

RESEARCH

Open Access



Classification of pavement climatic regions through unsupervised and supervised machine learnings

Qiao Dong^{1,2*} , Xueqin Chen³, Shi Dong⁴ and Jun Zhang⁵

Abstract

This study extracted 16 climatic data variables including annual temperature, freeze thaw, precipitation, and snowfall conditions from the Long-term Pavement Performance (LTPP) program database to evaluate the climatic regionalization for pavement infrastructure. The effect and significance of climate change were firstly evaluated using time as the only predictor and t-test. It was found that both the temperature and humidity increased in most States. Around one third of the 800 weather stations record variation of freeze and precipitation classifications and a few of them show significant change of classifications over time based on the results of logistic regression analyses. Three unsupervised machine learning including Principle Component Analysis (PCA), factor analysis and cluster analysis were conducted to identify the main component and common factors for climatic variables, and then to classify datasets into different groups. Then, two supervised machine learning methods including Fisher's discriminant analysis and Artificial Neural Networks (ANN) were adopted to predict the climatic regions based on climatic data. Results of PCA and factor analysis show that temperature and humidity are the first two principle components and common factors, accounting for 71.6% of the variance. The 4-means clusters include wet no freeze, dry no freeze, dry freeze and snow freeze. The best k-mean clustering suggested 9 clusters with more temperature clusters. Both the Fisher's linear discriminant analysis and ANN can effectively predict climatic regions with multiple climatic variables. ANN performs better with higher R square and low misclassification rate, especially for those with more layers and nodes.

Keywords: Climatic regions, Climate change, Machine learning, Principle component analysis (PCA), Factor analysis, K-means cluster analysis, Fisher's discriminant analysis, Artificial neural networks (ANN)

Introduction

Background

Climatic factors such as temperature and moisture have significant influence on the deterioration of both pavement structural capacity and pavement materials [1], and are key factors for pavement preservation and resilience analysis. Many countries have developed their own climatic region classifications for determining asphalt

binder grade, including the USA [2], China [3], Jordan [4], Italy [5], Thailand [6], Iran [7] and Yemen [8]. Not only are asphalt binder grade selected based on climatic regions, climatic factors are also critical for both flexural and rigid pavement design [9]. Yang et al. conducted a sensitivity analysis of the influence of climatic inputs on pavement distress development using the Mechanistic-Empirical Pavement Design Guide (MEPDG) software [10]. The MEPDG includes the Enhanced Integrated Climate Model (EICM) with historical hourly data from around 800 weather stations to model future climatic conditions for pavement performance prediction [11]. Basma et al. found that structural number of flexural pavement need to be adjusted to offset a reduction in

* Correspondence: qiaodong@seu.edu.cn

¹School of Transportation, Southeast University, 2 Southeast University Road, Nanjing 211189, Jiangsu Province, China

²National Demonstration Center for Experimental Road and Traffic Engineering Education (Southeast University), Nanjing 211189, Jiangsu Province, China

Full list of author information is available at the end of the article

the subgrade resilient modulus due to an increase in moisture content [12]. For pavement maintenance and preservation, Wang et al. found the same pavement preservation treatments used in different climatic regions performed significantly different based on the LTPP SPS-3 data [13]. Different pavement preservation strategies or techniques should be considered for different climatic regions.

Development of climatic regions

The climatic regions are usually determined based on the maximum and minimum temperature and rainfall. Geographic and environmental factors may also be considered. In the Long-term Pavement Performance (LTPP) program, the annual total amount of precipitation and the freezing index [14–16] are used to divide the United States into four climatic regions, including dry no Freeze, dry freeze, wet no Freeze and wet freeze. However, there are still several challenges for climatic regionalization for pavement infrastructure. Firstly, the number of climatic regions may need to be increased to better quantify the effects of climatic factors. For example, Wang et al. used 6 climatic regions by adding dry mild and wet mild to the original four climatic regions [13]. Bandara et al. classified the I-94 corridor in Michigan into four climatic regions [11]. Many States have microclimates by considering different geographical or environmental conditions and adding more weather stations or expanding the number of months of available data [11, 17].

Secondly, climatic factors include temperature, humidity, rainfall, snowfall, etc. and how to balance and consider all of those different factors is of importance. Bandara et al. used the average low temperatures in January, average high temperatures in July, average precipitation in January and average precipitation in July to classify the I-94 corridor in Michigan into 4 climatic regions [11]. Wang et al. used the number of days below 0 °C, the number of wet days, and the freeze–thaw cycles to classify climatic regions [13]. In agriculture studies, the Köppen climate classification considers only rainfall and temperature were usually used to determine climatic regions including tropical, arid, temperate, continental, and polar [18]. However, to further improve the accuracy and effects of climatic regionalization, more detailed information is required. It has been reported that the potential water balance of the soil over the growing cycle, heliothermal conditions over the growing cycle and night temperature during maturation have been used to build a multiple criteria climatic classification system for the grape-growing regions [19].

Moreover, the climatic changes poses threatens to transportation infrastructure, which may change the climatic regions, especially for those at the margins of

regions. Mills et al. evaluated the impact of climate change on flexible pavement design and performance based on 17 weather sites and found that low temperature cracking will be less problematic while rutting may cause earlier rehabilitation and reconstruction in Southern Canada [20]. Gudipudi et al. used 19 climate models to project future climatic and analyzed the impact of climatic change on pavement performance. It was found that projected climate changes are likely to cause greater distresses and/or earlier failure of the pavement including 2–9% more fatigue cracking and 9–40% more rutting at the end of 20 years [21, 22]. A rational procedure would be necessary to adjust current climatic regionalization based on observed or predicted climatic dataset.

Machine Learnings in climatic regionalization

The weather stations collect more detailed long-term climatic data that can be used to improve the climatic regionalization for pavement infrastructure. There have been several studies on the climatic regionalization of pavements based on the data collected from weather stations. Recently, the machine learning has been applied in many fields to identify relationships between variables and to provide highly accurate prediction or classification base on massive data samples and therefore has the potential for pavement climatic regionalization analysis. Yang et al. [23] adopted Principle Component Analysis (PCA) to identify three major factors including temperature, precipitation, and radiation for climatic regionalization of pavements and then the k-means cluster analysis to classify pavement climatic regions. The probabilistic neural network and Support Vector Machine (SVM) were also used to predict pavement climate regions and fairly high accuracy were obtained. There are mainly three types of machine learning algorithms, unsupervised learnings, supervised learnings and reinforced learnings. The unsupervised learnings identify key components or factors and classify unlabeled data based on their correlations while supervised learnings classify data through minimizing the misclassification or error of a model trained by labeled data. Therefore, unsupervised learnings can be used to find the optimal classification while supervised learnings can be used to predict the classification.

The reinforced learnings are to determine the actions in an environment to maximize the cumulative reward and are usually used for optimization in operation research.

In climatic regionalization studies, k-means and hierarchical clustering have been used to redefine the climate zones of Turkey based on temperature and precipitation data collected from 113 climate stations [24], to determine the climate regions in Argentine [25],

to determine rainfall regions in India [26], to identify regional climate change patterns [27], and to divide the European domain into regions of similar projected climate changes using predicted total temperature and precipitations [28]. PCA is usually used to identify key climatic components that can be used as criteria to determine climatic regions [29, 30]. One study reported that the annual variation in mean and minimum temperature, annual maximum temperature, and spring, summer, and fall precipitation are the five principle components for the climate regionalization in Puerto Rico [31]. Factor analysis can be used to identify the common factors for climatic data. The temperature, winter moisture and moisture factors, explaining 46%, 32% and 12% of the total variance, were identified based on factor analysis for the climatic regionalization of the Tibet, China [32]. ANN has been adopted to predict climatic regions in South America [33] and Puerto Rico [31]. A supervised classification with Mahalanobis distance was used to classify climate regions in China based on the data collected from 172 stations between 1984 and 2013 [34]. One study reported delineation of high resolution climate regions in Korean peninsula using the ANN, random forest, k-nearest neighbor, logistic regression, and SVM supervised learnings [35]. Both unsupervised and supervised learnings were adopted to delineate homogeneous climatic regions in Pakistan.

