indicated by Temp90). Additionally, sites with approximately similar precipitation may have very different cold weather conditions as indicated by FDD.

The Thornthwaite index is a moisture index that is determined based on both precipitation and potential evapotranspiration characteristics of a specific region (Yue & Bulut, 2014). The Thornthwaite index is a measure of moisture in pavement unbound layers, with a higher index value indicating higher internal moisture. Precipitation alone cannot determine whether a region has moist or dry pavement layers. Hence, an airport located in a region with high precipitation, but also a high evaporation rate, (such as MIA) does not exhibit high Thornthwaite index.

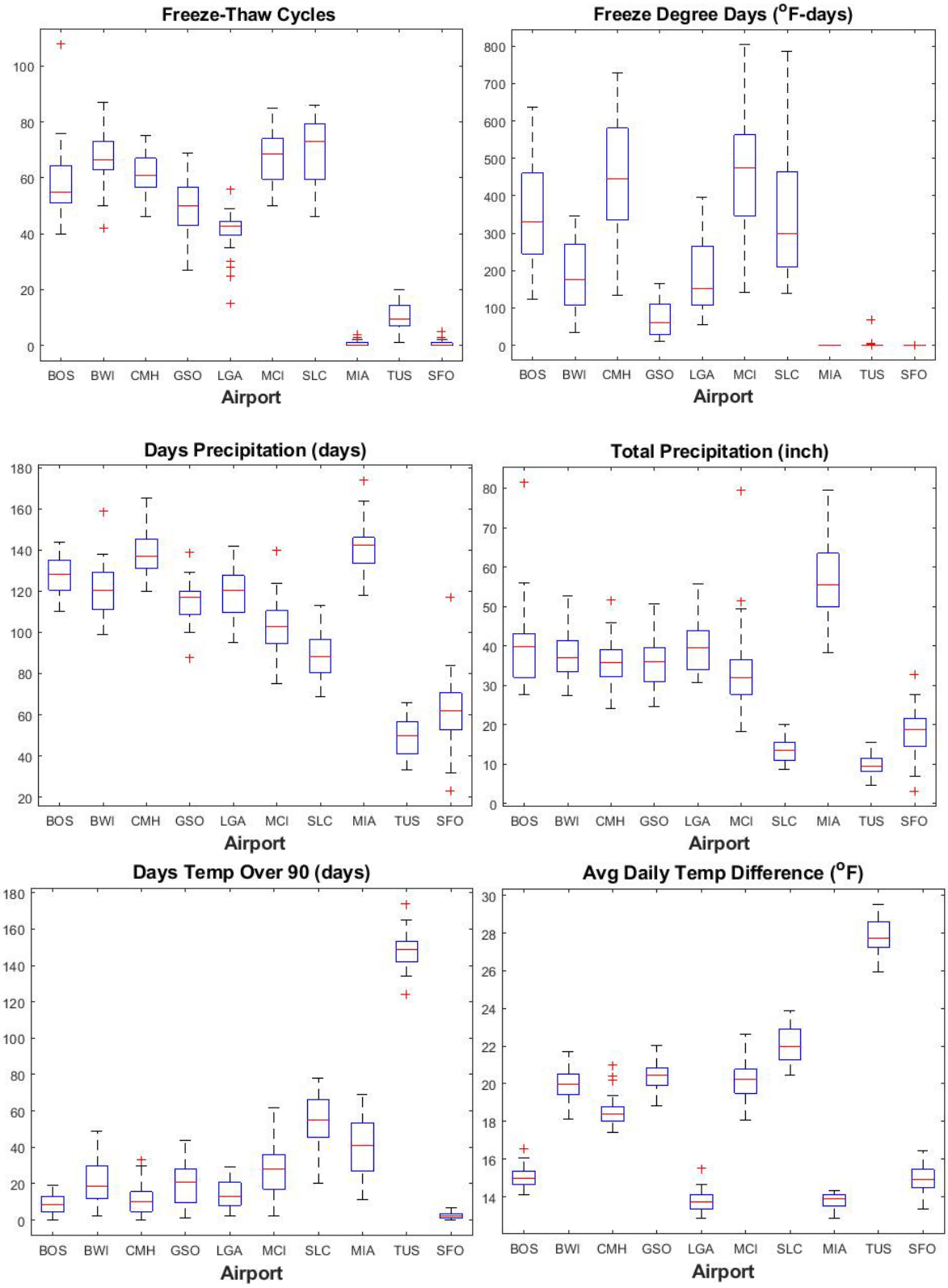


Figure 48. Variation of Environmental Variables for Flexible Pavements in a 26-Year Window

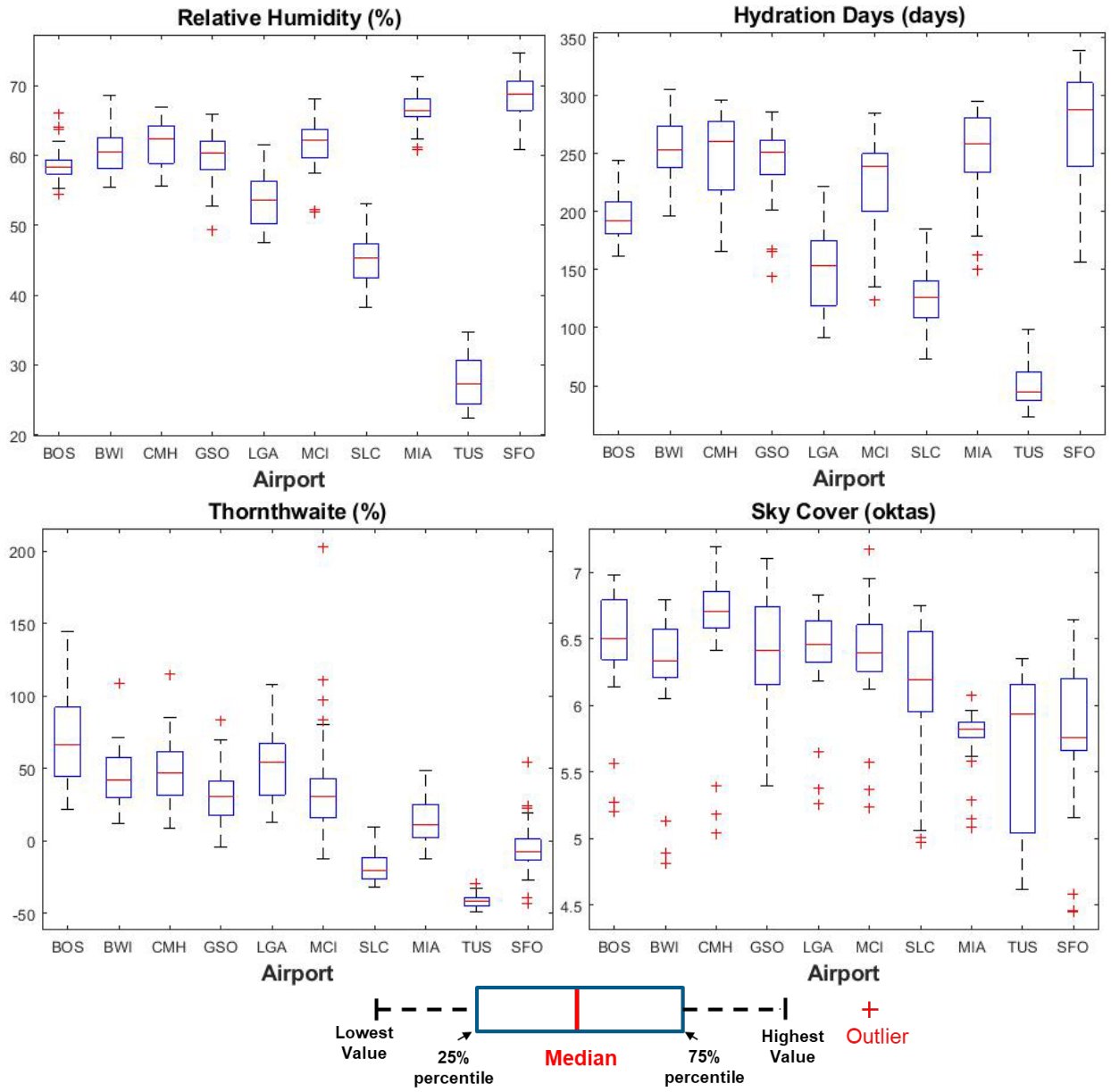


Figure 48. Variation of Environmental Variables for Flexible Pavements in a 26-Year Window (Continued)

Certain weather variables exhibited strong temporal behavior, i.e., their values fluctuated in a wide range over the years the performance data were collected. As an example, figure 49 shows the annual variation of FDD at BOS, where year-over-year ratios were as high as 5:1. This indicates that in some years the pavements were exposed to much more severe weather conditions than in other years, which itself could have accelerated pavement deterioration. However, the effect of individual weather events is anticipated to be residual, i.e., there is no immediate deterioration in the pavement condition as a consequence of an event.

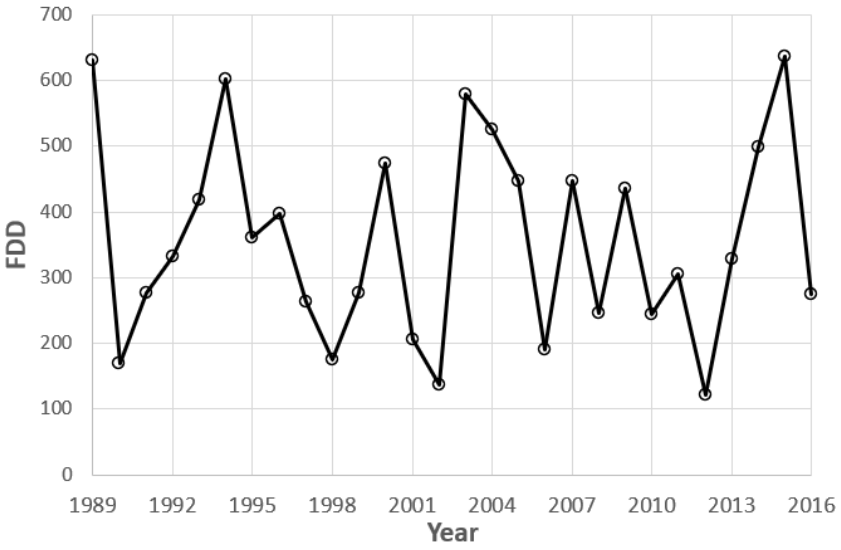


Figure 49. High Temporal Variations in FDD at BOS

4.3 CORRELATIONS AMONG ENVIRONMENTAL VARIABLES

A common problem in over-specified ML models occurs when two or more predictors are highly correlated. This phenomenon is called multicollinearity, or just collinearity. High collinearity between predictors means that variables in the collinear set share substantial amounts of information (Dormann et al., 2013). Collinearity exists in many real-world data such as climate/weather data. The collinear variables are often different manifestations of the same process. Collinearity also occurs when the whole set of information is described by relative quantities. For example, FDD and Temp90 are two extreme representations of temperature, which in some regions could have negative correlations. Collinearity may also be caused by inherent limitations in data, for example when the sample size is low and predictors are not identically distributed in the input domain. Perfectly correlated predictors lead to a mis-specified model. In this case, one or more of the perfectly correlated predictors can be omitted because no additional information is gained by including them (Guyon & Elisseeff, 2003). High collinearity may negatively impact the prediction performance of the model. However, the major problem with high collinearity is that it makes it difficult for the predictive model to assess the relative importance of the predictors with respect to the response. In that case, the model becomes sensitive to small changes in the data set, and the model becomes unstable due to a large inflation of the parameter variance. This makes it difficult to interpret the final model and to isolate the effects of collinear variables (Dormann et al., 2013). Another problem in a predictive model with high collinearity between the predictors is that the model will not be reliable for extrapolating the prediction beyond the range of sampled data. This is because the collinearity patterns are likely to change from one data set to another. For example, precipitation and moisture could be highly correlated in some climate regions, but not in other regions. Also, the correlation may change over the decades.

The current study involves many environmental variables, but the airports in the PA40 database were not sampled randomly across the states. Since the airport locations do not cover all geographic and climatic regions, a complete range of potential climate situations may not exist in

the data. Therefore, certain correlations that may appear redundant should not be eliminated without engineering judgment.

4.3.1 Pairwise Correlation

The most popular diagnostic of collinearity is the pairwise correlation coefficient (R). Spearman and Pearson pairwise correlations were considered in this study. The Spearman's correlation is a statistical measure of the strength of a monotonic relationship between paired data, i.e., it determines if one variable increases or decreases in relation to other variable. This method does not require a normal distribution of the data and is not limited to a linear relationship between the variables. Pearson measures the linear correlation between two variables, and its value is exact when both variables come from a normal distribution.

This study examines the temporal dependency among climate variables based on yearly fluctuations. The correlation matrices in figures 50 and 51, for the Pearson and Spearman correlation types respectively, illustrate the strength of pairwise linear relationships. The correlation plots in figures 50 and 51 use climate data spanning from 1989 to 2016 from 10 airports with flexible pavements. These graphs help to explain the correlations visually as a matrix of pairwise scatter plots. The diagonal of the matrix illustrates the histogram of each climate variable. The correlation coefficients (CC) are equal to the slopes of the displayed least-squares lines in those two figures. The CC values highlighted in red indicate a significant t-statistic or p-value. Establishing a threshold for collinearity assessment is a subjective matter and depends on the type of data. As a rule of thumb, pairwise correlations exceeding a threshold of 0.7 are considered "high collinearity." Based on this threshold value and the Pearson's correlations, the number of freeze-thaw cycles (FThC) is highly negatively correlated to the average daily temperature (Avg Temp) and positively correlated to FDD. Likewise, the variable Temp90 is negatively correlated to average RH and positively correlated to the average daily temperature difference (Temp Diff). The Spearman's correlation values indicate that Avg Temp is negatively correlated to FDD and FThC, and FDD is positively correlated to FThC. For both Pearson's and Spearman's correlations, the three variables DPrec, TPrec, and Thornthwaite are highly positively correlated to one another. RH is also highly correlated to the number of HydD according to both correlation types. Overall, the correlation results indicate that there are low and moderate strength pairwise correlations among most of the environmental variables, but only a few of the correlations can be considered high. Also, no environmental variable is correlated with all other variables.

4.3.2 Condition Number

While the correlation matrix provides one value per variable pair, other measures, such as condition number (CN), estimate a single value to describe the degree of collinearity. The CN is equal to the square root of the ratio of the largest to the smallest eigenvalue of the predictor matrix, and is often used in multivariate regressions to characterize the overall sensitivity of the model estimates to changes in the predictors. The calculated CN of the environmental variable matrix is 315.5, which is well above the benchmark of 30 (Belsley et al., 1980). This could suggest that prediction models may suffer from collinearity, and the prediction error may be high.

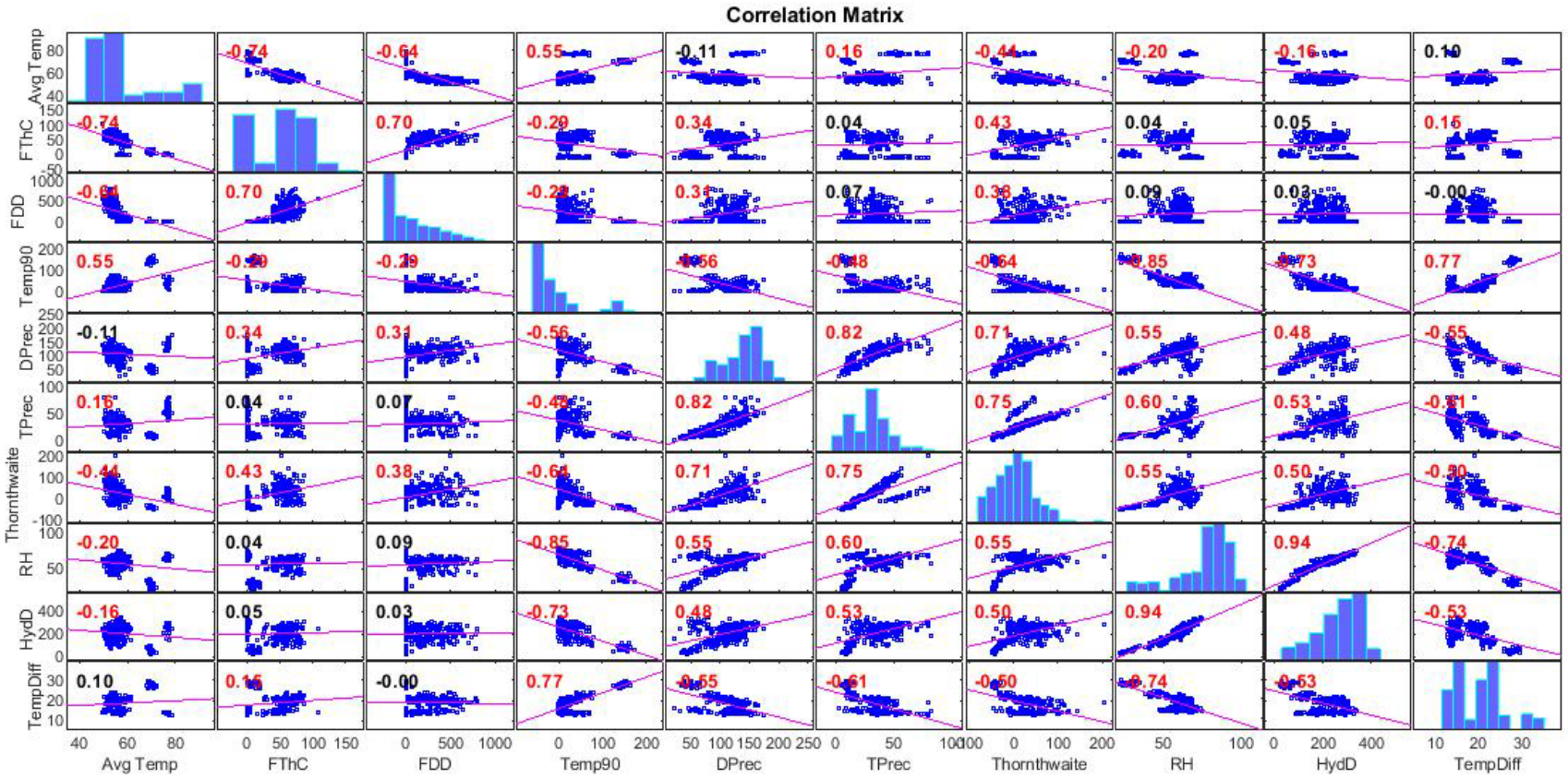


Figure 50. Pearson Correlation Matrix

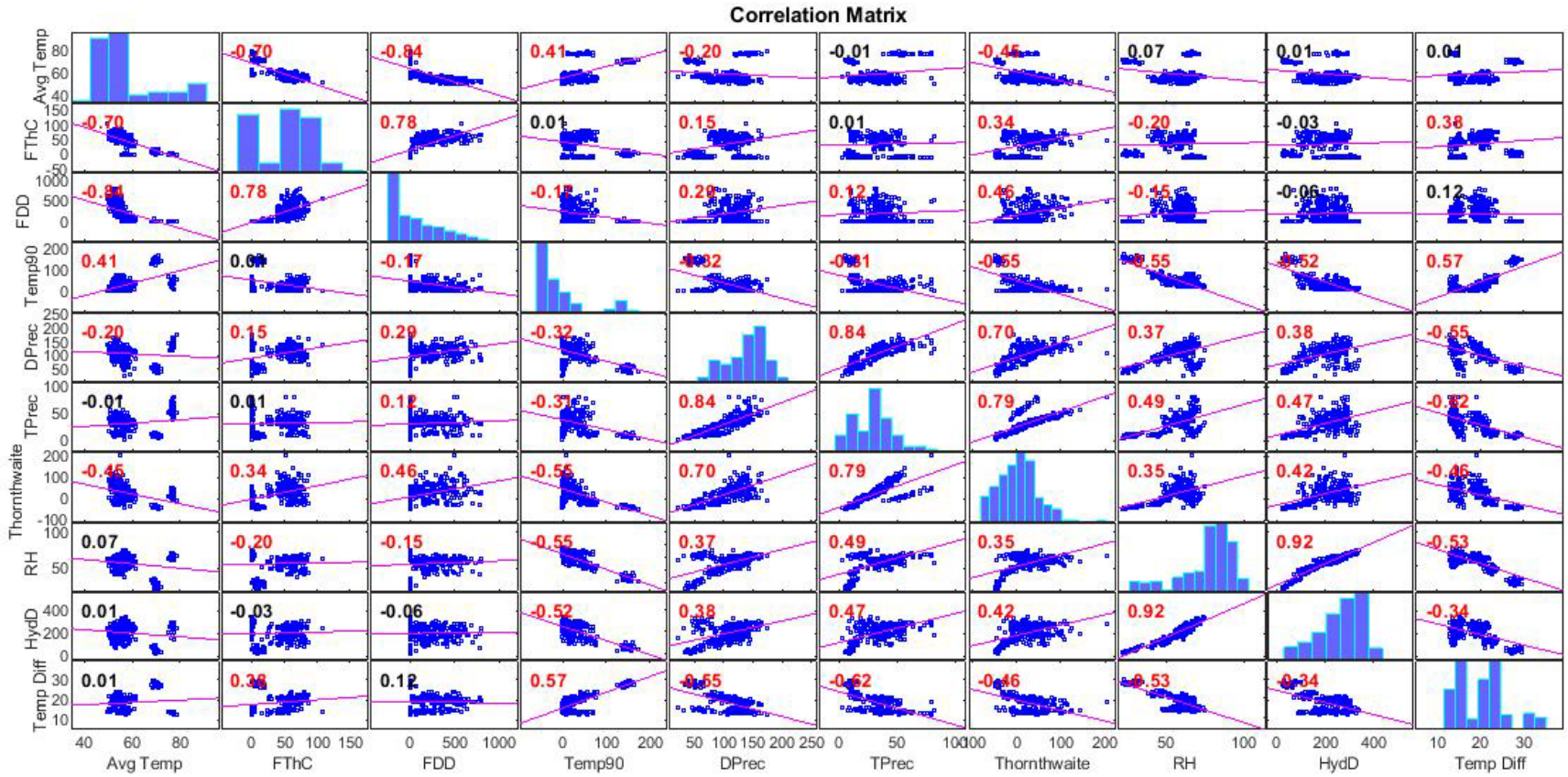


Figure 51. Spearman Correlation Matrix

4.3.3 Variance Inflation Factor

Another common collinearity diagnostic is the variance inflation factor (VIF). VIFs are the diagonal elements of the inverse of the correlation matrix (Booth et al., 1994). Calculated VIFs of the environmental variables are listed in table 11. A VIF equal to 1 indicates no collinearity condition. VIF values greater than 10 can be a sign of collinearity (Belsley, 1980; Hair et al., 1995).

