International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 07 | July 2018 www.irjet.net p-ISSN: 2395-0072

CRASH-PREDICTION ON NH44 (Previous NH1A) FROM LOWER MUNDA QAZIGUND TO KHANABAL, J&K

Mir Basit Mohmmad[1], Er Rajat Balyan[2], Dr Pooja Sharma[3], Burhan Nabi[4]

[1,4] M. Tech student , Civil Engineering Dept, Desh Bhagat University, Punjab, India
[2]Assistant Professor, Civil Engineering Dept , Desh Bhagat University, Punjab, India
[3]Head of Department, Civil Engineering Dept, Desh Bhagat University, Punjab, India
Civil Engineer ,NIT Srinagar, J&K.

***

Abstract - *The highway planning and design saw a paradigm shift when Henderson established that crashes are not only due to the negligence of the driver but is a combination of road infrastructure and vehicular characteristics. This and the several other related developments makes it possible to relate the crashes with infrastructure characteristics a study known as crash prediction models. Crash prediction model is used to access the safety of the planned and existing highways. The main objective of the study is to develop an infrastructure coefficient out of otherwise independent infrastructure characteristics using statistical techniques of principal component analysis and Analytic hierarchy Process that could reflect the overall quality of the highway and its related level of safety. The infrastructure characteristics include length of the segment, lane width, shoulder width, shoulder drop off, percent of highway with no passing zone etc. the infrastructure coefficient developed for the highway segment is used as a variable in crash prediction models and the relation between the two is calculated using regression analysis.*

Principal component analysis is a statistical technique to reduce the dimensionality to the visible number of dimensions in the present study from nine to two.

Analytic hierarchy process assigns the Weightage to the infrastructure characteristics (in present study four) based on their contribution to the crashes and hence to the safety of the highway segment using statistical techniques.

In the present study an IC value was calculated using both the techniques for each of the highway segment chosen and then two crash prediction models were developed.

Key Words: **Road Infrastructure, Crash Prediction Models, Principal Component Analysis, Analytic Hierarchy Process, Regression Analysis, Infrastructure Coefficient**

1. INTRODUCTION

Crashes usually result from a combination of four contributing elements viz the driver, the road, the vehicle, and the environment. Drivers are often involved in crashes not only because of their own errors, but also because they are affected by a combination of highway and/or vehicular characteristics. It is certainly not only the driver who bears ultimate responsibility for the occurrence of crashes. Henderson (1971) suggested that focusing too much on the driver as the cause of a crash often masked the ability to see other causes that could reduce crash rates and crash severity. Crash-prediction models enable highway engineers to provide an estimate of expected crash frequency as a function of traffic volume and roadway infrastructure characteristics over a highway segment. Such estimates are prerequisite for consideration of safety in highway planning and design. Crash modeling has attracted considerable research interest because of its wide variety of applications, important practical implications and most importantly human lives that are at risk during crashes.

Most of the previous work done including **Mayora and Rubio** on the development of crash prediction models concentrated on different regression methods, such as simple linear regression, simple quadratic regression and generalized linear models, including Poisson and negative binomial regression.

A bulk of the road accidents happen in cities, accounting for 43.2% of total the road accidents in the country every year. As per Transport research wing's 2016 report on Road Accidents in India, 2.98 lakh road accident took place on city roads, accounting for more than 80,000 deaths.

As per **National Crime Records Bureau (NCRB)** report **J&K** topped the list of "high accidental death- prone areas" with survival chances being only 36% as against of national average being 63.6%. As per J&K Police (Crime Department) data, Lower Munda-Khanabal stretch (NH44) has witnessed 80 deaths and 502 injuries in a total of 465 accidents during a period of five years between 2011 to 2015, therefore the study develops the **infrastructure coefficient** for the same.

2. Methods Used

2.1. Principal Component Analysis-Method to compute IC

Since the number of infrastructure variables involved were Nine (9) on Five (5) highway segments and to relate the infrastructure characteristics with the crash rates requires a Nine (9) dimensional space which provides the five (5) points in that space. With so much of dimensionality involved it becomes practically important to reduce the dimensionality to have a better representation of the data involved. A statistical approach called Principal Component Analysis was used to serve for the same. Principal Component Analysis (PCA), developed in the 1930s by Hotelling reduce the dimensionality to two (2). This technique helps us in correlating the infrastructure characteristics with the crash rates and hence the safety of a planned and/or existing highway.

