Paper No. 02-2160

Identifying Locations with Potential for Accident Reductions

Use of Direct Diagnostics and Pattern Recognition Methodologies

Jake Kononov

Safety performance functions reflect the complex relationship between exposure, usually measured in annual average daily traffic, and accident count for a unit of road section over a unit of time. One of the main uses of the safety performance functions is to identify locations that experience more accidents than expected, thus exhibiting a potential for accident reduction. Overrepresentation in the number of accidents above the expected or normal threshold predicted by the safety performance function is only one of many indicators of a potential for accident reduction. Accident type, severity, road condition, spatial distribution of accidents, and lighting conditions are only a few of the many important symptoms of the accident problem. Two methodologies are introduced for identification of locations with potential for accident reduction: direct diagnostics and continuous pattern recognition analysis. Use of these methodologies revealed that existence of accident patterns susceptible to correction may or may not be accompanied by the overrepresentation in accident frequency detected by the safety performance functions.

In his work *Physics and Philosophy* (*1*), Heisenberg observed: "Since the measuring device has been constructed by the observer . . . we have to remember that what we observe is not nature in itself but nature exposed to our method of questioning." In the study of road safety our best measuring device, at present, is a safety performance function (SPF) described by Hauer (*2*). SPFs reflect the complex relationship between exposure usually measured in annual average daily traffic (ADT) and accident count for a unit of road section over a unit of time. Use of SPFs calibrated for various facilities allows us to detect high hazard locations or "sites with promise" (*3*). Comparison of the accident history of a particular class of roads with some upper control limit predicted by the SPF allows us to detect such locations. If observed accident frequency on a segment of road is found to be consistently higher than upper control frequency predicted by the SPF, then this segment is classified as a high hazard location or as having a potential for accident reduction. This paper explores whether the SPF is the best measuring device for the identification of locations with potential for accident reduction.

DIRECT DIAGNOSTICS

Consider a hypothetical roadway segment, which is 1 mi long with the following accident history over a 3-year period: 10 accidents total, including 7 overturning, 2 rear-end, and 1 fixed object. Assume that overturning accidents on the average represent 22% of the total for this functional class. Considering that each accident can be viewed as an independent Bernoulli trial with 22% probability of overturning, the probability of observing 7 or more overturning accidents out of 10 can be computed as follows:

$$P(X \le x) = B(x, n; p) = \sum_{i=0}^{x} \frac{n!}{(n-i)!\,i!} p^i (1-p)^{n-i}$$

$$P(X \ge 7, 10; .22) = 1 - P(X \le 6, 10; .22)$$

$$P(X \ge 7) = 1 - \sum_{i=0}^{6} \frac{10!}{(10-i)!\,i!} .22^i (1-.22)^{10-i} = .0016$$

As can be seen from the preceding calculations, the probability that 7 accidents or more out of 10 will result in overturning as part of a normal statistical process is extremely low (.0016). Such low probability suggests that something in the roadway environment triggers overturning accidents. This element needs to be identified and corrected. In this hypothetical case history, an overturning accident problem has been identified by using direct diagnostics methodology.

CONTINUOUS PATTERN RECOGNITION ANALYSIS

Now consider a different situation in which that same segment is located within project limits of a 5-mi-long roadway improvement project. Figure 1 illustrates accident history within project limits segmented by 1-mi sections. As can be seen from this diagram, the accident history within project limits is as follows: total number of accidents = 50, overturning accidents = 15, and other accidents = 35. The probability of 15 or more overturning accidents out of 50 total using direct diagnostics can be computed as described in the previous example:

$$P(X \ge 15, 50; .22) = 1 - P(X \le 14, 50; .22)$$

$$= 1 - \sum_{i=0}^{14} \frac{50!}{(50-i)!\,i!} .22^i (1-.22)^{50-i}$$

$$= .12 \text{ (certainly a strong possibility)}$$

If only the direct diagnostics method is used to examine a 5-mi road improvement project for overturning frequency, it would be concluded that no overturning problem is present within project limits.