Objectives and scope

The LTPP recommends temperature and freeze data to classify climatic regions while neglect other information including raining days, sub-zero days, etc. Further, a long term climatic data are needed to determine the climatic regionalization due to the high variation of temperature and freeze condition in areas around the borders of the climatic regions. As the collection and accumulation of more detailed climatic data, it is interesting to include those detailed long-term climatic data to help determine climatic regionalization. The objectives of this study are to determine the main contribution factors of climatic data collected from the LTPP weather stations, to classify climatic regions through unsupervised machine learning methods, and to predict climatic regions through supervised machine learning methods. The general trend of climate change over time was evaluated and the significance was tested through linear regression and parameter t-tests. The unsupervised learning methods includes PCA, factor analysis and k-means cluster analysis. The supervised learning methods includes the Fisher's linear discriminant analysis and the Artificial Neural Network (ANN).

Data collection

Established in the 1980s, the LTPP has been collecting large quantities of pavement data from more than 2400

pavement sections in the USA and Canada. The LTPP collects climate data from the weather stations located near the test sections. The detailed hourly data are available and the daily, monthly and annual statistics such as maxim, minimum, average are calculated and stored in the LTPP. In this study, 21,666 annual climate data from 1948 to 2012 were collected from 800 weather stations in 62 States in the US, Canada etc. Table 1 summarizes the definitions and statistical descriptions of those data. Sixteen variables were collected, including temperature, humidity, precipitation, snowfall and freezing conditions. In the LTPP program, the wet/dry threshold is average annual precipitation of 508 mm and the freeze/no-freeze threshold is average annual freezing index 83.3 degree-Celsius days [36].

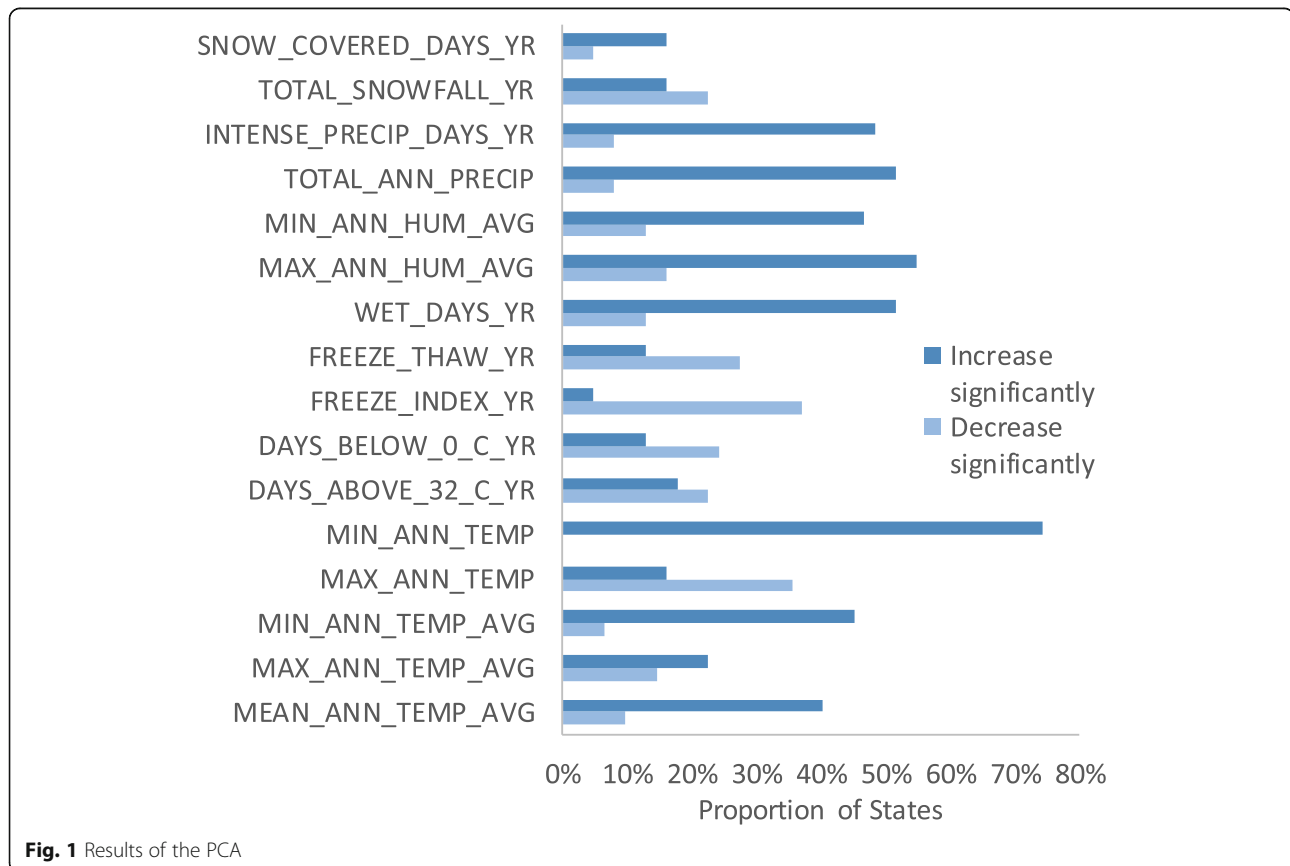
Climate change

Figure 1 shows the effects and significance of time on those climatic variables using time as the only predictor for simple linear regression. Positive and negative parameter estimates indicate increasing and decreasing trends, respectively. *P*-values lower than 0.05 are regarded as significant [37]. The horizontal axle is the proportion of States over the total 62 States. It can be seen that the mean, maximum and minimum of the average temperature in most States increase significantly, while the freeze index and freeze thaw cycles decrease significantly in the last 60 years with *p*-values less than 0.05. The minimum temperature in nearly 80% of the States increase significantly with time. Therefore, the global warming has been significant, as investigated in previous studies using historical data or climate models [21]. It is also noted that around 50% States show significant increasing precipitation and humidity while only 10–20% States show significant decreasing trend. The rising temperatures intensify the Earth's water cycle and increase evaporation, causing increased precipitation and flooding in some area close to the storm tracks. Satellite observations have found increased precipitation and total atmospheric water due to the increase of surface warming [38–40].

Figure 2 (a) shows the map of the climatic regions defined in LTPP. A general classification of each State is recorded in the LTPP database. However, it should be noted that the weather stations in one State may have different freeze or precipitation classifications. Actually, even for the same weather station, the freeze and precipitation classifications may change at different years. The freeze and precipitation classifications of each weather station at each year was calculated and the variation of freeze and precipitation classifications can be determined based on the criteria. For the 800 weather stations, 523 (65%) of them weather stations maintain the same freeze index classification and 564 (70%) of them

Table 1 Statistical description of the LTPP climatic data

Variables	Definition	Max.	Min.	Ave.	Std.
MEAN_ANN_TEMP_AVG	Average of daily mean air temperatures for a year (°C)	25.9	−4.1	13.3	4.7
MAX_ANN_TEMP_AVG	Average of daily maximum air temperatures for a year (°C)	32.6	1.8	19.8	5.0
MIN_ANN_TEMP_AVG	Average of daily minimum air temperatures for a year (°C)	22.1	−10	6.9	4.7
MAX_ANN_TEMP	Absolute maximum air temperature for a year (°C)	52.2	21.4	37.1	3.3
MIN_ANN_TEMP	Absolute minimum air temperature for year (°C)	10.1	−69.1	−16.8	9.3
DAYS_ABOVE_32_C_YR	Number of days when daily maximum air temperature was above 32.2 °C for a year	197	0	44.4	40.4
DAYS_BELOW_0_C_YR	Number of days when daily minimum air temperature was below 0 °C for a year	271	0	94.6	56.0
FREEZE_INDEX_YR	Sum of negative daily temperatures lower than 0 °C for a year	3369	0	252.3	326.4
FREEZE_THAW_YR	Number of freeze–thaw cycles for year	236	0	74.2	41.0
MAX_ANN_HUM_AVG	Average of daily maximum humidity for a year (%)	100	0	84.7	14.5
MIN_ANN_HUM_AVG	Average of daily minimum humidity for a year (%)	84	0	44.7	11.0
TOTAL_ANN_PRECIP	Total annual precipitation (mm)	3658.1	0	909.8	431.6
INTENSE_PRECIP_DAYS_YR	Number of days precipitation was greater than 12.7 mm for a year	84	0	21.4	12.1
WET_DAYS_YR	Number of days precipitation was greater than 0.25 mm for a year	261	0	124.3	40.8
TOTAL_SNOWFALL_YR	Total snowfall for a year (mm)	10,674	0	525.4	693.2
SNOW_COVERED_DAYS_YR	Number of days for which snow cover data were available for year	184	0	1.6	11.3