Table 11. Environmental Variable VIFs

Climate Variable	VIF
Avg temp	31.8
FThC	7.7
FDD	2.7
Temp90	15.1
DPrec	5.8
TPrec	35.7
Thornthwaite index	19.8
RH	37.6
HydD	21.2
Temp Diff	10.4

4.3.4 Belsley Collinearity Diagnostics

Belsley collinearity diagnostics is another technique for assessing the strength and sources of collinearity among variables. In the Belsley method, the variance decomposition proportions are calculated, which gives more specific information on the eigenvalues' contribution to collinearity (Belsley et al.,1980). A condition index in the Belsley formulation is defined as the ratio of the maximum singular value of the predictor matrix to the singular value of each predictor. High indices indicate separate near-dependencies in the data. Belsley formulated a method for identifying the specific predictors involved in each near-dependency. Belsley's method provides a measure of how important those dependencies are in prediction. The diagnostic simulations recommend that condition indices in the range of 5 to 10 demonstrate weak dependencies, and those in the range of 30 to 100 exhibit moderate to high dependencies. The results of the Belsley simulation are presented in table 12. The last two rows have a condition index larger than the tolerance of 30. In these rows, the Avg Temp, Temp Diff, TPrec, RH, and HydD variables have variance decomposition proportions exceeding the tolerance of 0.4. This suggests that these environmental variables exhibit multicollinearity. The tolerance for variance decomposition means that no variable attribute should be more than 0.4 to any one eigenvalue.

Table 12. Summary of Belsley Collinearity Diagnostics Simulation

Singular Value	Condition Index	Avg Temp	FThC	FDD	Temp90	DPrec	TPrec	Thornthwaite	RH	HydD	Temp Diff
2.7622	1	0	0.0006	0.0021	0.0002	0.0002	0.0002	0.0006	0	0.0001	0.0001
1.1541	2.3933	0	0.0004	0.0058	0.0142	0	0.0001	0.023	0	0	0.0003
0.7597	3.636	0.0001	0.0169	0.1726	0.001	0.0004	0.0021	0.0021	0.0001	0.0009	0
0.4949	5.5808	0	0.0002	0.0104	0.0532	0	0.0016	0.1377	0.0006	0.0048	0
0.3576	7.7238	0.0001	0.1868	0.5419	0.0021	0.0006	0.0058	0.0049	0	0.0001	0.0016
0.235	11.7553	0	0.0652	0.0261	0.0107	0.0709	0.0411	0.1876	0.0008	0.0272	0.009
0.1324	20.8659	0.0117	0.1004	0.0097	0.1871	0.0401	0.0891	0.1135	0.0038	0.1514	0.0407
0.0986	28.0092	0.0097	0.0969	0.0085	0.0945	0.787	0.2792	0.0545	0.0049	0.0628	0.0039
0.0733	37.6596	0.0086	0.1666	0.0101	0.433	0.0073	0.1581	0.1048	0.1039	0.0123	0.455
0.0253	109.0731	0.9698	0.366	0.2129	0.2041	0.0934	0.4227	0.3713	0.8859	0.7402	0.4894

4.4 IMPLEMENTATION OF DIMENSIONALITY REDUCTION METHODS ON ENVIRONMENTAL VARIABLES: FEATURE SELECTION

The focus of this section is on implementing dimensionality reduction techniques for those environmental variables initially considered as predictors for pavement performance. Feature selection methods are evaluated to decrease the dimensionality of the feature space. Feature selection is the process in ML that reduces the dimensionality of data by identifying and removing a subset of irrelevant or redundant variables that can decrease the model accuracy and quality.

4.4.1 Feature Selection Methods

There are two general feature selection approaches: feature ranking and subset selection (Bolón-Canedo et al., 2013). Feature ranking evaluates relevant variables based on their individual predictive power by assigning weights according to their degrees of relevance. Subset selection produces candidate feature subsets that are useful as predictors based on a certain search strategy. Ranking algorithms can be inefficient, particularly if some of the features are redundant. Alternatively, subset selection methods may exclude some redundant but relevant variables (Guyon & Elisseeff, 2003). Feature selection methods can also be separated into three categories: filter, wrapper, and embedded.

Filter methods select a subset of features based on their importance, considering their variance and relevance to the response. Selecting important features is part of the preprocessing step, and there is no learning algorithm involved in filter approaches. Wrapper methods use a learning algorithm to evaluate the relative benefit of adding or removing a feature from the feature subsets. The evaluation is performed based on a selection criteria by directly measuring the variation in model performance stemming from addition or removal of a feature. The algorithm is said to have achieved a good performance when its stopping criteria are met. Embedded methods perform feature selection as part of the model training process and are usually specific to given learning machines.

Filters are further divided into univariate and multivariate methods. Univariate methods consider each feature separately and ignore feature dependencies. Multivariate filter methods account for feature inter-dependency and are able to find relationships among the features.

The feature selection methods evaluated in this study employ a supervised learning scheme, i.e., a set of predictors is assessed against a target (in this case, anti-SCI). The following feature selection methods were used:

4.4.1.1 Filter Methods

- Pearson Correlation Coefficient (*R*) is a fast way for screening the relevance of an individual feature by measuring its correlation with the target. Pearson correlation is a univariate feature ranking method that can only detect linear dependencies between the variable and target.

- Correlation-based feature selection (CFS) is a subset selection multivariate filter algorithm. This algorithm evaluates the significance of a subset of features by examining the individual predictive ability of each feature alongside the degree of redundancy or correlation between them (Hall, 1999). Irrelevant features are disregarded due to their low correlation with the target. Redundant features are removed due to their high correlation with one or more of the remaining features.

- RReliefF is a modified version of the original Relief algorithm (Kira & Randell, 1992). The original Relief, which was proposed for classification problems, "calculates a proxy statistic to estimate feature … 'relevance' to the target" (Urbanowicz et al., 2018). The proxy statistic can be thought of as a weight or score. Relief randomly picks a sample of instances from the data and for each instance, it finds the nearest neighbor from the same and opposite class based on the Euclidean distance measure. The underlying assumption is that a useful feature should differentiate between data objects and different classes (Bolón-Canedo et al., 2013). ReliefF is an extension of Relief for classification and has the ability to deal with multiclass problems. ReliefF finds *k*, the nearest neighbor from every other class, and averages the weight update based on the prior probability of each class. RReliefF for regression is similar to ReliefF. RReliefF tries to penalize the predictors that give different values to neighbors with the same response values, and reward predictors that give different values to neighbors with different response values. The RReliefF algorithm does not assume the conditional independence of the features, and "can correctly estimate the quality of [features] in problems with strong dependencies between [features]" (Robnik-Šikonja & Kononenko, 2003).

4.4.1.2 Wrapper Methods

Wrapper methods were also implemented using two learners: linear regression and SVM. Cross-fold validation was used to estimate the benefits of adding or removing a feature from the feature subset and to evaluate the model accuracy. A forward stepwise selection approach was used in which the algorithm starts with an empty set of features and searches forward by adding features until the performance does not improve further. Variables that most improve the model are selected. A Gaussian kernel function was used for SVM learning.

4.4.2 Feature Selection Results

The performance data from keel sections of 10 runways with flexible pavements were used. Two approaches were used for implementation of feature selection on climate/weather variables. In the first approach, climate/weather variables were treated as sums, i.e., cumulative values from the date of construction/rehabilitation to the date of inspection. For each pavement performance record in the dataset, six cumulative weather variables were considered as predictors and the measured anti-SCI was considered as the target. Pavement age at the time of anti-SCI measurement and the previous anti-SCI measurement were also considered as input variables. In the second approach, weather variables were treated as averages, i.e., average values of weather variables over a given period of time (between the last rehabilitation and the inspection date, or between two inspections). Eleven weather variables were considered as the predictors, and the rate of anti-SCI deterioration (change in anti-SCI divided by the time between rehabilitation and inspection, or between two inspections) was considered as the target. Since the rate of anti-SCI deterioration is not constant, particularly for ages over 10 years, only performance data for pavements less than 10 years old were used in the second approach. This restriction was intended to isolate, at least in part, the influence of the pavement age on the deterioration rate when performing the feature selection analysis. Table 13 lists the input and output variables for both approaches.

Table 13. Input and Output Variables Used in Feature Selection Analysis

Approach	Input Variables	Target
Approach 1: Cumulative Climate Variables	FThC	Anti-SCI
	FDD	
	Temp90	
	DPrec	
	TPrec	
	HD	
	Age	
	Previous Anti-SCI	
Approach 2: Average Climate Variables	FThC	Rate of anti-SCI change
	FDD	
	Temp90	
	DPrec	
	RH	
	HydD	
	TempDiff	
	TPrec	
	Sky Cover	
	Wind	
	Thornthwaite Index	

The "Attribute Selection" module within the Waikato Environment for Knowledge Analysis (Weka) data mining and ML software program (Eibe et al., 2016), developed at the University of

Waikato in New Zealand, was used to perform the feature ranking. For Approach 1, table 14 summarizes the rank of each feature and the associated weights or scores as calculated by each ranking algorithm. Pavement age and previous anti-SCI are the most informative variables, while freezing degree days and freeze-thaw cycles were the least informative. Table 15 summarizes the subset of informative features as determined by several subset selection algorithms. The sequence in each column does not imply any ranking of variables. As discussed, these algorithms search for dependencies among the features, and remove redundant or correlated features that have minimal impact on the prediction performance. Pavement age, previous anti-SCI, and the number of days with a high temperature above 90ºF are all identified as informative variables in all methods. The wrapper methods retained freezing degree days, freeze-thaw cycles and total precipitation as relevant features.

Table 14. Feature Selection Using Ranking Algorithms in Approach 1

Filter Methods				Wrapper Methods			
Pearson Correlation		RReliefF		SVM—Gaussian Kernel		Linear Regression	
Input Variable	*R*	*Input Variable*	*Score*	*Input Variable*	*Score*	*Input Variable*	*Score*
Previous Anti-SCI	0.77	Previous Anti-SCI	0.040	Age	12.3	Previous Anti-SCI	5.7
Age	-0.77	Temp90	0.022	Previous Anti-SCI	10.5	Age	5.6
DPrec	-0.63	Age	0.016	Temp90	9.3	DPrec	3.5
Temp90	-0.59	TPrec	0.010	DPrec	9.1	Temp90	2.9
TPrec	-0.57	HydD	0.009	HydD	8.0	TPrec	2.8
HydD	-0.57	DPrec	0.009	TPrec	7.6	HydD	2.7
FThC	-0.54	FDD	0.006	FDD	7.1	FThC	2.4
FDD	-0.50	FThC	0.005	FThC	6.7	FDD	2.0

Table 15. Feature Selection Using Subset Selection Algorithms in Approach 1

Filter Method	Wrapper Methods	
CFS	SVM—Gaussian Kernel	Liner Regression
Age (year)	Age (year)	Age (year)
Previous Anti-SCI	Previous Anti-SCI	Previous Anti-SCI
Temp90	FThC	FThC
	FDD	FDD
	Temp90	Temp90
	TPrec	DPrec
	HydD	

For Approach 2, table 16 summarizes the rank and score of each weather variable. Table 16 shows that both wrapper methods and the Spearman correlation gave the highest rankings to the average daily temperature difference and precipitation variables, and lower rankings to freezing degree days and freeze-thaw cycles. The correlation coefficients (R) are relatively low and for some of the variables, the sign of the correlation contradicts engineering expectations. Moreover, the ranking scores are roughly the same for all the climate variables. This suggests that the rate of anti-SCI reduction is not significantly correlated to any of the climate/weather variables. In other words, no individual climate/weather variable can describe the changes in anti-SCI deterioration. In addition, the feature rankings obtained by filter and wrapper methods are inconsistent. One reason for this inconsistency is the different criteria used by the algorithm for selection of features. The inconsistency may also be due to the insignificant correlation between the features and the target.

Table 17 summarizes the subsets of relevant features as identified by several filter and wrapper methods. The CFS method identified the TempDiff as the only relevant feature. The wrapper methods identified FDD and DPrec as relevant features in addition to TempDiff.

Table 16. Feature Selection Using Ranking Algorithms in Approach 2

Filter Methods				Wrapper Methods			
Pearson Correlation		ReliefF		SVM—Gaussian Kernel		Linear Regression	
Input Variable	*R*	*Input Variable*	*Score*	*Input Variable*	*Score*	*Input Variable*	*Score*
Temp Diff	0.28	Sky Cover	0.023	Temp Diff	3.1	Temp Diff	0.1
DPrec	-0.21	Avg temp	0.019	TPrec	3.1	TPrec	0.0
TPrec	-0.20	RH	0.013	DPrec	3.1	DPrec	0.0
Temp90	0.16	Wind	0.013	Temp90	3.1	Temp90	0.0
RH	-0.15	FThC	0.011	RH	3.1	Sky Cover	0.0
FDD	0.11	TPrec	0.010	HydD	3.1	Thornthwaite	0.0
Wind	-0.09	Thornthwaite	0.009	Thornthwaite	3.1	Avg temp	0.0
HydD	-0.09	Temp Diff	0.009	Sky Cover	3.1	Wind	0.0
Thornthwaite	-0.07	DPrec	0.01	FDD	3.07	RH	-0.01
FThC	0.07	Temp90	0.01	Wind	3.06	HydD	-0.02
Avg temp	0.05	FDD	0.01	Avg temp	3.04	FDD	-0.02
Sky Cover	-0.05	HydD	0.00	FThC	3.04	FThC	-0.02

Table 17. Feature Selection Using Subset Selection Algorithms in Approach 2

Filter Method	Wrapper Methods	
CFS	SVM—Gaussian Kernel	Linear Regression
Temp Diff	Temp Diff	Temp Diff
	FDD	FDD
	DPrec	DPrec
	RH	
	HydD	
	Temp90	
	TPrec	
	Thornthwaite	

4.4.3 Discussion and Conclusion

Table 14 shows that the Pearson correlation coefficient between age and anti-SCI (-0.77) is greater than the correlation of any cumulative weather variable to anti-SCI. This means that none of the weather variables alone is a better predictor than age alone. Since the cumulative climate variables in approach 1 are dependent on pavement age, any high correlation with anti-SCI will not imply a cause of performance deterioration. On the other hand, a low correlation of a weather variable to performance does not imply that the variable is useless. For example, in table 14 cumulative FDD has the lowest correlation with anti-SCI, which goes against the common expectation. The reason for this low correlation is that FDD is a variable representing an extreme weather condition, making it a significant factor in cold weather regions but an insignificant factor in warmer regions. Figure 52 illustrates the scatter plot of anti-SCI versus cumulative FDD for each pavement section in the database. The figure shows two distinct clusters of data. One cluster includes the data from three airports in no-freeze regions (e.g., SFO, TUS, and MIA), and the other cluster represents the rest of the airports located in the northern and central parts of the United States. The FDD has a very low correlation with anti-SCI in those airports with nearly zero FDD, and a relatively high

correlation with anti-SCI for the rest of the airports with a higher probability of freezing events. Evaluation of the influence of weather inputs on pavement performance would be more meaningful if individual distress types were considered, as opposed to an overall surface condition indicator such as anti-SCI. The specific types of pavement distresses contributing to anti-SCI deterioration can differ greatly from one climate region to another. For example, while thermal cracking due to severe freezing is a common type of distress in a cold region, block cracking and rutting could be primary distresses in an airport with warm weather and high solar radiation. Therefore, it is anticipated that a climate variable such as FDD will have a high correlation with particular distress types, such as thermal cracking. In this case, FDD can be described as the primary factor causing thermal cracking.

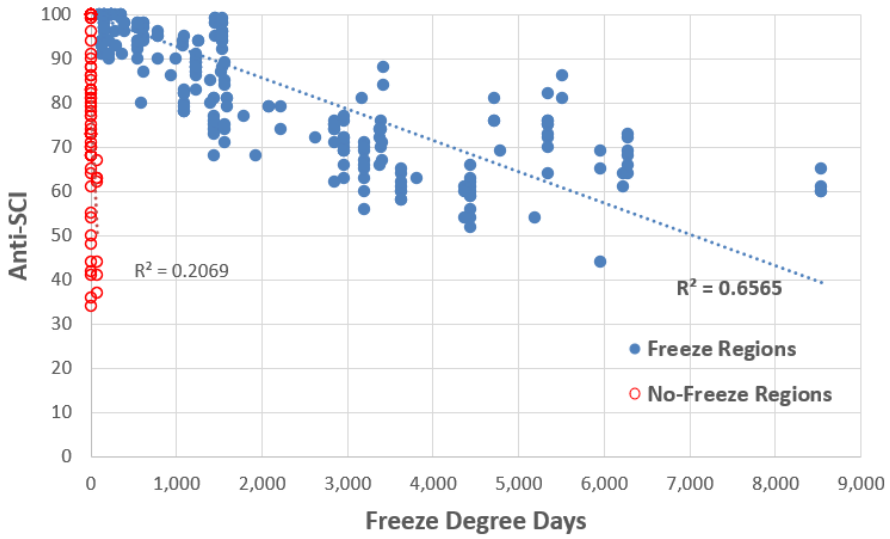


Figure 52. Correlation of Anti-SCI With Cumulative FDD in Northern and Southern Airports

Ranking algorithms that score variables individually and independently of each other are not able to determine which combination of variables would give the best performance. Ranking algorithms can give some information on how relevant each variable is to performance, but not enough to determine which variables should be kept or discarded for the prediction of pavement performance. The results of ranking algorithms on the average climate variables in approach 2 showed that the variation in the rate of anti-SCI deterioration cannot be related to any individual climate variable. However, it is likely that when all the climate variables are taken together, they can describe some variation in anti-SCI.

4.5 IMPLEMENTATION OF DIMENSIONALITY REDUCTION METHODS ON ENVIRONMENTAL VARIABLES: FEATURE CONSTRUCTION

Feature construction methods aim to reduce the dimensionality of the data by constructing new features from the original feature set. In this section, two unsupervised learning methods were employed for dimensionality reduction of the feature space by constructing or extracting new features from existing features. The first method is a clustering method based on the *k*-means algorithm, and the second method is PCA.

4.5.1 Unsupervised Cluster Analysis

Cluster analysis is an unsupervised learning technique used primarily to explore homogenous subsets of data objects within a dataset. In an ML model, clustering can be applied separately to both the input and output domains using only the information found in each domain, regardless of the dependencies between the inputs and the target. Clustering is the process of grouping data objects into clusters such that objects within each cluster have the highest similarity to one another, while also having the highest dissimilarity from objects in other clusters. Two data objects are grouped according to their mathematical similarity. For a dataset with *m* data objects (sample or observation) and *n* features or attributes, most clustering techniques convert the data matrix into a $m \times m$ "distance matrix" of inter-object similarities or dissimilarities (Milligan, 1996). Two common similarity measures are based on Euclidean distance or correlation (Pearson) between two objects. Two objects are considered to be close to each other when the dissimilarity is small. Choosing a correct similarity measure is a critical step in a clustering analysis. Distance measures are recommended when data samples are small, the magnitude of data is important, and the features are not correlated (Gower & Legendre, 1986). However, different similarity measures can be tested to determine which ones produce better results given the nature of data.

Another critical step is choosing a suitable clustering algorithm. Two popular types of clustering are partitional and hierarchical. A partitional clustering divides a set of data objects into mutually exclusive subsets (clusters) where each data object belongs to one subset only (Tan et al., 2013). Hierarchical clustering produces a nested hierarchy of similar groups of objects, according to a pairwise distance matrix of the objects (Keogh et al., 2001). The most commonly used partitional clustering algorithm is *k*-means.

4.5.2 Clustering Using *k*-means

An algorithm called *k*-means is an iterative partitioning algorithm that assigns subsets of data objects to exactly one of the *k* clusters defined by their centroids (David & Vassilvitskii, 2007). The number of *k* clusters is chosen before the algorithm starts. The *k*-means algorithm initializes centroids by randomly assigning data objects to centroids of each cluster and then assign data objects to a cluster such that the similarity within the cluster is at the maximum. The most commonly used similarity measure in a *k*-means algorithm is the Euclidean distance. The algorithm opts to minimize the objective function which is the Euclidean distance between the data objects and the cluster's centroids. The objective function is expressed in equation 32:

$$\min \sum_{i=1}^{m} \sum_{k=1}^{K} \left\| x^i - C_k \right\|^2 \tag{32}$$

where, m is the number of data objects, *k* is the number of pre-defined *k* clusters, x is the coordinates of data objects, and C is the centroid of each cluster. After each iteration, the centroid of each cluster is updated based on the data objects assigned to that cluster. The iterations stop when the centroids reach a steady state, i.e., there are only a few objects changing their clusters. While *k*-means can be applied simply to any type of data, it poses some disadvantages. One is that the number of clusters *k* must be determined as an input before starting the algorithm. The other limitation is that since the initial centroids are assigned randomly, the resulting clusters will be

different each time the algorithm is executed. Despite these limitations, *k*-means is well suited for the type of data in this study.

4.5.3 Clustering of Environmental Data Using *k*-means

Cluster analysis was conducted to find patterns in weather variables that are not intuitively obvious. Two approaches were implemented to group the performance data based on their climatic characteristics. In one approach, for each airport, the median or average values of the weather variables over the course of data collection were considered as the data objects for clustering. In this case, the outcome of the cluster analysis was assigning the airports to regions with similar average climate conditions. However, due to temporal behavior of the weather variables, it is possible that a weather variable in one geographical location has a significantly lower value than the average in that location. For example, despite the higher average number of freezing days in BOS than Baltimore/Washington International Thurgood Marshall Airport (BWI), in some years the pavement at BWI may have been exposed to more freezing degree days than at BOS. This suggests that the use of a single value (average or median) will not be a realistic representation of a weather variable with high temporal variations. In the second approach, the time-series weather variables in each airport were transformed into subsequences with various-sized sliding windows (Zolhavarieh et al., 2014). Subsequence in this context refers to segments of long time-series variables, for which the performance data were measured. The size of the sliding window is equivalent to the time between two pavement condition inspections or between the inspection and the last rehabilitation, for each performance record in the dataset. The data objects for the cluster analysis are the averages of each weather variable over the sliding windows. Several weather variables can be chosen simultaneously for grouping the datasets into clusters with similar climate characteristics. The choice of weather variable was determined based on engineering judgment and the collinearity between the variables. Highly correlated variables, as determined in section 4.3, were not considered for cluster analysis. Tables 18 and 19 illustrate the possible two- and three-way combinations of the weather variables used for clustering.

