

Research Article

Cluster-Based Pavement Deterioration Models for Low-Volume Rural Roads

V. Sunitha,¹ A. Veeraragavan,² Karthik K. Srinivasan,² and Samson Mathew¹

¹ Department of Civil Engineering, National Institute of Technology, Tiruchirappalli 620015, India

² Department of Civil Engineering, Indian Institute of Technology Madras, Chennai 600036, India

Correspondence should be addressed to V. Sunitha, sunitha@nitt.edu

Received 31 July 2012; Accepted 23 September 2012

Academic Editors: S. Easa and B. Uy

Copyright © 2012 V. Sunitha et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The management of low-volume rural roads in developing countries presents a range of challenges to road designers and managers. Rural roads comprise over 85 percent of the road network in India. The present study aims at development of deterioration models for the optimum maintenance management of the rural roads under a rural road programme namely Pradhan Mantri Gram Sadak Yojana (PMGSY) in India. Visual condition survey along the selected low-volume rural roads considers parameters like condition of shoulders, drainage features, cross-drainage structures, and camber, and pavement distresses, namely, potholes, crack area, and edge break, are collected for a period of three years. The deterioration models have a significant role in the pavement maintenance management system. However, the performance of a pavement depends on several factors. Cluster analysis can be used to group the pavement sections so that the performance of pavements in different clusters can be studied. Nonhierarchical clustering technique of *k*-means clustering was considered. Separate deterioration models have been developed for each of the clusters. A comparison of the models developed with and without clustered sections reveals that the clustering of pavement sections are preferred for the efficient rural road maintenance management.

1. Introduction

Rural roads which connect the villages comprise over 85 percent of the road network in India. The rural roads stimulate overall development by providing access to economic and social infrastructure and facilities. Most of these rural roads were neglected in the past, but with the rapid increase of Indian economy and public demand for better road infrastructure, the Government of India has taken up several road building programmes. For the development of these rural roads, a megaroad development program called Pradhan Mantri Gram Sadak Yojana (PMGSY) was launched in December 2000 by the Government of India to provide all-weather connectivity to unconnected rural habitations. Maintenance of these low volume roads is a routine work performed for upkeep of pavement, shoulder, and other facilities provided for road users. Lack of maintenance of the rural roads affects the people in the villages significantly as the time for access to markets and other social infrastructure

is increased. The implementation of different rural road projects improved the mobility and there is a need for further emphasis on preservation of the road assets, through timely and appropriate maintenance.

The scheduling of pavement maintenance and rehabilitation is a critical task. It is probably the key decision variable in any asset management system. This decision process relies upon the capability to predict the future pavement distress condition as a function of time. If the pavement performance prediction model can be developed to accurately reflect real pavement performance, the remaining service life for pavements can be more accurately estimated. Most of the pavement deterioration models use a single model or common model for all the pavement sections. It is not appropriate to have a single model for a variety of pavements. To solve this issue, the pavement sections are to be grouped into clusters of homogeneous groups. The clusters are homogeneous within clusters and heterogeneous between clusters. The main objective of this paper is to study

the performance of rural road sections grouped in different clusters and the development of deterioration models for each of the clusters.

Wang and Li [1] used a gray clustering-based methodology to evaluate the existing pavements following the Mechanistic Empirical Pavement Design guide framework. Gray clustering method is one pillar part of the gray system theory, which was proposed in the 1980s to solve problems with partially known information. Luo and Chou [2] introduced cluster-wise regression modelling for the deterioration of the pavement condition and recommended this method for improving the accuracy of the pavement condition predictions. Yu et al. [3] proposed the use of Linear Mixed Effects Model (LMEM) to predict future conditions of a specific pavement section by a weighted combination of the average deterioration trend of the family and the past conditions of the specific pavement. Luo and Yin [4] compared the proposed model of weighted regression function consisting of all clusters with conventional Markov probabilistic model. Analysis of the data collected during the study is an important aspect while developing pavement management system. Among the many analytical tools that are available for modelling, the cluster analysis is found to be important. Studies in the field of sampling of traffic count data and the traffic accidents using cluster analysis are available [5–7]. The defects in conventional methods of road section maintenance in practical work were analyzed and a more reasonable and operational method, which uses ordinal sample cluster to divide maintenance sections, was reported [8]. Though clustering has been adopted as an analyzing tool in various traffic studies, it has not been incorporated with Pavement Management System (PMS) for low volume rural roads. So the present study investigates the application of cluster analysis in the development of appropriate maintenance management for low volume rural roads in India.