**Fig. 1** Results of the PCA

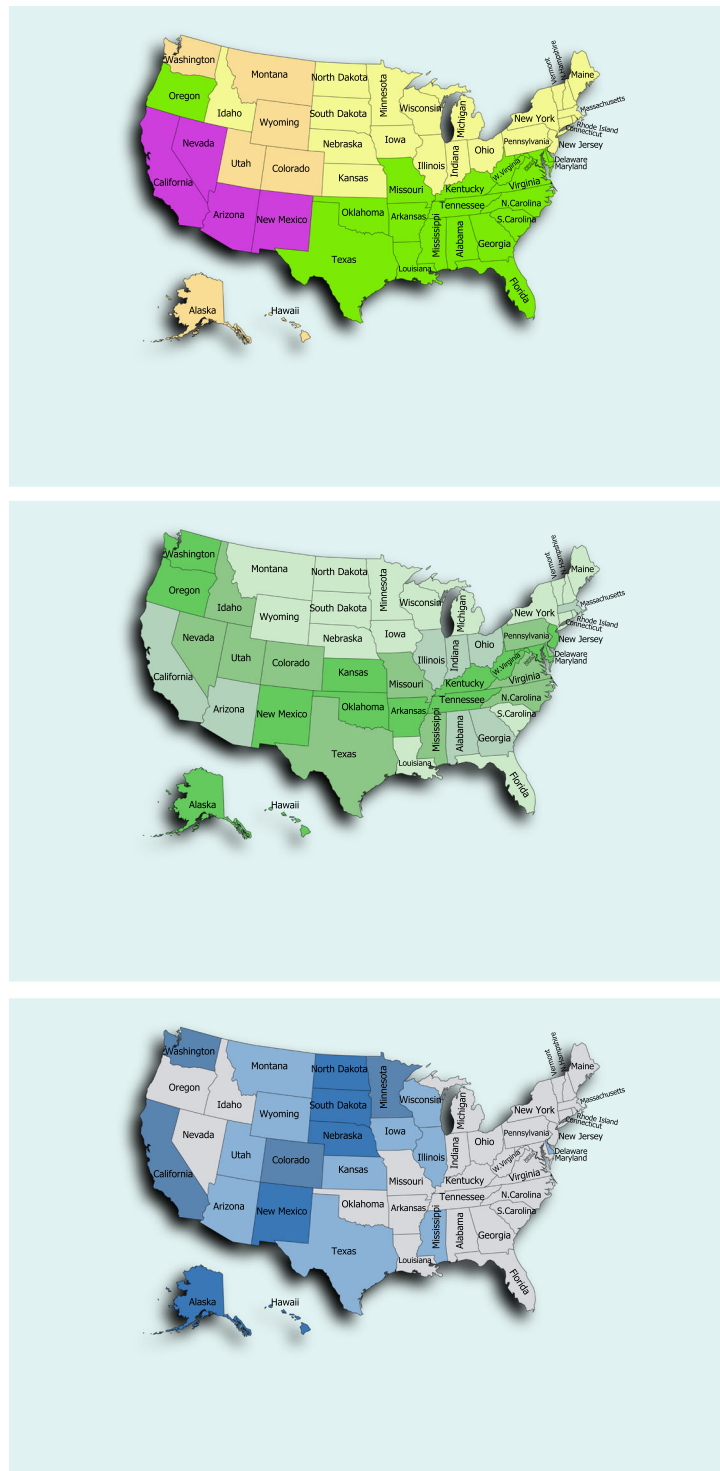


Fig. 2 Climatic regions and its variations. **a** Original four climatic regions in the LTPP. **b** Variation of freeze classifications. **c** Variation of precipitation classifications

maintain the same precipitation classification over the past 70 years. Figure 2 (b) and (c) show the variation of freeze and precipitation classifications, in which darker

color indicate high percentage of weather stations in this State recorded varying freeze and precipitation classifications. It can be seen from Fig. 2 (b) that the States at the

borders of north and south have high variations of freeze classifications.

To investigate if time is a significant factor for the probability of a weather station is freeze/no freeze or wet/dry, the logistic regression analysis using the freeze or precipitation classification as the target and time as the predictor was adopted to build the models for each of the weather station. Five hundred-thirteen logistic regression models were built and the parameter estimates of time as well as its P -value were obtained. Among the 277 (35%) weather stations recording variations of the freeze classifications, 196 weather stations shows increasing temperature and 16 of them are significant, while 81 shows decreasing temperature and none of them is significant. Among the 236 (30%) weather stations recording variations of the precipitation classification, 140 weather stations shows increasing precipitation and 7 of them are significant, while 96 shows decreasing precipitation and none of them is significant.

Methodology for classification

PCA

PCA is to convert a set of potentially correlated variables into a set of linearly uncorrelated variables. Each of the new variable is called principal component and is a linear combination of the original variables. As shown in Eq. (1), the first principal component F_1 is the linear combination of x_1, x_2, \dots, x_p that has maximum variance among all linear combinations and accounts for as much variation in the data as possible. The second principal component F_2 is the linear combination of x_1, x_2, \dots, x_p that accounts for as much of the remaining variation as possible, with the constraint that the correlation between F_1 and F_2 is 0. The third principal component F_3 is the linear combination of x_1, x_2, \dots, x_p that accounts for as much of the remaining variation as possible, with the constraint that the correlations between F_3, F_1 and F_2 are 0, and so on. a_{ij} is the loading coefficients of x_i on F_j , indicating the correlation of x_j on F_i . Either the covariance matrix or the correlation matrix of the variables can be used to calculate the components from their respective eigenvectors.

$$\begin{cases} F_1 = a_{11}x_1 + a_{21}x_2 + \dots + a_{p1}x_p \\ F_2 = a_{12}x_1 + a_{22}x_2 + \dots + a_{p2}x_p \\ \dots \\ F_p = a_{1p}x_1 + a_{2p}x_2 + \dots + a_{pp}x_p \end{cases} \quad (1)$$

The first several principal components can explain the major variation of the original dataset, and therefore can be used instead of the original dataset to reduce the dimensionality of a data set. In pavement engineering, PCA has been used to reduce the dimensionality of

dataset. Ghasemi used 5 principle components explaining 89.72% of the total variance of the original 17 asphalt mixture properties variables as the inputs for an ANN model predicting pavement permanent deformation [41]. Yao et al. reduced 21 traffic variables into 3 principle components for the pavement performance prediction [42]. In this study, PCA based on correlation matrix was firstly used to investigate the main components of the 16 climatic variables.

Factor analysis

Factor analysis has been widely used in psychology, sociology and economic studies to find the lower number of unobserved factors that can explain the variability among correlated variables. As shown in Eq. (2), each variable is a linear combination of common factors and an error term. μ is the average or intercept. a_{ij} is the factor loadings, indicating the contribution of common factors on the variance of the variable. f_1, f_2, \dots, f_m ($m \leq p$) are uncorrelated common factors. Factor analysis can be performed based on the orthogonal rotation technique of PCA or maximum likelihood method.

$$X_i = \mu + a_{i1}f_1 + a_{i2}f_2 + \dots + a_{im}f_m + \varepsilon_i \quad (2)$$

Factor analysis can be used to identify the common factors and to quantify the relationship between observed variables and the unobserved indicators. In pavement engineering, factor analysis has been adopted to evaluate the key factors of mixture properties and pavement performance. Tian et al. analyzed the 27 properties of asphalt mixture and find three common factors, including the permanent deformation factor, the shear resistance factor, and the moisture susceptibility factor [43]. Chen et al. used both single factor and multiple factor analysis to analyze the contributions of pavement performance measurements on the latent pavement performance factors including the roughness factor, the early age cracking factor and the aged severe damage factor [44]. In this study, factor analysis based on the principal method was used to identify the major common factors of the 16 climatic variables.