Table 18. Possible Combinations of Weather Features for Clustering—Two at a Time

Variable 1	Variable 2
FDD	TPrec
FDD	Avg Temp
FThC	Temp Diff
Temp90	TPrec

Table 19. Possible Combinations of Weather Features for Clustering—Three at a Time

Variable 1	Variable 2	Variable 3
FDD	Temp90	TPrec
FThC	Temp90	Thornthwaite

The *k*-means algorithm was used for clustering analysis with squared Euclidean as the distance metric (see section 4.5.2). The variables at the clusters should be similar within a group and should be different between the groups. The optimum number of clusters was found to be three or four

clusters depending on the combinations of variables used for clustering. The results of clustering are shown in figures 53 through 58.

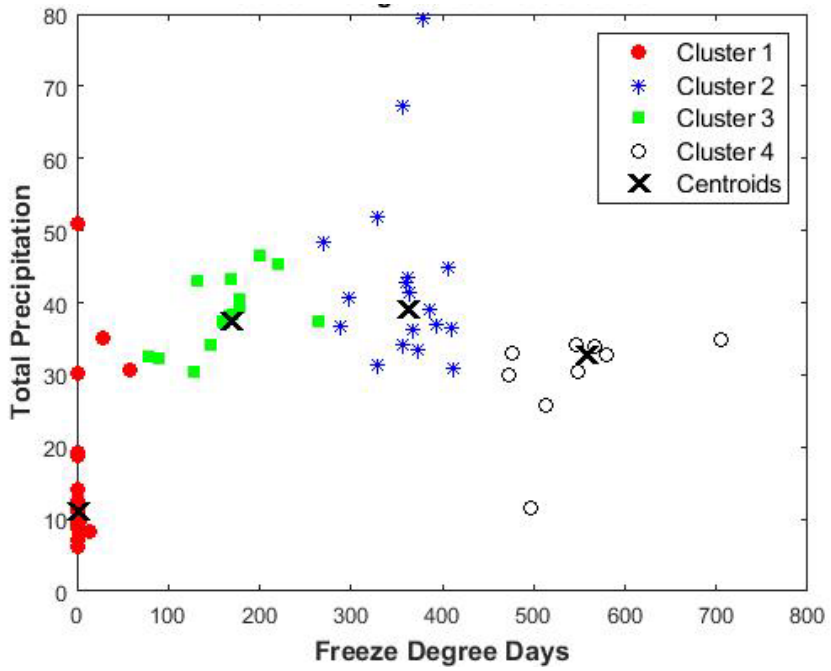


Figure 53. Cluster Assignments of the Performance Records Based on FDD and TPrec

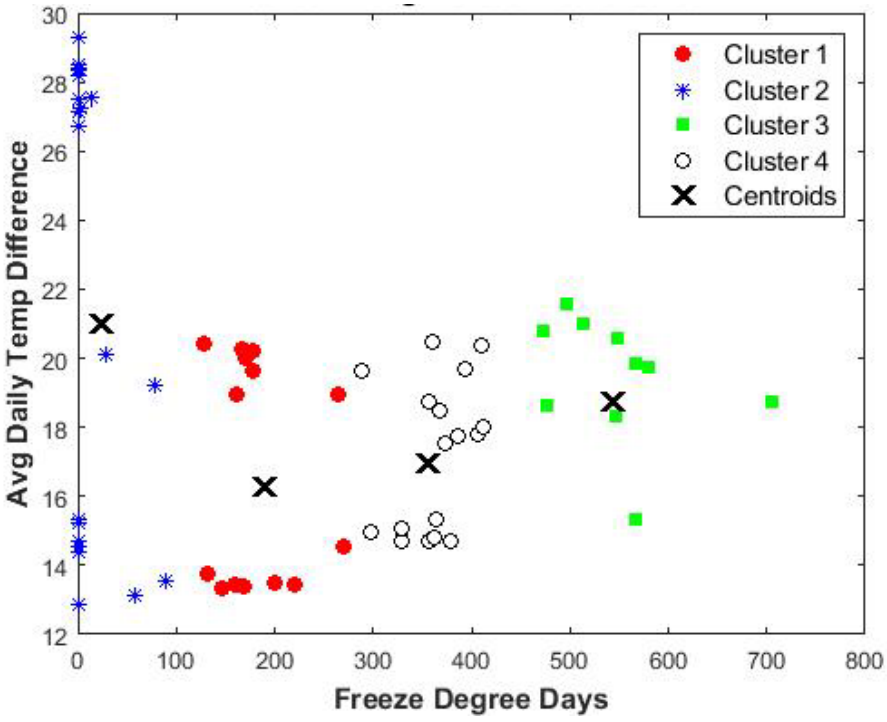


Figure 54. Cluster Assignments of the Performance Records Based on FDD and TempDiff

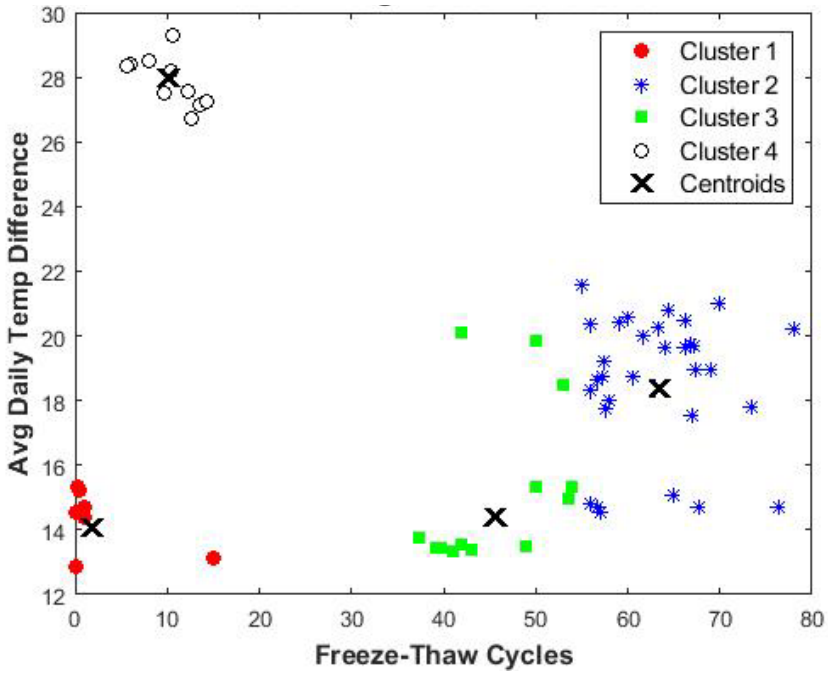


Figure 55. Cluster Assignments of the Performance Records Based on FThC and TempDiff

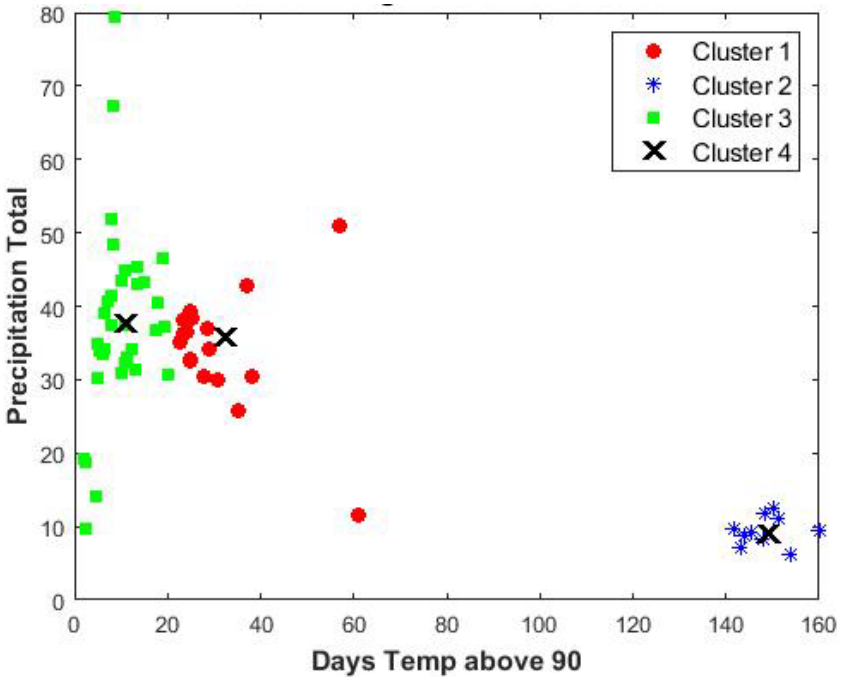


Figure 56. Cluster Assignments of the Performance Records Based on Temp90 and TPrec

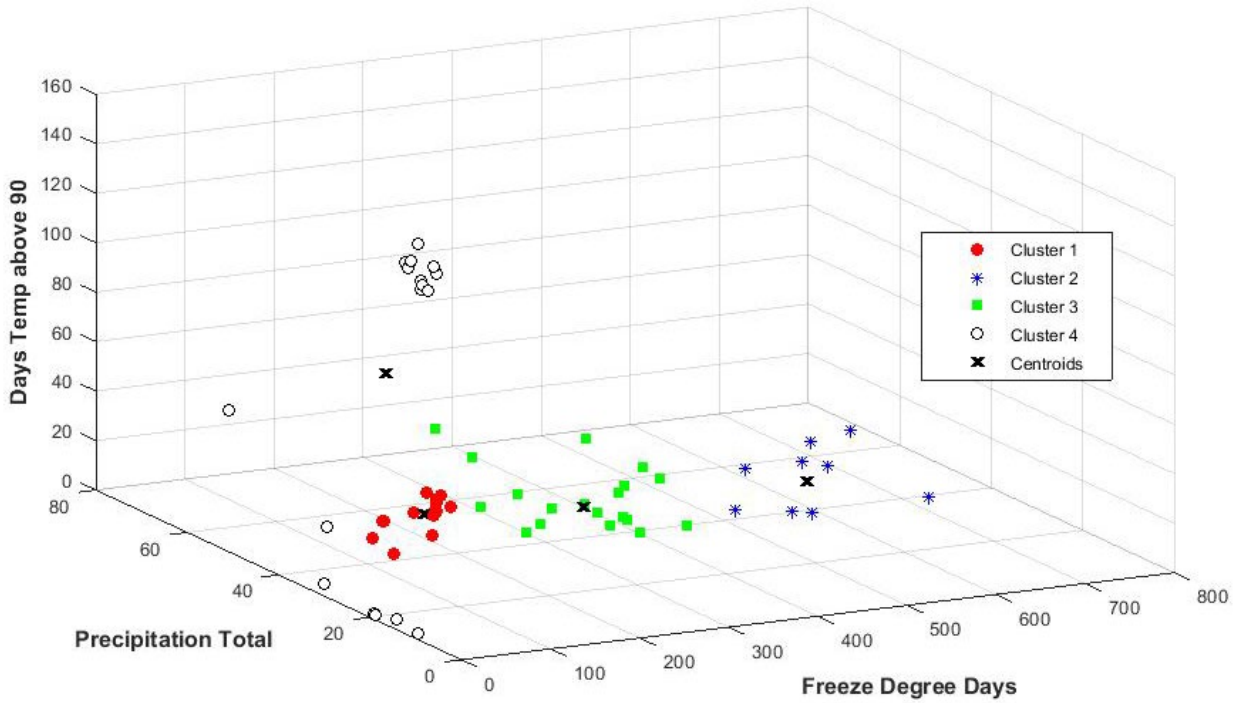


Figure 57. Cluster Assignments of the Performance Records Based on FDD, TPrec, and Temp90

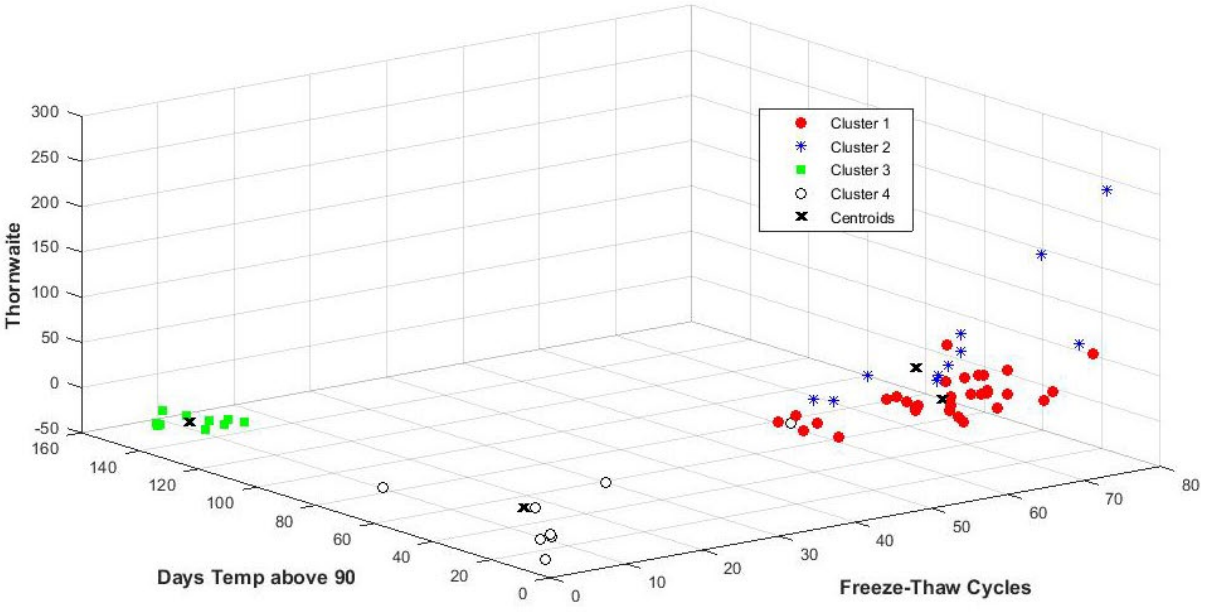


Figure 58. Cluster Assignments of the Performance Records Based on FThC, Temp90, and Thornthwaite Index

The outcome of cluster analysis can be used to divide the performance data records into clusters with homogeneous climatic properties. The high-dimensional climate feature space will be

reduced to only one feature, which includes the cluster assignments for each performance record. The assigned clusters can be treated as a categorical input variable for developing the ML predictive models. The optimal clustering strategies among those evaluated in this study (as illustrated in figures 53 to 58) will be the ones that yield the best prediction performance.

4.5.4 Principal Component Analysis

PCA is an unsupervised technique for dimensionality reduction of the multivariate feature (input) space. This reduction is achieved by transforming the data onto a set of dominant variables known as principal components (PC). PC are new uncorrelated variables constructed via linear combinations of all input variables (Jolliffe, 2002). The first few principal components account for the largest possible variance in the original data set. The dimension of the input space can be reduced by eliminating the components with low variance. Through decomposition of the covariance matrix of the original input data, the PCA computes the eigenvalue and eigenvectors. Each eigenvector is a unit vector in the direction of the new PC coordinate, and the PC with the highest eigenvalue explains the most variation. The PC with the highest variance will be used as inputs to the ML model under the assumption that they contain the most information. Unlike the feature selection approaches, PCA allows for incorporation of all input variables in the development of the ML model. Along with dimensionality reduction, PCA eliminates the intercorrelation between the features, making the model less likely to overfit. A major disadvantage of PCA is that the principal components do not have any physical meaning, making it difficult to interpret the influence of specific input variables.

The MATLAB "pca" function was used to perform the PCA. For an $m \times n$ matrix, where m is the number of observations and n is the number of features (input variables), the PCA returns an $n \times n$ coefficient matrix that contains coefficients of principal components. Each column of the latter matrix contains coefficients for one principal component, and the columns are in descending order of component variance. For each pavement performance record in the database, the PCA considered the following as input variables: eleven climate variables (evaluated as the average value since the previous measurement, or from the last rehabilitation); the pavement age; the previous anti-SCI; and the time since the previous anti-SCI. Since the ranges of input variables are all different, they were normalized to the range (-1, +1) prior to PCA, ensuring that each variable contributes equally to the analysis. Table 20 summarizes the PCA results. The first three PCs (highlighted in yellow) explain more than 95% of the total variance of the input variables and can be considered inputs to the ML model. The PCs are linear combinations of all 14 variables. Each column represents the coefficients of the PC.

Table 20. Principal Components and Coefficients of Each Input Variable

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14
Variance %	61.48	21.63	12.09	2.13	1.12	0.83	0.34	0.28	0.06	0.02	0.01	0.01	0.00	0.00
Age	0.158	-0.085	-0.023	0.175	0.035	0.003	0.103	0.502	-0.082	0.616	0.380	-0.350	-0.134	0.047
Previous Anti-SCI	0.280	0.061	0.509	-0.258	-0.189	-0.573	-0.218	-0.327	-0.090	0.203	0.131	-0.076	-0.016	0.027
Delta-Time	0.157	-0.088	-0.014	0.110	0.053	-0.083	-0.101	0.200	-0.403	0.156	-0.041	0.811	-0.215	-0.047
FThC	0.058	-0.031	0.202	0.279	-0.211	-0.101	0.860	-0.174	-0.067	-0.137	0.102	0.070	-0.121	-0.020
FDD	-0.572	-0.501	0.441	0.300	-0.278	0.050	-0.223	0.086	0.001	0.000	-0.008	0.002	0.013	0.004
Temp90	0.623	-0.342	0.089	0.386	0.013	0.352	-0.180	-0.293	0.264	0.040	-0.134	-0.024	-0.071	0.024
Dprec	0.074	0.215	0.536	-0.385	-0.030	0.639	0.101	0.124	-0.240	0.058	-0.132	-0.020	-0.007	-0.001
RH	0.138	0.075	0.201	-0.086	-0.131	-0.225	0.099	0.538	0.621	-0.040	-0.375	0.162	-0.005	0.080
HydD	-0.056	0.734	0.116	0.592	-0.193	0.039	-0.233	0.002	-0.021	-0.009	-0.001	-0.010	-0.001	-0.001
Temp Diff	0.214	-0.094	0.032	0.135	-0.033	-0.087	0.051	0.191	-0.256	-0.053	-0.088	-0.022	0.866	-0.226
TPrec	0.104	0.014	0.199	-0.033	0.172	0.083	-0.107	0.164	0.239	-0.443	0.758	0.188	0.064	-0.078
Sky Cover	0.159	-0.077	0.008	0.071	0.008	-0.096	-0.063	0.207	-0.339	-0.404	-0.066	-0.180	-0.046	0.768
Wind	0.159	-0.070	0.034	0.044	0.011	-0.142	-0.104	0.251	-0.259	-0.403	-0.182	-0.334	-0.399	-0.583
Thornthwaite	-0.138	0.036	0.338	0.204	0.867	-0.154	0.087	-0.058	0.018	0.065	-0.172	-0.045	0.016	0.022

4.6 ANALYSIS OF TRAFFIC DATA

Historical traffic data in the PA40 database are available for the flexible runways identified in table 21. The traffic data for the date range June 2014 to October 2019 were obtained from threaded track data (TTD) reports provided by the MITRE Corporation. Additional traffic data for years prior to 2014 were obtained directly from the airports (BWI and SFO). In PA40, runway usage data are searchable by date, aircraft type, arrival or departure, and runway end.

Table 21. Available Traffic Data in Flexible Runways

Airport	Runway	Available Years
BOS	4L-22R	June 2014 to Oct 2019
CMH	10L-28R	Jan 2014 to Oct 2019
CMH	10R-28L	Jan 2014 to Oct 2019
GSO	5L-23R	June 2014 to Sep 2019
LGA	4-22	June 2014 to Oct 2019
MCI	9-27	June 2014 to Sep 2019
MIA	12-30	June 2014 to Oct 2019
TUS	3-21	June 2014 to Oct 2019
TUS	11L-29R	June 2014 to Oct 2019
BWI	10-28	Jan 2011 to Oct 2019
SFO	10R-28L	Jan 1999 to Oct 2019

The objective of this section is to identify the most relevant traffic indices that influence pavement performance. Since the traffic data were collected from various sources, there is some inconsistency in data presentation. For example, some of the traffic data had information on the specific aircraft type for a given operation, whereas others broadly categorized the aircraft traffic based on type or operational use (general aviation versus military). In an effort to standardize the traffic data characteristics, the aircraft were classified into three broad categories based on the maximum takeoff weights (MTOW). Aircraft Category 1 (AC1) includes any aircraft with MTOW less than 100,000 lb. Aircraft Category 2 (AC2) includes aircraft with MTOW ranging from 100,000 lb to 350,000 lb. Aircraft with MTOW over 350,000 lb were identified as Aircraft Category 3 (AC3). MTOW for individual aircraft were assigned according to the FAA Aircraft Characteristics Database (FAA, 2018a).

Eight traffic indices were initially identified by considering the departure and arrival counts of each of the aircraft categories (AC) as summarized in table 22. Since heavier aircraft are likely to induce the most damage, a traffic count encompassing operations of aircraft with MTOW higher than 100,000 lb was also considered as a traffic index.

Table 22. Traffic Indices

Traffic Index	Description
DepAC2	Departures of Aircraft Category 2
DepAC3	Departures of Aircraft Category 3
ArrAC2	Arrivals of Aircraft Category 2
ArrAC3	Arrivals of Aircraft Category 3
DepAC2&3	Summation of Departures of Aircraft Categories 2 and 3
DepT	Total Departures
ArrT	Total Arrivals
TrafficT	Total Traffic

The annual values of the total traffic (TrafficT), departures of AC2 (DepAC2)and departures of AC3 (DepAC3) for some of the flexible runways are shown in figures 59 to 61. Most of the runways exhibited a relatively steady DepAC2 index. Runway 10-28 at BWI showed some fluctuation in DepAC2 traffic between 2011 and 2018, but an increasing trend can be observed after 2015. The DepAC3 index for SFO Runway 10R-28L declined from 2000 to 2006, then increased after 2012. The DepAC3 for the other runways showed small fluctuations. While a growth rate in fleet mix is sometimes assumed for the design, the variations of the annual traffic indexes do not indicate gradual growth in traffic for most of the runways. Given the limited data available for most of the runways (only 4 full years), it is not possible to reasonably determine a growth rate for the traffic indices.

