



Article

Principal Component Neural Networks for Modeling, Prediction, and Optimization of Hot Mix Asphalt Dynamics Modulus

Parnian Ghasemi ^{1,*}, Mohamad Aslani ² , Derrick K. Rollins ³ and R. Christopher Williams ¹¹ Department of Civil and Environmental Engineering, Iowa State University, Ames, IA 50011, USA² Department of Mathematics, Florida State University, Tallahassee, FL 32304, USA³ Department of Chemical Engineering, Iowa State University, Ames, IA 50011, USA

* Correspondence: pghasemi@iastate.edu; Tel.: +1-515-708-7692

Received: 14 June 2019; Accepted: 10 August 2019; Published: 17 August 2019



Abstract: The dynamic modulus of hot mix asphalt (HMA) is a fundamental material property that defines the stress-strain relationship based on viscoelastic principles and is a function of HMA properties, loading rate, and temperature. Because of the large number of efficacious predictors (factors) and their nonlinear interrelationships, developing predictive models for dynamic modulus can be a challenging task. In this research, results obtained from a series of laboratory tests including mixture dynamic modulus, aggregate gradation, dynamic shear rheometer (on asphalt binder), and mixture volumetric are used to create a database. The created database is used to develop a model for estimating the dynamic modulus. First, the highly correlated predictor variables are detected, then Principal Component Analysis (PCA) is used to first reduce the problem dimensionality, then to produce a set of orthogonal pseudo-inputs from which two separate predictive models were developed using linear regression analysis and Artificial Neural Networks (ANN). These models are compared to existing predictive models using both statistical analysis and Receiver Operating Characteristic (ROC) Analysis. Empirically-based predictive models can behave differently outside of the convex hull of their input variables space, and it is very risky to use them outside of their input space, so this is not common practice of design engineers. To prevent extrapolation, an input hyper-space is added as a constraint to the model. To demonstrate an application of the proposed framework, it was used to solve design-based optimization problems, in two of which optimal and inverse design are presented and solved using a mean-variance mapping optimization algorithm. The design parameters satisfy the current design specifications of asphalt pavement and can be used as a first step in solving real-life design problems.

Keywords: hot mix asphalt dynamic modulus; principal component analysis; linear regression modeling; artificial neural network; receiver operating characteristic; optimization

1. Introduction

The stress-strain relationship for asphalt mixtures under sinusoidal loading can be described by the dynamic modulus, $|E^*|$, a function of material's components properties, loading rate, and temperature [1,2]. The dynamic modulus is one of the primary design inputs in Pavement Mechanistic-Empirical (M-E) Design to describe the fundamental linear viscoelastic material properties [3–5], and is one of the key parameters used to evaluate rutting and fatigue cracking distress predictions in Mechanistic-Empirical Pavement Design Guide (MEPDG) [5,6]. Although $|E^*|$ has a significant role in pavement design, the associated test procedure is time-consuming and requires expensive equipments, so extensive effort has been extended to predict $|E^*|$ from hot mix asphalt (HMA) material properties [7–9].

Predictive modeling is a process of estimating outcomes from several predictor variables using data mining tools and probability theory. An initial model can be formulated using either a simple linear equation or a more sophisticated structure obtained through a complex optimization algorithm [10].

There are several well-known predictive models for dynamic modulus, some of them are regression models, and some more recent ones have used techniques that include Artificial Neural Networks (ANN) and genetic programming [11]. Andrei et al. [12], used 205 mixtures with 2750 data points and revised the original Witczak model, and the developed model has subsequently been reformulated to use binder shear modulus rather than binder viscosity [13]. Christensen et al. [14], developed a new $|E^*|$ predictive model based on the law of mixtures. The data base used for training the model contained 206 $|E^*|$ measurements from 18 different HMA mixtures. Jamrah et al. [15], attempted to develop improved $|E^*|$ predictive models for HMA used in the State of Michigan. They observed a significant difference between measured and fitted $|E^*|$ values, especially at high temperatures and low frequencies. Alkhateeb et al. [16], developed a new predictive model from the law of mixtures to be used over broader ranges of temperature and loading frequencies, including higher temperatures/lower frequencies. The predictor variables used in that model were Voids in Mineral Aggregate (VMA) and binder shear modulus (G^*).

Sakhaeifar et al. [17], developed individual temperature-based models for predicting dynamic modulus over a wide range of temperatures. The predictor variables used in their model were aggregate gradation, VMA, Voids Filled with Asphalt (VFA), air void (V_a), effective binder content (V_{eff}), G^* , and binder phase angel (δ).

The existing dynamic modulus predictive models in the literature typically use two or more predictors from the following list: aggregate gradation, volumetric properties, and binder shear properties. These predictor variables are not necessarily an independent set of variables and thus it may not be appropriate for use in developing models. Since cross-correlated inputs in a dataset can unfavorably affect the accuracy of a predictive model by unduly affecting the estimation of their causative effects on the response variable, a pre-processing step of data evaluation would be useful for studying the quality of the input variables and their pair-wise correlations [18]. Principal Component Analysis (PCA) is a multivariate statistical approach that not only reduces the dimensionality of the problem but also converts a set of correlated inputs to a set of orthogonal (pseudo-)inputs using an orthogonal transformation [19]. During such a transformation, PCA maximizes the amount of information of the original dataset \mathbf{X} by using a smaller set of pseudo-variables [20,21]. Another issue in all of the predictive models is extrapolation that can be risky because a model might behave differently outside of the convex hull that contains all of the data points used for its training. To avoid using points outside of this convex hull, a hyper-space containing all data points can be found and added as a constraint on the desired modeling problem.

Ghasemi et al. [22], developed a methodology for eliminating correlated inputs and extrapolation in modeling; they created a laboratory database of accumulated strain values of several asphalt mixtures and used the resulting framework to estimate the amount of permanent deformation (rutting) in asphalt pavement. Following their new PCA-based approach, this study focuses on developing a machine-learning based framework for predicting the dynamic modulus of HMA using orthogonal pseudo-inputs obtained from principal component analysis. Unlike most of the existing $|E^*|$ predictive models, the proposed framework uses different data sets for model training and performance testing. To avoid extrapolation, an n-dimensional hyperspace is developed and added as a constraint to the modeling problem. This study also claims to determine the optimal HMA design and design variables for a pre-specified $|E^*|$ by applying framework using an evolutionary-based optimization algorithm. It is worth pointing out that, unlike other predictive models, the proposed framework is not site-specific and also not limited to the materials used in the American Association of State Highway and Transportation Officials (AASHTO) road test, i.e., this framework can adjust itself based on the dataset presented to the framework. The need for a more robust and general framework for

performance prediction in asphalt pavement also stems from the availability of the vast amount of experimental data in this field. In this work, the developed framework operates in such a spirit and improves the accuracy of available models via machine learning-based approaches.

The remainder of the document is organized as follows: Section 2 presents material and methodology, followed by Section 3 that covers results and discussion. Two examples of the proposed framework's applications are discussed in Section 4, followed by conclusions presented in Section 5.

2. Material and Methodology

Twenty-seven specimens from nine different asphalt mixtures (three replicates for each mixture group) were used in this study. Using AASHTO TP 79-13 the dynamic modulus test was performed at three temperatures (0.4, 17.1, and 33.8 °C) and nine loading frequencies (25, 20, 10, 5, 2, 1, 0.5, 0.2, 0.1 Hz). The maximum theoretical specific gravity (G_{mm}), the bulk specific gravity (G_{mb}), and the effective binder content (V_{beff}) were determined and used to calculate other volumetric properties of the asphalt mixtures. Asphalt binder shear properties were obtained from a dynamic shear rheometer (DSR) test. Using ASTM D7552-09(2014) the test was performed over a wide range of temperatures (−10 to 54 °C) and frequencies (0.1 Hz to 25 Hz), the same test temperatures and loading frequencies used in the mixture dynamic modulus test. It is important to note that this study uses a consistent definition of frequency, and that in order to predict the dynamic modulus value of an asphalt mixture for example at 4 °C and 25 Hz, for example, one should use as a model input the complex shear modulus of asphalt binder, $|G^*|$, at 4 °C and 25 Hz. A summary of the nine different mixture properties is given in Table 1. Using the laboratory test results on 27 specimens, a database of 243 data points was created for use in further modeling.

Table 1. General Mixture Properties of Nine Asphalt Mixtures Used in this Study.

	Mix 1	Mix 2	Mix 3	Mix 4	Mix 5	Mix 6	Mix 7	Mix 8	Mix 9
Binder performance grade	58–28	58–28	58–28	58–34	58–34	58–34	64–28	64–34	64–28
% V_{beff}	4.20	4.10	4.10	3.90	3.50	4.30	4.20	4.00	4.60
% VMA	13.50	13.50	13.60	13.10	12.50	13.90	13.70	13.40	14.40
% VFA	70.30	70.40	70.60	69.60	68.10	71.20	70.80	70.20	72.30
G_{mb}	2.32	2.31	2.31	2.32	2.31	2.32	2.31	2.32	2.31
G_{mm}	2.41	2.46	2.51	2.48	2.64	2.46	2.48	2.51	2.44
% V_a	4.01	3.99	3.99	3.98	3.98	4.03	4	3.99	3.98
% passing 3/4"	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
% passing 1/2"	93.90	96.40	87.20	93.50	95.10	96.40	94.10	94.40	94.20
% passing 3/8"	77.50	84.60	73.70	76.40	83.10	87.30	83.40	82.00	80.90
% passing #4	49.80	53.10	48.40	52.20	52.20	60.90	63.80	48.20	58.60
% passing #8	34.40	38.40	35.10	43.60	38.80	46.90	47.10	34.90	46.00
% passing #30	16.70	18.70	17.90	20.90	18.80	23.40	21.70	19.20	25.90
% passing #50	10.30	10.80	10.90	11.40	9.90	12.40	11.90	11.80	13.80
% passing #100	6.10	5.90	6.40	5.80	5.40	6.10	6.60	6.10	7.20
% passing #200	3.60	3.30	6.20	3.30	3.50	3.40	4.00	3.10	4.00

2.1. Preliminary Processing Step: Input Variable Selection

A parsimonious set of input variables is required to develop a model [20]. For a common model structure, one can represent the expectation function of the response as $y_i = f_i(\mathbf{x}_i, \theta)$, where y_i is the expected response variable at the i th measurement, $i = 1, \dots, n$, \mathbf{x}_i is the input vector at the i th measurement, and θ is the vector of unknown model parameters with $\theta = [\theta_1 \dots \theta_q]^T$. It is assumed that the element in the i th row and j th column of the Jacobian matrix, \mathbf{J} , is $\frac{\partial \eta_i}{\partial \theta_j}$ i.e., $\mathbf{J} = \left\{ \frac{\partial \eta_i}{\partial \theta_j} \right\}$. Note that the j th column represents θ_j and its column vector reflects the variation in the response space as θ_j varies over a specific set of experimental conditions. If j and k are two orthogonal columns, their correlation coefficient (r) must be zero, meaning that the information used to estimate θ_j is independent from the information used to estimate θ_k and vice versa. The benefit of using orthogonal input variables

is that not only does it result in consolidation of causative effects of inputs on the output but it also maximizes parameter accuracy and therefore estimation accuracy of the predicted output.