2. Study Area

The study area is Tiruchirappalli District in Tamil Nadu, India, which extends over an area of 4,404 sq.km with a population density of 549 per sq.km. The district consists of 14 blocks and 507 villages. The blocks are shown in Figure 1. This district is situated in the geographical coordinates of north latitude between 10 and 11–30' and east longitude between 77–45' and 78–50'. The number of study roads and the minimum and the maximum lengths of the roads in each block are given in Table 1.

3. Data Collection

Pavement condition data is a prerequisite for the development of pavement management system. Extensive visual condition survey of the PMGSY roads that were constructed prior to 2008 was carried out for a period of three years. The age of the pavement sections varied from seven to nine years. The pro forma for the pavement data collection is shown in Figure 4. The different pavement distress data considered during the visual survey are (i) pothole, (ii)

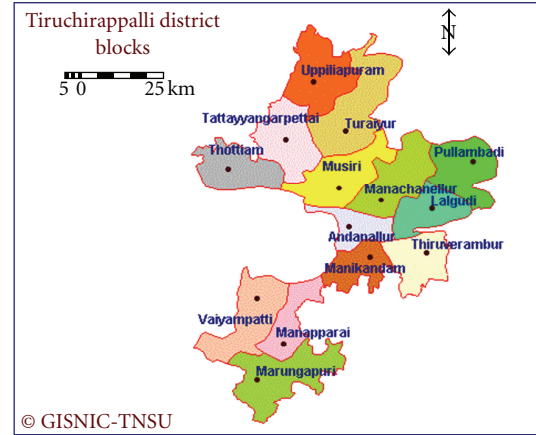


FIGURE 1: Block wise map of Tiruchirappalli District.

TABLE 1: Number of study roads selected in each block of Tiruchirappalli District.

Sl. no.	Block name	No. of roads	Length (min and max), km
1	Andanallur	8	0.5–1.5
2	Manikandam	19	0.5–3.3
3	Manapparai	21	1.0–2.6
4	Marungapuri	9	1.0–2.4
5	Vaiyampattin	9	0.8–3.21
6	Lalgudi	7	0.5–2.0
7	Pullampady	6	1.0–4.0
8	Manachanallur	3	1.2–2.4
9	Musiri	10	0.6–2.4
10	Thottiam	4	1.0–2.0
11	Thuraiyur	7	1.5–3.2
12	Uppiliapuram	5	1.6–3.4
13	Thiruverambur	7	0.65–2.9
14	Thathiangerpattin	9	0.6–3.0

camber, (iii) crack, and (iv) edge break. The condition of the shoulder is another factor that plays an important role on the performance. The rating for the shoulder is based on the condition and the slope of the shoulder as well as the vegetation present over the section. The condition of the side drainage, the shape and side slope of the drain, and the presence of silt in the side drain are taken into consideration while assigning the rating. The condition of the cross drainage structures has an effect on the condition of road. These cross drainage structures are rated based on the settlement and erosion of these structures as well as closure of openings of these structures due to silts and other debris. The data is collected for every 200 m length of pavement stretch. The brief description of the above parameters are given in the subsequent section and the list of indices used is given in Table 2.