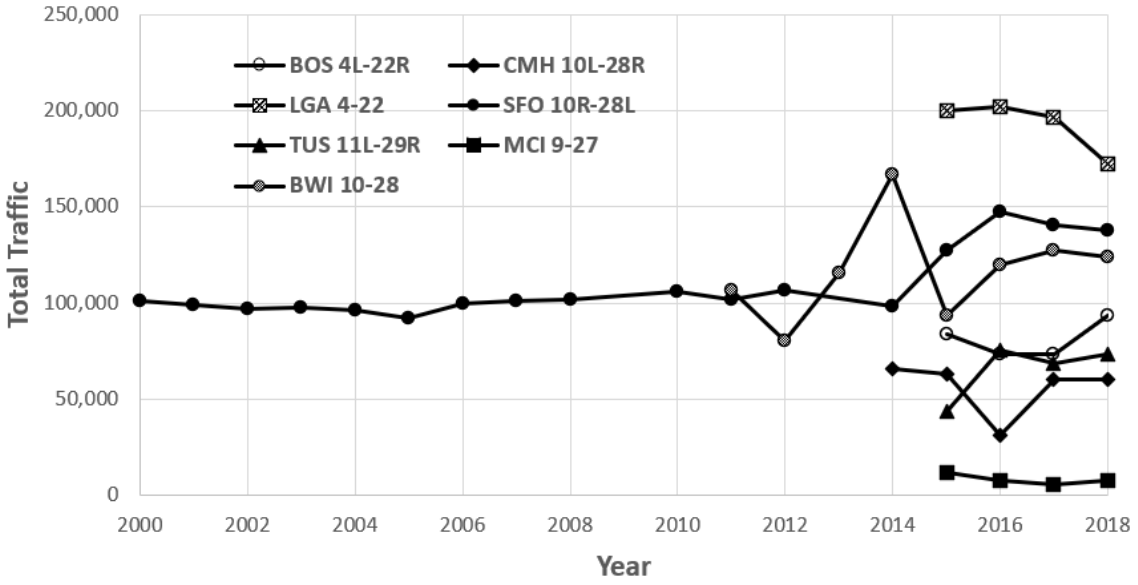


Figure 59. Variation of Total Traffic in Selected Runways Over the Course of Traffic Data Availability

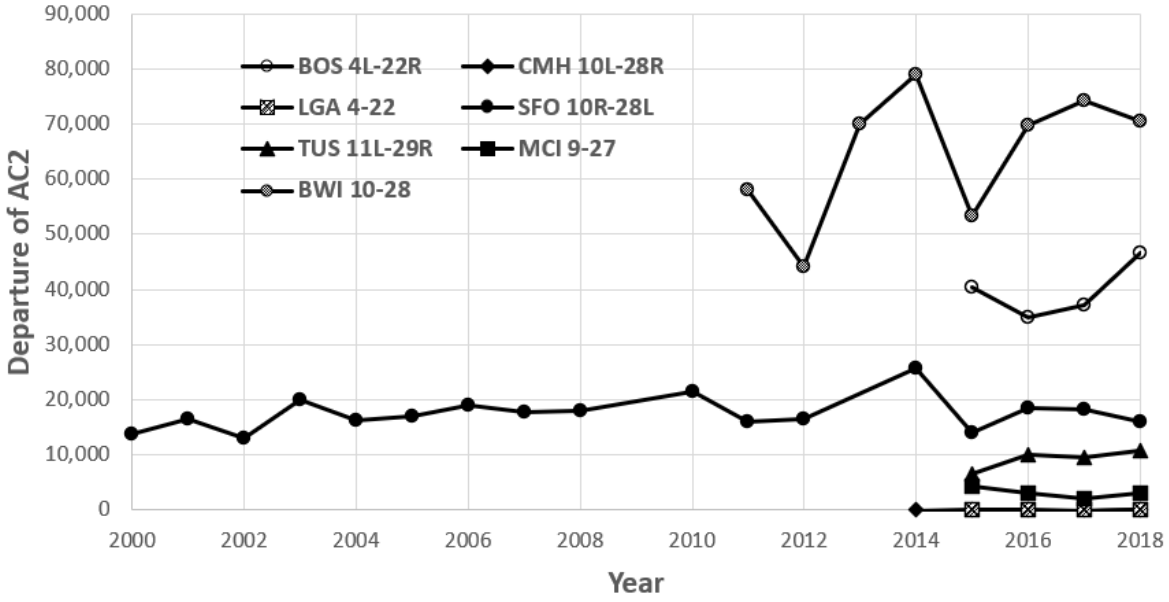


Figure 60. Variation of DepAC2 Index in Selected Runways Over the Course of Traffic Data Availability

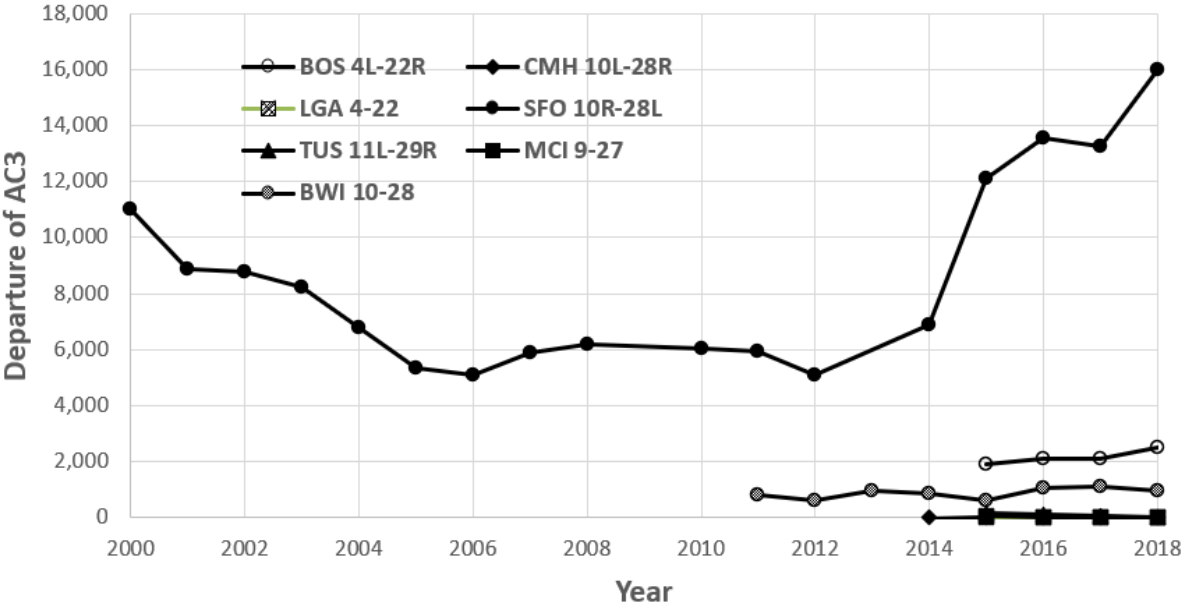


Figure 61. Variation of DepAC3 Index in Selected Runways Over the Course of Traffic Data Availability

For most of the runways in table 21, the majority of PCI survey data are for years prior to 2014, for which no traffic data are available. To evaluate the influence of traffic on pavement performance, available traffic data were back-projected to prior years, using the average annual traffic for the four years of available data. Since it is not possible to determine a growth rate for the traffic indices, the average values of each traffic index for years between 2015 and 2018 were

used to characterize the traffic indices for each runway. Where available (i.e., for SFO and BWI), the actual traffic indexes were used to evaluate the correlation to pavement performance. Figure 62 gives the average values of total departures and arrivals, and figure 63 gives the average distribution of departures of each aircraft category on each runway. As shown in figure 62, Runway 10-28 at BWI had the greatest number of departures followed by Runway 4L-22R at BOS and 4-22 at LGA. Figure 63 shows that Runway 10-28 at BWI and Runway 10R-28L at SFO had the greatest numbers of AC2 and AC3 departures, respectively.

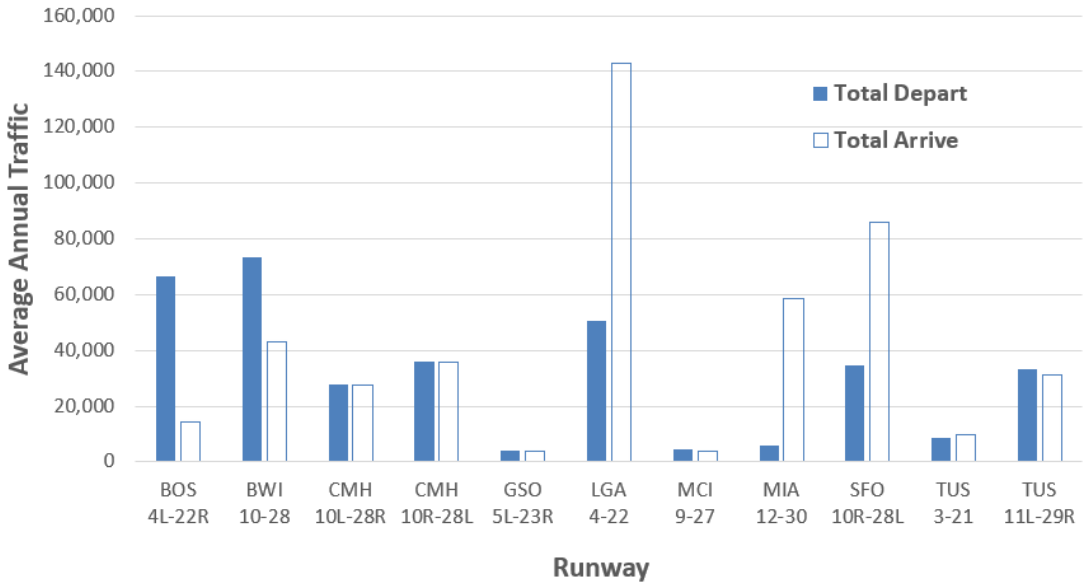


Figure 62. Average Annual Total Departures and Arrivals on Each Runway

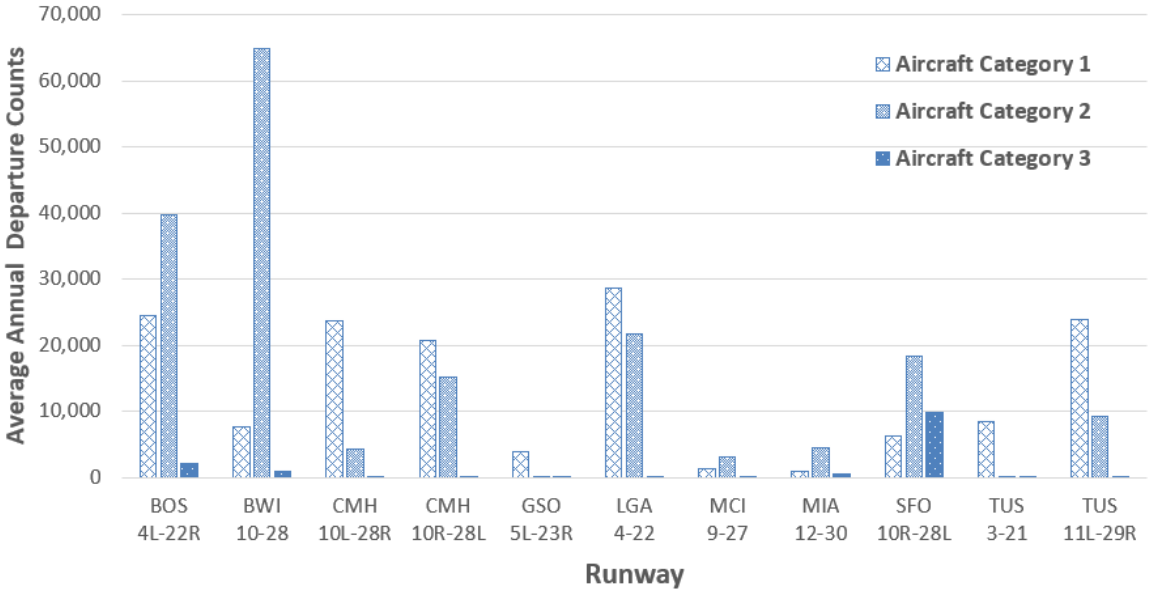


Figure 63. Average Annual Departure Based on Aircraft Weight on Each Runway

4.6.1 Pairwise Correlation Between Traffic Indices

Figure 64 shows the Pearson correlation matrix, which illustrates the strength of pairwise linear relationships between traffic indices. Pairwise correlations exceeding a threshold of 0.7 can be considered high collinearity. Based on this threshold value, DepT is highly correlated to DepAC2 and poorly correlated to DepAC3. Also, there is a low correlation between DepAC2 and DepAC3. The correlations in figure 64 are valid only for the evaluated traffic data in the current database and cannot be extended to other runways.

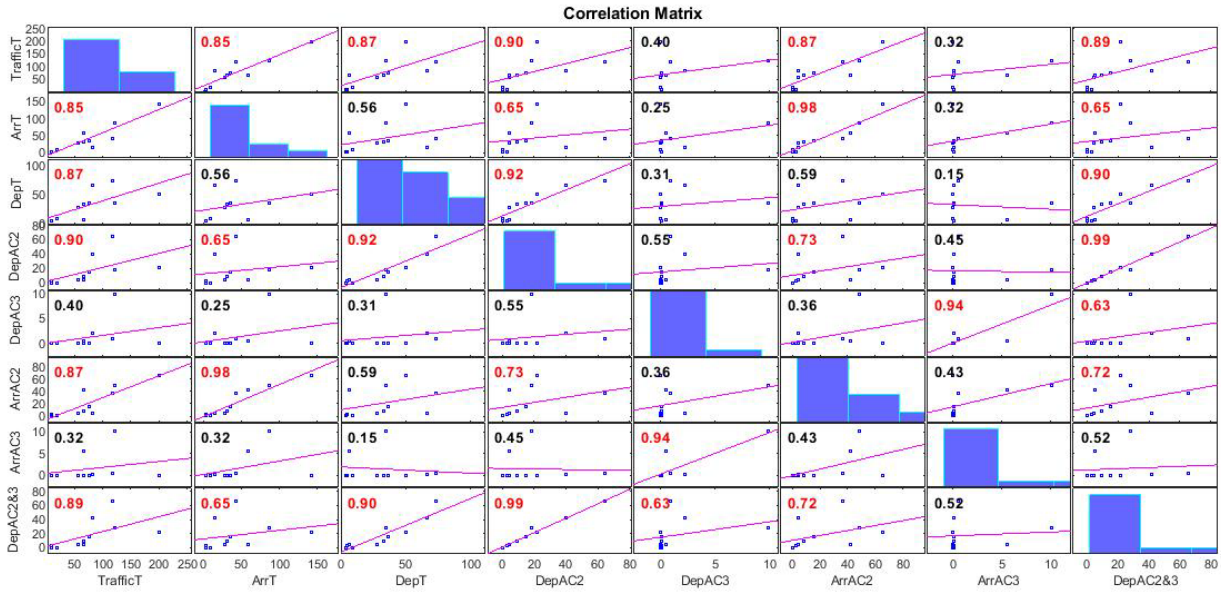


Figure 64. Pearson Correlation Between Traffic Indices

4.6.2 Correlation Between Traffic and Pavement Performance

This section examines the correlations between the runway traffic index and two pavement performance indices, PCI and SCI, both of which include load-related distresses. Since the traffic data concurrent with the performance are not available for most of the flexible runways, the annual average of each traffic index over the years for which traffic data are available (as identified in table 21) was used to calculate a representative traffic index at each runway. For each flexible pavement section in the runway keel areas, the accumulated traffic index was calculated from the last major rehabilitation/construction, and regressed against the performance index. The accumulated traffic index for each pavement section was calculated by multiplying the pavement age by the annual average traffic index value. For example, Section 9105C of RWY 4L-22R at BOS, PCI was inspected in 2012, approximately 8 years after a rehabilitation. For this runway, the average annual DepAC2 and DepAC3 are 39,790 and 2,160, respectively. Therefore, calculate the accumulated traffic index DepAC2&3 as 8 × (39,790 + 2,160) = 335,600. Regression analysis was used because only a single independent variable (traffic) was involved. PCI was plotted against the accumulated DepAC2&3 index, as shown in figure 65. While the PCI deterioration exhibits a downward trend, the R^2 resulting from fitting a logarithmic curve is close to zero. The variations of PCI with respect to accumulated DepAC2 and DepAC3 are also very low, as shown in figures 66 and 67. These results indicate that the evaluated traffic indices alone cannot explain most of the

PCI variations. The variation of PCI with pavement age is shown in figure 68 for comparison. Given that the R^2 from PCI-vs-Age graph ($R^2 = 0.67$) is much higher than that from the PCI-vs-Accumulated Traffic graphs, it may be concluded that pavement age is a much stronger predictor than the accumulated traffic alone. This suggests that it is not realistic to replace pavement age with just accumulated traffic when predicting PCI.

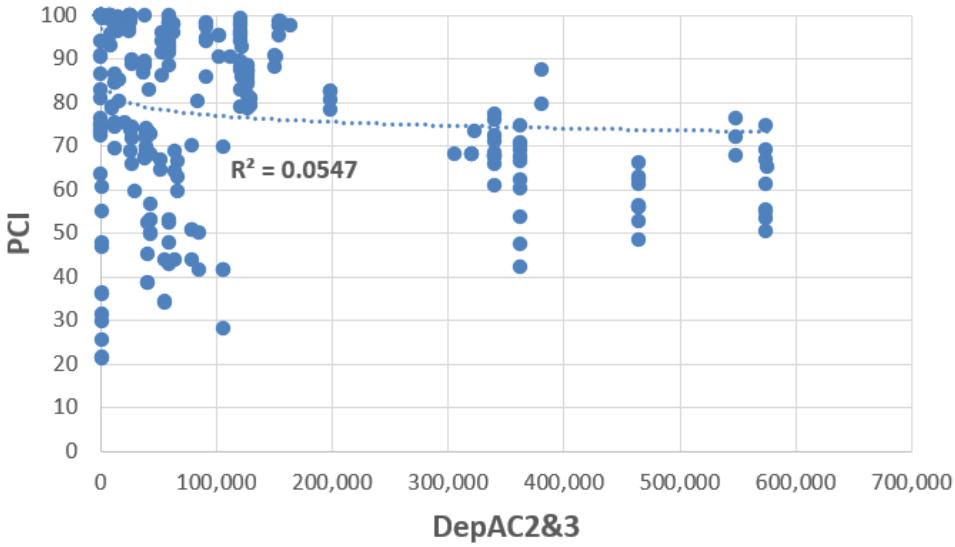


Figure 65. Variations of PCI With Accumulated DepAC2&3

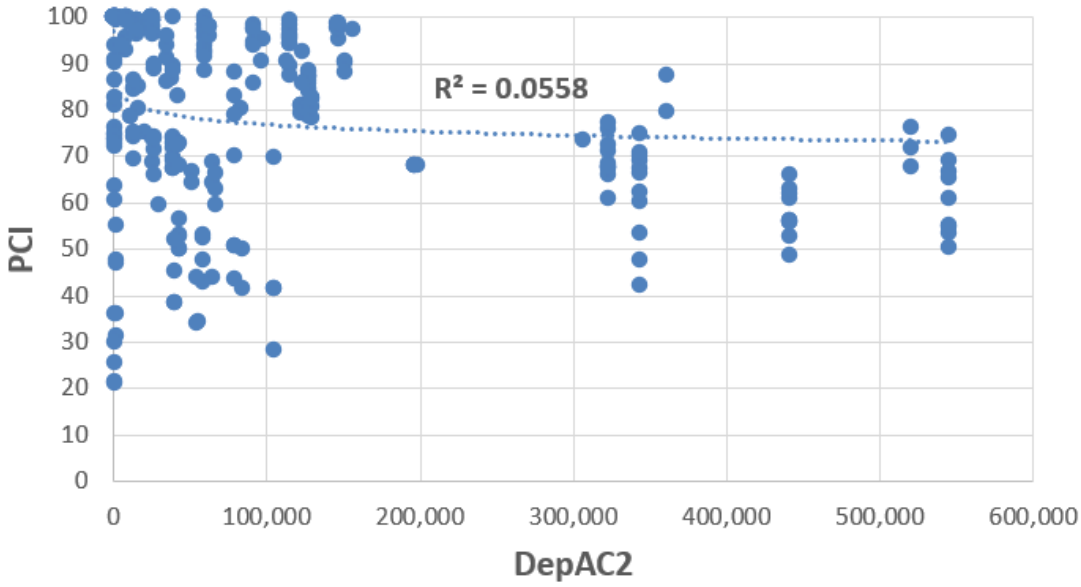


Figure 66. Variations of PCI With Accumulated DepAC2

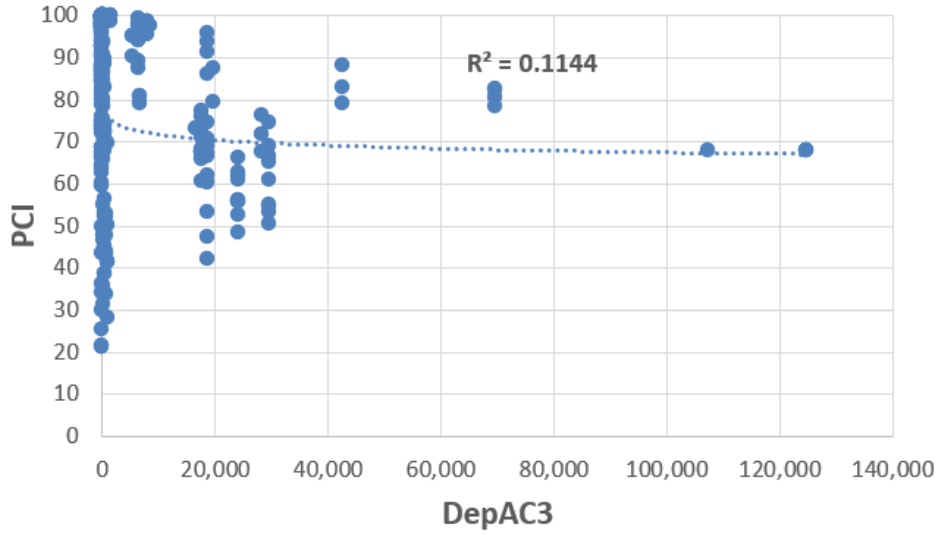


Figure 67. Variations of PCI With Accumulated DepAC3

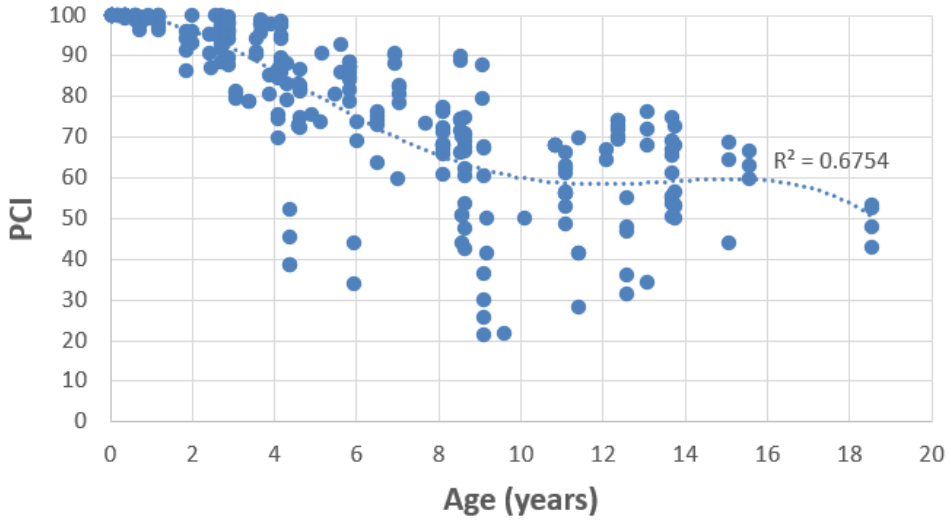


Figure 68. Variations of PCI With Pavement Age

The variation of SCI with accumulated DepAC2 and DepAC3 is shown in figures 69 and 70. The R^2 values resulting from fitting logarithmic curves are close to zero. Some pavement sections experienced low traffic but still exhibited poor SCI (e.g., TUS Runway 3-21). There are also some pavement sections with SCI near 100 with high accumulated traffic (e.g., SFO Runway 10R-28L). These observations indicate that poor SCI cannot be solely attributed to the high traffic level. One explanation for poor SCI in a low trafficked runway is that the SCI was compared to the cumulative traffic since the last major rehabilitation. This eliminates the contribution of the accumulated traffic since initial construction, but before rehabilitation, to the consumed fatigue life of the asphalt layer. While the rehabilitation activities bring the functional conditions back to a serviceable standard, they may not address structural deficiencies. In this case, for an old asphalt pavement that received multiple overlays spanning its lifetime, it is possible that the fatigue

damage that has been accumulated over the years at the bottom of the asphalt layer eventually manifests itself at the surface. One explanation for highly trafficked runways exhibiting negligible fatigue cracking (SCI close to 100) is the difference between the design traffic and actual traffic.