According to the literature [11–14,16,17,23,24], the stiffness characteristic of an asphalt mixture presented by a dynamic modulus can be estimated by its component properties. In this study, the input variables vector (x) defines the asphalt mixture's component properties. A summary of the selected input variables and their ranges in the dataset is presented in Table 2 with the x_i 's and y being the input and output variables, respectively.

Table 2. Selected Input Variables (x) And Output Variable (y) For Model Development.

Variable	Identity	Min.	Max.	Ave.	Std. Dev.
y	$\text{Log} E^* $	2.62	4.37	3.76	0.46
x_1	Cum. % retained on 3/4"	3.60	13.00	6.11	2.63
x_2	Cum. % retained on 3/8"	12.68	26.29	19.01	4.11
x_3	Cum. % retained on #4	36.20	51.76	45.86	5.319
x_4	Cum. % retained on #8	52.87	65.70	59.42	5.06
x_5	Cum. % retained on #30	74.06	83.30	79.63	2.76
x_6	Cum. % retained on #50	86.22	90.12	88.57	1.15
x_7	Cum. % retained on #100	92.81	94.59	93.83	0.48
x_8	% Passing from #200	3.07	6.18	3.81	0.89
x_9	$\text{Log} G^* $	−2.29	3.03	0.50	1.26
x_{10}	Phase angle (degree)	28.15	79.17	52.86	11.54
x_{11}	% V_{beff}	3.50	4.60	4.10	0.29
x_{12}	%VMA	12.50	14.40	13.51	0.49
x_{13}	%VFA	68.10	72.30	70.40	1.08
x_{14}	% V_a	3.98	4.01	3.99	0.01

Cross-correlation analysis is performed on the 14 selected predictor variables and the obtained pairwise correlation matrix is given in Table 3 along with the schematic heat map of the correlation matrix given by Figure 1. Correlation coefficients with absolute values of 0.5 or higher are displayed in bold red text. The corresponding cells in the correlation heat map are shown in dark blue and dark red as shown in Figure 1. According to Table 3, the absolute values of the 130 correlation coefficients are greater than 0.1, with 50 of them greater than 0.5, indicating that several of the input variables give an impression of being highly correlated. The correlation heat map also clearly indicates that a high level of correlation (dark blue and dark red cells) exists within the input variables. If the correlated input variables are detected, to enable accurate mapping of the inputs to the response variable, it would be useful to produce a smaller set of orthogonal pseudo-variables using the PCA method and use them in model development [20].

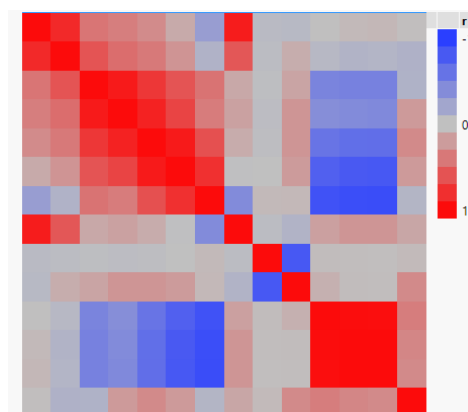


Figure 1. Heat map depict a visualization for the pairwise correlation matrix of the input variables. Each column i and row j shows correlation coefficient r_{ij} . Cells with colors otherthan gray indicate some level of correlation. As the colors get darker the level of correlation goes higher.

Table 3. Pairwise Correlation Matrix for the Selected Input Variables.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
x_1	1	0.832	0.412	0.366	0.294	0.119	−0.269	0.905	−0.044	−0.058	0.003	0.04	0.049	0.013
x_2	0.832	1	0.597	0.458	0.391	0.246	−0.109	0.583	−0.035	0.106	−0.061	−0.099	−0.089	−0.115
x_3	0.412	0.597	1	0.918	0.756	0.596	0.425	0.133	−0.019	0.154	−0.465	−0.485	−0.49	−0.111
x_4	0.366	0.458	0.918	1	0.87	0.687	0.375	0.169	−0.028	0.237	−0.388	−0.412	−0.424	0.212
x_5	0.294	0.391	0.756	0.87	1	0.919	0.618	0.112	−0.021	0.235	− 0.585	− 0.631	− 0.633	0.3
x_6	0.119	0.246	0.596	0.687	0.919	1	0.794	−0.009	0.003	0.203	− 0.741	− 0.796	− 0.806	0.209
x_7	−0.269	−0.109	0.425	0.375	0.618	0.794	1	−0.414	0.036	0.047	− 0.854	− 0.886	− 0.892	−0.087
x_8	0.905	0.583	0.133	0.169	0.112	−0.009	−0.414	1	−0.032	−0.102	0.179	0.238	0.238	0.142
x_9	−0.044	−0.035	−0.019	−0.028	−0.021	−0.003	0.036	−0.032	1	− 0.808	0.021	0.016	0.013	0.034
x_{10}	−0.058	0.106	0.154	0.237	0.235	0.203	0.047	−0.102	− 0.808	1	0.09	0.024	0.014	0.3
x_{11}	0.003	−0.061	−0.465	−0.388	− 0.585	− 0.741	− 0.854	0.179	0.021	0.09	1	0.988	0.985	0.372
x_{12}	0.04	−0.099	−0.485	−0.412	− 0.631	− 0.796	− 0.886	0.238	0.016	0.024	0.988	1	0.998	0.321
x_{13}	0.049	−0.089	−0.49	−0.424	− 0.633	− 0.806	− 0.892	0.238	0.013	0.014	0.985	0.998	1	0.301
x_{14}	0.013	−0.115	−0.111	0.212	0.3	0.209	−0.087	0.142	0.034	0.3	0.372	0.321	0.301	1

2.2. Orthogonal Transformation Using PCA

In multivariate statistics, PCA is an orthogonal transformation of a set of (possibly) correlated variables into a set of linearly uncorrelated ones, and the uncorrelated (pseudo-) variables, called principal components (PCs), are linear combinations of the original input variables. This orthogonal transformation is performed such that the first principal component has the greatest possible variance (variation within the dataset). This procedure is then followed for the second component, then the third component, etc. This means that each succeeding component in turn has the highest variance when it is orthogonal to the preceding components [25–28]. To help visualize the PCA transformation, a schematic dataset with three input variables is presented in Figure 2 (left). As shown in this figure, in conducting PCA the data points are transferred from the original 3D original input space (on the left) to a 2D principal component space (on the right).

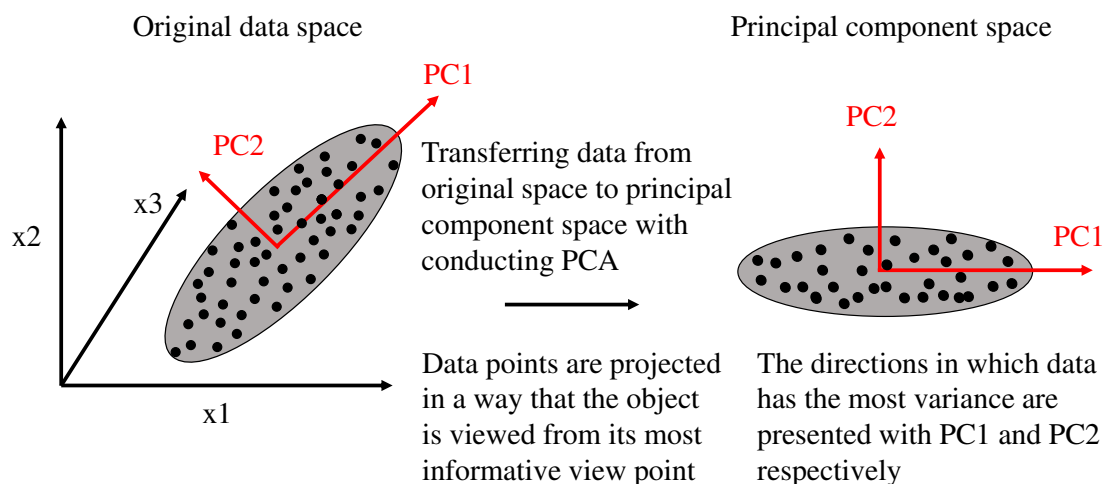


Figure 2. Schematic of the PCA transformation. Original data space presented on the left with 3 (input) variables transformed to a component space with lower dimension and pc_1 and pc_2 being the axes of the coordinate.

PCA can be performed either by eigenvalue decomposition of a data covariance (or correlation) matrix or by singular value decomposition. The process usually begins with mean centering the data matrix (and normalizing or using Z-scores) for each attribute as follows:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad \mathbf{Z} = \begin{bmatrix} \frac{x_{11}-\bar{x}_1}{s_1} & \frac{x_{12}-\bar{x}_2}{s_1} & \cdots & \frac{x_{1p}-\bar{x}_p}{s_p} \\ \frac{x_{21}-\bar{x}_1}{s_1} & \frac{x_{22}-\bar{x}_2}{s_2} & \cdots & \frac{x_{2p}-\bar{x}_p}{s_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{x_{n1}-\bar{x}_1}{s_1} & \frac{x_{n2}-\bar{x}_2}{s_2} & \cdots & \frac{x_{np}-\bar{x}_p}{s_p} \end{bmatrix} \quad (1)$$

where for $k = 1$ to n and $j = 1$ to p , x_{kj} is the k th measurement for the j th variable, \bar{x}_k is the sample mean for the k th variable, and s_k is sample standard deviation for the k th variable. As discussed in the previous section, highly correlated input variables lead to inflation of the standard error of estimate, negatively affecting the accuracy of the estimation. PCA will help us not only reduce the dimensionality of the modeling problem, but will also produce orthogonal pseudo-variables to be used in solving the problem. To perform PCA in this study we used eigenvalue decomposition of the correlation matrix of the data. The eigenvalues of the data correlation matrix are calculated, ranked, and sorted in descending order (representing their quota of the total variation within the dataset), as presented in Table 4. According to the eigenvalues, the first five PCs represent 95.8% of the existing variation within the dataset.

