

FERIT YAKAR, Ph.D.
E-mail: ferit.yakar@gop.edu.tr
Gaziosmanpasa University
60250 Tokat, Turkey

Safety and Security in Traffic
Review
Submitted: July 14, 2014
Approved: Nov. 24, 2015

IDENTIFICATION OF ACCIDENT-PRONE ROAD SECTIONS BY USING RELATIVE FREQUENCY METHOD

ABSTRACT

Assuming that traffic accident occurrence is determined by some road and environment-related factors, and that future traffic accidents will occur under the same conditions as past traffic accidents this study investigates the use of Relative Frequency Method (RFM) (also called frequency ratio method) in the determination of accident-prone road sections. In this method, the number of accidents or accident rate is not used directly; instead, a relationship between accident number and environmental properties of road sections is established and this relationship is used for accident-prone road section identification. Thus, limitations of raw risk estimators are minimized. Besides, the effect of local conditions to accident proneness of a road can be reflected in a better way. No inherent assumptions about the distribution of the accident data are needed to apply this method. The method was tested on a highway in Trabzon province of Turkey. At the end of the study, sensitivity and specificity values were calculated as 1.00 and 0.83, respectively, which reflects that the method identified all of the "accident-prone" sections (there is no false negative) and the method has very strong ability to distinguish "relatively safe" sections. The most useful property of this simple method is that, if accident data do not exist due to any reason for some part of the road, the method can be still used to identify accident-prone sections by using road properties.

KEY WORDS

Relative Frequency Method; Accident-Prone Road Sections; Road Safety;

1. INTRODUCTION

The first step in road safety improvement process is the identification of black-spots [1]. In a road network, if accident experience of a point or a segment is higher than the other segments of the network, most probably, the reason are the road and/or environmental conditions in that point or segment. Places like these are called accident-prone locations (also called hot spots, black spots, hazardous locations, sites with promise, risky sections, etc.).

There are two classic approaches for determining accident-prone locations. One is based on the observed number of crashes, and the other is based on regression analyses [2]. As Miranda-Moreno et al. [3] stated, the raw risk estimators have several limitations as discussed in a number of studies [4, 5]. Specifically, a ranking method relying on raw accident rates may produce large numbers of misclassifications (e.g. selecting relatively safe locations as accident-prone ones or vice versa) due to the random variation of traffic accidents from year to year [4, 6, 7, 8]. False positives (observation of elevated crash frequencies at a relatively safe site) lead to investment of public funds with little to no safety benefits. On the other hand, false negatives (an unsafe site that does not reveal elevated crash frequencies) lead to missed opportunities for effective safety investments [8].

Statistical models, such as Poisson or negative binomial regression models, have been employed to analyse vehicle accident frequency for many years. However, these models have their own model assumptions and pre-defined underlying relationship between dependent and independent variables. If these assumptions are violated, the model could lead to erroneous estimation of accident likelihood [9]. Recently some researchers have proposed 'distribution free' methodologies for the analysis of crash data. No inherent assumptions about the distribution of the crash frequency data are needed to apply these techniques, which are essentially driven by observed data. These data-driven techniques are powerful data analysis tools [10].

Many alternative ranking methods have been proposed for accident-prone location identification. Geurts and Wets [11] prepared a literature review about methods and techniques that are used to analyse black spots and black zones.

Elvik [12] studied the variables and methods used to identify hazardous road locations in eight European countries. The countries are Austria, Denmark, Flanders (the Flemish speaking region of Belgium), Ger-

many, Hungary, Norway, Portugal and Switzerland. It is stated that most of the approaches used in the countries surveyed were primitive and are likely to involve substantial inaccuracies. Most operational definitions of hazardous road locations were found: (1) not to refer to any population of similar sites, (2) to rely on a sliding window approach, and (3) identify hazardous road locations in terms of the recorded number of accidents.

Kwon et al. [13] evaluated the performances of three different methods for segmenting freeway sites to identify high collision concentration locations: Sliding Moving Window, Peak Searching and Continuous Risk Profile. For each of these three methods, the traffic collision data were used to estimate the excess expected average crash frequency with Empirical Bayes adjustment and the resulting lists were compared with the previously confirmed high collision concentration locations (or hot spots). The findings revealed that the Continuous Risk Profile method has the lowest false positive.

Montella [14] compared seven commonly applied Hot Spot Identification methods (crash frequency, equivalent property damage only crash frequency, crash rate, proportion method, empirical Bayes estimate of total-crash frequency, empirical Bayes estimate of severe-crash frequency, and potential for improvement) against four evaluation variables (the site consistency test, the method consistency test, the total rank differences test, and the total score test). He concluded that the empirical Bayes method should be the standard in the identification of hotspots.