3.1. Pothole. Failure of wearing course is due to the lack of proper mix design, improper gradation of aggregates, inadequate binder content, or an inferior type binder

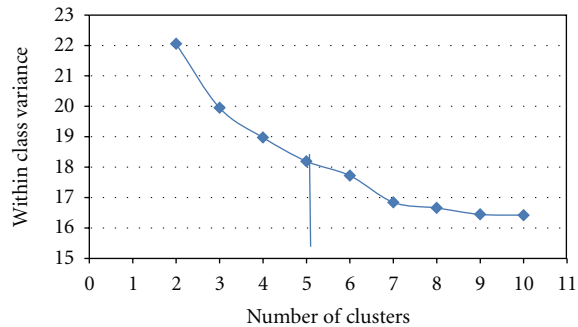


FIGURE 2: Number of clusters versus within class variance.

TABLE 2: List of indices.

Sl. no.	Indices	Description
1	LDS	Left drainage shape
2	LDI	Left drainage—silt and debris
3	LSV	Left shoulder condition—vegetation and slope
4	POD	Pothole
5	CAM	Camber
6	CRA	Crack
7	EDG	Edge break
8	RSV	Right shoulder condition—vegetation and slope
9	RDS	Right drainage shape
10	RDI	Right drainage—silt and debris
11	CDO	Cross-drainage structure—opening
12	CDS	Cross-drainage structure—settlement

TABLE 3: Rating for potholes.

Number	Rating
Nil	5
1–3	4
4, 5	3
6, 7	2
>7	1

TABLE 4: Rating for camber.

Camber provided	Rating
Very good	5
Good	4
Fair	3
Poor	2
Very poor	1

resulting in a poor bituminous surfacing which in turn leads to development of potholes. A rating system has been developed based on the number of potholes of 10 cm × 10 cm size per 200 m section and is shown in Table 3.

3.2. Camber. Camber is the cross slope provided to the road surface in the transverse direction to drain off the rain water from the pavement surface. Stagnant water can cause the bitumen to lose its adhesive property and make it to wear

TABLE 5: Rating for cracked area of paved surface.

Cracked area of paved surface	Rating
Nil	5
1–5%	4
6%–10%	3
11%–15%	2
>15%	1

TABLE 6: Rating for edge break.

Length of edge break in 200 m section of project road, m	Rating
Nil	5
1–10	4
11–20	3
21–30	2
>30	1

off. Proper drainage and quick disposal of water from the pavement surface is important for longevity of the pavement. The camber should be optimal. An excessive camber is a hindrance to driving and causes problems at intersections and to slow moving and heavily loaded vehicles like trucks. Table 4 shows the rating for the camber considered during the survey of the roads.

3.3. Crack. Wide and extensive cracks on pavement surface affects the riding quality of the pavement. Cracks can be of various types like alligator cracks, longitudinal cracks, cracks due to shear failure, and so forth. Table 5 indicates the rating provided based on the percentage of cracked area on the total paved surface.

3.4. Edge Break. Edge breaks are considered as a parameter affecting the performance of the pavement. Edge breaks contract the paved area and reduce the operating speeds. The rating scale, shown in Table 6, is based on the total cumulative length of the edge breaks in a pavement section of 200 m.

3.5. Condition of the Shoulder. Shoulders serve as an emergency lane for vehicle compelled to be taken out of pavement or roadway. While rating the condition of the shoulder, the slope of the shoulder is considered as an important factor as effective drainage depends on the cross slope of the shoulders. The presence of shrubs and vegetation also affects the sight distance in these roads. The rating for the shoulder is based on the condition of the shoulder as well as the vegetation present over the section. The rating adopted in the present investigation is described in Table 7.

3.6. Side Drainage. The side drains are an essential cross section element of any pavement. The drainage often consists of an open ditch parallel to the road carriageway with culverts at regular intervals to disperse the run-off to local watercourses. The rating of the side drains consists of two