Recalling the fact that the PCs are linear combinations of the original input variables, the PCs can be defined as in Equation (2):

$$pc_i = \sum_{j=1}^{14} \alpha_{ij} x_j + \beta_i \quad (2)$$

where $i = 1$ to 14 , the α_{ij} is the corresponding coefficients, the β_i are constants, and the x_j are the original input variables. Equation (2) can be stated in matrix notation as in Equation (3):

$$\mathbf{p} = \mathbf{M}\mathbf{z} + \mathbf{n} \quad (3)$$

where

$$\mathbf{p} = \begin{bmatrix} pc_1 \\ pc_2 \\ pc_3 \\ pc_4 \\ pc_5 \end{bmatrix} \quad \mathbf{M}^T = \begin{bmatrix} 0.03 & 0.19 & -0.08 & -0.06 & -0.09 \\ 0.03 & 0.11 & -0.04 & -0.05 & 0.04 \\ 0.06 & 0.04 & 0.00 & 0.00 & 0.10 \\ 0.06 & 0.05 & 0.02 & 0.04 & 0.07 \\ 0.13 & 0.06 & 0.05 & 0.09 & -0.01 \\ 0.33 & 0.01 & 0.09 & 0.16 & -0.20 \\ 0.71 & -0.54 & 0.05 & 0.13 & -0.20 \\ -0.03 & 0.52 & -0.20 & -0.07 & -0.53 \\ -0.01 & -0.05 & -0.4 & 0.43 & 0.12 \\ 0.00 & 0.01 & 0.06 & -0.02 & 0.00 \\ -1.26 & 0.64 & 0.52 & 0.61 & 0.67 \\ -0.75 & 0.37 & 0.21 & 0.28 & 0.31 \\ -0.34 & 0.17 & 0.09 & 0.12 & 0.14 \\ -4.41 & 18.38 & 47.08 & 76.24 & -40.48 \end{bmatrix} \quad (4)$$

$$\mathbf{n} = \begin{bmatrix} -55.95 \\ -58.54 \\ -218.20 \\ -352.79 \\ 174.78 \end{bmatrix}$$

Table 4. Eigenvalues of the Normalized Input Variables Matrix Sorted in Descending Order (Based on Their Contribution to Total Variation).

Number	Eigenvalue	Percent Variation	Cumulative Percent Variation
1	6.0225	43.018	43.018
2	3.2193	22.995	66.013
3	1.9746	14.104	80.118
4	1.4174	10.124	90.242
5	0.7850	5.607	95.848
6	0.3176	2.269	98.117
7	0.1091	0.779	98.896
8	0.0778	0.556	99.452
9	0.0549	0.392	99.844
10	0.0218	0.156	100

Further modeling efforts will be performed using the first five PCs.

2.3. Holdout Cross Validation

In prediction problems, cross validation will be used to estimate model accuracy. Cross validation is a model validation technique that can be used to prevent overfitting as well as to assess how the results of a statistical analysis can be generalized to an independent dataset. In this study, a holdout cross validation technique is used in which the given dataset is randomly assigned to two subsets, d_0 and d_1 , the training set and the test set, respectively. Since the training set contains 80% of the data points and the test set contains 20% of the data points, 80% of the data points are used to train the model and the remainder are used to evaluate the trained model's performance.

2.4. Principal Component Regression (PCR)

Linear regression attempts to model the relationship between response variables and explanatory variables by fitting a linear equation to observed data. In regression analysis, the least-squares method is used to calculate the best fitting line for the observed data by minimizing the sum of the squares of the residuals (differences between the measured responses and the fitted values by a linear function of parameters).

All possible regression structures were considered for representing the relationship between the response variable, $\log|E^*|$, and predictor variables (pc_1 , pc_2 , pc_3 , pc_4 , and pc_5). To estimate the values of the unknown coefficients in the model, the least-squares criterion of minimizing the sum of squared residuals (SSE) is used. Finally, after eliminating redundant terms, the reduced third order cubic and interaction terms were developed and selected as the best-fitted model. The developed model is called "Principal Component Regression (PCR)".

2.5. Principal Component Neural Network (PCNN)

A predictive model called "Principal Component Neural Network (PCNN)" is developed as briefly described in this section. ANNs are brain-inspired systems intended to replicate the way humans learn. Neural network structures consist of several layers, including input layers, output layers, and hidden layer(s), with nodes (neurons) in each layer [29–31]. A three-layer feed-forward neural network

is developed for this study. It consists of an input layer of five neurons (five input variables), a hidden layer of 10 neurons, and an output layer of one neuron (one response variable). A trial-and-error procedure of optimizing the computational time and cost function is used to choose the number of hidden neurons. In this study supervised learning is used in which a training dataset, including inputs and outputs, is presented to the network. The network adjusted its weights in such a way that the adjusted set of weights produces an input/output mapping resulting in the smallest error. This iterative process is carried on until the sum of square residuals (SSE) increases. After the learning or training phase, the performance of the trained network must be measured against an independent (unseen) testing data [29,32]. Let the input of each processing node be pc_i , the adjustable connection weight be w_{ij} , and let the bias at output layer be b_0 , so that the network transfer (activation) function is $f(\cdot)$. The j th output of the first layer can be obtained using Equation (5)

$$v_j = f_1(pc_i, w_{ij}), \quad i = 1, \dots, 5 \quad \text{and} \quad j = 1, \dots, 10 \quad (5)$$

and the response will be

$$\hat{y} = f_2(f_1(pc_i, w_{ij})) \quad (6)$$

If we assume that

$$f_2(v_j, w_{Hj}) = b_0 + \sum_j v_j w_{Hj}, \quad (7)$$

and for each j ,

$$f_1(pc_i, w_{ij}) = b_{Hj} + \sum_j pc_i w_{ij}, \quad (8)$$

then a feed-forward neural network can be formulated as follows:

$$\hat{y} = f_2 \left\{ b_0 + \sum_{j=1}^n \left[w_{Hj} \cdot f_1 \left(b_{Hj} + \sum_{i=1}^m pc_i w_{ij} \right) \right] \right\}, \quad (9)$$

where pc_i is pseudo input parameter i , w_{ij} is the weight of connection between input variable i (for $i = 1$ to 5) and neuron j of the hidden layer, b_0 is a bias at the output layer, w_{Hj} is the weight of connection between neuron j of the hidden layer and output layer neuron, b_{Hj} is a bias at neuron j of the hidden layer (for $j = 1$ to 10), and $f_1(t)$ and $f_2(t)$ are transfer functions of the hidden layer and output layer, respectively.

It should be pointed out that iteration proceeds until the convergence criterion is met. Thus, similar to the linear regression model, the validation set is not used. The Bayesian Regularization algorithm is used to achieve network training efficiency.

2.6. Effective Variable Space

It is widely known that the use of an empirical predictive model outside the convex hull containing the data points is prohibited. In this context, *effective variable space* is referred to the space where the uncertainty of the developed models is bound to their already calculated thresholds. In other words, outside of this region, the extrapolated behavior of the models may not be predictable. To guard against extrapolation, Ghasemi et al. [22] concluded that the space containing input data could be interpreted as a symmetrical convex space, then demonstrated how this space can be used in the design procedure.

Following the approach in [22], a normal distribution is assumed for each input variable (x_i), resulting in their joint distribution being bi-variate normal, and such distributions are usually represented in form of a contour diagram. Since a contour curve on such a diagram contains the points on a surface with the same distance from the $x_i x_j$ plane, these points have a constant density function [33] (see Figure 3 for an example of such distribution). The cross section is obtained by slicing a bi-variate normal surface at a constant distance from the $x_i x_j$ plane. As indicated in Figure 3,

the n -dimensional hyperspace is a hyper-ellipsoid with minimum volume (to avoid any gaps in the edges).

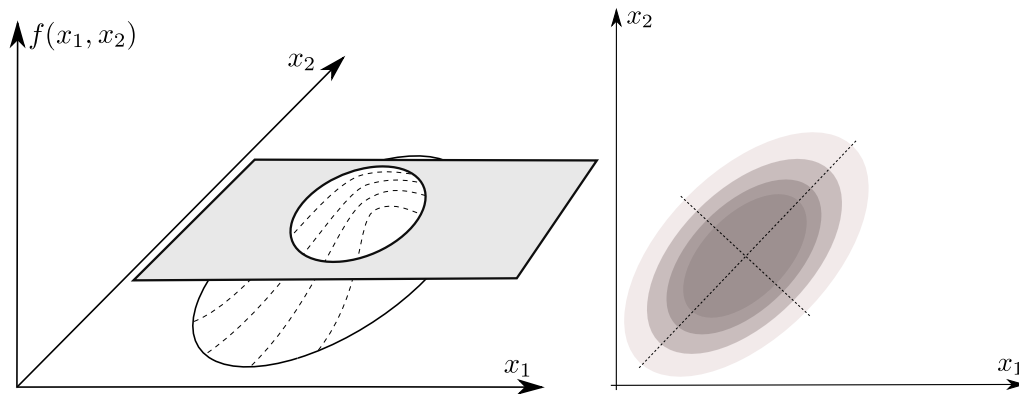


Figure 3. A schematic of a bivariate normal distribution (left) [33] and elliptic cross-section of this distribution (right). Projecting the cross-sectional area on the (x_1, x_2) plane results in an ellipsoid.