Colorado Department of Transportation, Region 6, 2000 South Holly, Denver, CO 80222.

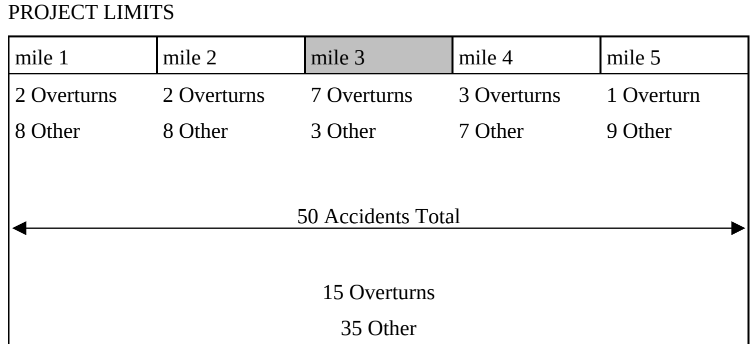

FIGURE 1 Accident history diagram segmented by 1-mi sections.

However, it is known that at least 1 mi out of 5 has a serious overturning problem. The question then becomes how can a "hidden" safety problem be identified within project limits? In other words, how can one systematically recognize a pattern of accidents within a roadway segment?

The problem involves detection of deviation outside the boundaries of the random Bernoulli process in the direction of reduced safety. This deviation is frequently confined to a very limited area and needs to be recognized (ferreted out) or classified as such through some form of propagation of continuous statistical testing. In order to make appropriate classification decisions, some amount of a priori knowledge is required about the expected system performance. This knowledge was compiled in an extensive data set describing various characteristics of accident distribution profiles endemic to specific classes of roads in Colorado. This data set was compiled for six classes of roads over a period of 8 years and contains 84 different parameters related to accident occurrence, such as accident type, severity, and roadway conditions. It represents a source of a priori knowledge base required for computing of a posteriori probabilities. One of the data sets representing two-lane, rural, mountainous roads is presented in Table 1.

In order to illustrate further the need for pattern recognition analysis, a case history can be examined involving a two-lane road in the mountainous area. Over 5 years of accident history, 142 accidents were recorded for the 7-mi road segment. SPF analysis reveals that accident frequency is well within the expected range for this type of facility. An SPF graph reflecting 6 years of accident history (averaged over 3-year periods) for the roadway segments (2 miles or longer) in the study is presented in Figure 2. Although accident frequency is well within the expected range, examination of the accident listing revealed unusual concentrations of nighttime accidents. Figure 3 shows the cumulative graph of nighttime accidents within study limits. One can test whether or not the overall number of nighttime accidents is overrepresented using a direct diagnostic approach, considering that 45 out of 142 accidents occurred under dark, unlighted conditions and that the Bernoulli probability of nighttime accidents on the two-lane unlighted rural road in the mountains is .3421 (Table 1).

$$
\begin{aligned}
P(X \geq 45, 142; .3421) &= 1 - P(X \leq 44, 142; .3421) \\
&= 1 - \sum_{i=0}^{44} \frac{142!}{(142 - i)!\,i!} .3421^{i}(1 - .3421)^{142-i} \\
&= .764
\end{aligned}
$$

Although a direct diagnostics test convincingly showed that overall nighttime accidents under dark, unlighted conditions are not overrepresented within project limits, one may still experience localized safety problems related to nighttime visibility. The cumulative graph of accidents when dark presented in Figure 3 reveals a sizable concentration of accidents between mileposts 8.5 and 9.2, where 15 out of 17 accidents occurred during unlighted conditions. One can test for the presence of a nighttime accident pattern between these mileposts using direct diagnostics:

$$
\begin{aligned}
P(X \geq 15, 17; .3421) &= 1 - P(X \leq 14, 17; .3421) \\
&= 1 - \sum_{i=0}^{14} \frac{17!}{(17 - i)!\,i!} .3421^{i}(1 - .3421)^{17-i} \approx 0
\end{aligned}
$$