The 5 highway segments, characterized by the 9 infrastructure variables, can be described by 5 points in 9 dimensional spaces. Principal Component Analysis (Cooley and Lohnes,1962) provides a method of reducing dimensionality to a visible number of dimensions, in this case from 9 to 2 dimensions. A two-dimensional plot has the following two advantages

1: Two-dimensional plot provides more perceptible display of the two clusters of poor and good roads than the three-dimensional plot.

2: The amount of variability explained by two components (a two-dimensional plot) was found to be 58% and that explained by three components (a three-dimensional plot) was 68%. This increase in the variability explained is less significant than the additional benefit resulting from better perception of the two dimensional plot. Principal Component Analysis computes the “distance” between each pair of points in the 9 dimensions. The distance may be zero if all 9 infrastructure components have the same value; the value increases with increased variability between components.

The purpose of Principal Component Analysis is to find 5 points in two dimensions (xi, yi) for highway i, such that the distances between these points are as similar as possible to the distances computed with the original 9 dimensions. Therefore, the distance between points (highways) in the two-dimensional plot represents the degree of similarity between infrastructure components: the closer the points, the more similar the highways. It is clear that axis rotation and shifting do not change the distances between the points. Since the data contains several variables with different units, we normalized all data designated for use in this analysis. The alphabet next to each data point is the road designation used as a convention in the research. It demonstrates that highways formed two groups based on their crash rate values:

1:- Lower crash-rate roads (safe roads with a crash rate equal to or less than 0.25 crashes per million vehicle-km)

2:-Higher crash-rate roads (i.e., dangerous roads with crash rates greater than 0.25 crashes per million vehicle-km).

This study differentiated higher crash-rate roads from lower crash-rate roads based on their infrastructure characteristics only and hence this finding shows that the crash rate is a function of the infrastructure characteristics.

. The “score” that a highway receives during the Principal Component Analysis on the Y-axis is, in fact, its infrastructure coefficient (ICPCA), which represents the overall infrastructure characteristic of that highway.

The Infrastructure Coefficient (ICPCA) is given in *Equation 1* as:

$$ICPCA = -0.094 + 0.7045 \times LW - 0.6894 \times NPZ + 0.1329 \times TOP + 0.1138 \times RC + 0.1253 \times$$

$$SW + 0.0108 \times SDR + 0.0365 \times RSS \quad \ldots\ldots\ldots\ldots\ldots \ (4)$$

Where

LW = Lane width (m.)

NPZ = Percentage of highway with a no-passing zone___

TOP = Topography

RC = Road consistency

SW – Shoulder width (m.)

SDR –Shoulder drop-off (cm.)

RSS – Road-Side Score

Note that one of the infrastructure elements is road consistency, this needs to be calculated separately according to the Polus et al. (2005) model as is discussed earlier. The importance of the ICPCA coefficient is that roadway engineers can rank the different roadway segments according to the resulting ICPCA, which represents the overall infrastructure characteristic of a highway segment and its level of safety i.e proneness to crashes.

2.2. Analytic Hierarchy Process-Method to compute IC

The Analytic Hierarchy process is an another method to compute Infrastructure Coefficient. The Analytic Hierarchy Process (AHP), first developed by Thomas Saaty (1980), is a mathematical decision-making technique that incorporates both qualitative and quantitative factors. The Analytic Hierarchy Process is used to rank highway infrastructures based on their Weightage to crashes and hence to the overall safety of the highway segment.

This was done by attributing a specific weight to each infrastructure characteristic. These weights are determined by the AHP method. The Infrastructure Coefficient (ICAHP) for a specific highway segment can be calculated by multiplying the weight of each infrastructure characteristic by its appropriate infrastructure-characteristic value for the specific segment and adding up the products. Highway segments having high ICAHP values represent a relatively good quality of highway design (with low crash rates) whereas segments with low ICAHP values represent a relatively poor quality highway design (with high crash rates).For using this method the actual infrastructure characteristics were converted to the corresponding variables.

The physical infrastructure characteristics were converted into ranges, separated by thresholds, and substituted the values with a score for each range. Low scores (such as 1) represented a poorly designed, seemingly dangerous infrastructure element, and higher scores (e.g., 7 for the road-side characteristics) an apparently safe and well designed infrastructure element. These elements received a particular nominal numerical score that represented the attributes of the infrastructure and its relative risk to drivers.