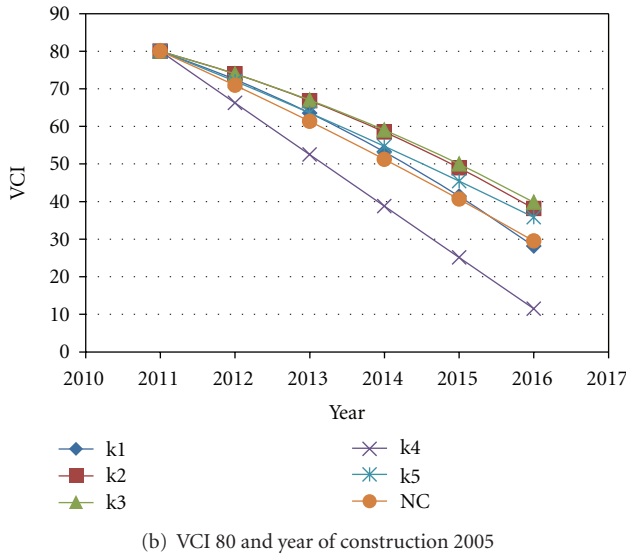
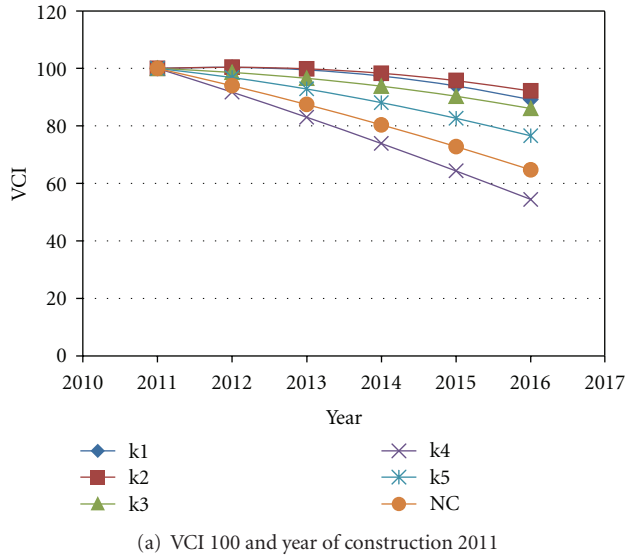


FIGURE 3: Deterioration models for with and without cluster models.

TABLE 7: Rating for shoulder condition.

Shoulder condition	Rating
Very good	5
Good	4
Fair	3
Poor	2
Very poor	1

factors. The first factor consists of the size, shape, and side slope of the drain. The second factor is the presence of silt in the side drain. Tables 8(a) and 8(b) show the rating scale adopted for the side drains.

3.7. Condition of the Cross Drainage Structure. Cross drainage structures like box or pipe culverts are common

TABLE 8: (a) Rating for shape of side drain and (b) rating for the amount of silts and debris.

(a)	
Shape and slope of side drain	Rating
Very good	5
Good	4
Fair	3
Poor	2
Very poor	1

(b)	
Percentage of silts and debris	Rating
Nil	5
1–25%	4
26%–50%	3
51%–75%	2
>75%	1

TABLE 9: (a) Rating for cross-drainage structure and (b) rating for the settlement of cross-drainage structure.

(a)	
Percentage of opening	Rating
Fully open	5
>75%	4
51%–75%	3
26%–50%	2
<26%	1

(b)	
Percentage of settlement	Rating
Nil	5
1–25%	4
26%–50%	3
51%–75%	2
>75%	1

element in these road sections. Condition of these cross drainage elements affects the condition of road. These cross drainage structures are rated based on the settlement and erosion of these structures as well as closure of the openings of these structures due to silts and other debris. The details are shown in Tables 9(a) and 9(b).