Cheng and Washington [8] used experimentally derived simulated data, and evaluated three hot spot identification methods observed in practice (simple ranking, confidence interval, and Empirical Bayes) in terms of percent false negatives and positives. Also, the effects of crash history duration are assessed. The results illustrate that the Empirical Bayes technique significantly outperforms the ranking and confidence interval techniques. False positives and negatives are inversely related, and three years of crash history appears, in general, to provide an appropriate crash history duration.

Qin et al. [2] stated that crash data often have skewed distributions and exhibit substantial heterogeneity. Changes at mean level do not adequately represent patterns present in the data and used quantile regression technique for identifying intersections with severe safety issues. They suggested that relative to other methods, quantile regression yields a sensible and much more refined subset of risk-prone locations.

Ghaffari et al. [1] compared the reliability analysis method with the commonly implemented Frequency and Empirical Bayesian methods using simulated data in identification of black -spots. The results indicated that the traditional methods can lead to an inconsistent prediction due to their inconsider-ation of the

variance of the number of crashes on each site and their dependence on the mean of the data.

Pirdavani et al. [15] developed a model to prioritize accident hotspots by using Multiple Criteria Decision-Making method, when traffic accident data are not available. The model is validated against an existing database of road sections and a sensitivity analysis is carried out on the proposed method.

Sadeghi et al. [16] represented a method to identify and prioritize accident-prone sections, which incorporates the segmentation procedure into data envelopment analysis technique.

In Turkey, the method used by General Directorate of Highways (GDH) for identifying black spots is called Rate-Quality-Control Method. This method is a statistical method and consists of calculating three different parameters (accident rate, accident frequency, and severity index) for each one kilometre long road section. Each of these values is compared with a critical value and if a certain road section has higher values than the critical ones for all these three parameters, the section is considered a black spot. This method is highly dependent on accident numbers, which can fluctuate from year to year and may lead to misclassifications. Also, the selection of the critical values are subjective judgments.

In internationally recognized iRAP methodology [17], the information taken from the road and the surroundings by a specially equipped vehicle is then analysed to assess the infrastructure risk factors. Star Rating Scores for road segments are calculated by using values taken from the tables. These tables were created by using previous studies and valid for all roads. On the basis of this analysis, the appropriate action plans are determined, in order to improve road safety [18].

In this study, Relative Frequency Method (RFM) (also called frequency ratio method, likelihood-frequency ratio method, likelihood ratio method, probabilistic-based frequency ratio method, etc.) is suggested for accident-prone road section determination. Road sections are the units of analysis in this method: Relative hazardousness of road sections are identified with respect to other sections of the road.

In this method, the number of accidents or accident rate is not used directly; instead, a relationship between accident number and environmental properties of road sections is established and this relationship is used for accident-prone road section identification. Thus, limitations of raw risk estimators (producing large numbers of misclassifications due to the random variation of traffic accidents from year to year) are minimized. Besides, by using accident history together with environmental properties of the road, the effect of local conditions (rules and regulations of the place, characteristics and traffic culture of the local people, etc.) to accident proneness of the road can be reflected better. Moreover, contrary to some statistical mod-

els such as Poisson or negative binomial regression, no inherent assumptions about the distribution of the accident data are needed to apply this method. One other advantage of the method is its simplicity: It is not necessary to use a special statistics software or deep knowledge of mathematics or probability. Even a simple calculator or an excel sheet is adequate to apply this method. Detailed information about RFM is given in Section "2.2. Relative Frequency Method".

2. MATERIALS AND METHODS

2.1 Study area and data

The study is conducted on 22 km (11 km west-east and 11 km east-west direction) long segment of D10 State Highway (generally known as Black Sea Coastal Highway), that is passing through Arsin and Yomra counties of Trabzon province of Turkey.

In Turkey, any accident involving deaths or injuries is officially reported and its location is recorded. This is not the case for material damage only accidents. Hence, the accidents under study are limited to those with deaths and/or injuries.

In the study, the data about accidents were obtained by one by one investigation of Traffic Accident Reports of 132 accidents with death and/or injury, belonging to 22 km study area for the years 2006-2010.

It is very important to determine the study period. From a purely statistical point of view, it is favourable to have as many accidents as possible. On the other hand, there should not be any changes at the spot (traffic flows or behaviour, geometry or surface, etc.) during the study period. This limits the size of the time-period. For these reasons, the length of the period used to identify black spots varies from 1 to 5 years and a period of 3 years is frequently used [12, 8]. In this study, since accident numbers are not so high in the study area, a 5-year period was used in order to obtain a balance between having a long period for getting many accidents and a short period so that the spot is not changed too much.