Scores for 5 of the 9 infrastructure elements are presented in *Table 4*, while *Table 3* presents the road-side scores. Some thresholds that we established in order to allocate the different infrastructure characteristics to representative ranges were based on engineering and common-sense judgment. Others were set by dividing the whole domain into an equal number of ranges. For example, shoulder width was categorized into four ranges. The first range-category included all highway segments with a shoulder width that was less than or equal to 0.9 m; this threshold was set, since part of the car would intrude into the through lane when a driver decides to stop on the shoulder; for example, in emergency situations. This is due to the fact that this shoulder width is less than the average width of a car. The second category contains shoulder widths of between 0.9 m and 1.8 m. In this case, most of the car's width would be within the shoulder; however, the shoulder width is still not enough to give a driver sufficient space to remain solely on the shoulder for a repair if needed. The third category (1.8 m – 2.4m) provides enough space for both the car and the driver's movement around the car; however, it is not enough for trucks. Lastly, category four (2.4 m – 3.0 m) provides sufficient shoulder width for trucks to safely park clear of the traffic lane. Shoulder drop-off is categorized into two levels. The first category includes highway segments with shoulder drop-offs of less than 5 cm. In this case, run-off-the-road instances generally do not cause loss of control. When the shoulder drop-off is greater than 5 cm, running off the shoulders onto a road-side area will in most cases result in a serious crash.

A similar approach was used to set the thresholds of road consistency, topography, and lane width. Thresholds of the remaining infrastructure characteristics were set by dividing the variables ranges into equal-size bars.

In order to use the Analytic Hierarchy Process, it was important to understand the relative safety importance of each infrastructure characteristic involved in the crashes. Several regression analyses of the correlation between crash rates and each infrastructure parameter were conducted prior to the analysis by the AHP method. For example, *Figure 1* shows the relationship between crash rate and road consistency, and *Figure 2* the relationship between crash rate and lane width. It can be observed that as road consistency improves and as lane width widens, crash rates decrease. The purposes of this preliminary study was

1: To find the relationships between individual infrastructure characteristic and crash rates (the relation observed agreed with engineering judgment and the results of previous researches).

2: To choose the infrastructure characteristics with the most significant relationship to road crash rates.

Table 6 shows the infrastructure characteristics chosen for the construction of the Infrastructure Coefficient (ICAHP) by Analytic Hierarchy Process, in descending order of their importance to crash rates. As depicted in, 9 infrastructure characteristics were chosen for analysis by AHP method whereas other remaining infrastructure characteristics, such as shoulder width, shoulder drop-off, and topography, were actually indirectly included in the road-side scale developed and taken into account in *Table 6.* The rest of the infrastructure characteristics were found to be correlated marginally and hence were excluded from further analysis.

It was difficult to identify a sufficiently large pool of experts in highway-safety design in order to obtain a reliable ranking of the relative importance to safety of each infrastructure characteristic. Therefore, the approach adopted was to determine the importance of each element based on the R-square results of the regression relationships found in the preliminary analysis. For example, because road consistency explains about 43%(R^2=0.43) of the crash-rate variance (see *Figure 1*), which was the highest among the infrastructure characteristics taken for analysis , it was considered in the analysis to be the most important road characteristic for safety and, therefore, given a rank of 4) In contrast, percent of highway with no passing zone is the least important to safety because of its lower R-square in the same analysis and hence was given a rank of 1 .

To find the relative importance of the infrastructure characteristics used in AHP, it was necessary to construct a matrix of pair wise comparisons for the infrastructure characteristics. The pair wise comparisons describe the relative safety importance of each two infrastructure characteristics. To achieve this, Saaty's scale (1980), which helps to determine pair wise judgments, was used. Saaty's scale consists of 7 levels, in which the mid-level equals 1, indicating that the two variables compared are of the same importance under a given condition. The highest level suggests that if objective i is much more important than objective j, then the pairwise-judgment value equals 8.

However, if objective i is much less important than objective j, then the pairwise-judgment value equals 1/8. Therefore, the possible values in descending order are as follows: 8, 4, 2, 1, 0.5, 0.25, and 0.125.