Khachiyan's work [34] formulates the problem of finding an approximate minimum volume enclosing ellipsoid (\mathcal{E}) given p data points in n -dimensions as an optimization problem. In Ghasemi et al. [22], the authors detailed the derivation of a procedure for solving this problem and obtaining its effective variable space. For brevity, the flowchart in Figure 4 summarizes this iterative method for finding the minimum volume enclosing ellipsoid. This algorithm was used to find two enclosing ellipsoids in the primary space (14-dimensional) and the pseudo space (5-dimensional) of the dataset. It should be pointed out that this space is independent of the predictive models and is used only to solve the optimal (and inverse) design problems.

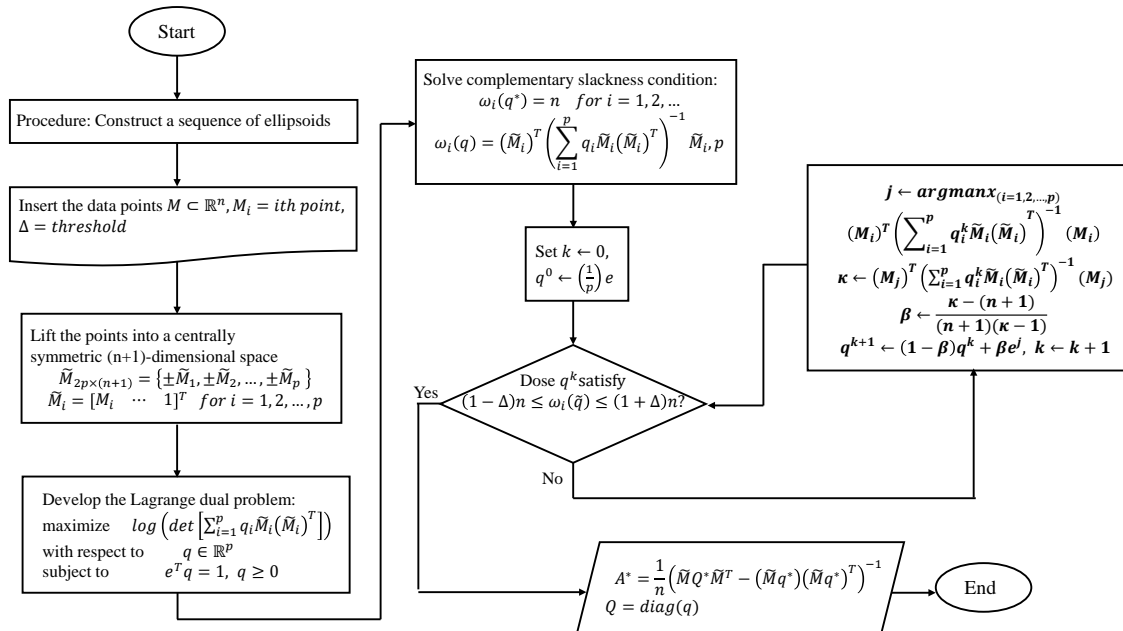


Figure 4. Iterative method to solve the problem of finding minimum volume enclosing ellipsoid [22].

2.7. Guideline for Implementation

A summary of the methodologies used to develop the framework is presented in Figure 5. The procedure begins with collecting experimental data from the laboratory, followed by the pre-processing step of input variable evaluation. The flowchart continues with the model development and the addition of a constraint on the n -dimensional input variable hyperspace to the modeling

problem. The developed models can then be used to predict pavement performance, solve design-based optimization problems, etc. There are a number of aspects of the proposed framework that can be achieved using free and commercially available software like *MATLAB*[®], *Python*, and *R packages*, and one may implement many parts of the framework in the language of their interest. For example, the algorithm to find the n-dimensional hyper-ellipsoid is very straightforward using the flowchart in Figure 4.

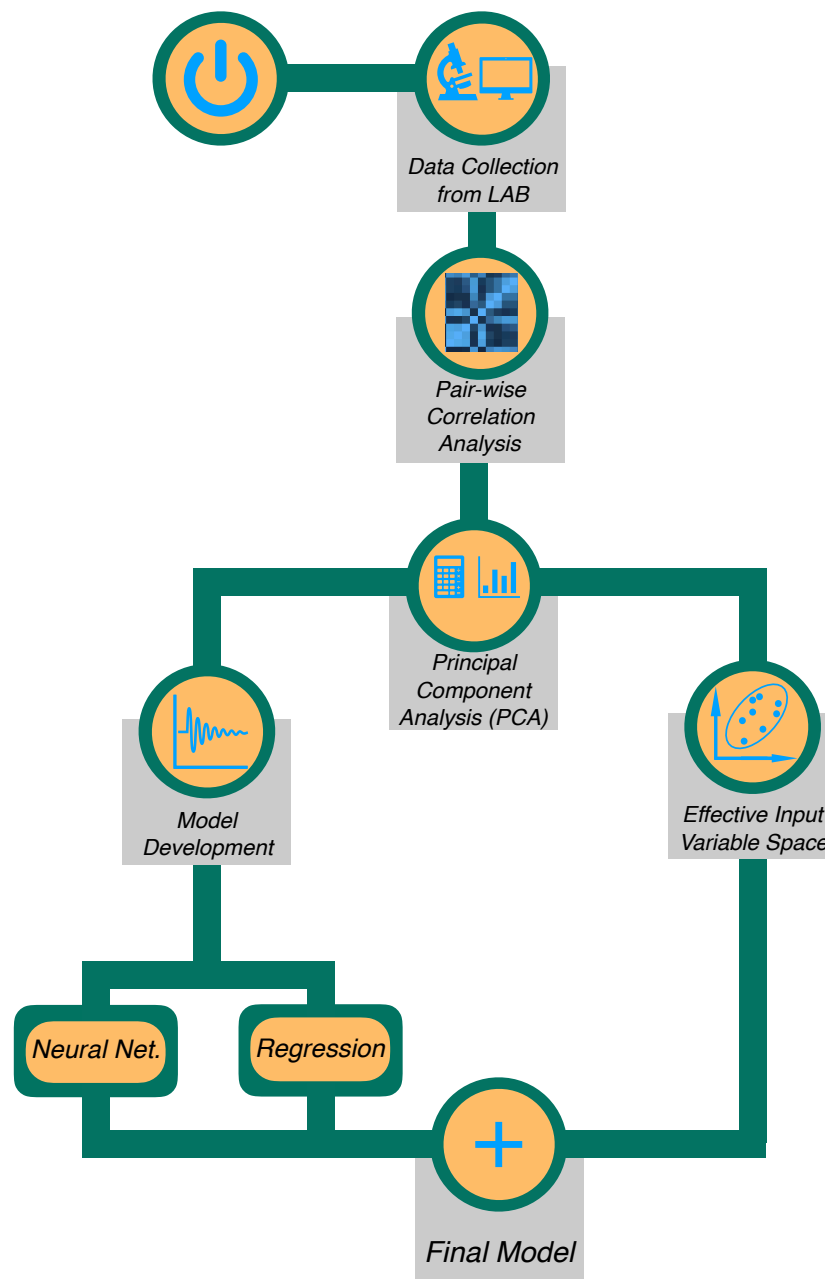


Figure 5. A summary of the methodologies used to develop the machine learning-based framework for predicting dynamic modulus (as an example of performance related property of asphalt mixture).

3. Developed Model Results, Performance, and Validation

The results produced by the developed models are presented in this section, and their capability to use empirical data to estimate the dynamic modulus of asphalt mixtures is evaluated.

3.1. Model Performance

The performance is first compared with the existing predictive models; modified Witczak, Hirsch, and Alkhateeb models selected from the literature are presented in Equations (10)–(13), respectively [13,14,16]

$$\begin{aligned} \log|E^*| = & -0.349 + 0.754(|G_b^*|^{-0.0052})(6.65 - 0.032\rho_{200} - 0.0027(\rho_{200})^2 + \\ & 0.011\rho_4 - 0.0001(\rho_4)^2 + 0.006\rho_{3/8} - 0.00014(\rho_{3/8})^2 - 0.08V_a - 1.06(\frac{V_{beff}}{V_{beff} + V_a})) \\ & + \frac{2.558 + 0.032V_a + 0.713(\frac{V_{beff}}{V_{beff} + V_a}) + 0.0124\rho_{3/8} - 0.0001(\rho_{3/8})^2 - 0.0098\rho_{3/4}}{1 + \exp(-0.7814 - 0.5785 \log|G_b^*| + 0.8834 \log \delta_b)} \end{aligned} \quad (10)$$

where $|E^*|$ is dynamic modulus in psi; $|G^*|$ is the binder shear modulus in psi; δ_b is the binder phase angle in degrees; $\rho_{3/4}$ is the cumulative percent aggregate retained on the 3/4" sieve (19 mm); $\rho_{3/8}$ is the cumulative percent aggregate retained on the 3/8" sieve (9.5 mm); ρ_4 is the cumulative percent aggregate retained on the No. 4 sieve (4.75 mm); ρ_{200} is the percent aggregate passing the No. 200 sieve (0.075 mm); V_a is the percent air void in the mix; V_{beff} is the effective asphalt content; VMA is the percent of voids in the mineral aggregate, and VFA is the percent voids filled with asphalt,

$$|E_m^*| = P_c \left(4,200,000(1 - \frac{VMA}{100}) + 3|G_b^*|(\frac{VFA \times VMA}{10,000}) \right) + \frac{(1 - P_c)}{\frac{1-VMA}{4,200,000} + \frac{VMA}{3|G_b^*|(VFA)}} \quad (11)$$

where

$$P_c = \frac{(20 + 3|G_b^*|(VFA)/(VMA))^{0.58}}{650 + (3|G_b^*|(VFA)/(VMA))^{0.58}} \quad (12)$$

and $|E_m^*|$ is dynamic modulus of HMA in psi; P_c is the aggregate contact volume; VMA is the percentage of mineral aggregate voids in compacted mixture; and VFA is the percentage of voids filled with asphalt in the compacted mixture,

$$|E_m^*| = 3(\frac{100 - VMA}{100}) \left(\frac{(90 + 1.45 \frac{|G_b^*|}{VMA})^{0.66}}{1100 + (0.13 \frac{|G_b^*|}{VMA})^{0.66}} \right) |G_g^*| \quad (13)$$