Cluster analysis

Cluster analysis is a widely used unsupervised machine learning method to classify data samples or variables into different groups based on their similarity. Distance metrics such as the Minkowski distance, Block distance and Euclidean distance are usually used to measure the similarity between samples. K-means clustering is the most frequent used cluster algorithm classifying n sample into k clusters based on the distance. As shown in Eq. (3), it uses selected k centroids as the beginning

points, and then performs iterative calculations to optimize the positions of the centroids by minimizing the distances within each cluster. The Cubic Clustering Criterion (CCC) can be used to estimate the number of clusters using k-means based on minimizing the within-cluster sum of squares through Monte Carlo methods. High CCC indicates good clustering. K-means cluster method has been used for pavement performance evaluation and pavement automatic evaluation data process. Wang et al. used a normalized cuts clustering to classify 35 pavement sections with 8 performance indicators into 5 clusters with different performance levels [45]. Li et al. used k-means clustering to identify the potential dipping in the groove measurement with laser profiling data [46]. In this study, k-means clustering was used to classify the 21,666 samples into different climatic regions.

$$d_{ij} = \min(\|x_i - z_j\|), x_i \in S, z_j \in Z \quad (3)$$

Discriminant analysis

The discriminant analysis is to classify samples into different groups based on its multiple characteristics. Different with cluster analysis which is an unsupervised learning, the discriminant analysis is a supervised machine learning and needs labeled classification. Frequently adopted discriminant algorithms include Bayesian discriminant, linear discriminant, etc. The linear discriminant is developed by Fisher in 1937 and is also called the Fisher discriminant. It uses a discriminant function maximizing the sum of squares between different groups and minimizing the sum of squares within a group. In 1987, Chou et al. built a discriminant model trained by historical data for pavement maintenance decision making [47]. A z value can be obtained from the model to determine if the pavement section needs an overlay treatment.

Ann

ANN is the most popular supervised learning algorithm for prediction and classification. In the ANN, the weights of nodes are trained during learning and an activation function is applied to the sum of weighted inputs to calculate outputs. The layers in ANN perform different transformations, enabling complicated non-linear calculation. A Deep Neural Network (DNN) is a type of ANN including multiple hidden layers and therefore can model very complex non-linear relationships. The training of the ANN is to find a set of weights that minimize the predictive error and the backpropagation is the most common training algorithm. ANN has already been extensively used in pavement material properties prediction and pavement performance modeling. Hussan

utilized nonlinear regression and ANN to predict rutting test results of asphalt mixture based on temperature, aggregate source, aggregate gradation, bitumen penetration values, and number of loading cycles [48]. Yao et al. used an DNN with two hidden layers and 64 nodes to predict pavement performance with 37 inputs [42].

Discussion of results

PCA

Figure 3 shows the PCA results for the 16 variables based on correlations. Figure 3 (a) is the scree plot showing the eigenvalue corresponding to each principal component in order from largest to smallest. The eigenvalues for the first two components are 7.7, and 3.8, respectively. Figure 3 (b) shows the portions of each component on the total variation and are scaled to sum to the number of variables. The first two components account for 47.5% and 24.1% of the total variance, respectively. It is rational to use the first two components to represent all of the 16 variables, accounting for 71.6% of the total variance.

Table 2 shows the loading matrix for the first five components. The i column of loadings is the i th eigenvector multiplied by the square root of the i th eigenvalue. Each component is the weight sum of the 16 variables with loadings as the weighting coefficients. High loading value indicates high correlation between the variable and the component. The loading values higher than 0.5 were bolded for better illustration. It can be seen that the first component is mainly related to the five temperature factors and four freeze condition factors. The second component is mainly related to the five precipitation and humidity factors. The third component is mainly related to two humidity factors, and the fourth is mainly related to snow covered days. The rest components have much less correlations with all the variables. Therefore, the temperature and humidity components can be used to represent the 16 climatic factors.

Factor analysis

Table 3 shows the rotated loading matrix of the first two factors based on the orthogonal rotation technique of PCA. The scree plot and the proportion of eigenvalues are the same as in Fig. 1. Each of the 16 variable can be expressed as the weight sum of the two common factors which explaining 71.6% of the total variance. It can be seen that the first common factor is the temperature factor and the 10 temperature related variables have large loading values. The second common factor is the humidity factor and the six humidity related variables have large loading values. It is noted that the snowfall and snow covered days are more related to the temperature factor. Further, the maximum annual temperature is related to both the temperature and the humidity factors

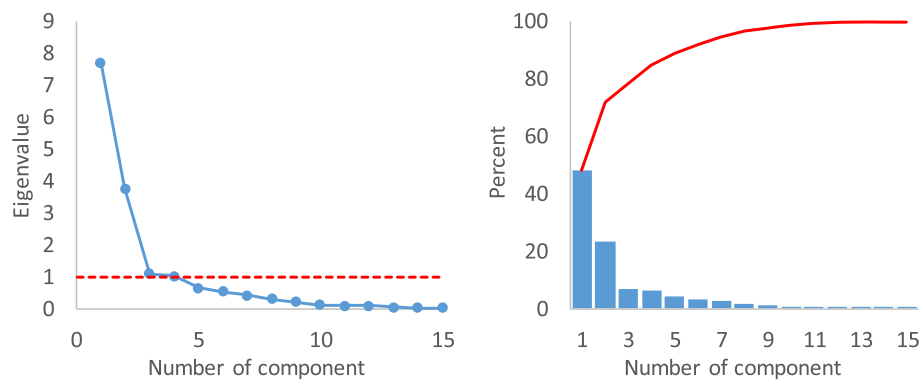


Fig. 3 Results of the PCA. **a** Scree plot. **b** Proportion of eigenvalues

and the loading value for humidity factor is negative (-0.61), indicating that the higher maximum annual temperature is usually related to lower precipitation level.

Cluster analysis

Firstly, the k-means clustering with four clusters were performed. Figure 6 shows the average of each of the 16 variables for the four clusters. When we classify all the samples into four groups, those are the center points that could achieve the minimum within cluster sum of squares. Based on temperature, precipitation and snowfall, we can estimate from Fig. 6 that cluster 1 is wet no freeze, cluster 2 is dry no freeze, cluster 3 is dry freeze and cluster 4 is snow freeze. The major difference between cluster 3 and 4 is not precipitation but the

snowfall. They are not exactly as the original four climatic regions defined by the LTPP.

Different number of clusters were also performed and the CCC values were shown in Fig. 4. It can be seen that the highest CCC is achieved at 9-mean clusters. Figure 5 shows the distribution of all the 21,666 sample for both 4-mean and 9-mean clustering, and the original four climatic regions defined by the LTPP in the coordinates of the first two principle components. The horizontal and vertical axes are the first and second principle components, representing temperature and humidity, respectively. For the 4-means clustering, it can be clearly seen from Fig. 5 (a) that cluster 1 and 2 are in high temperature region with high and low humidity, respectively. Cluster 3 and 4 are in low temperature region and cluster 4 has even lower temperature and higher