In this analysis, the objectives are the 4 infrastructure characteristics chosen and the criterion according to which the objectives are compared is road safety. For example, when comparing road consistency and percent of highway with no passing zone, there is a difference of 7 ranks between these characteristics (which is the maximum difference between any two infrastructure characteristic ranks – refer *Table* 6) because road consistency (Rank=8) is more important to safety than percent of highway with no passing zone (Rank=1). When calculated according to Saaty's scale, road consistency is much more important to safety than percent of highway with no passing zone; therefore, aij=8. Based on *Table 6* and Saaty's scale, we constructed the matrix of pair wise comparisons of infrastructure characteristics – Matrix "A" – which is presented in *Table 7*.

In Matrix "A," the number in the ith row and jth column gives the relative importance of the infrastructure feature in the ith row compared with the infrastructure feature in the jth column. The problem that remains is to correspond a set of weights, W_1,W_2. . ., Wn, from Matrix "A" for the objectives O_1, O_2,. On (infrastructure characteristics) following an understanding of how the pair wise comparisons aij convert to weights Wn. In this case, the largest eigen value of Matrix "A," shown in *Table 6*, is 5.096, resulting in a consistency index of 0.024, which is considered to be sufficiently close to 0. The corresponding eigen vector of the weights (normalized so that they add up to 1) is presented in *Table.*

Now, the Infrastructure Coefficient (ICAHP) can be computed for each roadway segment, and

the 5 roadway segments ranked by multiplying each of the weights by the appropriate infrastructure-characteristic value for each highway segment and then summing up the results as shown below

$$ICAHP= 0.26 \times LW + 0.14 \times NPZ + 0.45 \times RC + 0.15 \times RSS \ldots\ldots\ldots\ldots\ldots (5)$$

Where , the parameters involved have the same meaning as that in equation 4. The coefficients of *Equation 5* are the calculated weights, taken from collected data.

It is important to remember that for the purpose of this analysis and the statistical method used, it was preferable to convert the actual physical dimension of each infrastructure characteristic (lane width, road consistency, percent of highway with no- passing zone, etc.) to categorical variables.

Furthermore, since the coefficients of *Equation 5* are actually normalized weights, it was necessary to present the infrastructure characteristics in terms of the nominal variables of an equal number of categories (in our case, 4 categories) in each scale. Low scores on these scales indicate poor infrastructure quality (e.g., narrow lane width, bad consistency, etc.), and high scores a good infrastructure quality.

3. Crash- Prediction Models

The level of safety of planned or existing roadways can be estimated by relating its IC value to crash rates. The IC values were calculated using PCA and AHP method .This is important when assessing various alternatives and conducting an economic evaluation of cost benefit analysis, when it is necessary to allocate funds to the most cost- and safety-efficient projects. Roadway characteristics then could be converted to safety levels by using this IC coefficient. The relationship between crash rates (crashes per ten thousand vehicle kms.) and the infrastructure coefficients developed by PCA and AHP are shown in *Equation 6* and *Equation 7*:

$$CR\ (PCA) = 62.01 \times \exp\mathbf{-0.68 \times (ICPCA)} \ldots\ldots\ldots\ldots\ldots (6)$$

$$R^2 = 0.12$$

$$CR\ (AHP) = 13.39 \times \exp\mathbf{-0.33 \times (ICAHP)} \ldots\ldots\ldots\ldots\ldots (7)$$

$$R^2 = 0.13$$

The relationships between crash rate and the infrastructure coefficient according to both methods are presented in *Figure 6*. For each highway section, three values were calculated: the IC according to the PCA model (*Equation 4*); the IC according to the AHP model (*Equation5*) and crash rates. Based on these calculated values, data points are plotted in *Figure 6*, in which each highway segment appears twice. The alphabet next to each data point is the road designation adopted as convention. The calibrated models based on both statistical methods are also presented in *Figure 6*.