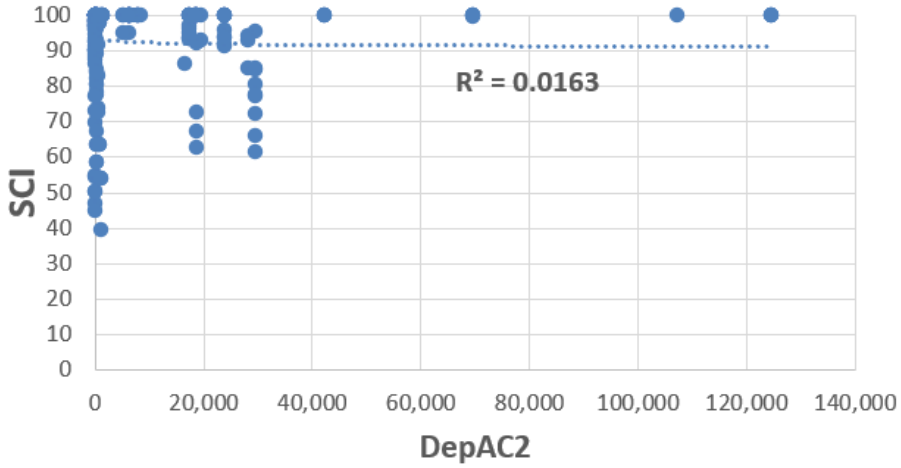


Figure 69. Variations of SCI With Accumulated DepAC2

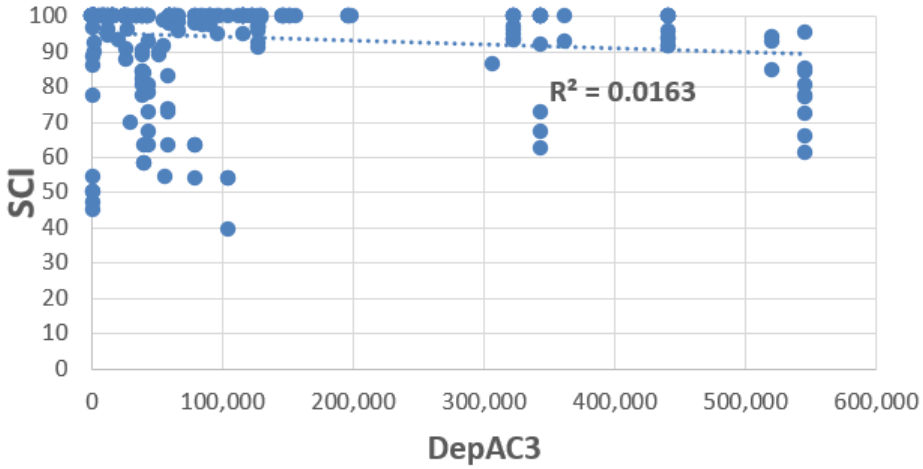


Figure 70. Variations of SCI With Accumulated DepAC3

4.6.3 Implementation of Feature Ranking on Traffic Indices

Two sets of feature ranking analysis were performed by considering PCI and SCI as the targets and eight traffic indices as predictors. Traffic indices were treated as cumulative values from date of construction/rehabilitation to the date of inspection. Pavement age at the time of PCI/SCI measurements was also considered as a predictor. Two filter methods, Pearson Correlation and ReliefF, and two wrapper methods (considering linear regression and SVM as the learner) were used for feature ranking analysis. Table 23 summarizes the rank of each feature with respect to PCI, and the associated weights or scores as calculated by each ranking algorithm. Note that, except for the Pearson correlation (defined on [-1, 1]), the absolute scores for different methods

are not bounded, nor are they on the same scale (i.e., the scores given by RRelief are not directly comparable to the scores given by SVM). All ranking methods identified pavement age as the most relevant predictor. Pavement age scores much higher than any of the traffic indices. Overall, DepT received high ranks and ArrT received low ranks among the traffic indices. However, the ranking methods were not in agreement in identifying a traffic index as the most informative predictor. Table 24 summarizes the rank of each feature with respect to SCI. All the ranking methods gave the highest rank to pavement age. DepT received a higher rank than other traffic indices. The Pearson correlation (R) indicates very low correlations between SCI and traffic indices, and for some indices the increase in traffic led to an increase in SCI, contradicting engineering expectations. The linear regression wrapper method gave nearly zero score to most of the traffic indices. Also, the SVM wrapper method gave roughly the same score to all traffic indices. These results suggest that the ranking algorithms were not able to identify any of the traffic indices as a relevant predictor of SCI.

Table 23. Results of Feature Ranking Analysis on Traffic Indices With PCI as Target

Filter Methods				Wrapper Methods			
Pearson Correlation		RReliefF		SVM Gaussian Kernel		Liner Regression	
Input Variable	*R*	*Input Variable*	*Score*	*Input Variable*	*Score*	*Input Variable*	*Score*
Age	-0.79	Age	0.024	Age	14.6	Age	7.52
DepT	-0.43	DepAC3	0.012	DepT	7.6	DepT	1.94
DepAC2	-0.32	ArrAC3	0.012	ArrAC2	7.0	DepAC2	1.04
DepAC2&3	-0.32	DepAC2&3	0.010	TrafficT	6.9	DepAC2&3	1.03
TrafficT	-0.28	TrafficT	0.010	DepAC3	6.8	TrafficT	0.71
DepAC3	-0.17	DepT	0.008	DepAC2	6.8	DepAC3	0.27
ArrAC2	0.10	DepAC2	0.007	DepAC2&3	6.6	ArrAC2	0.00
ArrAC3	-0.05	ArrAC2	0.007	ArrT	6.6	ArrAC3	0.00
ArrT	-0.02	ArrT	0.006	ArrAC3	6.5	ArrT	0.00

Table 24. Results of Feature Ranking Analysis on Traffic Indices With SCI as Target

Filter Methods				Wrapper Methods			
Pearson Correlation		RReliefF		SVM Gaussian Kernel		Liner Regression	
Input Variable	*R*	*Input Variable*	*Score*	*Input Variable*	*Score*	*Input Variable*	*Score*
Age	-0.47	Age	0.040	Age	14.4	Age	1.39
DepT	-0.19	TrafficT	0.013	DepT	13.7	DepT	0.20
DepAC2	-0.13	DepT	0.012	DepAC2	13.6	ArrAC2	0.08
ArrAC2	0.13	DepAC2	0.007	DepAC2&3	13.6	DepAC2	0.08
DepAC2&3	-0.12	ArrT	0.006	ArrAC3	13.4	DepAC2&3	0.04
ArrAC3	0.09	DepAC2&3	0.003	TrafficT	13.4	ArrT	0.00
TrafficT	-0.08	ArrAC2	0.001	DepAC3	13.4	DepAC3	0.00
ArrT	0.06	DepAC3	-0.015	ArrAC2	13.4	ArrAC3	0.00
DepAC3	0.03	ArrAC3	-0.016	ArrT	13.4	TrafficT	-0.04

4.6.4 Clustering of Traffic Data Using *k*-means

Runways listed in table 21 were grouped based on the similarities between two of their traffic indices by implementing *k*-means cluster analysis. Table 25 illustrates the combinations of the traffic indices used for clustering. The optimum number of clusters was three. The results of clustering are shown in figures 71 through 73. The outcome of cluster analysis can be used to group the performance data records from runways with similar traffic levels into the same clusters.

Table 25. Possible Combinations of Traffic Index for Clustering

Traffic Index 1	Traffic Index 2
DepT	DepAC3
DepAC2&3	DepAC3
DepAC2	DepAC3

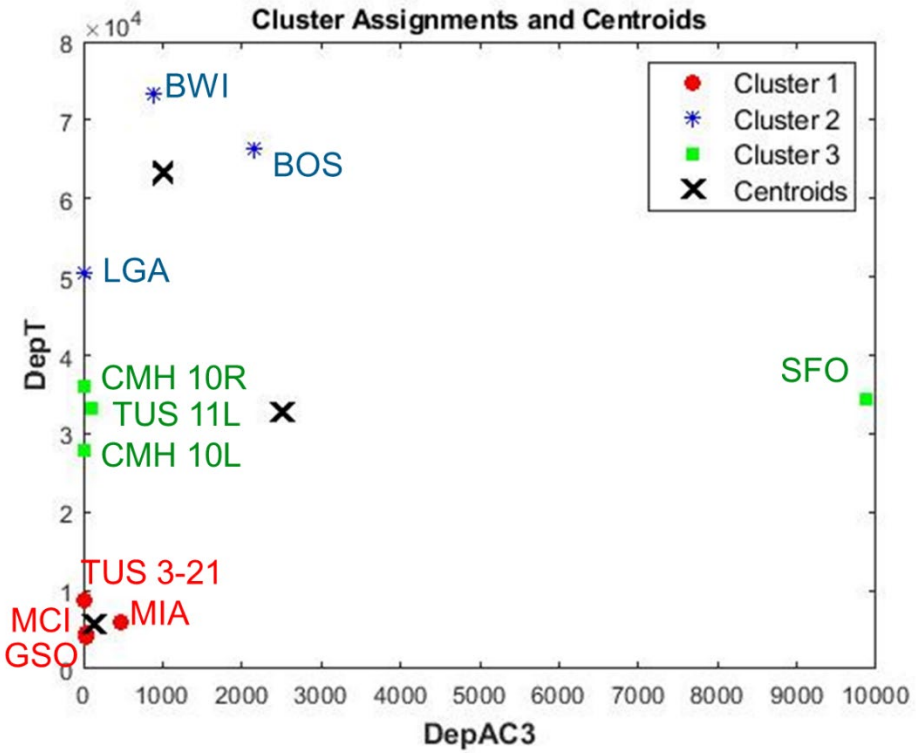


Figure 71. Cluster Assignments of the Runways' Traffic Level Based on DepT and DepAC3

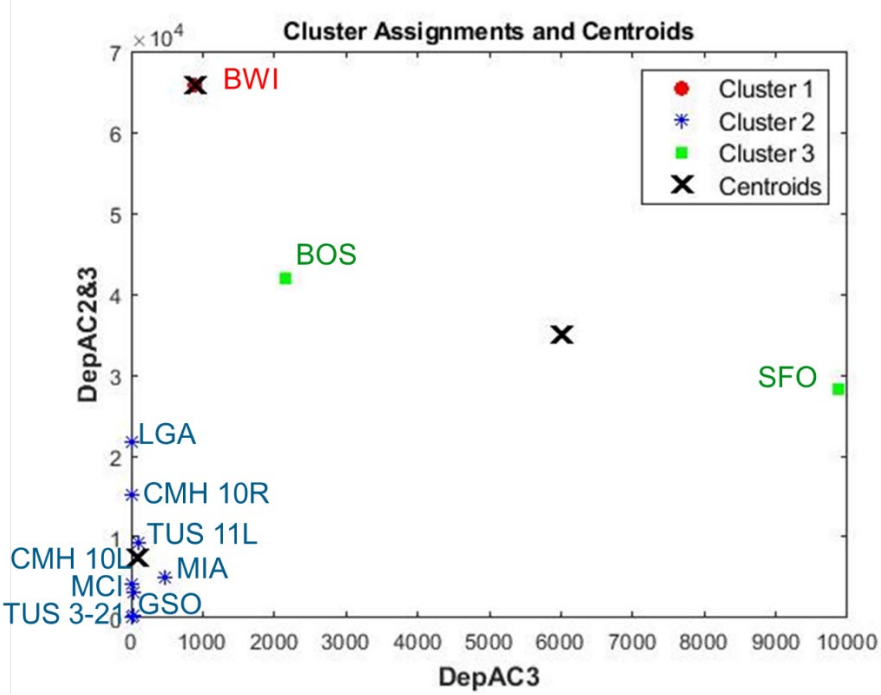


Figure 72. Cluster Assignments of the Runways' Traffic Level Based on DepAC2&3 and DepAC3

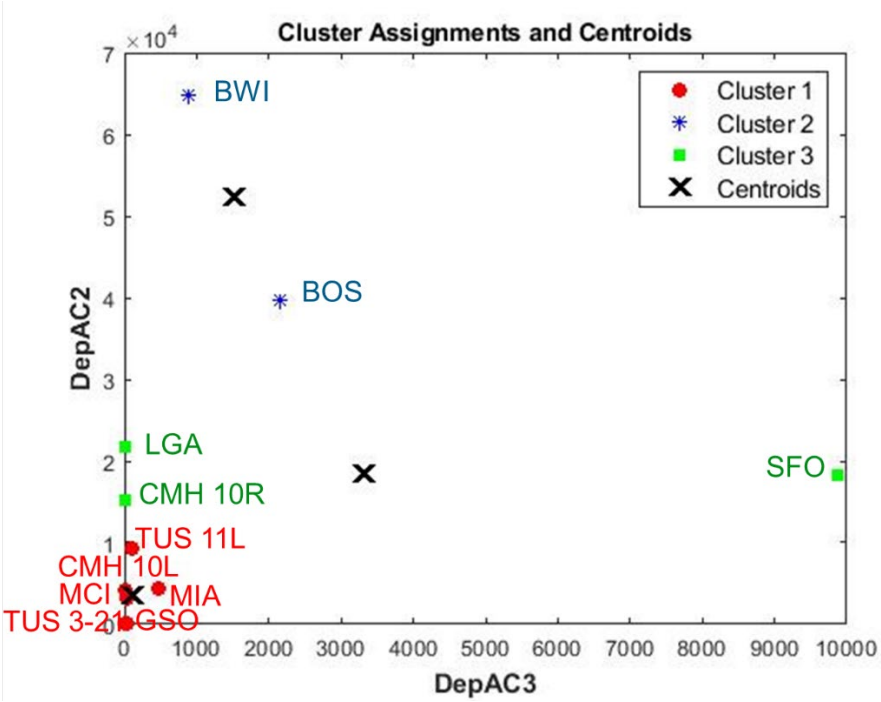


Figure 73. Cluster Assignments of the Runways' Traffic Level Based on DepAC2 and DepAC3

4.6.5 Conclusion of Traffic Analysis

Results of regression analyses indicated that PCI and SCI have poor correlations with accumulated traffic, i.e., traffic alone cannot explain all the variations in pavement performance. Runways with higher numbers of heavier aircraft (AC2 and AC3) did not show a higher rate of pavement deterioration. However, runways with lower traffic levels did exhibit load-related distresses, such as alligator cracking. This suggests that traffic alone is not adequate to explain the development of structural distresses in pavements. The presence of structural distresses can also be due to other factors such as inadequate pavement thickness and quality of construction materials. Feature ranking analysis identified pavement age as a more powerful predictor of PCI and SCI than any traffic index. This analysis did not identify any of the evaluated traffic indices as a highly relevant predictor. Total departure (DepT) received a higher rank, and total arrival (ArrT) received the lowest rank among the evaluated traffic indices. Nevertheless, low correlation between the traffic indices and PCI/SCI do not imply that traffic is not a contributor to pavement performance. Pavement design principles, along with findings from the feature selection analysis and pairwise correlation, can be used to identify sets of traffic indices to be used as the input variables for developing pavement performance predictive models. Three options are recommended for incorporating traffic as an input variable for predicting PCI and SCI:

- Consider using the DepAC2 and DepAC3 traffic indices. This approach is consistent with the airport pavement thickness design procedure in which only the departures of heavy aircraft are considered. Moreover, these traffic indices exhibited low correlation with one another, making them good candidates to discriminate between the runways' traffic levels.

- Use cluster analysis by assigning a cluster number to each performance data record. Assign the cluster numbers as a categorical input variable to differentiate the traffic level between the performance data records.

- Perform PCA by considering a set of traffic indexes (e.g., DepAC2 and DepAC3) in conjunction with other predictors, such as weather variables, layer thickness, and maintenance activities.

5. DEVELOPMENT OF PREDICTIVE MODELS FOR PAVEMENT PERFORMANCE

Sections 2 and 3 of this report showed that static and autoregressive approaches can be used for developing performance prediction models. Implementing both approaches for a simple case, where pavement age was the only predictor, indicated that ML models based on autoregressive approaches were more promising for predicting pavement performance. Section 4 identified the most informative features for predicting pavement performance. Collinearity analysis identified high correlations between some of the climate/weather variables. Feature selection analysis identified a subset of climate/weather variables that have relatively high correlations to the target (anti-SCI). In addition, cluster analysis indicated that performance data can be grouped into three or four clusters based on two or three climate characteristics. The dimension of the feature space can be reduced by replacing the weather variables with a cluster number assigned to each data record. PCA determined that the feature space (input domain) can be reduced to three principal components. In the following sections, findings from the previous sections are used to develop

predictive models for anti-SCI using three approaches: subset of features, PCA, and *k*-means clustering. The Weka suite (Eibe et al., 2016) was used for modeling all three approaches.

5.1 DEVELOPING PREDICTION MODELS FOR ANTI-SCI USING A SUBSET OF INFORMATIVE FEATURES

In this approach, those climate/weather variables that showed high collinearity were removed from the input space, and six remaining climate/weather variables were considered for model development. An autoregressive approach was used, whereby the previous anti-SCI was considered as a predictor. RF was employed as the base learner for developing the anti-SCI prediction model. Table 26 summarizes the input and output variables to the RF model. A complete dataset including the input and output variables was created using the pavement performance data from flexible runways. A total of 278 data records from 10 runways were used for training the RF model. One section from each runway was set aside for independent testing (38 data points). A bagged ensemble of regression trees was employed in the RF model for target prediction. The out-of-bag technique was used for cross-validation to tune the regression tree hyperparameters and improve the predictive performance. The RF hyperparameters were optimized using random search. Figures 74 and 75 show the comparison of RF-predicted and measured anti-SCI, for both training and testing subsets. Table 26 summarizes the prediction errors for both training and testing subsets. The accuracy of the RF model was acceptable in both training and testing subsets as determined by relatively low error. The model generalized well as determined by comparable errors between the training and testing subsets. Table 26 shows that the model accuracy at 10% error is 95% on the testing subset, i.e., the prediction error was less than 10% for 95% of the data. The accuracy measures presented in table 27 describe the average performance of the prediction model, but do not address validity across the input domain. The error analysis shows that the prediction errors are highest for the anti-SCI values below 50. One reason for this higher error is that limited data exist for pavements with anti-SCI less than 50, which makes the prediction in that region less reliable. One explanation for this data scarcity is that airport owners tend to prevent the pavement condition index from falling below 60 by replacing or rehabilitating the pavement.