Table 2 Loading matrix of the first five components

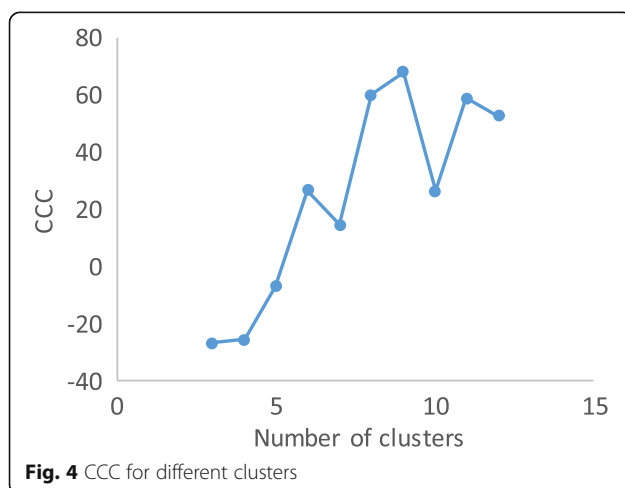
Variables	Prin1	Prin2	Prin3	Prin4	Prin5
MEAN_ANN_TEMP_AVG	0.978399	0.106364	-0.003164	0.042723	0.015729
MAX_ANN_TEMP_AVG	0.972456	-0.037692	0.003979	0.016051	0.083788
MIN_ANN_TEMP_AVG	0.940365	0.255010	-0.011688	0.068552	-0.058891
MAX_ANN_TEMP	0.661642	-0.509632	0.160909	0.017871	0.122109
MIN_ANN_TEMP	0.898253	0.167616	-0.044046	0.045169	0.013632
DAYS_ABOVE_32_C_YR	0.813625	-0.329665	0.068207	0.187960	-0.047039
DAYS_BELOW_0_C_YR	-0.935344	-0.236707	0.004261	-0.102506	0.142964
FREEZE_INDEX_YR	-0.824088	-0.103101	0.094111	0.207272	-0.368696
FREEZE_THAW_YR	-0.730925	-0.306496	-0.056753	-0.299598	0.491193
TOTAL_ANN_PRECIP	0.123616	0.895597	-0.302086	-0.117232	0.062065
INTENSE_PRECIP_DAYS_YR	0.190140	0.846556	-0.297667	-0.150064	0.100138
WET_DAYS_YR	-0.349331	0.770524	-0.226374	-0.017618	-0.069989
MAX_ANN_HUM_AVG	-0.182615	0.654580	0.665804	0.081985	0.195059
MIN_ANN_HUM_AVG	-0.323228	0.757478	0.506772	0.126679	-0.028140
TOTAL_SNOWFALL_YR	-0.745707	0.010550	-0.217644	0.198152	-0.217309
SNOW_COVERED_DAYS_YR	-0.269808	0.048246	-0.244001	0.855057	0.348518

Table 3 Rotated loading matrix of the first two factors

Variables	Factor 1	Factor 2
MEAN_ANN_TEMP_AVG	0.982994	−0.047980
MAX_ANN_TEMP_AVG	0.954591	−0.189334
MIN_ANN_TEMP_AVG	0.968678	0.104785
MAX_ANN_TEMP	0.573785	−0.606849
MIN_ANN_TEMP	0.913414	0.025054
DAYS_ABOVE_32_C_YR	0.752046	−0.452869
DAYS_BELOW_0_C_YR	−0.960855	−0.087493
FREEZE_INDEX_YR	−0.258940	0.089854
FREEZE_THAW_YR	−0.769869	−0.188397
WET_DAYS_YR	0.262178	0.865239
TOTAL_ANN_PRECIP	0.320213	0.806396
INTENSE_PRECIP_DAYS_YR	−0.200770	0.798712
MAX_ANN_HUM_AVG	−0.077982	0.675087
MIN_ANN_HUM_AVG	−0.224511	0.815680
TOTAL_SNOWFALL_YR	−0.734878	0.127059
SNOW_COVERED_DAYS_YR	−0.830072	0.027067

humidity, causing the high snowfall as shown in Table 4. Compared with Fig. 5 (c), the cluster borders are more distinct.

For the optimal 9-means clustering with the lowest within cluster sum of squares, we can see from Fig. 5 (b) that the data points are more centralized, especially for the temperature principle component. In addition to the freeze and no freeze clustering, the model suggest four to five temperature clustering. We could use cold, cool, mild, warm, and hot instead of the original freeze and non freeze temperature classification and could obtain nine climatic regions including wet hot, wet warm, wet mild, wet cool, wet cold, dry hot, dry warm, dry mild and dry cool. This finding agrees with Wang's recommendation to add wet mild and dry mild regions to the

**Fig. 4** CCC for different clusters

original four climatic regions [13]. In summary, PCA and factor analysis can identify the main component and common factors for the 16 climatic variables and cluster analysis can be used to classify data samples or weather stations to help determining climatic regions.

Discriminant analysis

With known climatic regions, the supervised machine learning algorithm can be used to determine the regionalization based on collected climatic data from a new weather station or from the climate change. There are 477 samples in the original sample with no climatic regions labeled and therefore the rest 21,189 samples were used for the following supervised learning analyses. 66% of randomly selected samples were used as training set and the rest were used as the testing dataset. The model parameters are firstly trained with the training dataset and then the model is tested with testing dataset.

The Fisher's linear discriminant analysis was conducted first. Table 5 shows classification matrix for training and validation datasets. The sum of each row of the two datasets is 100%. In the training dataset, for the wet no freeze climatic region, 85% of the classification are correct and the majority (12%) of the misclassifications are classified as wet freeze. In the testing dataset, for the wet no freeze climatic region, 81% of the classification are correct and the majority (15%) misclassifications are classified as wet freeze. It can be seen that the classification matrix of the testing dataset are close to those of the training datasets, indicating there is no overfitting. Overfitting means the model is only valid for the training dataset but not work for the testing dataset and therefore the robustness the discriminant model is validated.

Figure 6 shows the distribution of classified samples in the coordinates of the first two principle components. It can be seen from Fig. 6 that results of the Fisher's discriminant analysis are very close as the original four climatic regions shown in Fig. 5 (c). The misclassification rate for the training and testing datasets are 13.6% and 14.4%, respectively, which is fairly good since it is classified based on the value of the linear combination of predictors. Therefore, the discriminant model can be used to classify climatic regions based on the 16 climatic variables while the classification accuracy could be further improved with proper supervised learning algorithms.

Ann

Due to the large volume of model parameters and non-linear transformations capability, ANN has been proved to obtain higher prediction accuracy in machine learnings. As shown in Fig. 7, two ANN models were established. One has one hidden layer with five nodes, and the other one has two hidden layers with 10 nodes in