where $|E_m^*|$, $|G_b^*|$, and $|G_g^*|$ (the complex shear modulus of binder in the glassy state, assumed to be 10^9 Pa.) are in Pa. Equation (14) shows the best reduced third-order (linear) regression model (PCR) fitting the measured response:

$$\begin{aligned} \hat{y} = & c_0 + c_1pc_1 + c_2pc_2 + c_3pc_3 + c_4pc_4 + c_5pc_5 \\ & + c_6pc_1pc_2 + c_7pc_1pc_3 + c_8pc_1pc_4 + c_9pc_1pc_5 \\ & + c_{10}pc_2pc_3 + c_{11}pc_2pc_4 + c_{12}pc_2pc_5 + c_{13}pc_3pc_4 \\ & + c_{14}pc_3pc_5 + c_{15}pc_4pc_5 + c_{16}pc_1pc_2pc_3 + c_{17}pc_1pc_2pc_4 \\ & + c_{18}pc_1pc_2pc_5 + c_{19}pc_1pc_3pc_4 + c_{20}pc_1pc_3pc_5 + c_{21}pc_2pc_3pc_4 \\ & + c_{22}pc_1pc_4pc_5 + c_{23}pc_2pc_4pc_5 + c_{24}pc_3pc_4pc_5 \end{aligned} \quad (14)$$

where, $c_0 = 6.59$; $c_1 = 2.58$; $c_2 = 4.4$; $c_3 = -0.36$; $c_4 = 0.49$; $c_5 = 1.93$; $c_6 = -0.33$; $c_7 = -0.77$; $c_8 = -1.69$; $c_9 = 0.15$; $c_{10} = -1.65$; $c_{11} = -4.68$; $c_{12} = 4.81$; $c_{13} = 0.7$; $c_{14} = -0.85$; $c_{15} = -1.58$; $c_{16} = -0.17$; $c_{17} = -0.79$; $c_{18} = 1.83$; $c_{19} = 0.04$; $c_{20} = 0.18$; $c_{21} = 0.42$; $c_{22} = 0.05$; $c_{23} = 0.32$; $c_{24} = 0.06$. The trained three-layer ANN (PCNN) presented in Equation (9) contains the following connection weights and biases:

$$\mathbf{W}^T = \begin{bmatrix} -0.511 & 0.134 & 0.654 & -1.064 & -0.267 \\ -0.315 & -0.147 & -0.267 & 0.177 & -1.047 \\ -0.060 & -1.266 & 0.759 & -1.248 & -0.331 \\ -0.075 & 0.022 & 0.208 & 0.015 & 0.167 \\ -0.074 & 0.022 & 0.206 & 0.015 & 0.165 \\ 0.103 & -0.177 & 1.253 & -1.045 & 0.535 \\ 0.078 & -0.020 & -0.231 & -0.014 & -0.172 \\ 0.238 & 0.070 & -0.885 & 0.848 & 0.943 \\ 0.123 & 0.456 & -0.387 & 1.547 & -0.017 \\ -0.079 & 0.020 & 0.213 & 0.014 & 0.173 \end{bmatrix}$$

$$\mathbf{W}_H = \begin{bmatrix} 0.869 \\ -0.886 \\ 0.632 \\ -0.291 \\ -0.288 \\ -0.859 \\ 0.299 \\ 0.556 \\ 0.971 \\ -0.299 \end{bmatrix}, \mathbf{B}_H = \begin{bmatrix} 0.162 \\ 0.710 \\ 0.319 \\ -0.008 \\ -0.009 \\ -0.570 \\ 0.007 \\ 0.290 \\ -0.373 \\ -0.007 \end{bmatrix}, \mathbf{B}_0 = [0.148]$$

Figure 6 presents the performance of the developed models in terms of measured values of dynamic modulus versus the fitted dynamic modulus values. The measured and fitted values are fairly close to the line of equality, indicating that the fitted values are highly correlated with the measured ones.

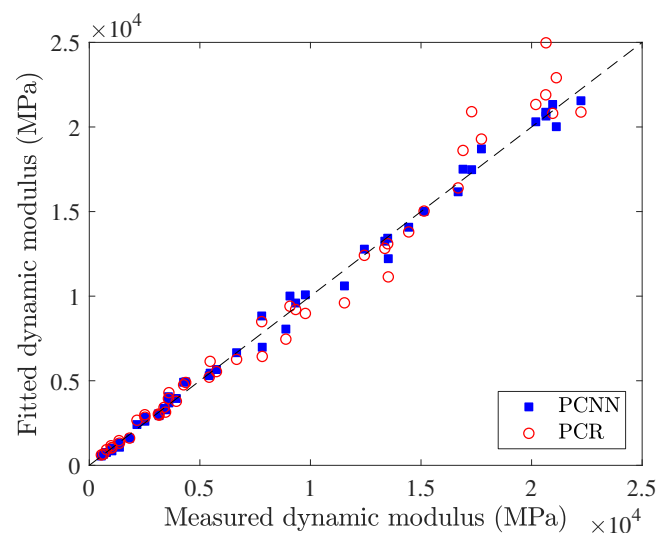


Figure 6. Measured values of dynamic modulus versus fitted values by PCR and PCNN.

Comparisons of PCR and PCNN performance to that of the existing predictive models are conducted based on three statistics: average difference (AD), average absolute difference (AAD), and correlation between measured and fitted values of response (r_{fit}). A summary of the definitions of these statistical components and their formulas is presented in Table 5. In the formulas presented in Table 5, y_i is the i th measured response, \hat{y}_i is the i th fitted response, and n is the number of data points.

The results of the comparison are presented in Table 6. According to the values of r_{fit} in Table 6, the estimated dynamic modulus values obtained from PCR and PCNN models are highly correlated

with measured values according to the values, showing that the both PCR and PCNN performed well in terms of modeling the response variable.

Table 5. Statistics which are used to compare model performance.

Statistical Component	Formula	Definition
Average difference (AD)	$AD = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)$	An estimate of systematic model bias
Average absolute difference (AAD)	$AAD = \frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $	Average closeness of the fitted and measured values of response
r_{fit}	$r_{fit} = \frac{n \sum_{i=1}^n y_i \hat{y}_i - (\sum_{i=1}^n y_i)(\sum_{i=1}^n \hat{y}_i)}{\sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2} \sqrt{n \sum_{i=1}^n \hat{y}_i^2 - (\sum_{i=1}^n \hat{y}_i)^2}}$	Correlation of the measured and fitted values of response
Coefficient of determination (R^2)	$R^2 = 1 - \frac{SS_{res}}{SS_{total}}$	Portion of the response variation elucidated by regressors in the fitted model in linear models

Although the corresponding values of r_{fit} for modified Witczak, Hirsch, and Alkhateeb models are 0.93, 0.95, and 0.95, respectively, the average difference and average absolute difference with respect to the measured response are significantly higher than those of PCR and PCNN. This means that the fitted values by the modified Witczak, Hirsch, and Alkhateeb models are not close as those fitted by PCR and PCNN to the response value. In other words, r_{fit} , which reflects the correlation between response and estimated response (if one goes up the other one goes up), could be biased, and in this situations other statistics (AD, and AAD) could be used to evaluate the goodness of fit. The dynamic modulus measured and predicted values are presented in Figure 7 for four asphalt mixtures. According to the presented master curves the current study (PCNN model) provides the closest values of E^* to the measurements for all of three test temperatures, while, the conventional models either overestimate or underestimate the response variable.

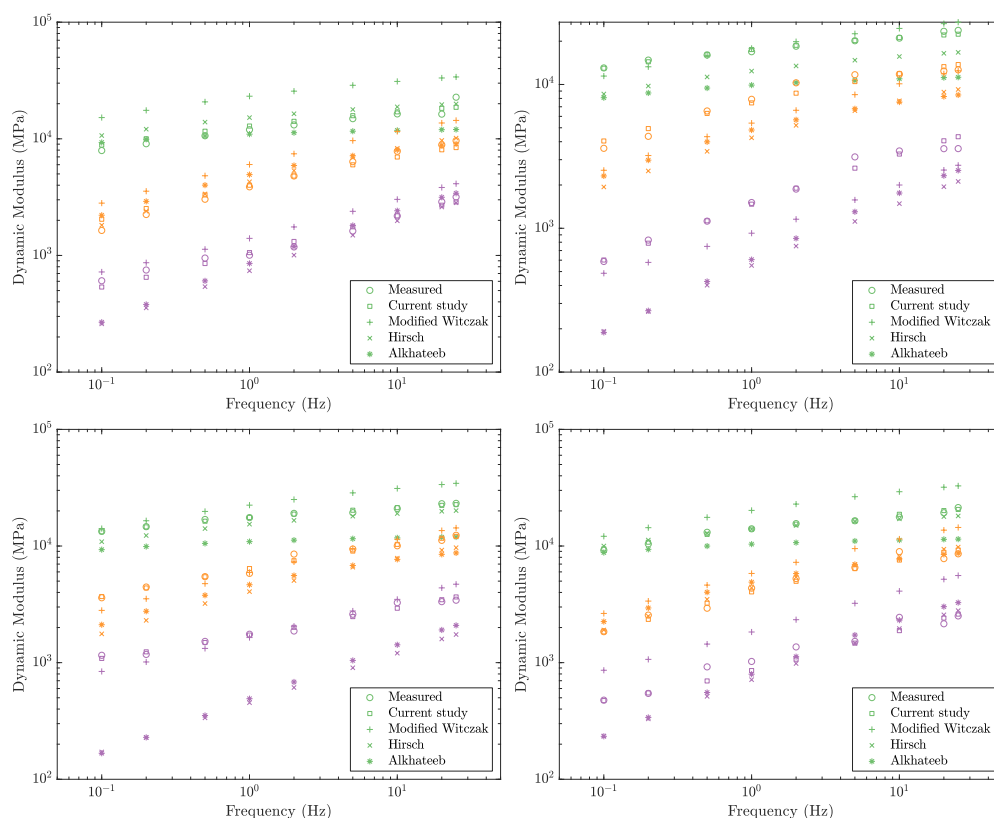


Figure 7. Comparing the measured and predicted (current study and conventional models) dynamic modulus. Results are presented at three temperatures. Green color indicates 0.4 °C, Orange color indicates 17.1 °C, and Purple color indicates 33.8 °C.