Only properties related to road itself and its environment were handled in the study. Other properties related to drivers, vehicles, weather conditions and time of the day were not handled. The properties of road and its environment were summarized from both inventory files of GDH, which is the responsible institution, and Traffic Accident Reports. These data are then confirmed by on-site investigation.

The study showed that road and environment properties may be different for each direction of the divided highway. For example, if there is a merging road in one direction, it only affects that direction and does not affect the other one. Therefore, in this study, each direction of the divided road was handled as a different road segment.

Accident-prone location determination studies should use as many factors as possible that are known to influence road safety. At the beginning of the study 17 variables were planned to be used. However, when data were collected, it was seen that the effect of some of these variables (road type, lighting conditions, sidewalk existence, shoulder existence, type of the pavement, lane width, speed limit, and the number of lanes) on accident occurrence cannot be observed in this study area, since variable value is the same for all along the road. These variables were cancelled and the remaining nine variables that are listed in the second column of *Table 1* were used in the study. In this table, the first column shows the variable number, and the third column shows the possible values for the variables. The fourth and fifth columns are used in the calculation of RF values given in the last column. These last three columns are explained in the next section (Section 2.2. Relative Frequency Method (RFM)).

The determination of the section length is very important in accident-prone location determination studies. Use of a constant length is almost compulsory because the interpretation of accident data would be more complicated for sections of variable length. On the other hand, there is no clear indication of what the best length of a dangerous road segment should be, nor or whether an optimal length can be defined [11]. If the selected section length is too long, it will be hard to ensure homogeneity in the sections. On the other hand, if the selected section length is too short, the precision of the location data may become inadequate. All of these subjects should be simultaneously taken into consideration in determination of the section length. In this study, the section lengths were decided as one km long at the beginning. However, the difficulty of ensuring the homogeneity of the section properties along 1 km was realized and this length was decided to be shortened. Then, a trial was made with a 100 m section length, but this time it was found that the precision of location data makes it impossible. Consequently, 500 m section length was chosen as an optimum length which can both ensure the section homogeneity and take into account the location data precision.

The total length of the road under study is 22 km and when this length is divided by 500 m long sections, a total of 44 sections was obtained. The sections 1 to 22 were on the west-east direction and 23 to 44 were on the east-west direction. The values of the variables on these sections as well as the number of accidents in these sections are given in *Table 2* (due to lack of space, only a small portion of that table is given). For example, on the first section, the value for the first variable (first variable is settlement variable) is 1, which indicates that this section is passing through a settlement area (as it can be seen from the third column of *Table 1*).

Table 1 - Variables, variable classes, and Relative Frequencies of variable classes

Var #	Variable	Variable Classes	# of sections in the class (C)	# of "sections with accident" in the class (A)	Relative Frequency of Class (RF)
1	Settlement situation	1 - Settlement area	13	2	2.26
		2 - Non-settlement area	31	1	0.47
2	Horizontal curvature	1 - Straight	28	2	1.05
		2 - Slight curve	16	1	0.92
		3 - Sharp curve with fences	0	0	0
		4 - Sharp curve without fences	0	0	0
3	Vertical grade	1 - Flat	32	3	1.38
		2 - Slight slope	12	0	0
		3 - Steep slope	0	0	0
		4 - Hill top	0	0	0
4	Intersections	1 - Rotary intersection (signalized)	2	1	7.33
		2 - Rotary intersection (unsignalized)	5	0	0
		3 - Other intersection (signalized)	0	0	0
		4 - Other intersection (unsignalized)	4	0	0
		5 - No intersection	33	2	0.89
5	Passages	1 - Underpass	16	1	0.92
		2 - Overpass	6	0	0
		3 - No passage	22	2	1.33
6	Presence of collector roads	1 - Collector way exists	8	1	1.83
		2 - Collector way does not exist	36	2	0.81
7	Significant buildings /places	1-Facilities for vehicles (oil station, truck park, etc.)	2	0	0
		2 - Public buildings (harbour, etc.)	5	1	2.93
		3 - Social, sport-related or religious facilities (park, sport area, mosque, etc.)	5	1	2.93
		4 - Other places	32	1	0.46
8	Bridges	1 - Bridge	8	0	0
		2 - No bridge	36	3	1.22
9	AADT	8,000	18	2	1.63
		9,500	22	1	0.67
		16,000	4	0	0