Table 26. Input and Output Variables to RF Model—Subset of Feature Approach

Variable Type	Variables	Descriptions
Input Variables	Age (t)	Time since last major rehabilitation
	Anti-SCI (t-1)	Previous anti-SCI measurement
	Delta-Time	Time since previous anti-SCI measurement
	Asphalt Thickness	Thickness of asphalt at the time of anti-SCI
	FThC	Average climate variables between two anti-SCI measurements
	FDD	
	Temp 90	
	TempDiff	
	TPrec	
	Thornthwaite Index	
Target	Anti-SCI (t)	Measured anti-SCI

Table 27. Performance of the RF Model—Subset of Features Approach

	Training Subset	Testing Subset
R^2	0.89	0.90
RMSE	5.16	4.55
RRSE	9.5%	9.0%
Accuracy (5% Error)	69%	68%
Accuracy (10% Error)	90%	95%

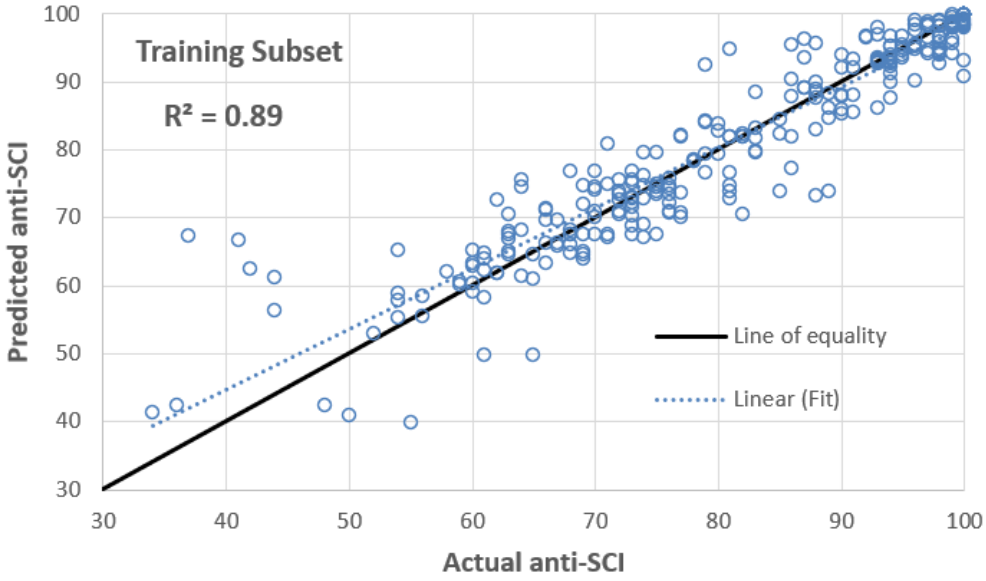


Figure 74. Comparison of Predicted and Actual Anti-SCI for Training Subset (subset of feature approach)

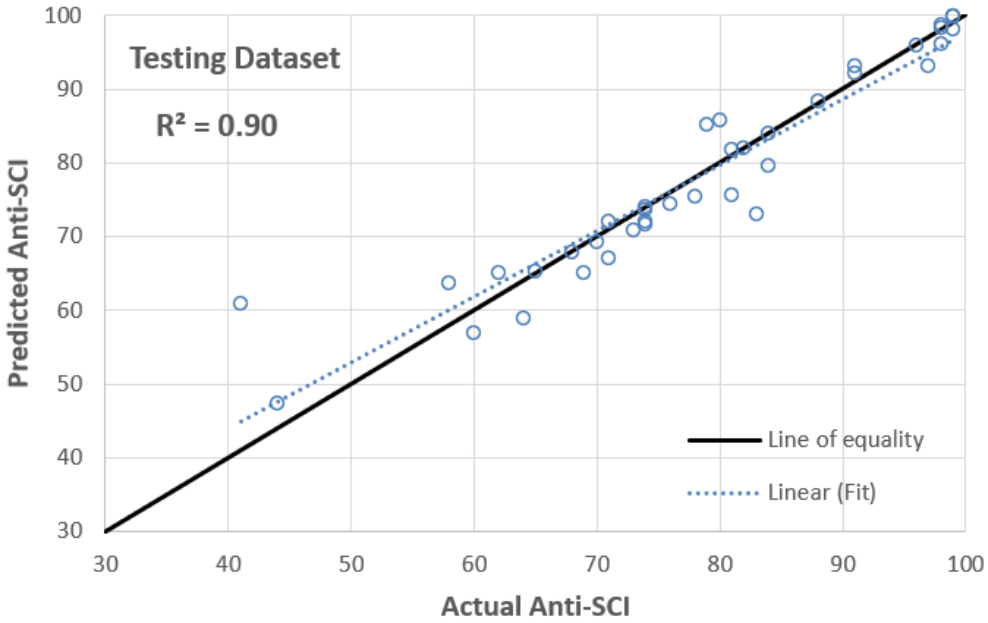


Figure 75. Comparison of Predicted and Actual Anti-SCI for Testing Subset (subset of feature approach)

5.2 DEVELOPING PREDICTION MODELS FOR ANTI-SCI USING PRINCIPAL COMPONENT ANALYSIS

Similar to the previous approach, an autoregressive model was used to predict anti-SCI. The input domain consisted of: pavement age, previous anti-SCI, time since previous anti-SCI, and 11 climate/weather variables (calculated as average values since the previous anti-SCI measurement), as identified in table 20. The input data were transformed into principal components. Three principal components that explain 95% of the variance of the input domain were considered as input variables for predicting anti-SCI, as shown in table 28. RF was employed as the base learner for developing the anti-SCI prediction model. A total of 278 data records from 10 runways were used for training the RF model, and one section from each runway was set aside for independent testing (38 data points). Table 29 summarizes the prediction errors for both the training and testing subsets. The model generalizes well as determined by comparable errors between the training and testing subsets. The estimation error using PCA is similar to the first approach, in which a subset of weather variables was considered. Figures 76 and 77 show the comparison of RF-predicted and measured anti-SCI, for both the training and testing subsets.

Table 28. Input and Output Variables to RF Model—PCA Approach

Variable Type	Variables
Input Variables	PC1
	PC2
	PC3
Target	Anti-SCI (t)

Table 29. Performance of the RF Model—PCA Approach

	Training Subset	Testing Subset
R^2	0.89	0.90
RMSE	5.11	4.81
RRSE	9.0%	7.3%
Accuracy (5% Error)	74%	67%
Accuracy (10% Error)	88%	90%

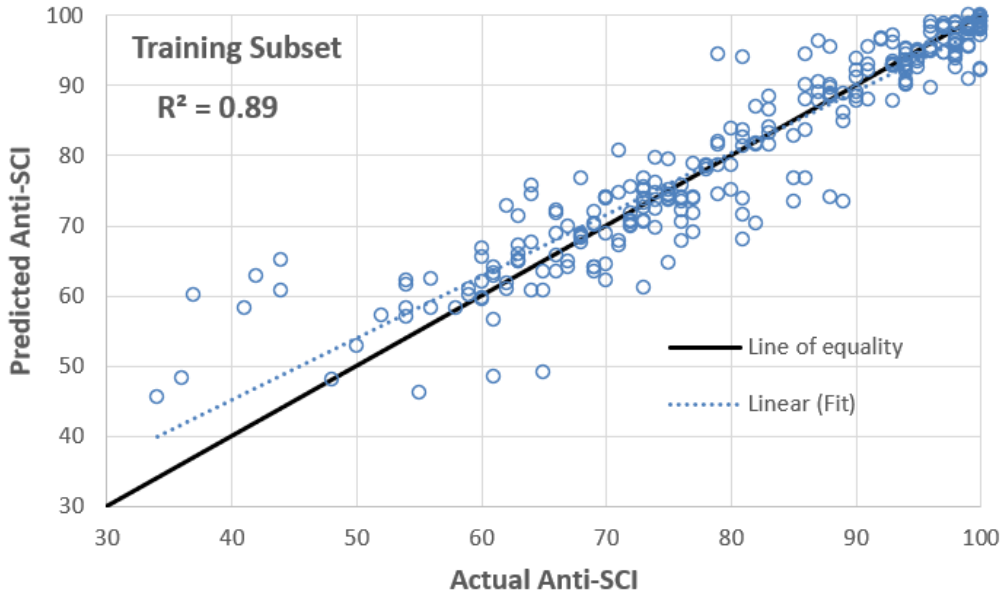


Figure 76. Comparison of Predicted and Actual Anti-SCI for Training Subset in the PCA Approach

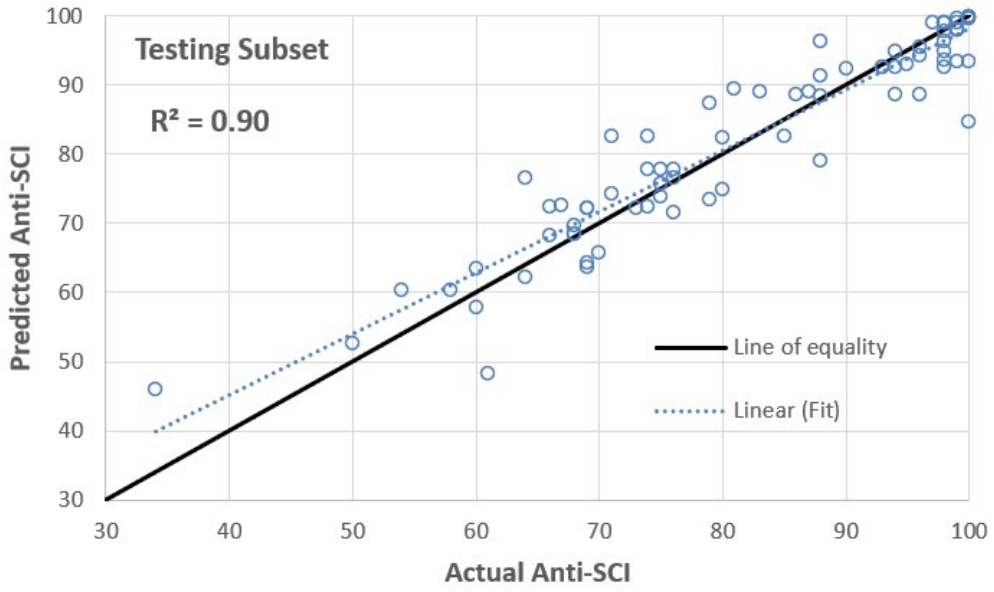


Figure 77. Comparison of Predicted and Actual Anti-SCI for Testing Subset in the PCA Approach

5.3 DEVELOPING PREDICTION MODEL FOR ANTI-SCI USING FEATURES FROM CLUSTER ANALYSIS

In this section, data records with similar climate characteristics were grouped into clusters using the result of *k*-means cluster analysis as presented in section 4.5. For each data record, the time since the last performance measurement was considered as a sliding window, and the average of each weather variable was calculated over that window. The data records were then grouped into clusters based on similarities of two or three of their average climate variables. After examining different combinations of climate variables for cluster analysis, it was determined that clustering based on three climate variables, namely: (1) FDD, (2) Temp90, and (3) TPrec, yielded the best model performance. The performance data records were then grouped into four clusters (numbered 1 through 4) based on their climate characteristics. Table 30 lists the average weather/climate variables at the centroid of each cluster. The numbers of data records assigned to each cluster are also shown in table 30.

Table 30. Average Climate Variables at the Centroid of Each Cluster

Climate Variables	Cluster 1	Cluster 2	Cluster 3	Cluster 4
FDD	108	330	549	2
Temp90	21	12	22	149
TPrec	35	40	30	9
Instances	117 (36%)	126 (39%)	31 (10%)	51 (16%)

Similar to the previous approaches, an autoregressive model was used for anti-SCI prediction and RF was used as the base learner. Table 31 illustrates the input and output variables in the RF model used in the cluster approach.

Table 31. Input and Output Variables in the RF Model—Cluster Approach

Variable Type	Variables	Descriptions
Input Variables	Age (t)	Time since last major rehabilitation
	Anti-SCI (t-1)	Previous anti-SCI measurement
	Delta-time	Time since previous anti-SCI measurement
	Asphalt thickness	Thickness of asphalt at the time of anti-SCI
	Climate Group	Cluster 1, 2, 3, or 4
Target	Anti-SCI (t)	Measured anti-SCI

The same training and testing subsets as in the PCA and the subset of variable approaches were used again in the cluster approach. Table 32 summarizes the prediction errors for both the training and testing subsets. The model generalized well, as shown by comparable errors between the training and testing subsets. Figures 78 and 79 show comparisons of the RF-predicted and measured anti-SCI, for both the training and testing subsets.

Table 32. Performance of the RF Model—Cluster Approach

	Training Subset	Testing Subset
R^2	0.88	0.93
RMSE	5.41	3.86
RRSE	10.2%	6.3%
Accuracy (5% Error)	69%	68%
Accuracy (10% Error)	90%	92%

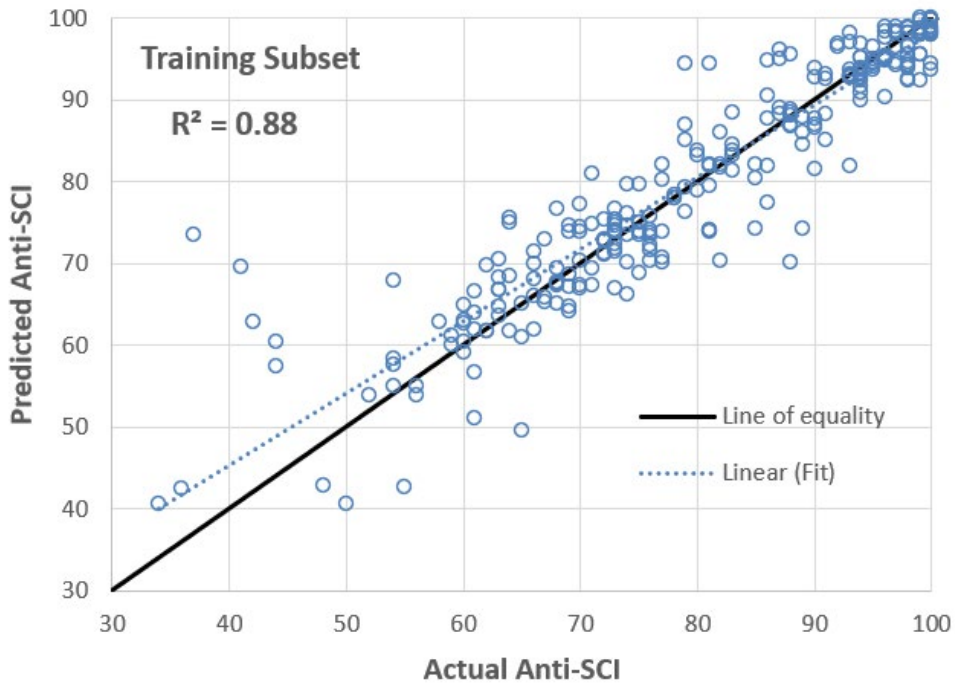


Figure 78. Comparison of Predicted and Actual Anti-SCI for Training Subset for the Cluster Approach

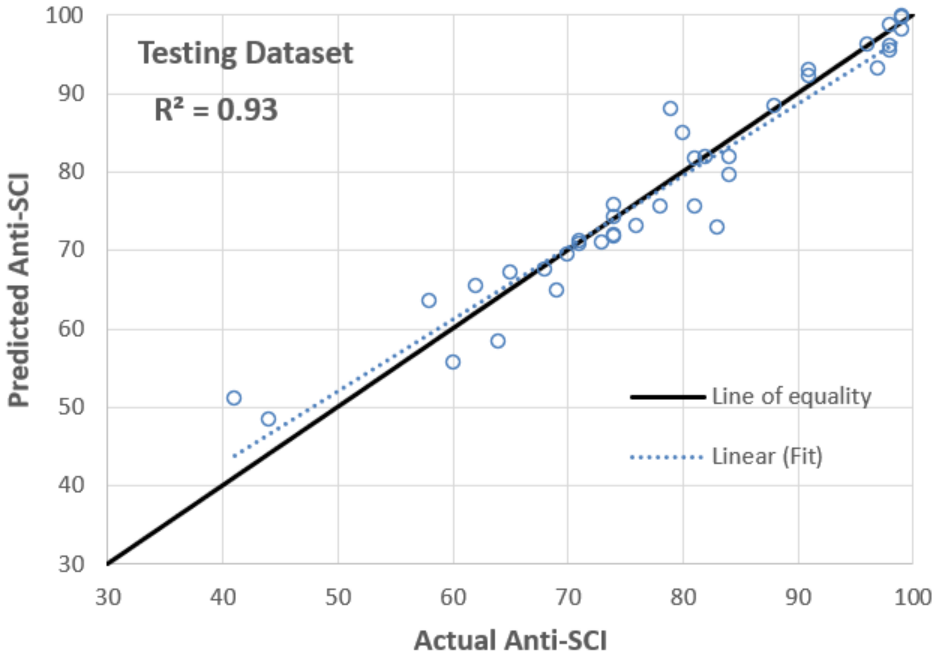


Figure 79. Comparison of Predicted and Actual Anti-SCI for Testing Subset for the Cluster Approach

Figure 80 compares the anti-SCI deterioration curves for each data cluster. Cluster 1 has the slowest deterioration rate. Cluster 1 represents generally mild climate conditions (e.g., lower average number of FDD compared to clusters 2 and 3, lower average number of days with a high temperature above 90ºF compared to cluster 4). Likewise, data in cluster 4 have the highest number of days with a high temperature above 90ºF and exhibit the fastest deterioration rate. However, one should be circumspect about drawing a general conclusion based on these observations, since factors other than climate conditions could have contributed to the deterioration.

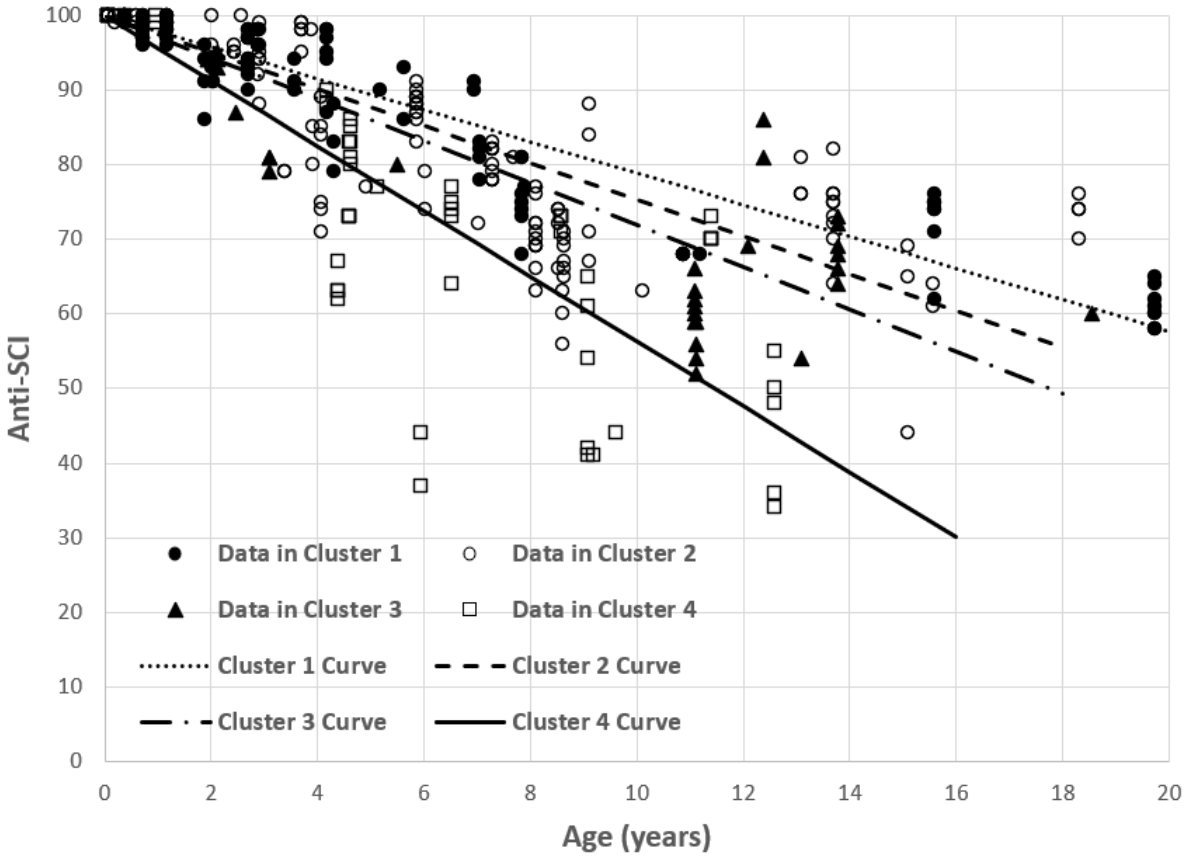


Figure 80. Anti-SCI Deterioration Rate at Different Climate Clusters

5.4 CONCLUSION OF ANTI-SCI PREDICTIVE MODELING

Three modeling approaches were implemented in this section for predicting anti-SCI. All approaches were based on a similar autoregressive RF algorithm, but they used different strategies for incorporating weather variables. In the first approach, a subset of informative climate variables with low collinearity were considered. In the second approach, 11 climate variables along with other inputs were transformed into PC. In the third approach, the performance data records were grouped into four clusters based on their climatic characteristics. Results indicated that all three models are capable of predicting anti-SCI with relatively high accuracy. The third approach showed lower errors on testing subsets compared to other approaches. The prediction errors were higher for anti-SCI values less than 50. The developed ML models showed promising generalization capability, as the model performance on testing subsets outperformed the

performance on training subsets. Comparisons between the models developed in this section and the autoregressive model in section 3.3.7 indicate that taking into account the weather variables improved the prediction performance. However, the improvement in performance was not significant. This suggests that pavement age and previous anti-SCI are the most significant predictors for anti-SCI.

6. RESEARCH FINDINGS AND RECOMMENDATIONS

6.1 SUMMARY AND FINDINGS

This study applies ML techniques for prediction of pavement performance to data from the FAA's PA40 database. A majority of previous efforts focused on the development of predictive models for pavement performance indicators such as IRI and PCI, or predictive models for particular distress types such as cracks, rutting, and faulting, by using data from the LTPP and the state highway performance data. Based on the literature, ANN, SVM, and RF were potentially feasible ML techniques for modeling pavement performance. Section 1 of this report provides a concise introduction to these ML methods.

To validate the application of each ML method in predicting pavement performance, simplified prediction models were developed for two pavement performance indicators (anti-SCI and PCI) by considering the pavement age as the primary predictor. The researchers implemented two modeling approaches (static and autoregressive) for each ML method. In the static approach, the pavement functional age was the only predictor. In the autoregressive approach, the previous condition measurement, and the time lapse between the two measurements, were considered as additional predictors. Variant architectures and schemes from each ML model were tested and compared to identify the best model for performance indicator prediction. Model performance was evaluated based on accuracy (minimizing the prediction error), generalization (similar performance on new data) and conformance to engineering principles. The anti-SCI and PCI prediction curves obtained from the ML models were compared to prediction curves from fourth-degree polynomial regression models. The following are key findings from the study:

- The R^2 values of the ML models were all above 0.7 and were comparable with the regression model, indicating their suitability for predicting pavement performance. However, the accuracy of the static models was not high, suggesting that pavement age alone cannot explain all the variation in PCI and anti-SCI.

- All models exhibited comparable performance on the independent testing subsets, meaning that they yielded good generalization performance.

- While the static RF model produced a lower prediction error compared to the static ANN and SVM, the predicted deterioration curve did not exhibit a monotonically descending trend. This suggests that quantitative evaluation (e.g., accuracy and generalization) alone is not adequate to determine the overall performance of the model and qualitative evaluation (e.g., conforming to engineering principles) must also be considered.

- Autoregressive models generally had smaller errors and higher R^2 than Static models and yielded similar generalization performance, making them a more suitable modeling

approach for the PA40 data. Autoregressive models enable predicting future performance in desired increments when performance is provided at a given age. Overall, the autoregressive RF model exhibited the best prediction performance.