4. Visual Condition Index

From the above visual survey data, 550 sections of 200 m each were selected for further analysis. The data collected for three year period was used in the analysis. The pavement sections are usually rated based on Pavement Condition Index (PCI), Pavement Serviceability Index (PSI), and so forth. In the present study, as the parameters considered are different from the usual parameters required for PCI or PSI calculations, an index called Visual Condition Index (VCI) is formulated by the authors using Factor Analysis. From the

Condition and maintenance needs assessment
BT/WBM/gravel road

District _____ Block _____ Class: MDR/ODR/VR
 Road chainage: from.....to..... Name (from-to) _____
 Road length _____ km
 Year of construction/renewal: _____ Month _____ Year _____
 GPS coordinates: _____ Start Ch. Lat _____ Long _____

Start km.	Left drain		Left Shoulder	Carriageway condition				Right Shoulder	Right drain		Cross-drainage	
	Shape/rain cut/side slope	Silt	Vegetation	Pothole	Camber	Crack area	Edge breaks	Vegetation	Shape/rain cut/side slope	Silt	Opening and condition	Settlement and erosion
0–200												
200–400												
400–600												
600–800												
800–1000												

Remarks, if any:

FIGURE 4: Pro forma for pavement data collection.

factor analysis results, the four factors were identified which logically associate the twelve measures of pavement condition presented in Table 2. The factor loadings (correlation between the original variables and the factors) are shown in Table 10. For clarity, the highest factor loading for each indicator across all factors is shown in bold letters. Data on 12 indicators of pavement conditions were factor analyzed (Principal components, Varimax rotation) and four factors were extracted. These four factors were identified as distress factor (factor I), side drainage factor (factor II), Shoulder vegetation factor (factor III), and cross drainage factor (factor IV). A total variance of 77.8 percent was explained by these factors. The index numbers were standardized to the scale of 0 to 100 for ease of comparison across all sections and named as Visual Condition Index (VCI). The VCI of zero indicates poor condition and VCI 100 indicates good condition. The detailed procedure of computation of VCI is beyond the scope of this paper. The paper on VCI formulation by the same authors is in press [9].

TABLE 10: Rotated factor loadings.

Sl. no.	Index	Factors			
		I	II	III	IV
1	LDS	0.156	0.650	0.511	0.013
2	LDI	0.194	0.902	0.025	0.115
3	LSV	0.178	0.119	0.896	0.012
4	POD	0.821	0.128	0.229	0.070
5	CAM	0.764	0.082	0.115	0.034
6	CRA	0.821	0.284	0.012	0.070
7	EDG	0.811	0.157	0.208	0.062
8	RSV	0.197	0.135	0.867	0.043
9	RDS	0.179	0.694	0.457	0.038
10	RDI	0.189	0.905	0.032	0.117
11	CDO	0.071	0.105	0.015	0.902
12	CDS	0.082	0.080	0.038	0.907
Percentage of variance		23.351	22.627	17.808	14.02
Total variance explained			77.806		

5. Cluster Analysis-Multivariate Data Analysis Tool

While studying the output from the different visual condition survey conducted in the Tiruchirappalli District, a large number of basic data units (pavement sections) was encountered. These basic data units have to be classified on the basis of some homogeneity or similarity of various

attributes chosen. Cluster analysis is a class of techniques used to classify cases into groups that are relatively homogeneous within themselves and heterogeneous between each other, on the basis of a defined set of variables [10]. Its objective is to group data units and/or variables into clusters such that the elements within a cluster have a high degree of “natural association”, while the clusters themselves are

TABLE 11: Within class variance and between class variance.

Number of clusters	Within class variance	Between class variance	Minimum no. of sections
2	22.052	8.090	245
3	19.949	10.336	145
4	18.973	11.738	50
5	18.195	12.090	28
6	17.722	12.998	20
7	16.845	13.440	18
8	16.660	13.625	6
9	16.449	13.836	6
10	16.424	13.861	6

relatively distinct. Two basic philosophies followed while considering cluster analysis are as follows:

Hierarchical Clustering. In this system, the data are not partitioned into a particular number of clusters in a single step. Data units are consecutively grouped or divided to form new clusters.

Nonhierarchical Clustering. This method is designed to cluster data units into a single classification of k clusters where k is either specified a priori or determined as a part of the method.