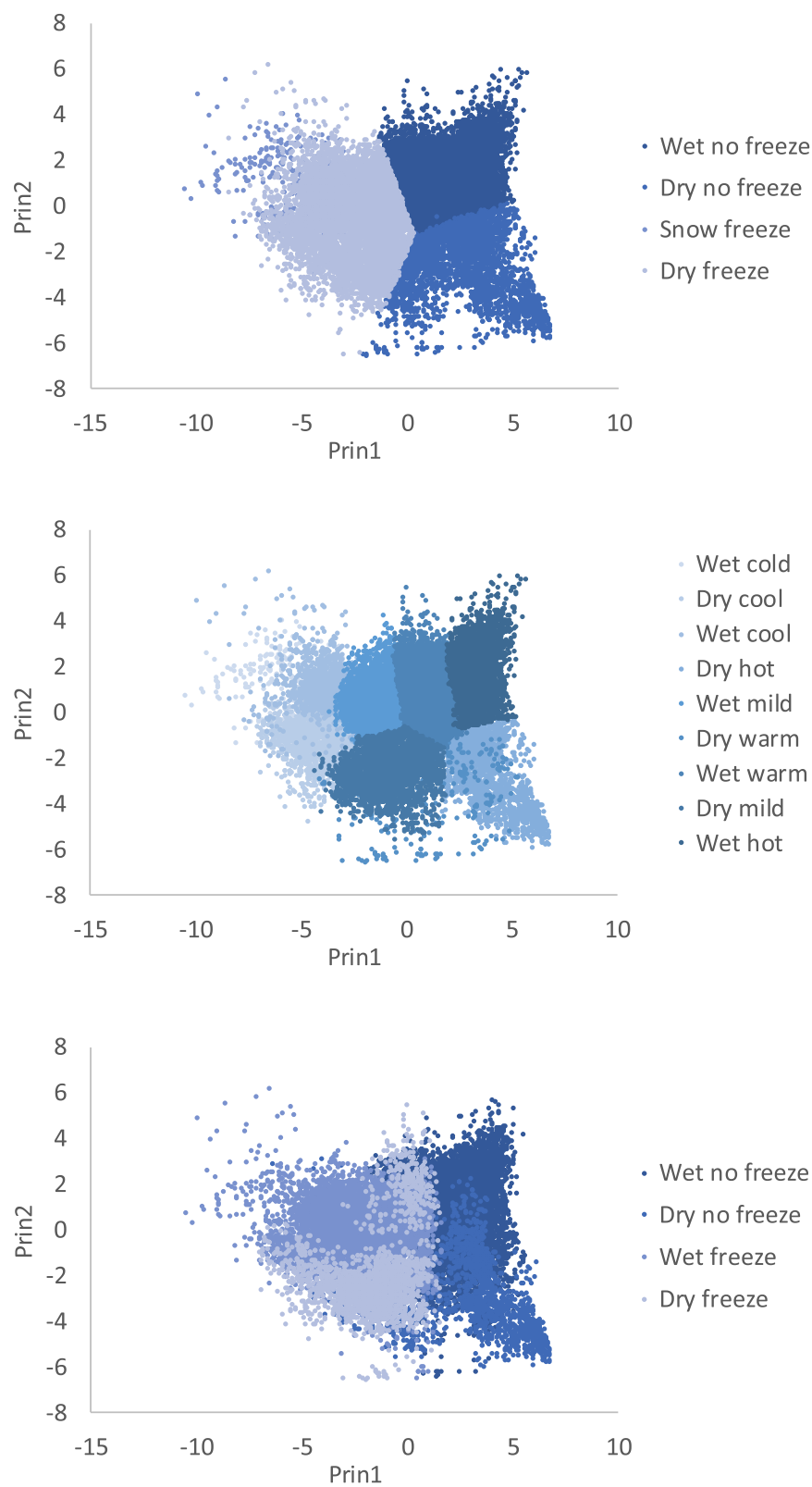


Fig. 5 Results of cluster analysis. **a** Four clusters. **b** Nine clusters. **c** Original four climatic regions by LTPP

Table 4 Average of climatic variables for each cluster

Clusters	4 clusters			
	1(wet no freeze)	2(dry no freeze)	3(dry freeze)	4(snow freeze)
Sample number	8601	3537	9099	429
MEAN_ANN_TEMP_AVG	16	17	9	5
MAX_ANN_TEMP_AVG	22	25	15	11
MIN_ANN_TEMP_AVG	10	10	2	0
MAX_ANN_TEMP	37	41	35	32
MIN_ANN_TEMP	−10	−10	−25	−29
DAYS_ABOVE_32_C_YR	50	93	14	4
DAYS_BELOW_0_C_YR	55	57	148	170
FREEZE_INDEX_YR	49	49	582	983
FREEZE_THAW_YR	50	53	101	95
MAX_ANN_HUM_AVG	91	68	87	90
MIN_ANN_HUM_AVG	49	30	48	53
TOTAL_ANN_PRECIP	1254	473	814	974
INTENSE_PRECIP_DAYS_YR	31	11	18	20
WET_DAYS_YR	144	76	141	174
TOTAL_SNOWFALL_YR	130	142	1127	2218
SNOW_COVERED_DAYS_YR	0	0	1	95

the first layer and 8 nodes in the second layer. Table 6 shows the classification matrixes for both training and validation datasets of the two ANNs. For the one layer ANN, the generalized R square of the training and testing datasets are 94% and 93% respectively, which are very close and indicate no overfitting. In the training dataset, for the dry freeze climatic region, 92% of the classification are correct and the majority (5%) of the misclassifications are classified as wet freeze. In the testing dataset, for the dry freeze climatic region, 92% of the classification are correct and the majority (6%) misclassifications are classified as wet freeze. It is noted that the overall misclassification rate are 9.6% and 9.7% respectively, much higher than the Fisher's linear discriminant analysis. This is because the nonlinear calculation capability of ANN. For the two layer ANN, the generalized R square of the training and testing datasets are 97% and 96% respectively, and the overall misclassification rate are 6% and 6.6% respectively, indicating that more model

parameters could significantly improve the model accuracy.

Conclusions and future research

In this study, the climatic data were used to investigate the climate regionalization for pavement infrastructure. Firstly, 16 historical climatic data variables including annual temperature, freeze thaw, precipitation, and snowfall conditions were extracted from the LTPP database and the effect and significance of climate change were evaluated. The unsupervised machine learning including PCA, factor analysis and cluster analysis were firstly conducted to identify the main component and common factors for climatic variables, and then to classify datasets into different groups. Then, the Fisher's discriminant analysis and ANN models were built to predict the climatic regions based on climatic data. The benefit of unsupervised machine learnings is to identify the key factors and find the optimal clustering of climatic

Table 5 Classification matrix for training and validation datasets of the discriminant analysis

Climatic region	Training				Testing			
	Wet no freeze	Dry no freeze	Wet freeze	Dry freeze	Wet no freeze	Dry no freeze	Wet freeze	Dry freeze
Wet no freeze	85%	1%	12%	2%	81%	2%	15%	2%
Dry no freeze	12%	86%	0%	1%	12%	87%	0%	1%
Wet freeze	9%	0%	85%	6%	9%	0%	86%	5%
Dry freeze	4%	3%	5%	88%	4%	3%	6%	86%

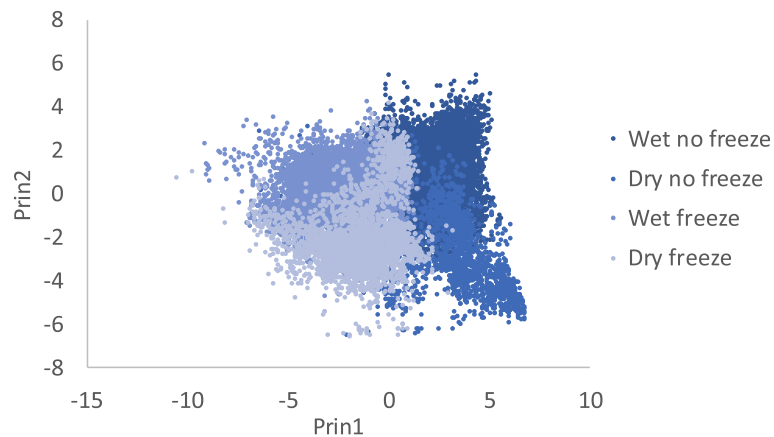


Fig. 6 Results of Fisher's discriminant analysis

conditions based on the similarities among data samples, while supervised machine learnings could provide a more accurate classification based on the data.

Investigation on the LTPP annual climatic data shows that the mean, maximum and minimum of the average temperature in most States increase significantly while the freeze index and freeze thaw cycles decrease significantly. In addition, around 50% States show significant increasing precipitation and humidity while only 10–20% States show significant decreasing trend. The rising temperatures increase evaporation, causing increased precipitation. Around one third of the 800 weather stations record variation of freeze and precipitation classifications and a few of them show significant change of

classifications over time based on the results of logistic regression analyses.

Results of PCA show that the first two components, which are highly correlated with temperature and humidity respectively account for a total of 71.6% of the variance and can be used to reduce the dimensionality of the original climatic variables. Results of factor analysis show temperature and humidity are the two common factors, and the snowfall and snow covered days are more related to the temperature factor. The 4-means clusters include wet no freeze, dry no freeze, dry freeze and snow freeze. The 9-means cluster model with highest CCC suggest 4 and 5 temperature clusters for dry or wet conditions.

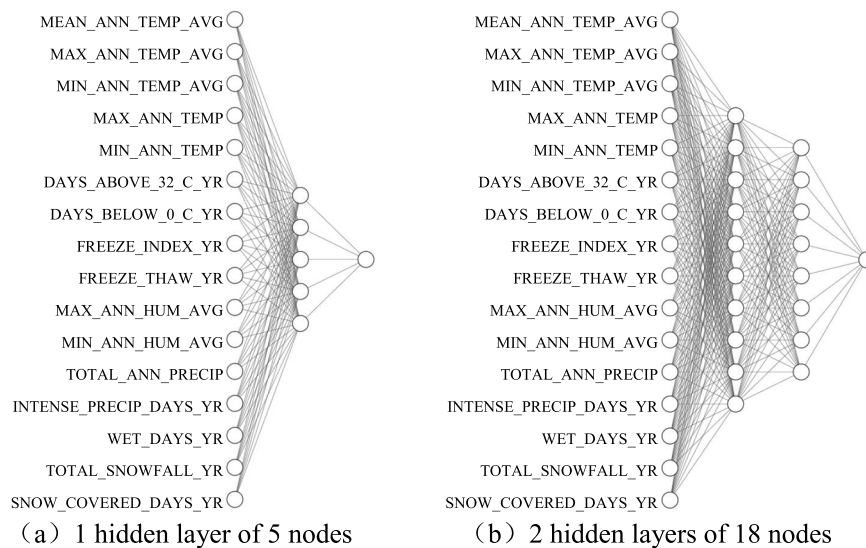


Fig. 7 Structure of the ANNs

Table 6 Classification matrix for training and validation datasets of the two ANNs