Table 2 - Variable values and the number of accidents in the sections

Section #	Start km	End km	Variable #									# of accidents
			1	2	3	4	5	6	7	8	9	
1	9.00	9.50	1	1	1	4	2	2	4	2	16,000	1
2	9.50	10.00	1	1	2	5	3	2	4	1	16,000	4
3	10.00	10.50	1	2	1	1	1	2	2	2	9,500	14
4	10.50	11.00	2	1	1	5	1	2	3	2	9,500	0
-	-	-	-	-	-	-	-	-	-	-	-	-
42	10.50	10.00	2	2	1	1	1	2	4	2	9,500	5
43	10.00	9.50	2	1	2	5	3	2	4	1	16,000	2
44	9.50	9.00	1	1	1	5	2	2	4	2	16,000	0

2.2 Relative Frequency Method (RFM)

The RFM is an easily applied probability model, and it is widely used especially in landslide susceptibility mapping literature [19, 20, 21, 22, 23, 24]. The main assumption of these studies is "In general, it is necessary to assume that land slide occurrence is determined

by land-slide-related factors, and future landslides will occur under the same conditions as past landslides" [25]. This model uses ratios of the area where the landslide occurred to the total area, normalized to have an average value of one. A ratio above one indicates a relatively higher correlation between a parameter and the

occurrence of a landslide in that area, whereas values less than one imply a lower correlation [19].

Similar to main assumption of these studies, it can be assumed that "traffic accident occurrence is determined by some road and environment related factors, and future traffic accidents will occur under the same conditions as past traffic accidents". Therefore, RFM method can be adapted for determination of accident-prone road sections. The steps of the study are as follows:

- 1) The first step is the determination of variables related to road itself and its environment, which may influence accident occurrence. A long list can be obtained from the past studies and technical reports. However, every variable found from literature cannot be used in every study due to several reasons. For example, data about a variable may not be available (due to any reason) or that variable may have the same value all along the road. Or, variables to be used may change according to the type of the road (for example, one of the most important variables for a road passing through settlements may be the existence of pedestrian facilities, whereas this variable will not have any meaning for an access controlled road). Therefore, variables should be determined carefully for each study. The possible values of each variable (that is, variable classes) should also be defined carefully.
- 2) The second step is the determination of sections. Ensuring that the comparison is possible, the section lengths should be equal. The subjects that should be taken into consideration while determining the section lengths were discussed in section "2.1. Study area and data". After determining the section length, beginning from the first point and adding the section length each time, the sections should be determined and numbered.
- 3) The third step is the determination of section properties. For each section, the value of variables should be determined. For example, while evaluating "the horizontal curvature" variable (2nd variable, as seen in *Table 1*), the values of this variable (1- straight, 2- slight curve, 3- sharp curve with fences, 4- sharp curve without fences) should be determined for all sections. For this purpose, the inventory files of GDH, Traffic Accident Reports, or field surveys may be used.
- 4) In the fourth step, a threshold is determined and sections having more accidents than this threshold are defined as "section with accident". The determination of this threshold is explained in Section "2.2.1. Determination of accident threshold and the number of risk classes".
- 5) In the fifth step, relative frequencies of variable classes are calculated by using the formula in Equation 1.

$$RF = (A/B) / (C/D) \quad (1)$$

where:

- RF – relative frequency of variable class;
- A – the number of "sections with accident" belonging to related class of a variable;
- B – the total number of "sections with accident";
- C – the number of sections belonging to the related class of a variable;
- D – the total number of sections.

The *B* and *D* values in the formula are the same for all of the variables: *D* is determined in the second step, and *B* is determined in the fifth step. However, *C* and *A* values should be calculated separately for each class of different variables. The relative frequency values calculated by Equation 1 represent the relative contribution of that class to the accident occurrence. Relative frequencies higher than 1 mean high correlation and relative frequencies lower than 1 mean low correlation.

- 6) By adding relative frequency values for all variables, the total relative frequency values are obtained for each section. The total relative frequency value of a section represents relative accident risk of that section: the higher the total relative frequency value, the higher is the accident risk.
- 7) After calculating the total relative frequency values for all sections, the task is to determine the accident-prone sections. This decision should be made by creating risk classes according to total relative frequencies of the road under investigation. Sections may be divided into two groups: "accident-prone" and "relatively safe" sections. Similarly, risk classes 3, 4, or 5 may also be created. Determining the number of classes is explained in detail in Section "2.2.1. Determination of accident threshold and the number of risk classes".