Table 6. Performance Comparison of the Developed and Existing Models using Statistical Analysis (na*: not applicable. R^2 is applicable to models with linear parameters and thus it is not applicable to PCNN due to the non-linear nature of its parameters.).

		Average Difference (MPa)	Average Absolute Difference (MPa)	r_{fit}	R^2
PCR	Training	3.9	575.3	0.996	0.99
	Testing	−162.3	718.9	0.995	na
PCNN	Training	13.2	380.7	0.997	na
	Testing	9.7	337.5	0.997	na
Modified Witzack		−2460	3152.1	0.93	0.88
Hirsch		1241.6	1785.7	0.95	0.91
Alkhateeb		2844.5	2984.5	0.95	0.90

A graphical comparison of the PCR and PCNN performance and that of the existing models is presented in the following section.

3.2. Receiver Operating Characteristic Analysis (ROC)

A receiver operating characteristic (ROC) graph is a technique for visualizing, organizing, and selecting classifiers based on their performance. ROC graphs are widely used in medical decision-making as well as in machine learning and data-mining research [35]. True ROC curves plot the false positive rate (probability of false alarm) on the x -axis and the true positive rate (probability of detection) on the y -axis. A classifier is said to perform well if the ROC curve climbs rapidly towards the upper left-hand corner. The more the curve deviates from $y = x$ behavior, the more accurate the prediction is [36]. We can borrow from the concept of ROC curve to obtain a measurement of fit for the competing models and a ROC graph for this study is presented in Figure 8 for this study. As described in Equation (15), the x -axis indicates the standardized residuals ordered from the lowest to the highest (e_i^{**}). Residuals are sorted in ascending order and divided by the largest one that belongs to the Hirsch model.

$$e_i^{**} = \frac{|e_i^*|}{|e_{i_{MAX}}^*|} \quad (15)$$

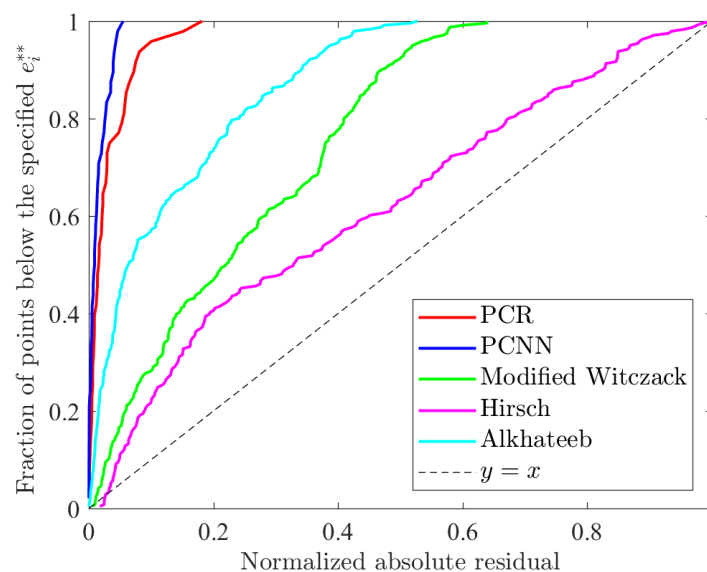


Figure 8. ROC curves for the developed and the existing predictive models for the dynamic modulus. The curves indicate the goodness of the fit provided by the models with regard to a pre-specified residual. The PCNN model showed the highest performance (farthest from the $y = x$ line), and the Hirsch model showed the lowest (closest to the $y = x$ line).

The y -axis indicates the fraction of points whose standardized residuals are less than e_i^{**} [37]. Although the curves obtained for all of the models are monotonically non-decreasing and climb towards the upper left-hand corner (the desired situation shows that the predictive models perform well), PCNN and PCR curves were the highest, proving their better performance to be better than that of existing predictive models.

A convenient global measure of the goodness-of-the-fit is the Area Under the Curve (AUC). To compare classifiers, it is more desirable to reduce ROC performance, to a single scalar value representing expected performance. Since the AUC is a portion of the area of the unit square, its value will always lie between 0 and 1. The AUC values for the PCNN, PCR, Alkhateeb, modified Witczak, and Hirsch models are 0.9864, 0.9717, 0.8746, 0.7609, and 0.6320, respectively. One can use the ROC and AUC analysis results and rank the predictive models according to their performances. In this study, the PCNN model reflected the highest performance in predicting the dynamic modulus value, while the Hirsch model ranked the lowest among all the models.

3.3. Model Validation

The current regression model is presented in the following general form as

$$y_i = f_i(\mathbf{Z}_i, \boldsymbol{\theta}) + e_i^* \quad (16)$$

In the above equation f_i is the i th expectation function, $\boldsymbol{\theta}$ is the vector of parameters, and e_i^* is a random deviation of y_i from f_i . This term is assumed to be independent and normally distributed with a mean of zero and unknown variance σ^2 for $i = 1, \dots, n$, where n is the number of input vectors. If the above assumptions are violated, the results of the analysis could be misleading or erroneous. These assumptions can be testified by examining residuals as defined by

$$e_i^* = y_i - \hat{y}_i \quad (17)$$

The assumption of independency holds when the residuals plot does not reflect a trivial pattern. The normality assumption is assessed by creating a normal probability plot of the residuals. When the error has a normal distribution, this plot will appear as a straight line [10]. These assumptions were checked for PCR and PCNN, as presented in Figure 9. The assumption of equal variances does not appear to be violated because there are no trivial pattern in this plot. Figure 9 presents the normal probability of the residuals in which it can be seen that the data points are close to the straight line and the normality assumption is validated.

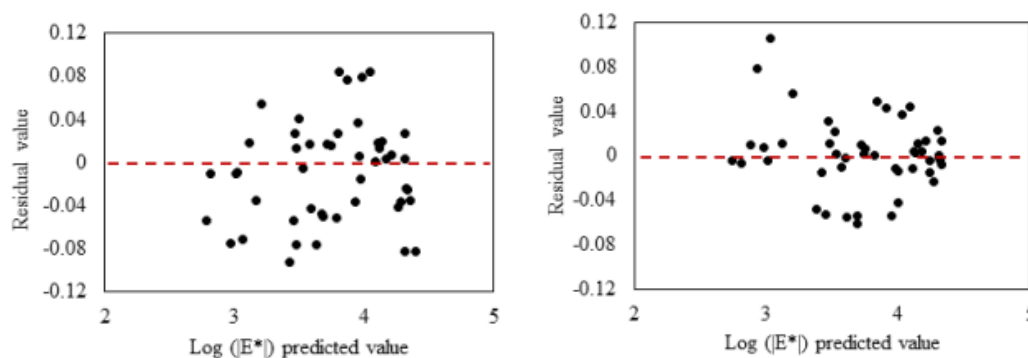


Figure 9. Cont.

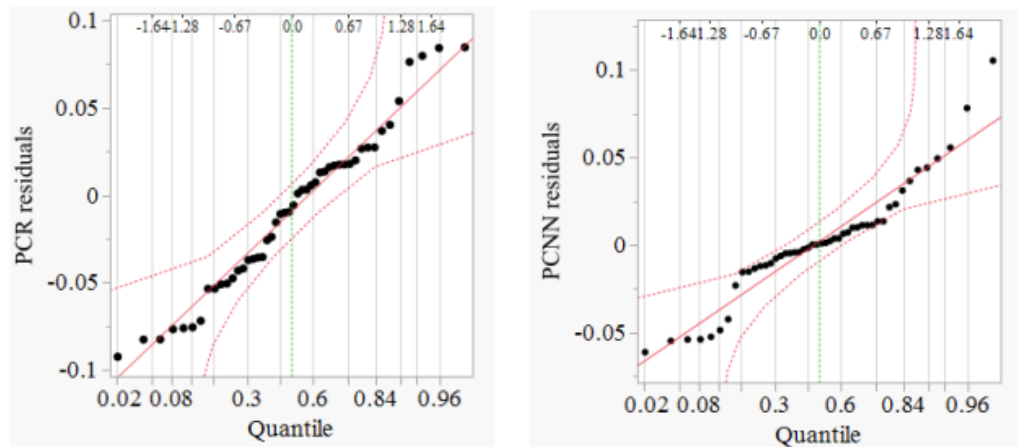


Figure 9. Checking the assumptions of independency using residual plot for PCR (left) and PCNN (right) models and the normality using normal probability plot of the residuals for PCR (left) and PCNN (right) models.

4. Application of the Framework: Flexible Pavement Design and Optimization

The above framework is used along with an optimization algorithm to answer the following two central questions:

- what design parameters result in the maximum $|E^*|$?
- what design parameters result in a pre-specified $|E_0^*|$?

One can see that the first item corresponds to the optimal design problem while the second one corresponds to the so-called inverse design.

Since it was shown through multiple statistical measurements that PCNN had the best prediction capability, this model is used in the following section to solve the optimization problems. The ANN used in PCNN is essentially an interconnected nonlinear function, and this necessitates the application of a global optimizer. Moreover, the effective variable space enters the problem as a series of constraints and further restricts the available algorithms. The optimal design problem is formulated as follows:

$$\begin{aligned}
 & \text{maximize} && |E^*| = F_{ANN}(\mathbf{x}) \\
 & \text{with respect to} && \mathbf{x} = (x_1, \dots, x_{14}) \\
 & \text{subject to} && (\mathbf{x} - \mathbf{v})^T A (\mathbf{x} - \mathbf{v}) \leq 1, \\
 & && (\mathbf{x}_{pca} - \mathbf{v}')^T A' (\mathbf{x}_{pca} - \mathbf{v}') \leq 1,
 \end{aligned} \tag{18}$$

where the vector of fourteen variables is \mathbf{x} , and $(\mathbf{x} - \mathbf{v})^T A (\mathbf{x} - \mathbf{v}) \leq 1$ are the enclosing ellipsoid constraint equations for the original and PCA-based variables. A penalty function approach is used to convert the above constrained problem to an unconstrained one [38]. In this case, when the penalty function is active, it decreases (increases) the objective function when the problem is one of maximization (minimization), and the degree of penalty is based on the closeness of the solution to the corresponding constraint.