5.1. Nonhierarchical Clustering. Non-hierarchical clustering technique was used in the present study. Over the years, a number of researchers have investigated methods to support identification of clusters in data. Due to its simplicity, k -means algorithm is one of the most extensively used methods. K -means is an algorithm based on the concept of choosing a preliminary set of centroids and assigning each point to the nearest centroid. Once the initial clusters are determined, new centroids are calculated and the points are again assigned to the nearest centroid. This process is repeated until optimal boundaries for each cluster are determined [11]. In the present study, Squared Euclidean distance is used as the measure of dissimilarity. The inverse of the distance can be presented as a measure of similarity or proximity. Euclidean distance can be defined as the sum of the squared differences over all of the variables, that is, D_{ij} between two cases, i and j with variable values $(x_{i1}, x_{i2}, \dots, x_{ik})$ and $(x_{j1}, x_{j2}, \dots, x_{jk})$ is defined by

$$D_{i,j} = \sqrt{\sum_{k=1}^n (x_{ki} - x_{kj})^2}. \quad (1)$$

The objective of k -means is to minimise the squared Euclidean distance between the data points and the centroid of the cluster. The advantage of the k -means approach is that more specific grouping hypotheses can be evaluated and that the overall procedure can be performed more quickly than agglomerative approaches. In terms of computer programmes, cluster analysis is such a long-standing and popular technique that it is rare to find a commercial multivariate statistical package that does not include it in some form.

These ranges from inexpensive plug-ins for MS-Excel (e.g., UNISTAT, XL-STAT, StatistiXL) to sophisticated stand-alone packages. In this work, all calculations were carried out using an add-on function to Excel found in the XL-STAT.

5.2. Data for K-Means Clustering. The data for the cluster analysis is the twelve pavement condition indices listed in Table 2. The data were collected for three years. For each section, 36 attributes (12 indices \times 3 years) are considered. For clustering of the 550 sections, these 36 attributes are used. The attributes are weighted by the rotated factor loadings (specified in bold letters) shown in Table 10.

5.3. Number of K-Means Cluster. One of the assumptions in k -means clustering is deciding the number of clusters. In k -means clusters, the numbers of clusters are to be specified a priori. In this work, the numbers of clusters are decided based on two criteria.

- (i) Minimising within class variance or maximising between class variance of clusters.
- (ii) Minimum number of sections in any class should not be less than 5% of the total sample.

The k -means clustering was done for a number of classes varying from 2 to 10. The within class variance, the between class variance, and the lowest number of sections in the clusters are given in Table 11. The variation of within class variance for the different cluster groups is shown in Figure 2. The minimum number of sections in any cluster is fixed as 28, which is 5% of the total sections. Considering the above two criteria, the number of clusters in k -means is considered as 5 for further analysis. The within class variance and the between class variance for the cluster group 5 are 18.195 and 12.09, respectively, and the minimum number of section is 28.

5.4. Results of K-Means Clustering. In the 5 cluster group, the clusters are numbered as 1 to 5 according to the size of the cluster. Among the five cluster groups, cluster 1 consists of 150 sections, cluster 2 with 148 sections, 3 with 127 sections, 4 with 97 sections, and 5 with 28 sections. The within group variance of these five clusters varies from 12.312 to 28.872. The average and the maximum distance from the centroid of these clusters are also given in Table 12.

TABLE 12: Cluster composition.

Cluster	1	2	3	4	5
Within-groups variance	20.414	12.670	12.312	28.872	25.515
Minimum distance from centroid	2.077	1.835	2.140	3.203	2.673
Average distance from centroid	4.357	3.453	3.382	5.218	4.651
Maximum distance from centroid	7.534	6.184	6.618	8.213	9.468
Size	150	148	127	97	28

TABLE 13: Cluster centroids.