The range (the difference between maximum and minimum total relative frequency values) is divided by the number of risk classes in order to find the risk class widths. According to this width, a table of risk classes is created. By using this table, the risk class of each section is determined. The sections corresponding to the first class are determined as accident-prone sections.

2.2.1 Determining the accident threshold and the number of risk classes

Accident threshold is the threshold that determines whether a section is defined as a "section with accidents" or a "section without accidents". It is not possible to determine an accident threshold that will

Table 3 – Possible combinations (for different values of accident threshold and the number of classes), and Sensitivity – specificity values (and their sums) for these combinations

Combination #	Acc. Threshold	Class #	Sensitivity	Specificity	Sensit.+Spec.
1	0	2	0.85	0.64	1.49
2	0	3	0.15	0.91	1.06
3	0	4	0.06	1	1.06
4	0	5	0.03	1	1.03
-	-	-	-	-	-
31	7	4	0.4	1	1.4
32	7	5	0.4	1	1.4
33	8	2	1	0.83	1.83

be valid for all studies. Instead, this threshold should be determined according to the data period length and the properties of the road handled. For example, this threshold may be determined as 2 for a study using a data period of one year, whereas a threshold may be determined as 6 if the data period is three years. Similarly, this threshold will be high on the roads with high AADT value, since the number of accidents will be higher on these roads.

The number of risk classes has also great influence on the results. In the study, the sections may be divided into two, as "accident-prone" and "relatively safe" ones; however, depending on the aim of the study, more than two classes (3, 4, 5, or more) may also be used (for example, if 5 classes are used, the sections can be identified as very risky – little risky – neutral – little safe – safe).

According to the different values of the above-explained two factors, several combinations may come out. These combinations are given in the second and third columns of Table 3 (due to the lack of space, a part of the table is not given). In order to determine the best combination, the sensitivity and specificity values, which are explained in section 2.2.2, should be calculated.

2.2.2 Calculation of sensitivity and specificity

Sensitivity and specificity are statistical measures of the performance of a binary classification test, also known in statistics as "classification function". There are four possible results of a test (Table 4).

Table 4 – Possible results of a test

Test Result	Actual	
	Positive	Negative
Positive (Accident-prone)	TP (3)	FP (7)
Negative (Relatively safe)	FN (0)	TN (34)
Total	TP+FN (3)	FP+TN (41)

In Table 4:

TP (True Positive) – Actually "accident-prone", test result is also "accident-prone";

FP (False Positive) – Actually "relatively safe", but test result is "accident-prone";

FN (False Negative) – Actually "accident-prone", but test result is "relatively safe";

TN (True Negative) – Actually "relatively safe", test result is also "relatively safe".

Sensitivity (also called the true positive rate) measures the proportion of actual positives which are correctly identified as such.

$$Sensitivity = TP / (TP + FN) \tag{2}$$

Specificity (sometimes called the true negative rate) measures the proportion of negatives which are correctly identified as such.

$$Specificity = TN / (TN + FP) \tag{3}$$

A perfect predictor would be described as having 100% sensitivity and 100% specificity; however, theoretically it is very rare. Therefore, the test method is better if it has high sensitivity and specificity values. In this study, the sum of these two measures (the last column in Table 3) is used for the selection of the best combination.

2.2.3 Performance of the method

One of the methods that can be used to measure the performance of the results is the simple overlaying method. Ayalew et al. [26] used this method in order to measure the performance of their landslide sensitivity map created by using logistic regression technique. In their study, they overlaid the sensitivity map by landslide inventory map and investigated how many "cells with landslide" are present in each of their sensitivity map classes. They stated that, in high performance studies, it is desired that the percent of cells identified as "risky cells" should be as low as possible, whereas the percent of the landslide coinciding with these cells should be as high as possible.

If the same approach is adapted to this study, the number of sections in each of the risk classes and the number of accident-prone sections in these risk classes can be determined in order to measure the perfor-

mance of the method. In high performance studies, it is desired that the percent of sections identified as "accident-prone" should be as low as possible whereas the percent of the "sections with accidents" coinciding with these sections should be as high as possible. Results are given in Section "3. Results".

3. RESULTS

The procedure explained in section "2.2 Relative Frequency Method" was applied to the data. The variables and variable classes were determined and given in *Table 1*. Then, sections were determined and given in *Table 2*, and thus, the total number of sections (D value in Equation 1; for this study 44) was obtained. Then, the section properties were determined, that is, for each section, it was determined which class is valid for each variable and given in *Table 2* (due to the lack of space, a part of the table is not given). The number of accidents occurred in those sections was also given in the last column of the same table (*Table 2*).