The relationships in *Equation 6* and *Equation 7* are a function of the infrastructure coefficients developed and presented in *Equation 4* and *Equation 5.* Equations 6 and 7 can be used to predict crash rates on new or existing two-lane highways, based on their infrastructure characteristics. The linear correlation coefficients between each two infrastructure features were examined. Some infrastructure features were strongly correlated for example, lane width with road consistency (0.75), road-side score with lane width (0.65) and shoulder widths with lane width (0.85). Some of these correlations were expected based on engineering judgment. For example, guardrails would be used more in an area with mountainous terrain, which also has less design consistency and more no-passing zones. Furthermore, roads with high standard design elements often, although not always, have good quality elements in all their geometric features, and these are correlated. These correlations, however, may preclude the use of the models presented (*Equations 6* and *7*) to identify the exact contribution of a specific individual element to expected crash rates. Nevertheless, the use of the models to estimate the crash rates of roads based on their infrastructure coefficient is still valid. Other possible effects are included in the "error term" as is often done in regression analysis. It is not claimed that they do not exist, however they are not considered in these models. The Analytic Hierarchy Process method, discussed earlier, was used to identify the parameters that contribute the most, and relative weights were given to the most important infrastructure parameters i.e., those that most reflect a relative importance to safety.

It was found that there is a significant similarity between the two models (notice *Figure 6* – the two lines are very close to each other). Part of this similarity is caused by the fact that the AHP method was based on the correlation results (R-square values presented in *Table* an approach that is similar to that used by the PCA method. Even though there is a greater similarity between the two models yet the Analytic Hierarchy process based on four infrastructure characteristics seems to be the practical solution for assessing the highway safety as against the Principal Component Analysis based on 9 Infrastructure characteristics because of its simplicity to use and better explanation of crash rate variability (R^2 being 0.138 as against 0.105 in PCA).