To expand the performance models beyond pavement age, the climate/weather and traffic attributes were considered as predictors. A series of weather variables were identified and then various supervised and unsupervised learning techniques were applied to determine the key climate parameters that impact pavement performance. The problem with having large numbers of insignificant predictors is that prediction variance increases and makes it difficult to identify the influence of a given feature. Various collinearity tests were performed to explore the pairwise correlation between the climate variables. Feature selection methods were applied in an attempt to reduce the dimensionality of feature (input) space by identifying and removing a subset of irrelevant and redundant variables that can decrease the model's accuracy and quality. Two general FS strategies were considered: feature ranking and subset selection. In these evaluations, for each pavement section, two approaches were used: in the first approach, the average values of stored climate variables over the time period between two inspections were computed and compared to the rate of anti-SCI deterioration; the second approach compared anti-SCI to the cumulative values of stored climate variables from the time of rehabilitation to the time of inspection. Also, feature construction methods were implemented to reduce the dimensionality of the feature space by constructing new features from the original feature set. The FC methods included two unsupervised learning methods: cluster analysis (based on the *k*-means algorithm) and PCA. The key findings of the environmental data analysis are as follows:

- There are low- and moderate-strength pairwise correlations among most of the climate variables, but only a few of the correlations can be considered high. The pairwise correlation between the climate variables can change from region to region.

- Pavement age, previous anti-SCI, and cumulative Temp90 received higher rank, and cumulative FDD and FThC received the lowest rank among the variables in all ranking algorithms. Since the cumulative weather variables are dependent on pavement age, the high correlation with anti-SCI does not imply a direct cause of performance deterioration.

- The correlation coefficient between each cumulative weather variable and anti-SCI was less than the correlation between age alone and anti-SCI. This suggests that none of the weather variables alone is a better predictor than age for predicting anti-SCI.

- The rate of anti-SCI reduction is not significantly correlated to any of the weather variables, meaning that no individual weather variable can describe the changes in anti-SCI deterioration.

- The evaluation of the impact of climate on pavement performance would have been more meaningful if individual distress types were considered as opposed to an overall surface condition indicator such as anti-SCI.

- Low correlation between a weather variable and performance does not imply that the variable has no value. It is likely that when all the weather variables are taken together, they can describe some variation in anti-SCI.

- Cluster analysis indicated that the performance dataset can be grouped into three to four clusters with homogenous climatic characteristics based on two or three weather variables. The high-dimensional climate feature space can be reduced to only one feature that includes the cluster assignments for each performance record.

- The PCA indicated that three principal components can explain 95% of the total variance of input variables.

Eight traffic indices were identified to examine the influence of traffic on two pavement performance indicators: PCI and SCI. Pairwise correlation analysis and feature ranking analysis were performed to identify the most relevant traffic indices for predicting PCI and SCI. A series of *k*-means cluster analysis were also performed to group the runways with similar traffic conditions into the same clusters. The key findings of the traffic data analysis are as follows:

- PCI and SCI have poor correlations with accumulated traffic indices. Differences between design traffic and actual traffic may account for this poor correlation. Traffic alone cannot explain all the variations in pavement performance. Runways with higher numbers of heavier aircraft did not show a higher rate of pavement deterioration. Conversely, runways with lower traffic levels exhibited load-related distresses such as alligator cracks.

- The low correlations could be due to inherent limitations in traffic data or calculation of cumulative traffic.

- Age is a much stronger predictor than any of the evaluated traffic indices.

- The low correlations between traffic indices and PCI/SCI do not imply that there is no value in incorporating traffic for performance prediction. Pavement design principles can be used to select appropriate traffic index.

- Three options were recommended for incorporating traffic as input for developing PCI and SCI predictive models: 1) use two traffic indices, namely DepAC2 and DepAC3; 2) assign cluster number to the data records; and 3) perform PCA by incorporating sets of traffic index, and other variables such as age, climate, and pavement structure.

Findings from the environmental data analysis were used to develop predictive models for anti-SCI using an autoregressive approach and RF as the base learner. Three different approaches for considering the input variables were implemented. In the first approach, a subset of informative weather variables with low collinearity were considered. In the second approach, 11 weather variables along with other inputs were transformed into principal components, and in the third approach, the performance data records were grouped into four clusters based on their climatic characteristics. The key findings are as follows:

- All the models were capable of predicting anti-SCI with relatively high accuracy. The third approach showed lower errors on testing subsets as compared to other approaches.

- Despite the acceptable average performance of the models, the models were not robust across the input domain. The models yielded larger inaccuracies for anti-SCI values less

than 50. One reason for this higher error is the scarcity of performance data for airport pavements with anti-SCI of less than 50.

- While taking into account that the weather variables improved the prediction performance, the improvement in performance was not significant. Pavement age and previous anti-SCI were still the most significant predictors for anti-SCI.

6.2 RECOMMENDATIONS FOR FUTURE RESEARCH

Based on investigations performed under this report, the following additional research is recommended:

1. Expand the traffic data analysis by distributing the traffic to the runway section. The pairwise correlation analysis, feature ranking analysis, cluster analysis, and PCA can be reiterated.

2. Use a similar ML process to develop predictive models for PCI suitable for immediate use by industry. The impact of traffic can be incorporated according to the recommended methods in this report.

3. Complete the preliminary analysis by investigating climatic influence on individual pavement distress types, such as block cracking, longitudinal/transfers cracking, and alligator cracking, using ML methods.

4. Expand the developed anti-SCI prediction models by considering additional features, such as traffic, maintenance activity, pavement material properties, subgrade type and drainage conditions, in an attempt to improve the prediction accuracy.

5. Develop ML prediction models for the components of SL index: FOD, roughness indices, and friction indices. This family of models will output an estimated value for each performance index.

6. Develop an ML solution to combine individual performance indexes and enhance the SL index. The developed SL model will allow practitioners to determine whether the runway pavement is serviceable given certain performance indices. The model also enables the projection of runway SLs that can be used to strategize for maintenance.

7. REFERENCES

Alharbi, F. (2018). *Predicting pavement performance utilizing artificial neural network (ANN) models* [Doctoral dissertation, Iowa State University] Iowa State University Digital Repository Graduate Theses and Dissertations. https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=7710&context=etd

Amin, S. R., & Amador-Jiménez, L. E. (2017). Backpropagation Neural Network to estimate pavement performance: dealing with measurement errors. *Road Materials and Pavement Design*, *18*(5), 1218–1238. https://doi.org/10.1080/14680629.2016.1202129

Arthur, D., A., & Vassilvitskii, S. (2007, January 7-9). K-means++: The advantages of careful seeding. In H. Gabow (Chair), *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* [Symposium]. SODA 2007,New Orleans, LA, United States. https://dl.acm.org/doi/10.5555/1283383.1283494

Ashtiani, A.Z., Shirazi, H., Murrell, S., Speir, R. & Brill, D.R. (2019, July 21-24). *Performance Model Development for Extended Airport Pavement Life.* International Airfield and Highway Pavements Conference, Chicago, IL, USA. https://doi.org/10.1061/9780784482476.025

ASTM D5340-20 (2020), *Standard Test Method for Airport Pavement Condition Index Surveys*, ASTM International, West Conshohocken, PA. www.astm.org

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: identifying influential data and sources of collinearity*. Wiley. https://doi.org/10.1002/0471725153

Ben-Hur, A., & Weston, J. (2010). A user's guide to support vector machines. In O. Carugo & F. Eisenhaber (Eds.), *Data mining techniques for the life sciences. Methods in molecular biology (Methods and Protocols)* (Vol. 609, pp. 223-239). Humana Press. https://doi.org/10.1007/978-1-60327-241-4_13

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. *Knowledge and Information Systems An International Journal, 34*, 483-519. https://doi.org/10.1007/s10115-012-0487-8

Bontempi, G., Ben Taieb, S., & Le Borgne, Y. A. (2013). Machine learning strategies for time series forecasting. In M. A. Aufaure & E. Zimányi (Eds.), *Business Intelligence. eBISS 2012. Lecture notes in business information processing* (Vol. 138, pp. 62-77). Springer. https://doi.org/10.1007/978-3-642-36318-4_3

Booth, G. D., Niccolucci, M. J., Schuster, E. G., & Intermountain Research Station (Ogden, Utah). (1994). *Identifying proxy sets in multiple linear regression: an aid to better coefficient interpretation*. U.S. Department of Agriculture, Forest Service, Intermountain Research Station.

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis, forecasting and control*. Holden-Day.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cristianini, N., & Schölkopf, B. (2002). Support vector machines and kernel methods: The new generation of learning machines. *AI Magazine*, *23* (3), 31-41. https://doi.org/10.1609/aimag.v23i3.1655

Dietterich, T. G. (2000). Ensemble methods in machine learning. In: *Multiplier Classifier Systems. MCS 2000. Lecture Notes in Computer Science* (Vol. 1857, pp. 1-15). Springer. https://doi.org/10.1007/3-540-45014-9_1

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., & Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography. 36*(1), 27-46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Frank, E., Hall, M. A., & Witten, I. H. (2016). The WEKA workbench—online appendix for *Data Mining: Practical Machine Learning Tools and Techniques*, Fourth Edition. Morgan Kaufmann. https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf

Gower, J. C. & Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification, 3*, 5-48. https://doi.org/10.1007/BF01896809

Guyon, I. & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157-1182. https://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis:with readings* (4th ed.). Prentice-Hall.

Hall, M. A. (1999). *Correlation-based feature selection for machine learning* [Doctoral dissertation]. University of Waikato. The University of Waikato Department of Computer Science Archive. https://www.cs.waikato.ac.nz/~mhall/thesis.pdf

Harvey A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press. https://doi.org/10.1017/CBO9781107049994

Jolliffe, I. T. (2002). *Principal component analysis, second edition*. Springer-Verlag.

Kargah-Ostadi, N. (2013). *Enhancing analytical toolboxes of pavement management systems via integration of computational intelligence* [Doctoral dissertation, Pennsylvania State University]. Penn State Electronic Theses and Dissertations for graduate school. https://etda.libraries.psu.edu/files/final_submissions/8919

Karush, W. (1939). *Minima of functions of several variables with inequalities as side conditions* [Master's thesis, University of Chicago]. http://pi.lib.uchicago.edu/1001/cat/bib/4111654

Keogh, E., Chakrabarti, K., Pazzani, M., & Mehrotra, S., (2001). Dimensionality reduction for fast similarity search in large time series databases. *Journal of Knowledge and Information Systems*, *3*, 263-286. https://doi.org/10.1007/PL00011669

Kira, K., & Rendell, L. A. (1992). The feature selection problem: traditional methods and a new algorithm. In P. Rosenbloom & P. Szolovits (Chairs), *Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI'92)* (pp. 129-134). Association for the Advancement of Artificial Intelligence (AAAI). https://www.aaai.org/Papers/AAAI/1992/AAAI92-020.pdf

Kotsiantis, S. B., Kanellopoulos, D., & Pintelas P. E. (2006). Data preprocessing for supervised leaning. *International Journal of Computer Science*, *1*(1), 111-117. https://www.academia.edu/12362275/Data_preprocessing_for_supervised_learning

Kumar, A., McCann, R., Naughton, J., & Patel, J. M. (2016). Model selection management systems: the next frontier of advanced analytics. *ACM SIGMOD Record*, *44*(4), 17-22. https://doi.org/10.1145/2935694.2935698

Li, N., Xie, W. C., & Haas, R. (1996). Reliability-based processing of Markov chains for modeling pavement network deterioration. *Transportation Research Record: Journal of the Transportation Board*, *1524*(1), 203-213. https://doi.org/10.1177/0361198196152400124

MacKay, D. J. C. (1992). Bayesian interpolation. *Neural Computation, 4*(3), 415–447. https://doi.org/10.1162/neco.1992.4.3.415

Marcelino, P., de Lurdes Antunes, M., Fortunato, E., & Gomes, M. C. (2019). Machine learning approach for pavement performance prediction. *International Journal of Pavement Engineering*, *22*(3), 341-354. https://doi.org/10.1080/10298436.2019.1609673

Milligan, G. W. (1996). Clustering validation: results and implications for applied analyses. In P. Arable, L. J. Hubert, & G. De Soete (Eds.), *Clustering and Classification* (pp. 341-375). World Scientific. https://doi.org/10.1142/9789812832153_0010

Ozbay, K., & Laub, R. (2001). *Models for pavement deterioration using LTPP* (Report number FHWA-NJ-1999-030). Federal Highway Administration, U.S. Department of Transportation, Washington DC. http://cait.rutgers.edu/files/FHWA-NJ-1999-030.pdf

Robnik-Šikonja, M., & Kononenko, I. (2003). Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning*, *53*, 23-69. https://doi.org/10.1023/A:1025667309714

Samarasinghe, S. (2007). *Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition*. Auerbach Publications.

Shekharan, A. R. (2000, August 11-12). Pavement performance prediction by artificial neural network. In N. O. Attoh-Okine (Ed.), *Computational Intelligence Applications in Pavement and Geomechanical Systems. Proceedings of the 2nd International Workshop on Artificial Intelligence and Mathematical Methods in Pavement and Geomechanical Systems* (pp. 89-98). A.A. Balkema.

Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P. (2003). Random forest: a classification and regression tool for compound classification and QSAR modeling. *Journal of chemical information and computer sciences*, *43*(6), 1947-1958. https://doi.org/10.1021/ci034160g

Tabatabaee, N., Ziyadi, M., & Shafahi, Y., (2013). Two-stage support vector classifier and recurrent neural network predictor for pavement performance modeling. *Journal of Infrastructure Systems*, *19*(3), 266-274. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000132

Tan, P. N., Steinbach, M., & Kumar, V. (2013). *Introduction to data mining* (1st ed.). Pearson.

Tsagris, M., Lagani, V., & Tsamardinos, I. (2018). Feature selection for high-dimensional temporal data. *BMC Bioinformatics*, *19*, Article 17. https://doi.org/10.1186/s12859-018-2023-7

Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J.H. (2018). Relief-based feature selection: Introduction and review. *Journal of Biomedical Informatics, 85*, 189-203. https://doi.org/10.1016/j.jbi.2018.07.014

Yang, J., Lu, J. J., & Gunaratne, M. (2003). *Application of neural network models for forecasting of pavement crack index and pavement condition rating* (Final Report for contract number BC353-13). Florida Department of Transportation. http://ftp.fdot.gov/file/d/FTP/FDOT%20LTS/CO/research/Completed_Proj/Summary_SMO/FDOT_BC353_13rpt.pdf

Yue, E., & Bulut, R. (2014, February 23-26). *Evaluation of the climatic factors for the classification of Oklahoma pavement* [Technical paper]. Geo-Congress 2014, Atlanta, GA, United States. https://doi.org/10.1061/9780784413272.392

Ziari H., Maghrebi, M., Ayoubinejad, J., & Waller, S. T. (2016). Prediction of pavement performance: Application of support vector regression with different kernels *Transportation Research Record: Journal of the Transportation Research Board, 2589*(1)*,* 135–145. https://doi.org/10.3141/2589-15

Zolhavarieh, S., Aghabozorgi, S., & Teh, Y. W. (2014). A review of subsequence time series clustering. *The Scientific World Journal*, *2014*, Article ID 312521. https://doi.org/10.1155/2014/312521

APPENDIX A—DEVELOPMENT OF HISTORICAL CLIMATE AND WEATHER DATA LINKS BETWEEN PA40 AND EXISTING FEDERAL AVIATION ADMINISTRATION DATABASES

A.1 INTRODUCTION

This appendix presents the content of an unpublished technical note authored by Russell Gorman, Scott Murrell, and Timothy Parsons of Applied Research Associates, Inc. The technical note describes the research effort that determined how weather events were summarized and implemented in the software tool, PA40.

A.1.1 BACKGROUND

The Federal Aviation Administration (FAA) Office of Airport Safety and Standards (AAS) has requested that methodologies be developed to extend the expected life of large hub runway construction from the current standard of 20 to 40 years. The development of life extension methodologies is being executed using a three-pronged approach consisting of: (a) comprehensive data collection for a selection of runways at large- and medium-hub U.S. airports, including construction, performance, material property, traffic and environmental data; (b) the development of advanced performance models based on the collected data relating various performance measures to traffic, structural, and environmental inputs; and (c) the development of a new definition of airport pavement life that is consistent with the assumptions of life cycle cost analysis (LCCA), e.g., that the end of pavement life occurs when the annual cost of maintaining continued service at a safe operational level exceeds the annualized cost of replacement.

Data collection in support of extended airport pavement life began in 2012. To date, data have been collected on 28 runways at 22 U.S. airports. Field visits were made to a subset of these airports and additional field data collected on 13 of these runways. To facilitate analysis of the collected data for the development of advanced performance models the FAA has developed a project specific software tool—PA40.

PA40 is a standalone implementation of the FAA PAVEAIR pavement management system. The PA40 data warehouse contains tables and functions not currently implemented in the public version of FAA PAVEAIR. PA40 is not accessible through the FAA PAVEAIR website. To facilitate the future study of the influence of climate/weather on the long-term performance of airfield pavements, the FAA directed that new remote database access and calculation functionality be added to PA40.

A.1.2 SCOPE

This document details the weather and climate data search functions in PA40, including the justifications for and planned use of the particular data elements selected and the source data for each element.

The FAA databases contain National Oceanic and Atmospheric Administration (NOAA) maintained Automated Weather Observing System (AWOS) and Automated Surface Observing Systems (ASOS) information. The scope of this research includes working with to these data sets

from within PA40. Specific required weather data include: temperature, wind speed, precipitation, relative humidity (or dew point), solar radiation or sky cover. PA40 was also modified to compute derived weather fields from the basic data. The weather data were considered network-level (airport-level) data in PA40, and made available as hourly readings for each airport in the database. Climate is distinct from weather and refers to long-term environmental trends expressed by statistics, such as average annual precipitation or average annual high/low temperatures. The scope also includes creating new functions as needed within the PA40 graphical user interface to query, return, and display the weather and climate data for all airports in the study.

A.2 APPROACH

First, climate/weather data elements to link to PA40 were assessed to determine which derived climate/weather fields should be computed and displayed. Then, the available climate/weather information in the FAA databases was determined to communicate what was needed and the format and to assist the FAA in the development of the web service. The FAA climate/weather databases contain NOAA reported data measured by AWOS and ASOS at the Extended Airport Pavement Life Study (EAPL) airports. The solar radiation data originates from the National Solar Radiation Database maintained by the National Renewable Energy Laboratory (NREL) of the Department of Energy (DOE). While the web service was being implemented by others, the web page was developed and server-side processing calculations were coded. The relationship of the three main components of the Climate/Weather Link are shown in figure A-1.

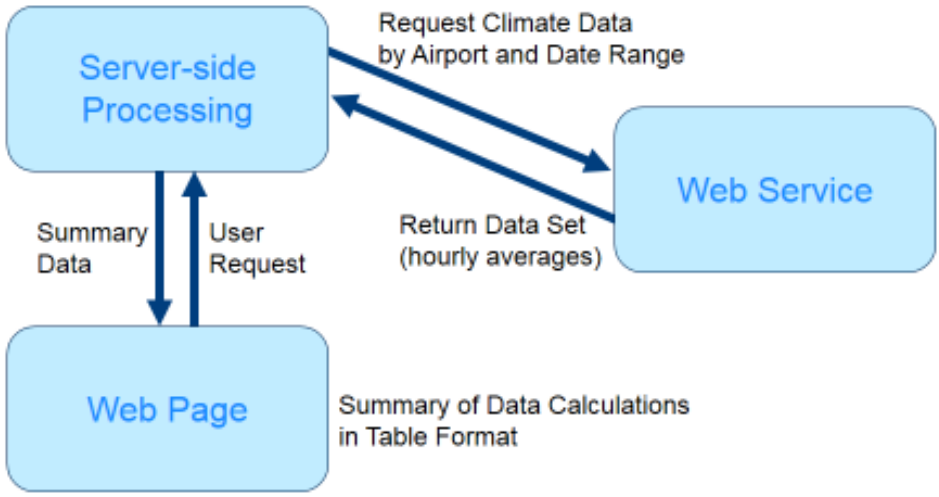


Figure A-1. Climate Weather Link Components and Their Relationships

The user can select a particular site and date/time range on the web page user interface. Then, the data for that date/time range and site are retrieved from the web service. Those data are then received by the server-side processing component and the report calculations are processed. After all processing is complete, the results are fed back to the web page and displayed to the user in tabular format.

The following sections describe the calculations performed by the server-side software component on the raw data that result in the report that is presented to the user.

A.2.1 HIGH TEMPERATURE (DEGREES FAHRENHEIT)

Each hourly average temperature received by the web service is reviewed. The highest hourly average temperature in the data is retrieved.

A.2.2 LOW TEMPERATURE (DEGREES FAHRENHEIT)

Each hourly average temperature received by the web service is reviewed. The lowest hourly average temperature in the data is retrieved.

A.2.3 AVERAGE TEMPERATURE (DEGREES FAHRENHEIT)

Each hourly average temperature received by the web service is stored. The totaled result is used to calculate the average hourly average temperature in the entire data set (i.e., every hourly average temperature contained in the user selected date/time range).

A.2.4 AVERAGE DAILY TEMPERATURE RANGE (DEGREES FAHRENHEIT)

The temperature range is calculated daily by calculating the difference between the high and low values found each day. The average of all daily temperature ranges is then calculated and reported to the user in degrees Fahrenheit.

A.2.5 NUMBER OF FREEZE THAW CYCLES (NUMBER OF CYCLES)

Each hourly average temperature received by the web service is reviewed. The number of times the temperature goes from a value of less than 32 degrees Fahrenheit to a value greater than 32 degrees Fahrenheit is recorded.

A.2.6 FREEZE DEGREE DAYS (DEGREES FAHRENHEIT)

Each hourly average temperature received by the web service is reviewed and the average daily temperature is calculated for each day that falls within the date/time range selected by the user. Any day where the average daily temperature is less than 32 degrees Fahrenheit is considered a freeze degree day.

For each freeze degree day that is found, an overall running total of the number of degrees the daily average temperature of each freeze degree day falls below 32 degrees Fahrenheit is maintained. The value of the running total of these temperature deltas is the freeze degree days value that is presented to the user.

A.2.7 DAYS TEMPERATURE OVER 90 DEGREES FAHRENHEIT (DAYS)

For the contiguous United States, temperatures above 90ºF are typically considered hot, and NOAA and other agencies keep track of the occurrences across the U.S. The New York City Panel on Climate Change defined an extreme temperature event as a day when the temperature reached 90ºF or hotter.[1] Higher than normal temperatures are known to have a negative impact on pavement performance, which warrants further study under the EAPL. Although 90ºF does not have the same inherent physical significance as 32ºF (the freezing point of water), it nevertheless serves as a convenient high-temperature threshold.