Then, accident threshold and the number of risk classes were determined. For this purpose, sensitivity and specificity values, as well as their sums, were calculated for various combinations of accident threshold and the number of risk classes. These values were given in *Table 3*. In this Table, the combination making the sum of sensitivity and specificity values maximum (1.83) was selected, which was the combination 33 (the accident threshold is 8 and the number of risk classes 2).

By using the selected accident threshold value of 8, "sections with accident" were determined from *Table 2*. That is, sections having more than 8 accidents were identified as "section with accident", and these

sections were written with bold characters in *Table 2*. Thus, the total number of "sections with accident" (B value in Equation 1; for this study, 3) was obtained.

In the following step, the numbers of "sections with accident" belonging to each class of each variable (A values in Equation 1) and the number of sections belonging to each class of each variable (C values in Equation 1) were obtained from *Table 2*. For example, for "horizontal curvature" variable (second variable, as can be seen from *Table 1*), the number of sections belonging to "slight curve" class (second class) will be denoted as C, and the number of "sections with accident" belonging to that class (slight curve class) will be denoted as A. A and C values, which are calculated for each class of each variable were given in the fourth and fifth columns of *Table 1*. By using this numbers and Equation 1, the RF values of each class of each variable were calculated and given in the last column of *Table 1*.

RF values were assigned to sections for all variables and the Total RF values were calculated for each section by adding all RF values. RF and Total RF values were given in *Table 5* (due to the lack of space, a part of the table is not given).

After calculating the Total RF values for all sections, the range, that is, the difference between maximum and minimum total relative frequency values (18.44 – 4.20 = 14.24), is divided by the number of risk classes (for this study 2) in order to find the risk class width and by using this width, a table of risk classes is created (*Table 6*).

By using *Table 6*, the risk class of each section is determined as shown in the last column of *Table 5*. Ten sections corresponding to the first class are identified as "accident-prone".

Table 5 – RF and Total RF values for all sections

Sec. #	Settlement	Hor. Curv.	Ver. Grade	Intersection	Passages	Collector roads	Sig. Build.	Bridges	AADT	Total RF	Risk Class
1	2.26	1.05	1.38	0.00	0.00	0.81	0.46	1.22	0.00	7.17	2
2	2.26	1.05	0.00	0.89	1.33	0.81	0.46	0.00	0.00	6.80	2
3	2.26	0.92	1.38	7.33	0.92	0.81	2.93	1.22	0.67	18.44	1
4	0.47	1.05	1.38	0.89	0.92	0.81	2.93	1.22	0.67	10.34	2
-	-	-	-	-	-	-	-	-	-	-	-
42	0.47	0.92	1.38	7.33	0.92	0.81	0.46	1.22	0.67	14.18	1
43	0.47	1.05	0.00	0.89	1.33	0.81	0.46	0.00	0.00	5.02	2
44	0.26	1.05	1.38	0.89	0.00	0.81	0.46	1.22	0.00	8.06	2

Table 6 – Risk classes table

Risk class	Minimum RF value	Maximum RF value	The # of sections	The % of sections	The # of accident-prone sections	The % of accident-prone sections
1	11.32	18.44	10	22.72	3	60
2	4.20	11.32	34	77.28	2	40
Total			44	100.00	5	100

4. DISCUSSION

Traffic accident occurrence is determined by some road- and environment-related factors, and future traffic accidents will occur under the same conditions as past traffic accidents. With this assumption, the use of RFM in the determination of accident-prone road sections is investigated in this study. The RFM is widely used especially in landslide susceptibility mapping literature. Since the study area has two dimensions in landslide studies, the basic units are the cells with two dimensions. However, since the handled property is a linear engineering structure (can be assumed as one dimension) in this study, the basic units are the road sections with 1 dimension.

Contrary to many other accident-prone road section determination methods, no inherent assumptions about the distribution of the accident data are needed to apply this method.

The most important advantage of the method is its simplicity. It is not necessary to use a special software or deep knowledge of mathematics or probability. Even a simple calculator or an excel sheet may be adequate for the application. On the other hand, in order to speed up and ease the procedure, it may be beneficial to utilize a software in big studies having so many sections and so many variables. This study used a computer program that was prepared by using MATLAB software. By using this program, it becomes also possible to make several trials by changing the section lengths, the number of risk classes, or the accident threshold, very quickly.