List of Figures

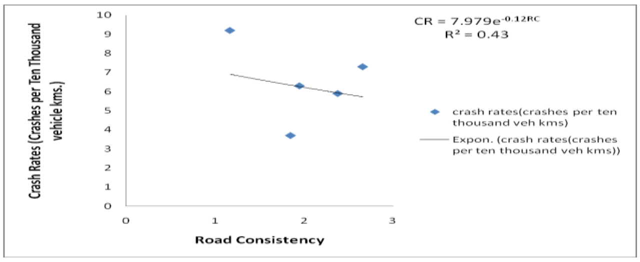

Figure 1 Crash Rates vs. Road Consistency

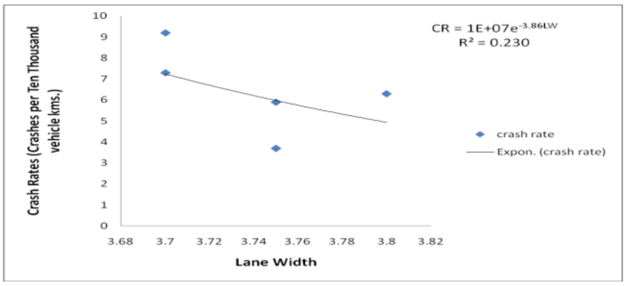

Figure 2 Crash Rates vs. Lane Width

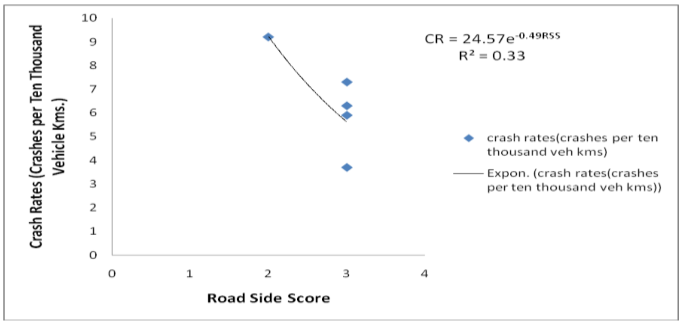

Figure 3 Crash Rates vs. Road Side Score

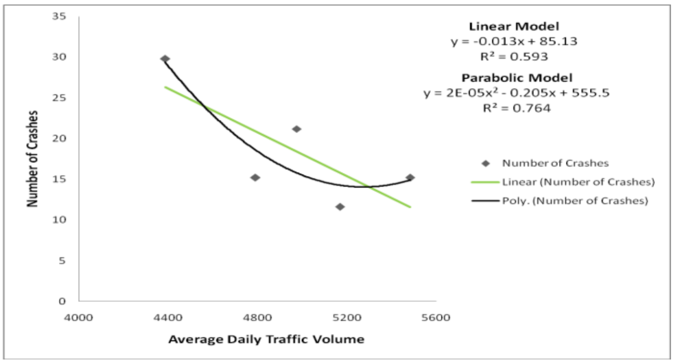

Figure 4 Relationship between Number of Crashes and Average Daily Traffic Volume

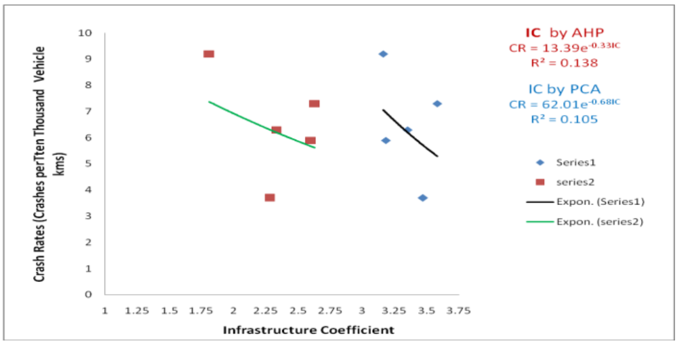

Figure 5 Relationship between Crash Rates and Infrastructure Coefficients

3. Results and conclusions

The purpose of this research was to develop an Infrastructure Coefficient (IC) that represents the overall characteristic of a highway and to develop models by two different methods that correlate this IC with crash rates on two-lane rural highways. The two statistical methods used: Principal Component Analysis (PCA) and Analytic Hierarchy Process (AHP).

The two Infrastructure Coefficients developed succeeded in distinguishing between lower crash-rate roads and higher crash-rate roads by the difference in their overall infrastructure characteristics. Furthermore, these

Infrastructure Coefficients enable highway planners and safety auditors to predict crash rates based on the infrastructure features of the entire highway. These coefficients can be used when evaluating several alternatives for a new highway or when rehabilitating and upgrading existing highways in order to improve their overall safety features.

REFERENCES

1) National Crime Records Bureau 2013 survey and 2017 report

2) RAMKY INFRASTRUCTURE PRIVATE LTD. Hyderabad based company currently working on NH44 fron Banihal to Srinagar . Road Geometrics

3) COOLEY, W.W. and LOHNES, P.R. (1962). Multivariate Procedures for the Behavioural Sciences. Wiley, New York.

4) J&K Traffic Police (Crime Department Qazigund). Crash Data From 2011 to 2015

5) HADI, M.A., ARULDHAS, J., CHOW, L.F., and WATTLEWORTH, J.A. (1993). Estimating Safety Effects of Cross-Segment Design Various Highway Types Using Negative Binomial Regression. Transportation Research Record 1500, p. 169

6) HENDERSON, M. (1971). Human Factors in Traffic Safety: A Reappraisal. Traffic Accident Research Unit, Department of Motor Transport, New South Wales, Australia.

7) HOTELLING, H. (1933). Analysis of a Complex Series of Statistical Variables into Principal Components. Journal of Educational Psychology 24, p. 417, p. 498.

8) 8. JOSHUA, S.C and GARBER, N.J. (1990). Estimating Truck Accident Rate and Involvement Usin Linear and Poisson Regression Models. Transportation Planning and Technology 15,p. 41.

9) KARLAFTIS, M.G. and GOLIAS, I. (2002). Effects of Road Geometry and Traffic Volume on Rural Roadway Accident Rates. Accident Analysis and Prevention 34(3), p. 357.

10) MAYORA, J.M.P. and RUBIO, R.L. (2003). Relevant Variables for Crash-Rate Prediction on Spain's Two-Lane Rural Roads. Transportation Research Board 82nd Annual Meeting. Washington, DC.

11) POLUS, A., POLLATSCHEK, M. A., MATTAR-HABIB, C., and JARROUSH, J. (2005).

12) An Enhanced Integrated Design-Consistency Model

13) for Both Level and Mountainous

14) Highways and Its Relationship to Safety. Road and

15) Transport Research 14(4), p. 13

16) SAATY, T.L. (1980). The Analytic Hierarchy Process, McGraw-Hill, New York.

17) ZEGEER, C. V., REINFURT, D. W., HUMMER, J., HERF, L., and HUNTER, W. (1988).

18) Safety Effects of Cross-Section Design for Two-Lane

19) Roads. Transportation Research

BIOGRAPHIE

"The author has done B.Tech in civil Engineering from Punjab Technical university and has Pursued M.Tech in Transportation Engineering From Desh Bhagat University. Author is a Resident of Bumthan MirBazar Anantnag, J&K.