Climatic region	Training				Validation			
	Wet no freeze	Dry no freeze	Wet freeze	Dry freeze	Wet no freeze	Dry no freeze	Wet freeze	Dry freeze
1 layer ANN								
Wet no freeze	80%	2%	13%	4%	82%	3%	11%	4%
Dry no freeze	8%	88%	0%	4%	10%	86%	0%	4%
Wet freeze	3%	0%	92%	5%	3%	0%	92%	6%
Dry freeze	2%	1%	5%	92%	1%	1%	6%	92%
2 layer ANN	1	2	3	4	1	2	3	4
Wet no freeze	86%	2%	8%	4%	84%	2%	9%	5%
Dry no freeze	5%	93%	0%	2%	5%	92%	0%	3%
Wet freeze	2%	0%	94%	4%	2%	0%	93%	5%
Dry freeze	1%	1%	5%	93%	1%	1%	5%	93%

Both the Fisher's linear discriminant analysis and ANN can effectively predict climatic regions with multiple climatic variables. ANN performs better with higher R square and low misclassification rate, especially for those with more layers and nodes. This study focused on using multiple climatic data. In future study, the geological and solar radiation data could be included to potentially improve the clustering and prediction.

Abbreviations

ANN: Artificial Neural Networks; CCC: Cubic Clustering Criterion; DNN: Deep Neural Network; EICM: Enhanced Integrated Climate Model; LTPP: Long-term Pavement Performance; MEPDG: Mechanistic-Empirical Pavement Design Guide; PCA: Principle Component Analysis; SVM: Support Vector Machine

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s43065-021-00020-7>.

Additional file 1:

Acknowledgements

This study is supported by the Natural Science Foundation of Jiangsu Province under Grant No. BK20181279 and BK20200468, the Fundamental Research Funds for the Central Universities, CHD under Grant No. 300102341508, and the Science and Technology Project of Zhejiang Provincial Department of Transport under Grant No. 2020045 and No. 2020053, to which the authors are very grateful.

Authors' contributions

QD prepared the introduction part and performed the data analysis. XC performed the data analysis, and was a major contributor in writing the manuscript. SD performed the data analysis and help prepare the manuscript. JZ helped download and prepare the data. All authors read and approved the final manuscript.

Funding

Natural Science Foundation of Jiangsu Province under Grant No. BK20181279.
Natural Science Foundation of Jiangsu Province under Grant No. BK20200468.
Fundamental Research Funds for the Central Universities, CHD under Grant No. 300102341508.

Science and Technology Project of Zhejiang Provincial Department of Transport under Grant No. 2020045

Science and Technology Project of Zhejiang Provincial Department of Transport under Grant No. 2020053

Availability of data and materials

The datasets analyzed during the current study are available in the Long-term Pavement Performance (LTPP) database repository, <https://infopave.fhwa.dot.gov/>

Declarations

Competing interests

The authors have declared that no competing interests exists.

Author details

¹School of Transportation, Southeast University, 2 Southeast University Road, Nanjing 211189, Jiangsu Province, China. ²National Demonstration Center for Experimental Road and Traffic Engineering Education (Southeast University), Nanjing 211189, Jiangsu Province, China. ³Department of Civil Engineering, Nanjing University of Science and Technology, 200 Xiaolingwei Street, Nanjing 210094, Jiangsu Province, China. ⁴College of Transportation Engineering, China Engineering Research Center of Highway Infrastructure Digitalization, Ministry of Education of PRC, Chang'an University, Middle-section of South Erhuan Road, Xi'an 710064, Shaanxi Province, China. ⁵Louisiana Transportation Research Center, 4101 Gourrier Ave, Baton Rouge, LA 70808, USA.

Received: 3 February 2021 Accepted: 11 March 2021

Published online: 23 March 2021

References

- Sol-Sánchez M, Moreno-Navarro F, García-Travé G, Rubio-Gámez MC (2015) Laboratory study of the long-term climatic deterioration of asphalt mixtures. *Constr Build Mater* 88:32–40. <https://doi.org/10.1016/j.conbuildmat.2015.03.090>
- Chinowsky PS, Price JC, Neumann JE (2013) Assessment of climate change adaptation costs for the US road network. *Glob Environ Chang* 23(4):764–773. <https://doi.org/10.1016/j.gloenvcha.2013.03.004>
- Zhang H, Gong M, Huang Y, Miljković M (2020) Study of the high and low-temperature behavior of asphalt based on a performance grading system in Northeast China. *Constr Build Mater* 254:119046. <https://doi.org/10.1016/j.conbuildmat.2020.119046>
- Asi IM (2007) Performance evaluation of SUPERPAVE and Marshall asphalt mix designs to suite Jordan climatic and traffic conditions. *Constr Build Mater* 21(8):1732–1740. <https://doi.org/10.1016/j.conbuildmat.2006.05.036>
- Viola F, Celauro C (2015) Effect of climate change on asphalt binder selection for road construction in Italy. *Transp Res Part D: Transp Environ* 37: 40–47. <https://doi.org/10.1016/j.trd.2015.04.012>
- Jitsangiam P, Chindapasirt P, Nikraz H (2013) An evaluation of the suitability of SUPERPAVE and Marshall asphalt mix designs as they relate to Thailand's