In order to measure the performance of the method, the number of sections and the number of accident-prone sections in each of the risk classes were determined. As it can be seen from Table 6, 10 out of 44 sections (22.72%) exist in the 1st class (that is, accident-prone). On the other hand, 3 out of 5 "sections with accident" (60%) exist in the 1st class. That is, a large percent of "sections with accident" was captured in a small percent of sections.

The sensitivity and specificity values were calculated as 1.00 and 0.83, respectively. The 1.00 value for sensitivity is a very good value and it reflects that the method identifies all of the "accident-prone" sections and there is no false negative. The 0.83 value for specificity is also a good value and it shows that the method has very strong ability to distinguish "relatively safe" sections.

5. CONCLUSION

It is possible to make use of the Relative Frequency Method in various ways. For example, accident-prone sections of a road can be identified very easily and quickly in order to determine the sections to be rehabilitated. The most useful property of the method is

that, if accident data do not exist due to any reason for some part of the road, the method can be still used to identify accident-prone sections by using the road (and environmental) properties. In order to validate this property, a test was applied for measuring the performance of the method in cases where accident data are absent for some part of the road. In the test, accident data belonging to sections 15-19 were deleted and the method was applied to the remaining data. In this case, the combination making the sum of sensitivity and specificity values maximum ($0.75 + 0.98 = 1.73$) was the combination 22 (the accident threshold is 5 and the number of risk classes is 3) and for this combination, 26.52% of accidents were captured in 9.09% of sections.

Note that this property is valid for sections belonging to the same road. This method identifies the relatively risky sections of the road and can make comparisons only in the handled road sections. RF values obtained from different roads cannot be used to compare the accident-proneness of different roads.

This study dealt with accidents involving fatalities and injuries together, but in the future studies more weight can be placed on the accidents with fatalities by simply multiplying the number of accidents with fatalities with a number greater than 1. Also, the future studies may investigate the effect of section length, the effect of data period, or the effect of the number of classes.

FERIT YAKAR, Ph.D.

E-mail: ferit.yakar@gop.edu.tr

Gaziosmanpaşa Üniversitesi

60250 Tokat, Türkiye

ÖZET

Bu çalışmada, trafik kazalarının yol ve çevresiyle ilgili bazı faktörlere bağlı olduğu ve gelecekteki trafik kazalarının da geçmiştekilerle aynı şartlar altında meydana geleceği varsayımıyla, kazaya meyilli yol kesimlerinin belirlenmesinde Bağıl Frekans Yönteminin (BFY) (frekans oranı yöntemi de denir) kullanımı araştırılmıştır. Bu yöntemde, kazaya meyilli yol kesimlerinin belirlenmesinde kaza sayısı veya kaza oranı değerleri doğrudan kullanılmamakta, bunun yerine kaza sayısı ile yol kesiminin çevresel özellikleri arasında bir ilişki kurulmakta ve kullanılmaktadır. Böylece, ham risk tahminleyicilerinin sınırlılıkları en aza indirilmektedir. Üstelik, yerel koşulların yolun kazaya meyilliliğine etkileri daha iyi yansıtılabilmektedir. Bu yöntemin uygulanabilmesi için kaza verilerinin dağılımıyla ilgili bir ön şart yoktur. Yöntem, Trabzon'daki bir yolda denenmiştir. Çalışma sonunda, duyarlılık ve özgüllük değerleri sırasıyla 1.00 ve 0.83 olarak hesaplanmıştır ki bu da yöntemin bütün kazaya meyilli kesimleri tespit ettiğini (hatalı negatif bulunmamakta) ve kazaya meyilli olmayan kesimleri ayırt etmekte de çok güçlü olduğunu göstermektedir. Bu basit yöntemin en kullanışlı özelliği, yolun bazı kesimlerine ait kaza verilerinin herhangi bir sebeple mevcut olmaması durumunda dahi yolun çevresel özellikleri kullanılarak kazaya meyilli kesimlerin tanımlanabilmesidir.

ANAHTAR KELIMELER

Bağlı Frekans Yöntemi; Kazaya Meyilli yol Kesimleri; Yol Güvenliği;