Since the inverse design problem aims at finding the specification of a predefined goal, it is defined as a minimization problem as follows:

$$\begin{aligned}
 & \text{minimize} && \text{error} = ||E^*| - |E_0^*|| \\
 & \text{with respect to} && \mathbf{x} = (x_1, \dots, x_{14}) \\
 & \text{subject to} && (\mathbf{x} - \mathbf{v})^T A (\mathbf{x} - \mathbf{v}) \leq 1, \\
 & && (\mathbf{x}_{pca} - \mathbf{v}')^T A' (\mathbf{x}_{pca} - \mathbf{v}') \leq 1,
 \end{aligned} \tag{19}$$

where $|E_0^*|$ is the desired (goal) dynamic modulus. Although a similar penalization method can also be used to address the constraints in this case, for the above problem the constraints will penalize the objective function when they are active.

Reliable solution of the above problems requires the application of a gradient-free optimization algorithm. Gradient-based optimization algorithms are not applicable in this case because of the network-based nature of the ANNs. Evolutionary-based algorithms are potentially easy-to-use algorithms in the above problems. Novel algorithms have been used to solve complex optimization problems in recent years [39,40], and in this case, Mean-Variance Mapping Optimization (MVMO), an in-house optimization algorithm based on the work by Elrich et al. [41,42], is used. The constraints are handled using the approach described in Aslani et al. [43], in which, the convergence rate of a constrained MVMO was compared to the already-developed methodologies using benchmark structural problems. Authors in [22] indicated that a constrained MVMO is capable accurately identifying an optimal value with a minimum number of simulations. It should be noted that the choice of optimization algorithm is not the principal focus of this study.

Figure 10 (left) depicts the convergence achieved for the first design problem by the constrained MVMO algorithm. The initial data points are random making it heavily penalized, and then the objective function increases as the algorithms evolves. Exploration-exploitation behavior is achieved using adaptive strategies in the course of optimization for MVMO. $\delta = 0.05$ is used as the threshold in Figure 4. Solving the maximization problem resulted in $|E_{max}^*| = 53,703$ MPa. The optimal design parameters are presented in the first column of Table 7.

To find the maximum amount of dynamic modulus one could design for without low temperature failure in the asphalt binder, the maximization problem was solved one more time with an additional constraint of $G^* \sin \delta \leq 5000$, resulting in $|E_{max}^*| = 36,307$ MPa. Corresponding design parameters are presented in the second column of Table 7 as the optimal design 2.

Figure 10 (right) shows the convergence of the algorithm for the inverse design problem after starting randomly from three different initial points, with the algorithm is terminated when the error reaches about 10^{-9} . A pre-specified $|E_0^*|$ of 20,417 MPa is considered and the inverse problem of finding the corresponding design parameters is solved. Because of non-linearity of the function, the problem has no unique solution. Three of the possible solutions are presented as designs 1 to 3 in Table 7.

Finally, the five sets of design parameters are compared with current design specification, with the results shown in Table 7. The percentage of aggregate passing by each sieve size is within the acceptable range of the gradation specification. Gradation charts are presented in Figure 11. The obtained percentages of air voids are 4%, which is the target value in the design specification. The obtained values for VMA are slightly less than 14% for a nominal maximum aggregate size (NMAS) of 12.5 mm because the VMA values of the nine mixtures used to train the PCNN are slightly less than 14% (see Table 1). The acceptable range for VFA varies with the amount of traffic load measured in million Equivalent Single Axle Loads (ESALs) as follows:

- traffic loading $< 0.3 \rightarrow 70 < \text{VFA} < 80$
- $0.3 < \text{traffic loading} < 3.0 \rightarrow 65 < \text{VFA} < 78$
- traffic loading $> 3.0 \rightarrow 65 < \text{VFA} < 75$

The VFAs obtained for all of the five sets of design are satisfied for all of the traffic categories.

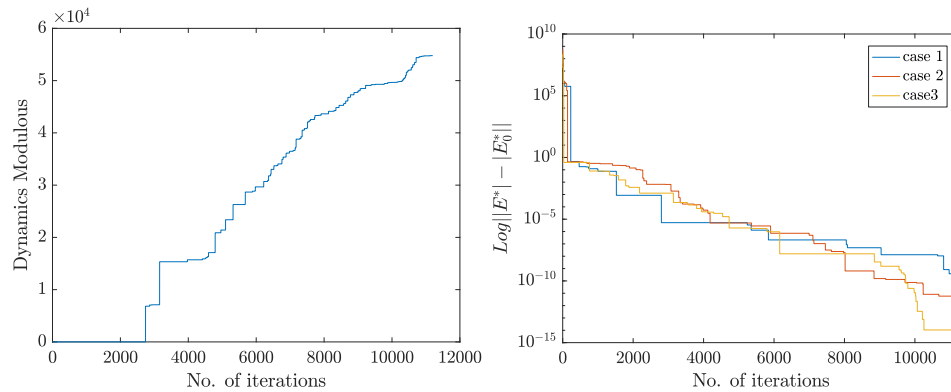


Figure 10. Convergence for the optimal design problem (left) and inverse (right) design of dynamic modulus.

Table 7. The Corresponding Design Parameters Obtained from Solving Optimal Design and Inverse Design Problems.

Identity	Optimal Design 1	Optimal Design 2	Design 1	Design 2	Design 3	Design Specification			
						Control Points		Restricted Zone	
						Lower	Upper	Lower	Upper
%Passing from 3/4"	100	100	100	100	100	-	100	-	-
%Passing from 1/2"	93.38	94.03	92.25	91.88	91.80	90	100	-	-
%Passing from 3/8"	81.74	81.72	79.57	79.92	80.70	-	90	-	-
%Passing from #4	53.00	53.90	55.36	55.23	54.39	-	-	-	-
%Passing from #8	39.56	40.51	41.37	41.08	40.92	28	58	39.1	39.1
%Passing from #30	20.75	20.68	21.02	20.87	20.83	-	-	19.1	23.1
%Passing from #50	11.66	11.60	12.08	11.81	12.02	-	-	15.5	15.5
%Passing from #100	6.22	6.21	6.52	6.38	6.40	-	-	-	-
%Passing from #200	4.10	3.85	4.38	4.58	4.56	2	10	-	-
G* (Mpa)	103.13	7.81	133.51	30.20	11.82	-	-	-	-
Phase angle (degree)	35.71	39.60	47.69	47.27	44.77	2	8	-	-
Vbeff%	4.11	4.18	4.02	4.06	4.05	-	-	-	-
VMA	13.47	13.56	13.41	13.45	13.44	-	-	-	-
VFA	70.29	70.50	70.11	70.24	70.24	-	-	-	-
Va%	4.00	4.00	3.99	4.00	4.01	4	-	-	-

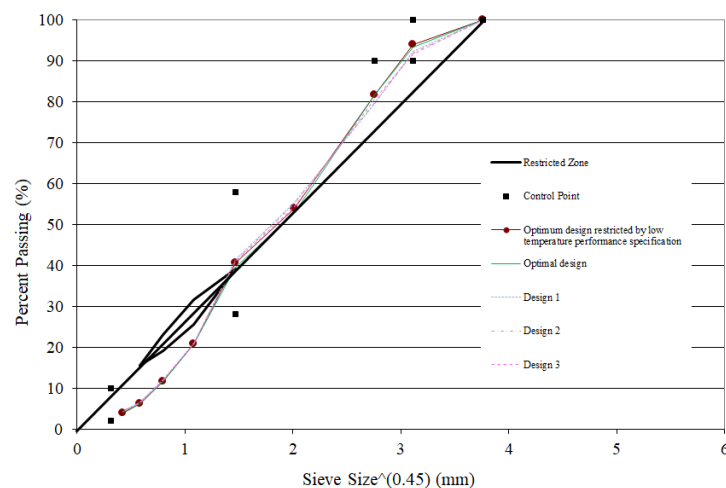


Figure 11. Aggregate gradation graphs with 12.5 mm NMAS particle size distribution obtained from PCNN.

5. Conclusions

This study used the HMA dynamic modulus data and focused to evaluate the quality of predictor variables to be used in a procedure of model development. Correlation analysis is performed to identify cross-correlated input variables, and correlated inputs are replaced by orthogonal pseudo-inputs (PCs)

obtained using PCA. Two separate models are developed using multivariate regression and ANN (called PCR and PCNN, respectively). Extrapolation in empirical modeling is addressed by adding the constraint of an n-dimensional enclosing ellipsoid to the modeling problem. Performances of the proposed models were compared to existing predictive models using both statistical analysis and ROC analysis. The models developed satisfactorily estimated the dynamic modulus value, with PCNN indicating remarkably better performance when fitted to the test data than the existing predictive models from the literature. These PCA-based approaches are thus highly recommended as precise modeling strategies in this application. Moreover, these methodologies appear to be capable of modeling other material properties and future investigation in this regard is recommended. To determine this framework's application in pavement design, two optimization problems including optimal design and inverse design have been presented and solved using a mean-variance mapping optimization algorithm. The results for the two problems are in a good agreement with the HMA mix design specification and thus could be a reasonable starting point in solving real-life design problems. Although, the developed models as well as obtained optimal design parameters are based on the empirical database created in this study, the suggested framework has the capability of being re-trained and adjusted to fit new data. For obtaining more reliable and applicable results, a larger empirical database would be required.