This calculation is similar to the Freeze-Degree-Day calculation, except degree-days above 90 degrees Fahrenheit are summed instead of degree-days below freezing. Each hourly average temperature received by the web service is reviewed. The number of days in which the temperature exceeded 90 degrees Fahrenheit is recorded.

A.2.8 DAYS PRECIPITATION (DAYS)

Each hourly average precipitation amount received by the web service is reviewed. The number of days in which the total daily precipitation exceeded 0 inches is recorded.

A.2.9 DURATION PRECIPITATION (HOURS)

Each hourly average precipitation amount received by the web service is reviewed. Using this data, each precipitation event will be calculated (in hours) and recorded. A precipitation event starts when precipitation is first detected (i.e., precipitation value is greater than 0 inches) in a given hour of data. The precipitation event continues as long as precipitation is also detected in the next contiguous hour or hours. The precipitation event stops upon the first contiguous hour in which there is no precipitation recorded (i.e., precipitation value equals 0 inches). A single precipitation event may span across one or more days as long as the consecutive hours all have experienced precipitation.

The number of precipitation events is recorded, and the average precipitation event duration in hours is calculated.

A.2.10 PRECIPITATION TOTAL (INCHES)

Each hourly average precipitation amount received by the web service is summed. The total number of inches of precipitation is recorded.

A.2.11 FREEZE PRECIPITATION DAYS (DAYS)

Each hourly average precipitation amount and each hourly average temperature received by the web service is reviewed. If at least one average hourly temperature in a particular day is below 32 degrees Fahrenheit and at least one average hourly precipitation value in the same day is greater

[1] Climate Risk Information 2013, Observations, Climate Change Projections and, Maps; New York City Panel on Climate Change, June 2013.

than 0 inches, then this is considered a freeze-precipitation day. Note that the "freeze hour" and the "precipitation hours" are not necessarily the same hour. That is, one hour can contain precipitation, and a different hour can have an average freezing temperature. However, it is also acceptable if they are the same hour. The total count of days meeting these criteria is reported.

A.2.12 WIND SPEED AVERAGE (MILES PER HOUR)

Each hourly average wind speed value received by the web service is summed. The totaled result is used to calculate the average hourly wind speed in the entire data set (i.e., every hourly average wind speed contained in the user selected date/time range).

A.2.13 SKY COVER (OKTAS)

Each hourly average sky cover value received by the web service is summed. The totaled result is used to calculate the average hourly sky cover in the entire data set (i.e., every hourly average sky cover contained in the user selected date/time range).

Table A-1 shows the list of possible integer values for sky coverage layer code provided by the web service and each corresponding text description. Values and meanings are defined in the document written by NOAA/National Centers for Environmental Information, titled "Federal Climate Complex Data Documentation for Integrated Source Data," dated January 12, 2018.

Table A-1. Sky Coverage Layer Codes

Sky Coverage Code	Description	Amount of Sky Covered (oktas)
0	CLR (clear)—no obscuration	0/0
1-2	FEW	$0<X \leq 2$
3-4	SCT (scattered)	$2< X \leq 4$
4-7	BKN (broken)	$4< X \leq 8$
8	OVC (overcast)	8
9	Obscured	8
10	Partial obscuration	NA

A.2.14 THORNTHWAITE MOISTURE INDEX (%)

The Thornthwaite Moisture Index (Thornthwaite, 1948) (TMI) is a measure of moisture available in the soil. It can be generally described as reflecting the aridity or humidity of the soil and climate, calculated from the collective effects of precipitation, evapotranspiration, soil water storage, moisture deficit and run off. Although it was developed as an agricultural tool, it has also been used for engineering applications in which soil moisture is a consideration, including foundation design and pavement performance (Philip & Taylor, 2012). In TMI calculations, precipitation and evapotranspiration are equally important climatic factors. Thornthwaite described the climatic cycle by balancing the rainfall, potential evapotranspiration, and soil water holding capacity. Computations of the Thornthwaite Moisture Index (TMI) are based upon monthly rainfall, mean temperature data, an estimate of the water holding capacity of the soil and the site location. Moist

climates have positive values of TI and dry climates have negative values. Table A-2 shows the climate types corresponding to TMI ranges.

Table A-2. Climate Types Corresponding to TMI Ranges

Thornthwaite Climate Type		Thornthwaite Moisture Index (%)
A	Perhumid	>100
B4	Humid	80 to 100
B3	Humid	60 to 80
B2	Humid	40 to 60
B1	Humid	20 to 40
C2	Moist subhumid	0 to 20
C1	Dry subhumid	-20 to 0
D	Semi-arid	-40 to -20
E	Arid	-60 to -40

The publication titled "Review of Calculation Procedures of Thornthwaite Moisture Index and its Impact on Footing Design" (Karunarathne et al., 2016) was used as a guide for the approach on how to calculate the TMI. Method 2 documented in this paper was used as the basis of the calculations and is described below.

There are 3 major steps in this calculation:

Step 1:
Cycle through each day's worth of data and calculate the following:

- total precipitation per month (P)
- the average temperature per month (t)
- the monthly heat index value (i)

$$i = (0.2 \times t)^{1.514}$$

Step 2:
Calculate the following:

- Determine the Total Heat Index (I) by calculating the summation of all monthly heat index values that were calculated in Step 1
- Determine the value of (a), which will be used in upcoming calculations in Step 3

$$a = 6.75 \times 10^{-7} \times I^3 - 7.771 \times 10^{-5} \times I^2 + 0.01792 \times I + 0.49239$$

Step 3:
Cycle through the monthly calculations from Step 1 and use the values calculated in Step 2 to calculate the following below:

- Calculate potential evaporation (PE) for each month (assume 12-hour daylight each day)

$$PE = 1.6 \times \left(\frac{10 \times t}{I}\right)^{a}$$

- Calculate soil moisture storage (S)

Note that P is the total monthly precipitation.

Assume that the initial soil moisture value and the maximum soil moisture store are both 10cm.

S = MIN (MAX(Si-1 + (P – PE), 0), Smax)

- Runoff ®

R = MAX((Si-1 + (P – PE) – Smax), 0)

- Moisture Deficit (D)

D = -MIN (Si-1 + (P – PE), 0)

Then after completing the cycle through each month of data calculate the following:

- Total Runoff is the summation of all monthly runoff values.
- Total Deficit is the summation of all monthly moisture deficit values.
- Total potential evaporation is the summation of all monthly potential evaporation values.
- Calculate Ia:

$$Ia = 100 \times \frac{D}{PE}$$

- Calculate Ih:

$$I_h = 100 \times \frac{R}{PE}$$

- Calculate TMI:

$$TMI = I_h - 0.6 \times I_a$$

A minimum of 12 months of data is required to calculate TMI. TMI is calculated for each year in the data set and then the annual TMI values are averaged to determine the value reported. When a query is for a number of months not simply divisible by 12, the remaining fraction of a year is disregarded.

A.2.15 RELATIVE HUMIDITY HIGH (%)

The following calculation determines the relative humidity. This calculation is used when processing the daily data to determine the overall high and low humidity values and the average humidity.

$$Relative\ Humidity = 100 \times \frac{e^{\left(\frac{17.625*\text{dewpoint}}{243.04+\text{dewpoint}}\right)}}{e^{\left(\frac{17.625*\text{temperature}}{243.04+\text{temperature}}\right)}}$$

Each hourly average dewpoint and temperature value received by the web service is reviewed. The humidity is calculated for each hour, and the highest value found across the entire data set is reported.

A.2.16 RELATIVE HUMIDITY LOW (%)

Each hourly average dewpoint and temperature value received by the web service is reviewed. The humidity is calculated for each hour, and the lowest value found across the entire data set is reported.

A.2.17 RELATIVE HUMIDITY AVERAGE (%)

Each hourly average dewpoint and temperature value received by the web service is reviewed. The humidity is calculated for each hour, and the average humidity across the entire data set is calculated and reported.

A.2.18 HYDRATION DAYS (NUMBER OF DAYS)

Each hourly average dewpoint and temperature value received by the web service is reviewed. The humidity is calculated for each hour, and then a count of the number of days where at least one hour's relative humidity reading exceeds 80% is determined. The 80% value was selected based on the relative humidity in the pores required for the Portland Cement Hydration reaction to occur, as reported in Mehta and Monteiro (2006).

A.2.19 SOLAR RADIATION (WATTS PER METER SQUARED)

Each hourly solar radiation value received by the web service is reviewed. The total radiation for every hour is stored and reported to the user.

A.3. RETRIEVING AND DISPLAYING CLIMATE/WEATHER DATA USING PA40

The software flow is as follows: first, a user selects a particular site and date/time range on the web page user interface and then selects the "Climate Report" button, as shown in figure A-2. This action results in a request sent by the server-side processing component to the web service. The request to the web service contains the site name and the start and end date range. Note that the "Filter" button and field are not functional in the current version of the software, but is a placeholder for future use.

The following data are returned by the web service for each hour in the requested date/time range:

- The respective hour
- The average temperature for that hour in degrees Fahrenheit

- The average wind speed in miles per hour
- The dew point average in degrees Fahrenheit
- The total precipitation in inches and hundredths of an inch
- The solar radiation average in watts per meter squared
- Average sky cover in oktas

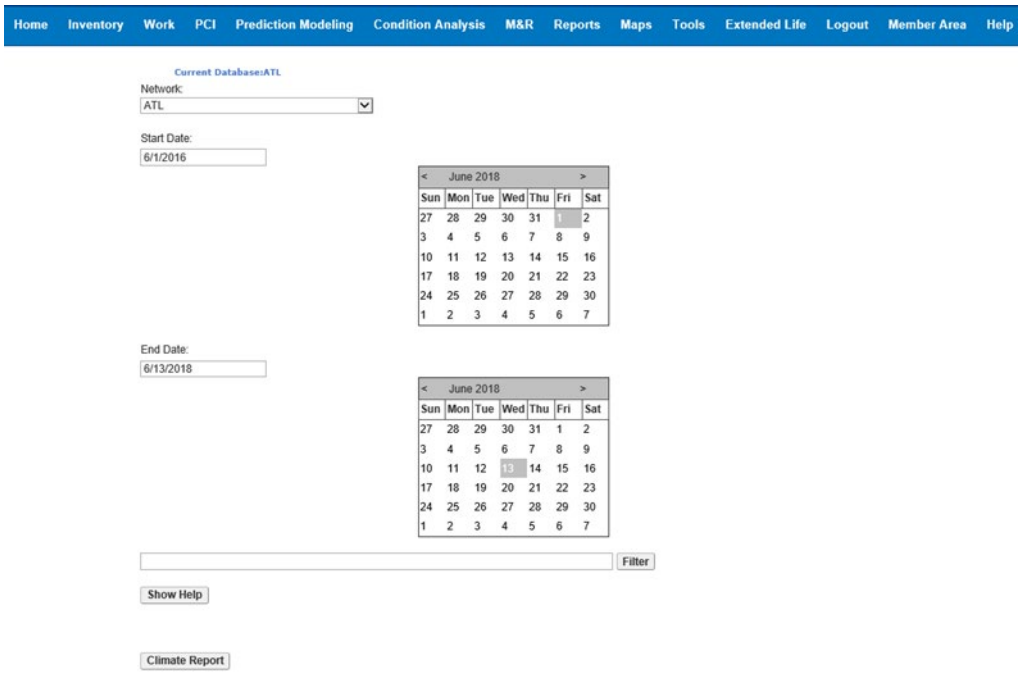


Figure A-2. Web Page User Interface Example

After the server-side component receives this data back from the web service, it then processes all the data and returns it back to the web page, where it is displayed to the user, as illustrated in figure A-3.

Network Name	ATL
Start Date	Friday, January 1, 1999
End Date	Thursday, January 1, 2004
Completeness	36,286 of 61,343 hours contained data (i.e., completeness percentage - 59.15%)
Temp High	97.00 F
Temp Low	8.00 F
Temp Average	61.68 F
Average Daily Temperature Range	17.00 F
Freeze Thaw	109 freeze-thaw cycles
Freeze Degree Days	85.35 F
Days Temperature over 90	47 days
Days Precipitation	462 days
Average Duration of Precipitation Event	3.22 hours
Precipitation Total	274.41 inches
Thornthwaite	65 %
RHumidity High	85.40 %
RHumidity Low	4.98 %
RHumidity Avg	62.78 %
Hydration Days	1025 days
Sky Cover	6.19 oktas
Solar Radiation	12,657,832.00 watts/m^2
Wind Average	8.79 mph
Freeze Precipitation Days	12 days

Figure A-3. Results Table Displayed on Web Page

In addition to the climate/weather data processing results, missing data are reported as expected hours versus actual hours for which data were returned, and the percentage of completeness.

A.4 PLANNED USE OF CLIMATE AND WEATHER DATA

An approach to determine climate and/or weather's influence on pavement performance is described in the following steps:

- A review of the available distress data stored in PA40 will be performed with the goal of identifying airports with relatively complete data sets covering an extended period. A number of sections will exhibit distresses, and data will be available from trafficked (keel) and lightly trafficked (outside keel) sections. Sections with multiple inspections will be reviewed to determine if the distresses exhibit progressive behavior, that is, a distress that increases in quantity and/or severity over time.

- For the airports with complete distress data sets, climate and weather data will be retrieved from the FAA web service using PA40. Data will be retrieved and calculations performed for the periods coinciding with the available distress data.

- This step includes performing various analyses (analysis of variance [ANOVA], main effect, liner regression, t-Tests or others) of climate and weather against the distress data, to look for correlation between the variables. Some previously reported relationships that can be further investigated include:

  PCC PAVEMENTS
  - Average Temperature/Humidity/Wind Speed/Solar Radiation and Shrinkage Cracking/Map Cracking.
  - Freeze precipitation days and Durability Cracking.
  - Precipitation/Humidly and Alkali Silica Reactivity (ASR).
  - Precipitation and Faulting/Corner Breaks.
  - Temperature and Joint Spall/Corner Spall
  - Freeze Thaw Cycles and Joint Spall/Corner Spall
  - Days above 90^0F and Blowups.
  - Average Temperature and Longitudinal, Transverse, and Diagonal (LTD) Cracking

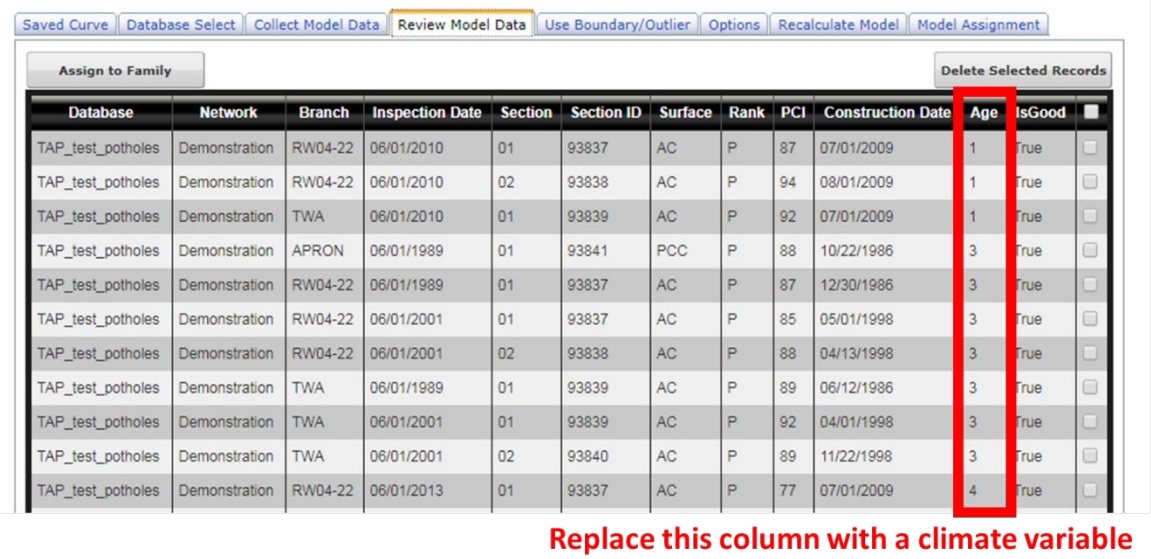
  ASPHALT PAVEMENTS
  - Average Temperature and Rutting
  - Average Temperature and Shoving
  - Average Temperature and Longitudinal &Transverse (L&T) Cracking
  - Freeze Thaw Cycles and L&T Cracking
  - Freeze Thaw Cycles and Block Cracking
  - Freeze Thaw Cycles and Swell
  - Solar Radiation and Weathering/Raveling
  - Precipitation and Depressions
  - Precipitation and L&T Cracking

- In additional to a detailed analysis of individual distresses an assessment of Climate/Weather Data vs anti-structural condition index (Anti-SCI) (a non-load-related component of pavement condition index [PCI]), foreign object debris (FOD) Index, Roughness and PCI can also be made to determine if a relationship exists.

A.5 SUMMARY AND DISCUSSION

Previous research indicates that the presence and severity of some distresses defined in ASTM D5340 correlate to the climate zone in which a pavement is located (dry-freeze, dry-non-freeze, wet-freeze, wet non-freeze). PA40 has been modified to further explore this relationship by providing the capability to associate specific weather events experienced by a pavement with the distresses and condition of that pavement at the time of inspection. The system focuses on the historical moisture and temperature conditions of a pavement as most climate-related distresses are believed to be affected by these two factors. Solar radiation is also included in the system in order to determine if a correlation can be established between weathering/raveling and solar radiation input.

The expected use of the system is to produce data sets similar to those used for generating family deterioration curves in a standard pavement management system, but allowing the use of data other than age as the independent variable. Once the existence of a correlation between a condition and a climate variable is established, future work can add the capability to generate these model types in PA40, as shown in figures A-4 and A-5.



Database	Network	Branch	Inspection Date	Section	Section ID	Surface	Rank	PCI	Construction Date	Age	IsGood
TAP_test_potholes	Demonstration	RW04-22	06/01/2010	01	93837	AC	P	87	07/01/2009	1	True
TAP_test_potholes	Demonstration	RW04-22	06/01/2010	02	93838	AC	P	94	08/01/2009	1	True
TAP_test_potholes	Demonstration	TWA	06/01/2010	01	93839	AC	P	92	07/01/2009	1	True
TAP_test_potholes	Demonstration	APRON	06/01/1989	01	93841	PCC	P	88	10/22/1986	3	True
TAP_test_potholes	Demonstration	RW04-22	06/01/1989	01	93837	AC	P	87	12/30/1986	3	True
TAP_test_potholes	Demonstration	RW04-22	06/01/2001	01	93837	AC	P	85	05/01/1998	3	True
TAP_test_potholes	Demonstration	RW04-22	06/01/2001	02	93838	AC	P	88	04/13/1998	3	True
TAP_test_potholes	Demonstration	TWA	06/01/1989	01	93839	AC	P	89	06/12/1986	3	True
TAP_test_potholes	Demonstration	TWA	06/01/2001	01	93839	AC	P	92	04/01/1998	3	True
TAP_test_potholes	Demonstration	TWA	06/01/2001	02	93840	AC	P	89	11/22/1998	3	True
TAP_test_potholes	Demonstration	RW04-22	06/01/2013	01	93837	AC	P	77	07/01/2009	4	True

Figure A-4. Recommended Method to Integrate a Climate Variable Into the PA40 Modeling System

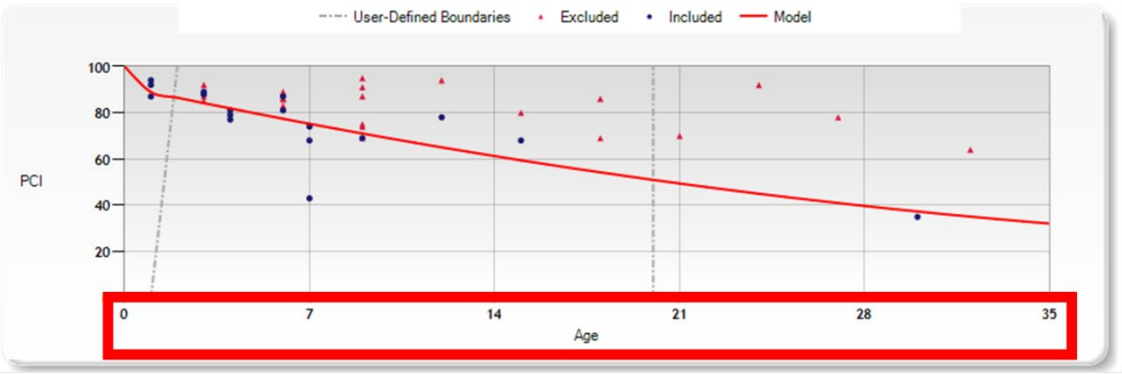


Figure A-5. Recommended Climate Modeling Capability for PA40 After the Existence of a Correlation is Established

A.6 REFERENCES

Karunarathne, A. Gad, E. Disfani, M. M., Sivanerupan, S. & Wilson, J. L. (2016). Review of calculation procedures of Thornthwaite moisture index and its impact on footing design, *Australian Geomechanics*, *51*, 85-95. https://australiangeomechanics.org/papers/review-of-calculation-procedures-of-thornthwaite-moisture-index-and-its-impact-on-footing-design/

Mehta, P. K., & Monteiro, P. J. M. (2006). *Concrete: microstructure, properties, and materials* (3rd ed.), p. 61, McGraw-Hill.

National Oceanic and Atmospheric Administration, National Centers for Environmental Information, & U.S. Air Force–14th Weather Squadron. (2018, January). *Federal climate complex data documentation for integrated surface data (ISD)*. Asheville, NC. https://www.ncei.noaa.gov/data/global-hourly/doc/isd-format-document.pdf

New York City Panel on Climate Change. (2013, June). *Climate Risk Information 2013: Observations, Climate Change Projections, and Maps*. C. Rosenzweig and W. Solecki (Eds.), NPCC2. City of New York Special Initiative on Rebuilding and Resiliency, New York, New York. http://www.nyc.gov/html/planyc2030/downloads/pdf/npcc_climate_risk_information_2013_report.pdf

Philp, M., & Taylor, M. (2012, October). *Beyond agriculture: exploring the application of the Thornthwaite moisture index to infrastructure and possibilities for climate change adaptation* [Discussion paper—node 2]. Australian Climate Change Adaptation Research Network for Settlements and Infrastructure. https://www.accarnsi.unsw.edu.au/sites/accarnsi/files/uploads/PDF/Discussion/Beyond%20Agriculture%20-%20Exploring%20the%20application%20of%20the%20Thornthwaie%20Moisture%20Index.pdf

Thornthwaite, C. W. (1948). An approach toward a rational classification of climate. *Geographical Review*, *38*, 55-94.