REFERENCES

- [1] Ghaffari A, Kashani AT, and Moghimdarzi S. Identification of Black Spots Based on Reliability Approach. *Promet – Traffic&Transportation*. 2013; 25, 525-532.
- [2] Qin X, Ng M, Reyes PE. Identifying crash-prone locations with quantile regression. *Accident Analysis and Prevention*. 2010; 42, 1531–1537.
- [3] Miranda-Moreno L F, Labbe A, Fu L. Bayesian multiple testing procedures for hotspot identification. *Accident Analysis and Prevention*. 2007; 39, 1192–1201.
- [4] Hauer E. *Observational Before-After Studies in Road Safety*. Pergamon, Tarrytown, NY, 1997.
- [5] Miaou SP, Song JJ. Bayesian ranking of sites for engineering safety improvement: decision parameter, treatability concept, statistical criterion and spatial dependence. *Accident Analysis and Prevention*. 2005; (37), 699–720.
- [6] Persaud BN, Hauer E. Comparison of two methods for de-biasing before and after accident studies. *Transport. Res. Record*. 1984; 975, 43–49.
- [7] Persaud BN. Safety migration, the influence of traffic volumes, and other issues in evaluating safety effectiveness. *Transport. Res. Record*. 1986; 1086, 33–41.
- [8] Cheng W, Washington S. Experimental evaluation of hotspot identification methods. *Accident Analysis and Prevention*. 2005; 37, 870–881.
- [9] Chang LY, and Chen WC. Data mining of tree-based models to analyze freeway accident frequency. *Journal of Safety Research*. 2005; 36(4), 365–375.
- [10] Abdel-Aty M, and Pande A. Crash data analysis: Collective vs. individual crash level approach. *Journal of Safety Research*. 2007; 38, 581–587.
- [11] Geurts K, Wets G. Black Spot Analysis Methods: Literature Review. 2003. <https://doclib.uhasselt.be/dspace/bitstream/1942/5004/1/17.pdf>. 22.05.2014.
- [12] Elvik R. A survey of operational definitions of hazardous road locations in some European countries. *Accident Analysis and Prevention*. 2008; 40, 1830–1835.
- [13] Kwon OH, Park MJ, Yeo H, Chung K. Evaluating the performance of network screening methods for detecting high collision concentration locations on highways. *Accident Analysis and Prevention*. 2013; 51, 141–149
- [14] Montella A. A comparative analysis of hotspot identification methods. *Accident Analysis and Prevention*. 2010; 42, 571–581.
- [15] Pirdavani A, Brijs T, Wets G. A Multiple Variables Decision-Making Approach for Prioritizing Accident Hotspots in the Absence of Crash Data, *Transport Reviews: A Transnational Transdisciplinary Journal*. 2010; 30(1), 97-113.
- [16] Sadeghi A, Ayati E, and Neghab MP. Identification and Prioritization of Hazardous Road Locations by Segmentation and Data Envelopment Analysis Approach. *Promet – Traffic&Transportation*. 2013; 25, 127-136.
- [17] <http://www.irap.org/en/about-irap-3/methodology>.
- [18] Bellos, E., Efstathiadis, S., Gkremos, I., Leopoulos, V. (2014). Road infrastructures' risk assessment: a valuable tool for investments' decisions, In conference proceedings of People, Buildings and Environment 2014, an international scientific conference, Kroměříž, Czech Republic, pp. 67-78, ISSN: 1805-6784.
- [19] Akgun A. A comparison of landslide susceptibility maps produced by logistic regression, multi-variables decision, and likelihood ratio methods: a case study at İzmir, Turkey. *Landslides*. 2012; 9, 93–106.
- [20] Yılmaz I. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat–Turkey), *Computers & Geosciences*. 2009; 35, 1125–1138.
- [21] Lee S, Sambath T. Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environ Geol*. 2006; 50(6):847–855.
- [22] Demir G, Aytekin M, Akgun A, İközler SB, Tatar O. A comparison of landslide susceptibility mapping of the eastern part of the North Anatolian Fault Zone (Turkey) by likelihood-frequency ratio and analytic hierarchy process methods. *Nat Hazards*. 2013; 65:1481–1506.
- [23] Pradhan B, Latif ZA, Aman SNA. Application of airborne LiDAR derived parameters and probabilistic-based frequency ratio model in landslide susceptibility mapping. *Appl Mech Mater*. 2012; 225, 442–447.
- [24] Regmi AD, Devkota KC, Yoshida K, Pradhan B, Pourghasemi HR, Kumamoto T, Akgun A. Application of frequency ratio, statistical index, and weights-of-evidence models and their comparison in landslide susceptibility mapping in Central Nepal Himalaya. *Arab J Geosci*. 2013; 7, 725-742.
- [25] Lee S, Talib JA. Probabilistic landslide susceptibility and factor effect analysis. *Environmental Geology*. 2005; 47, 982–990.
- [26] Ayalew L, Yamagishi H, Marui H and Kanno T. Landslides in Sado Island of Japan: Part II. GIS-based susceptibility mapping with comparisons of results from two methods and verifications. *Engineering Geology*. 2005; 81(4) 432-445.