- climatic conditions. *Constr Build Mater* 40:961–970. <https://doi.org/10.1016/j.conbuildmat.2012.11.011>
7. Aflaki S, Tabatabaee N (2009) Proposals for modification of Iranian bitumen to meet the climatic requirements of Iran. *Constr Build Mater* 23(6):2141–2150. <https://doi.org/10.1016/j.conbuildmat.2008.12.014>
 8. Hussain GMA, Abdulaziz MAG, Xiang ZN, Al-Hammadi MA (2020) Climate zones of the asphalt binder performance for the highway pavement design. *Civil Eng J* 6(11):2220–2230. <https://doi.org/10.28991/cej-2020-03091613>
 9. Li R, Schwartz CW, Forman B (2013) Sensitivity of predicted pavement performance to climate characteristics. In *Airfield and Highway Pavement 2013. Sustain Efficient Pav*:760–771
 10. Yang X, You Z, Hiller J, Watkins D (2017) Sensitivity of flexible pavement design to Michigan's climatic inputs using pavement ME design. *Int J Pav Eng* 18(7):622–632. <https://doi.org/10.1080/10298436.2015.1105373>
 11. Bandara N, Henson S, Klieber K (2014) Creating a climate zone map for mechanistic empirical pavement designs. In *T&DI Congress 2014. Planes, Trains, and Automobiles*: 682–691
 12. Basma AA, Al-Suleiman TI (1991) Climatic considerations in new AASHTO flexible pavement design. *J Transp Eng* 117(2):210–223. [https://doi.org/10.1061/\(ASCE\)0733-947X\(1991\)117:2\(210\)](https://doi.org/10.1061/(ASCE)0733-947X(1991)117:2(210))
 13. Wang Y, Wang G, Ahn YH (2012) Impact of climate conditions on effectiveness of asphalt pavement preservation techniques. *Transp Res Rec* 2292(1):73–80. <https://doi.org/10.3141/2292-09>
 14. Schwartz, C. W., Elkins, G. E., Li, R., Visintine, B. A., Forman, B., Rada, G. R., Groeger, J. Evaluation of long-term pavement performance (LTPP) climatic data for use in mechanistic-empirical pavement design guide (MEPDG) calibration and other pavement analysis; Turner-Fairbank Highway Research Center: 2015
 15. Coffey S, Park S, McCarthy LM (2018) Sensitivity analysis of the mainline travel lane pavement service life when utilizing part-time shoulder use with full depth paved shoulders. *Int J Pav Res Technol* 11(1):58–67. <https://doi.org/10.1016/j.ijprt.2017.09.003>
 16. Jackson, N., Puccinelli, J., Effects of multiple freeze cycles and deep frost penetration. Long-Term Pavement Performance Data Analysis Support: National Pooled Fund Study TPF-5, 2006
 17. Hasan MA, Tarefder RA (2018) Development of temperature zone map for mechanistic empirical (ME) pavement design. *Int J Pav Res Technol* 11(1): 99–111. <https://doi.org/10.1016/j.ijprt.2017.09.012>
 18. de Sá Júnior A, de Carvalho LG, Da Silva FF, de Carvalho Alves M (2012) Application of the Köppen classification for climatic zoning in the state of Minas Gerais, Brazil. *Theoretical Appl Climatol* 108(1):1–7. <https://doi.org/10.1007/s00704-011-0507-8>
 19. Tonietto J, Carboneau A (2004) A multicriteria climatic classification system for grape-growing regions worldwide. *Agric For Meteorol* 124(1–2):81–97. <https://doi.org/10.1016/j.agrformet.2003.06.001>
 20. Mills BN, Tighe SL, Andrey J, Smith JT, Huen K (2009) Climate change implications for flexible pavement design and performance in southern Canada. *J Transp Eng* 135(10):773–782. [https://doi.org/10.1061/\(ASCE\)0733-947X\(2009\)135:10\(773\)](https://doi.org/10.1061/(ASCE)0733-947X(2009)135:10(773))
 21. Gudipudi PP, Underwood BS, Zalgout A (2017) Impact of climate change on pavement structural performance in the United States. *Transp Res Part D: Transp Environ* 57:172–184. <https://doi.org/10.1016/j.trd.2017.09.022>
 22. Qiao Y, Zhang Y, Zhu Y, Lemkus T, Stoner AM, Zhang J, Cui Y (2020) Assessing impacts of climate change on flexible pavement service life based on falling weight Deflectometer measurements. *Phys Chem Earth* 120:102908. <https://doi.org/10.1016/j.pce.2020.102908>
 23. Yang Y, Qian B, Xu Q, Yang Y (2020) Climate Regionalization of Asphalt Pavement Based on the K-Means Clustering Algorithm. *Advances in Civil Engineering*. <https://doi.org/10.1155/2020/6917243>
 24. Unal Y, Kindap T, Karaca M (2003) Redefining the climate zones of Turkey using cluster analysis. *Int J Climatol* 23(9):1045–1055. <https://doi.org/10.1002/joc.910>
 25. Aliaga VS, Ferrelli F, Piccolo MC (2017) Regionalization of climate over the Argentine pampas. *Int J Climatol* 37:1237–1247. <https://doi.org/10.1002/joc.5079>
 26. Mannan A, Chaudhary S, Dhanya C, Swamy A (2018) Regionalization of rainfall characteristics in India incorporating climatic variables and using self-organizing maps. *ISH J Hydraulic Eng* 24(2):147–156. <https://doi.org/10.1080/09715010.2017.1400409>
 27. Mahlstein I, Knutti R (2010) Regional climate change patterns identified by cluster analysis. *Clim Dyn* 35(4):587–600. <https://doi.org/10.1007/s00382-009-0654-0>
 28. Carvalho M, Melo-Gonçalves P, Teixeira J, Rocha A (2016) Regionalization of Europe based on a K-means cluster analysis of the climate change of temperatures and precipitation. *Physics Chem Earth, Parts A/B/C* 94:22–28. <https://doi.org/10.1016/j.pce.2016.05.001>
 29. Comrie AC, Glenn EC (1998) Principal components-based regionalization of precipitation regimes across the Southwest United States and northern Mexico, with an application to monsoon precipitation variability. *Clim Res* 10(3):201–215. <https://doi.org/10.3354/cr010201>
 30. Darand M, Daneshvar MRM (2014) Regionalization of precipitation regimes in Iran using principal component analysis and hierarchical clustering analysis. *Environ Processes* 1(4):517–532. <https://doi.org/10.1007/s40710-014-0039-1>
 31. Malmgren BA, Winter A (1999) Climate zonation in Puerto Rico based on principal components analysis and an artificial neural network. *J Clim* 12(4):977–985. [https://doi.org/10.1175/1520-0442\(1999\)012<0977:CZIPRB>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<0977:CZIPRB>2.0.CO;2)
 32. Leber D, Holawe F, Häusler H (1995) Climatic classification of the Tibet autonomous region using multivariate statistical methods. *GeoJournal* 37(4): 451–472. <https://doi.org/10.1007/BF00806934>
 33. Boulanger J-P, Martinez F, Segura EC (2006) Projection of future climate change conditions using IPCC simulations, neural networks and Bayesian statistics. Part 1: temperature mean state and seasonal cycle in South America. *Clim Dyn* 27(2–3):233–259. <https://doi.org/10.1007/s00382-006-0134-8>
 34. Yang L, Lyu K, Li H, Liu Y (2020) Building climate zoning in China using supervised classification-based machine learning. *Build Environ* 171:106663. <https://doi.org/10.1016/j.buildenv.2020.106663>
 35. Park S, Park H, Im J, Yoo C, Rhee J, Lee B, Kwon C (2019) Delineation of high resolution climate regions over the Korean peninsula using machine learning approaches. *PLoS One* 14(10):e0223362. <https://doi.org/10.1371/journal.pone.0223362>
 36. Chatti, K., Buch, N., Haider, S., Pulipaka, A., Lyles, R. W., Gilliland, D., Desaraju, P. LTPP data analysis: Influence of design and construction features on the response and performance of new flexible and rigid pavements; 2005; pp 20–50
 37. Washington, S., Karlaftis, M., Mannering, F., Anastasopoulos, P., Statistical and econometric methods for transportation data analysis. 3rd ed.; Chapman and Hall/CRC: 2020
 38. Wentz FJ, Ricciardulli L, Hilburn K, Mears C (2007) How much more rain will global warming bring? *Science* 317(5835):233–235. <https://doi.org/10.1126/science.1140746>
 39. Hui Z, Zhang J, Ma Z, Li X, Peng T, Li J, Wang B (2018) Global warming and rainfall: lessons from an analysis of mid-Miocene climate data. *Palaeogeogr Palaeoclimatol Palaeoecol* 512:106–117. <https://doi.org/10.1016/j.palaeo.2018.10.025>
 40. Tabari H (2020) Extreme value analysis dilemma for climate change impact assessment on global flood and extreme precipitation. *J Hydrol* 125932
 41. Ghasemi P, Aslani M, Rollins DK, Williams R (2019) 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* 59(4):1335–1353. <https://doi.org/10.1007/s00158-018-2133-x>
 42. Yao L, Dong Q, Jiang J, Ni F (2019) Establishment of prediction models of asphalt pavement performance based on a novel data calibration method and neural network. *Transp Res Rec* 2673(1):66–82. <https://doi.org/10.1177/0361198118822501>
 43. Tian P, Shukla A, Nie L, Zhan G, Liu S (2018) Characteristics' relation model of asphalt pavement performance based on factor analysis. *Int J Pav Res Technol* 11(1):1–12. <https://doi.org/10.1016/j.ijprt.2017.07.007>
 44. Chen X, Dong Q, Zhu H, Huang B, Burdette EG (2019) Contributions of condition measurements on the latent pavement condition by confirmatory factor analysis. *Transportmetrica A* 15(1):2–17. <https://doi.org/10.1080/23249935.2017.1369195>
 45. Wang W, Wang S, Xiao D, Qiu S, Zhang J (2018) An Unsupervised Cluster Method for Pavement Grouping Based on Multidimensional Performance Data. *Journal of Transportation Engineering, Part B: Pavements* 144(2): 04018005. <https://doi.org/10.1061/JPEODX.00000030>

46. Li L, Luo W, Wang KC, Liu G, Zhang C (2018) Automatic groove measurement and evaluation with high resolution laser profiling data. *Sensors* 18(8):2713. <https://doi.org/10.3390/s18082713>
47. Chou C-P, McCullough BF (1987) Development of a distress index and rehabilitation criteria for continuously reinforced concrete pavements using discriminant analysis. *Trans Res Record*:1117
48. Hussan S, Kamal MA, Hafeez I, Ahmad N, Khanzada S, Ahmed S (2020) Modelling asphalt pavement analyzer rut depth using different statistical techniques. *Road Mat Pav Design* 21(1):117–142. <https://doi.org/10.1080/14680629.2018.1481880>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)