Author Contributions: Conceptualization, P.G. and M.A.; methodology, P.G., M.A., and D.K.R.; software, P.G. and M.A.; validation, P.G., M.A., D.K.R., and R.C.W. formal analysis, P.G., M.A., D.K.R., and R.C.W.; investigation, P.G., M.A., D.K.R., and R.C.W.; resources, P.G., M.A., D.K.R., and R.C.W.; data curation, P.G., M.A., D.K.R., and R.C.W.; writing—original draft preparation, P.G., M.A., D.K.R., and R.C.W.; writing—review and editing, P.G., M.A., D.K.R., and R.C.W.; visualization, P.G., M.A., D.K.R., and R.C.W.; supervision, P.G., M.A., D.K.R., and R.C.W.; project administration, R.D.W.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank anonymous reviewers for their fruitful comments on the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ghasemi, P. Performance Evaluation of Coarse-Graded Field Mixtures Using Dynamic Modulus Results Gained from Testing in Indirect Tension Mode of Testing. Graduate Theses and Dissertations 16717. Available online: <https://lib.dr.iastate.edu/etd/16717> (accessed on 1 August 2019).
2. Birgisson, B.; Roque, R.; Kim, J.; Pham, L.V. *The Use of Complex Modulus to Characterize the Performance of Asphalt Mixtures and Pavements in Florida*; Technical Report; Florida Department of Transportation: Tallahassee, FL, USA, 2004.
3. Arabali, P.; Sakhaeifar, M.S.; Freeman, T.J.; Wilson, B.T.; Borowiec, J.D. Decision-making guideline for preservation of flexible pavements in general aviation airport management. *J. Transp. Eng. Part B Pavements* **2017**, *143*, 04017006. [CrossRef]
4. Bozorgzad, A.; Lee, H.D. Consistent distribution of air voids and asphalt and random orientation of aggregates by flipping specimens during gyratory compaction process. *Constr. Build. Mater.* **2017**, *132*, 376–382. [CrossRef]
5. Guide, N.D. *Guide 1-37A, Guide for Mechanistic-Empirical Design of New and Rehabilitated Pavement Structures*, National Cooperative Highway Research Program; Transportation Research Board, National Research Council: Washington, DC, USA, 2004.
6. AASHTO, A. *Mechanistic-Empirical Pavement Design Guide: A Manual of Practice*; AAoSHaT Officials (American Association of State Highway and Transportation Officials): Washington, DC, USA, 2008.
7. Nobakht, M.; Sakhaeifar, M.S. Dynamic modulus and phase angle prediction of laboratory aged asphalt mixtures. *Constr. Build. Mater.* **2018**, *190*, 740–751. [CrossRef]
8. Peng, C.; Feng, J.; Feiting, S.; Changjun, Z.; Decheng, F. Modified two-phase micromechanical model and generalized self-consistent model for predicting dynamic modulus of asphalt concrete. *Constr. Build. Mater.* **2019**, *201*, 33–41. [CrossRef]

9. Shu, X.; Huang, B. Dynamic modulus prediction of HMA mixtures based on the viscoelastic micromechanical model. *J. Mater. Civ. Eng.* **2008**, *20*, 530–538. [\[CrossRef\]](#)
10. Devore, J.L. *Probability and Statistics for Engineering and the Sciences*; Cengage Learning: Boston, MA, USA, 2011.
11. El-Badawy, S.; Abd El-Hakim, R.; Awed, A. Comparing Artificial Neural Networks with Regression Models for Hot-Mix Asphalt Dynamic Modulus Prediction. *J. Mater. Civ. Eng.* **2018**, *30*, 04018128. [\[CrossRef\]](#)
12. Andrei, D.; Witczak, M.; Mirza, M. Development of a revised predictive model for the dynamic (complex) modulus of asphalt mixtures. In *Development of the 2002 Guide for the Design of New and Rehabilitated Pavement 451 Structures*; NCHRP: Washington, DC, USA, 1999.
13. Bari, J.; Witczak, M. New predictive models for viscosity and complex shear modulus of asphalt binders: For use with mechanistic-empirical pavement design guide. *Transp. Res. Rec. J. Transp. Res. Board* **2007**, *2001*, 9–19. [\[CrossRef\]](#)
14. Christensen, D., Jr.; Pellinen, T.; Bonaquist, R. Hirsch model for estimating the modulus of asphalt concrete. *J. Assoc. Asph. Paving Technol.* **2003**, *72*, 97–121.
15. Jamrah, A.; Kutay, M.E.; Ozturk, H.I. *Characterization of Asphalt Materials Common to Michigan in Support of the Implementation of the Mechanistic-Empirical Pavement Design Guide*; Technical Report; Transportation Research Board: Washington, DC, USA, 2014.
16. Al-Khateeb, G.; Shenoy, A.; Gibson, N.; Harman, T. A new simplistic model for dynamic modulus predictions of asphalt paving mixtures. *J. Assoc. Asph. Paving Technol.* **2006**, *75*, 1254–1293.
17. Sakhaeifar, M.S.; Richard Kim, Y.; Garcia Montano, B.E. Individual temperature based models for nondestructive evaluation of complex moduli in asphalt concrete. *Constr. Build. Mater.* **2017**, *137*, 117–127. [\[CrossRef\]](#)
18. Ghasemi, P.; Aslani, M.; Rollins, D.K.; Williams, R.C.; Schaefer, V.R. Modeling rutting susceptibility of asphalt pavement using principal component pseudo inputs in regression and neural networks. *Int. J. Pavement Res. Technol.* **2018**. [\[CrossRef\]](#)
19. Ren, R.; Han, K.; Zhao, P.; Shi, J.; Zhao, L.; Gao, D.; Zhang, Z.; Yang, Z. Identification of asphalt fingerprints based on ATR-FTIR spectroscopy and principal component-linear discriminant analysis. *Constr. Build. Mater.* **2019**, *198*, 662–668. [\[CrossRef\]](#)
20. Fodor, I.K. *A Survey of Dimension Reduction Techniques*; Technical Report; U.S. Department of Energy: Washington, DC, USA, 2002.
21. Johnson, R.A.; Wichern, D.W. *Applied Multivariate Statistical Analysis*; Prentice-Hall: Upper Saddle River, NJ, USA, 2014; Volume 4.
22. Ghasemi, P.; Aslani, M.; Rollins, D.K.; Williams, R. Principal component analysis-based predictive modeling and optimization of permanent deformation in asphalt pavement: Elimination of correlated inputs and extrapolation in modeling. *Struct. Multidiscip. Optim.* **2018**, *59*, 1335–1353. [\[CrossRef\]](#)
23. Kim, Y.R.; Underwood, B.; Far, M.S.; Jackson, N.; Puccinelli, J. *LTPP Computed Parameter: Dynamic Modulus*; Technical Report; Federal Highway Administration Research and Technology: Washington, DC, USA, 2011.
24. Sakhaeifar, M.S.; Richard Kim, Y.; Kabir, P. New predictive models for the dynamic modulus of hot mix asphalt. *Constr. Build. Mater.* **2015**, *76*, 221–231. [\[CrossRef\]](#)
25. Rollins, D.K.; Zhai, D.; Joe, A.L.; Guidarelli, J.W.; Murarka, A.; Gonzalez, R. A novel data mining method to identify assay-specific signatures in functional genomic studies. *BMC Bioinf.* **2006**, *7*, 377. [\[CrossRef\]](#)
26. Jolliffe, I.T. Principal Component Analysis, Second Edition. *Encycl. Stat. Behav. Sci.* **2002**, *30*, 487.
27. Kuźniar, K.; Waszczyszyn, Z. Neural networks and principal component analysis for identification of building natural periods. *J. Comput. Civ. Eng.* **2006**, *20*, 431–436. [\[CrossRef\]](#)
28. Hua, X.; Ni, Y.; Ko, J.; Wong, K. Modeling of temperature–frequency correlation using combined principal component analysis and support vector regression technique. *J. Comput. Civ. Eng.* **2007**, *21*, 122–135. [\[CrossRef\]](#)
29. Kartam, N. Neural Networks in Civil Engineering: Systems and Application. *J. Comput. Civ. Eng.* **1994**, *8*, 149–162.
30. Sanabria, N.; Valentin, V.; Bogus, S.; Zhang, G.; Kalhor, E. *Comparing Neural Networks and Ordered Probit Models for Forecasting Pavement Condition in New Mexico*; Technical Report; Transportation Research Board: Washington, DC, USA, 2017.

31. Gong, H.; Sun, Y.; Mei, Z.; Huang, B. Improving accuracy of rutting prediction for mechanistic-empirical pavement design guide with deep neural networks. *Constr. Build. Mater.* **2018**, *190*, 710–718. [\[CrossRef\]](#)
32. Cheng, B.; Titterton, D. Neural networks: A review from a statistical perspective. *Stat. Sci.* **1994**, *9*, 2–30. [\[CrossRef\]](#)
33. Kutner, M.H.; Nachtsheim, C.; Neter, J. *Applied Linear Regression Models*; McGraw-Hill/Irwin: Chicago, IL, USA, 2004.
34. Todd, M.J.; Yildirim, E.A. On Khachiyan's algorithm for the computation of minimum-volume enclosing ellipsoids. *Discret. Appl. Math.* **2007**, *155*, 1731–1744. [\[CrossRef\]](#)
35. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [\[CrossRef\]](#)
36. Bi, J.; Bennett, K.P. Regression error characteristic curves. In Proceedings of the 20th International Conference on Machine Learning (ICML-03), Washington, DC, USA, 21–24 August 2003; pp. 43–50.
37. Marti-Vargas, J.R.; Ferri, F.J.; Yepes, V. Prediction of the transfer length of prestressing strands with neural networks. *Comput. Concr.* **2013**, *12*, 187–209. [\[CrossRef\]](#)
38. Rahami, H.; Kaveh, A.; Aslani, M.; Asl, R.N. A hybrid modified genetic-nelder mead simplex algorithm for large-scale truss optimization. *Int. J. Optim. Civ. Eng.* **2011**, *1*, 29–46.
39. He, J.; Yao, X. From an individual to a population: An analysis of the first hitting time of population-based evolutionary algorithms. *IEEE Trans. Evol. Comput.* **2002**, *6*, 495–511.
40. Cai, Z.; Cai, Z.; Wang, Y. A Multiobjective Optimization-Based Evolutionary Algorithm for Constrained Optimization. *IEEE Trans. Evol. Comput.* **2006**, *10*, 658–675. [\[CrossRef\]](#)
41. Rueda, J.L.; Erlich, I. Testing MVMO on learning-based real-parameter single objective benchmark optimization problems. In Proceedings of the 2015 IEEE Congress on Evolutionary Computation (CEC), Sendai, Japan, 25–28 May 2015; pp. 1025–1032.
42. Erlich, I.; Venayagamoorthy, G.K.; Worawat, N. A mean-variance optimization algorithm. In Proceedings of the IEEE Congress on Evolutionary Computation, Barcelona, Spain, 18–23 July 2010; pp. 1–6.
43. Aslani, M.; Ghasemi, P.; Gandomi, A.H. Constrained mean-variance mapping optimization for truss optimization problems. *Struct. Des. Tall Spec. Build.* **2018**, *27*, e1449. [\[CrossRef\]](#)



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).