Cluster	1	2	3	4	5
LDS_1	3.860	4.243	4.724	3.247	3.286
LDI_1	4.513	4.027	4.945	3.763	3.536
LSV_1	3.007	4.088	4.315	3.103	3.071
POD_1	2.847	4.649	4.740	4.165	1.107
CAM_1	4.187	4.595	4.732	4.216	2.571
CRA_1	4.773	4.966	4.992	4.907	2.036
EDG_1	3.060	4.453	4.591	3.794	1.429
RSV_1	3.173	4.115	4.362	3.237	3.143
RDS_1	3.820	4.257	4.748	3.196	3.429
RDI_1	4.440	3.946	4.945	3.773	3.464
CDO_1	4.913	4.939	4.874	4.443	4.929
CDS_1	5.000	5.000	5.000	4.928	5.000
LDS_2	3.760	4.182	4.654	3.144	3.143
LDI_2	4.393	3.973	4.890	3.608	3.500
LSV_2	2.933	4.034	4.181	3.031	3.036
POD_2	2.493	4.405	4.598	3.629	1.000
CAM_2	4.013	4.493	4.654	3.959	2.321
CRA_2	3.980	4.784	4.827	4.309	1.536
EDG_2	2.553	4.250	4.378	3.237	1.250
RSV_2	3.153	4.068	4.299	3.206	3.071
RDS_2	3.733	4.216	4.622	3.155	3.286
RDI_2	4.313	3.885	4.827	3.598	3.429
CDO_2	4.867	4.905	4.764	4.072	4.821
CDS_2	4.953	4.973	4.921	4.474	4.929
LDS_3	3.487	3.791	4.394	3.103	2.750
LDI_3	3.167	3.189	4.291	2.784	2.857
LSV_3	3.220	3.784	3.945	3.072	2.857
POD_3	2.173	3.797	4.102	2.979	1.000
CAM_3	3.893	4.486	4.386	3.887	2.643
CRA_3	3.160	4.047	4.205	3.536	1.607
EDG_3	2.187	3.635	3.724	2.670	1.214
RSV_3	3.287	3.736	3.969	3.041	2.357
RDS_3	3.187	3.480	4.315	3.010	2.857
RDI_3	3.120	3.162	4.323	2.856	2.786
CDO_3	4.393	4.682	4.346	3.371	4.357

The cluster centroids for each cluster group varies for all the attributes. The centroid of the clusters for each of these attributes is shown in Table 13. As already mentioned, there are 36 variables in the cluster analysis. The variable LDS_1 indicates left drainage shape in the first year of survey, that is, in 2009. Similarly, LDS_2 and LDS_3 represent the left

TABLE 14: Distances between the cluster centroids.

Cluster	1	2	3	4	5
1	0.000	5.016	6.212	3.597	6.431
2	5.016	0.000	2.969	4.458	10.316
3	6.212	2.969	0.000	6.304	11.668
4	3.597	4.458	6.304	0.000	7.893
5	6.431	10.316	11.668	7.893	0.000

drainage shape in second year (2010) and third year (2011) of surveys, respectively. Similarly the other attributes are also represented. Apart from this, the distance between the cluster centroids are mentioned in Table 14.

6. Pavement Deterioration Models

The prediction of pavement condition is important in any pavement management system. Many pavement prediction models like regression, Markov prediction models, stochastic models, and so forth, are generally available. In this work, regression models are used. Separate models are developed for each of the identified clusters. To compare the effect of clustering on the deterioration models, a model is developed for all the sections without clustering. The models would help to optimize the scheduling of the rehabilitation activities and to determine the funding level required to achieve a predetermined level of performance.

The rate of deterioration of the pavement depends on the age of the pavement and the condition of the pavement. Hence, the parameters considered for projecting the condition of the pavement are the age of the pavement and the condition of the pavement in the previous year. Multiple linear regression analysis is used where there is one dependent variable and two or more independent variables. Separate linear regression models were developed for different cluster groups. The general form of the equation is

$$VCI_{in} = \beta_0 + \beta_1 \text{age}_{in} + \beta_2 VCI_{i(n-1)}, \quad (2)$$

where VCI_{in} = Visual Condition Index of the pavement section i in year n , $VCI_{i(n-1)}$ = Visual Condition Index of the pavement section i in year $(n - 1)$, and age_{in} = age of the pavement section i in year n .

The coefficients of (2) obtained for different clusters and without cluster model are given in Table 15.

To illustrate the effect of cluster models, a section with PCI 100 with the year of construction as 2011 is assumed. The deterioration of the section if it is in different clusters is shown in Figure 3(a). A graph is also drawn for a particular

TABLE 15: Deterioration model coefficients for K -means clusters.

Cluster	β_0	β_1	β_2	R^2
1	-0.395	-1.247	1.020	0.724
2	-1.524	-0.972	1.029	0.823
3	-7.055	-0.566	1.062	0.737
4	2.155	-1.225	0.908	0.737
5	5.159	-1.043	0.927	0.813
Without cluster	-5.635	-0.503	1.001	0.801

value of VCI for the without cluster model in Figure 3(a). Similarly, Figure 3(b) shows the deterioration models with clusters and without clusters for the VCI value of 80 in 2011 with the year of construction as 2005. In the Figures 3(a) and 3(b), the notations $k1$ to $k5$ indicate the graphs for the models of clusters 1 to 5, respectively, and NC represents without cluster model. If the projection of the pavement sections is done according to the without cluster model, then the prediction of the pavement condition will be either higher or lower than the original VCI value.

From the above models given in Table 15 and from the graphs in Figures 3(a) and 3(b) it is well established that grouping all the pavement sections in one basket and developing a single model for all the road sections will result in either underestimation or overestimation of the pavement condition.

7. Conclusions

In this paper the pavement sections are clustered using k -means clustering technique. The attributes considered are the condition of shoulders, drains, cross drainage structures, and camber, and pavement distresses, namely, potholes, crack area, and edge break, collected at every 200 m section. Five clusters were chosen considering the criteria of minimum number of sections and the variance of the clusters. Separate models were developed for each of these clusters and also one model is developed for the entire road sections without clustering. From the models and the examples shown in above sections, the need and the importance of clustering of road sections are highlighted.

References

- [1] K. C. P. Wang and Q. Li, "Gray clustering-based pavement performance evaluation," *Journal of Transportation Engineering*, vol. 136, no. 1, Article ID 004001QTE, pp. 38–44, 2010.
- [2] Z. Luo and E. Y. Chou, "Pavement Condition Prediction using Clusterwise Regression," in *Proceedings of the Transportation Research Record Annual Meeting CD-ROM*, 2006.
- [3] J. Yu, E. J. Chou, and Z. Luo, "Development of linear mixed effects models for predicting individual pavement conditions," *Journal of Transportation Engineering*, vol. 133, no. 6, pp. 347–354, 2007.
- [4] Z. Luo and H. Yin, "Probabilistic analysis of pavement distress ratings with the clusterwise regression method," *Transportation Research Record*, no. 2084, pp. 38–46, 2008.
- [5] D. A. Niemeier, J. M. Utts, and L. Fay, "Cluster analysis for optimal sampling of traffic count data: air quality example," *Journal of Transportation Engineering*, vol. 128, no. 1, pp. 97–102, 2002.
- [6] Y. Luo, X. Guo, H. Li, and X. Zhu, "The traffic safety study based on cluster analysis and sampling theory," in *Proceedings of the International Conference on Intelligent Computation Technology and Automation (ICICTA '08)*, pp. 564–568, October 2008.
- [7] M. N. Postorino and G. M. L. Sarne, "Cluster analysis for road accidents investigation," in *Proceedings of the International Conference on Advanced Computation Methods*, vol. 1, pp. 451–469, 2006.
- [8] W. Jia and T. Tingting, "Research on ordinal cluster division for maintenance road section," in *Proceedings of the 2nd International Conference on Intelligent Computing Technology and Automation (ICICTA '09)*, pp. 707–710, October 2009.
- [9] V. Sunitha, A. Veeraragavan, K. K. Srinivasan, and S. Mathew, "Application of factor analysis in maintenance management of low volume roads," *International Journal of Pavement Research and Technology*. In press.
- [10] B. S. Everitt, S. Landau, and M. Leese, *Cluster Analysis*, Oxford University Press, New York, NY, USA, 4th edition, 2001.
- [11] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297, 1967.