Such low cumulative probability of observing 15 or more accidents under dark, unlighted conditions strongly suggests that there is something in the roadway environment between mileposts 8.5 and 9.2 that triggers nighttime accidents. A nighttime field visit revealed that curve signing on numerous curves between mileposts 8.5 and 9.2 had no nighttime retroreflectivity and no curve delineation was installed. Such a problem is generally susceptible to correction through installation of new curve signing with high-grade retroreflective sheeting in combination with installation of chevrons and delineators. If these countermeasures do not prove effective, installation of permanent lighting may be considered.

The findings involving safety of this two-lane mountainous road segment can be summarized as follows:

1. Analysis using SPFs shows that accident frequency is well within the expected range.
2. Direct diagnostics analysis of nighttime collisions shows no abnormalities.
3. There is, however, a significant nighttime safety problem between mileposts 8.5 and 9.2 related to the lack of retroreflectivity of curve signing and inadequate delineation.

What does it all mean? It means that there is a significant yet correctable safety problem, which is not detectable either through SPF analysis or the direct diagnostics method, which makes a compelling case for the pattern recognition analysis.

ANALYTICAL FRAMEWORK FOR CONTINUOUS PATTERN RECOGNITION

Two previous examples demonstrated that SPF and direct diagnostics analysis alone are not sufficient in detecting safety problems hidden within roadway segments. In order to recognize a "hidden"

TABLE 1 Normative Values of Various Accident Characteristics

Rural Mountainous 2-Lane Undivided Highway			
Description	**Accidents**	**Percent**	
Property Damage Only	6,143	58.32%	
Injury	4,186	39.74%	
Fatality	204	1.94%	100.00%
Persons Injured	6,345	N/A	
Persons Killed	232	N/A	
Single Vehicle Accidents	8,009	76.04%	
Two Vehicle Accidents	2,235	21.22%	
Three or more Vehicle Accident	284	2.70%	
Unknown Number of Vehicles	5	0.05%	100.00%
On Road	4,086	38.79%	
Off Road Left	2,572	24.42%	
Off Road Right	3,858	36.63%	
Off Road at Tee	5	0.05%	
Off Road in Median	1	0.01%	
Unknown Road Location	11	0.10%	100.00%
Overturning	2,345	22.26%	
Other Non Collision	225	2.14%	
School Age Pedestrians	2	0.02%	
All Other Pedestrians	16	0.15%	
Broadside	73	0.69%	
Head On	473	4.49%	
Rear End	754	7.16%	
Sideswipe (Same Direction)	146	1.39%	
Sideswipe (Opposite Direction)	488	4.63%	
Approach Turn	97	0.92%	
Overtaking Turn	149	1.41%	
Parked Motor Vehicle	113	1.07%	
Railway Vehicle	1	0.01%	
Bicycle	24	0.23%	
Motorized Bicycle	1	0.01%	
Domestic Animal	118	1.12%	
Wild Animal	1,015	9.64%	
Unknown Accident Type	6	0.06%	100.00%
Total Fixed Objects	4,124	39.15%	
Total Other Objects	363	3.45%	
Daylight	6,071	57.64%	
Dawn or Dusk	556	5.28%	
Dark - Lighted	147	1.40%	
Dark - Unlighted	3,603	34.21%	
Unknown Lighting	156	1.48%	100.00%
No Adverse Weather	7,816	74.20%	
Rain	414	3.93%	
Snow or Sleet or Hail	2,015	19.13%	
Fog	28	0.27%	
Dust	1	0.01%	
Wind	101	0.96%	
Unknown Weather	158	1.50%	100.00%
Dry Road	6,221	59.06%	
Wet Road	724	6.87%	
Muddy Road	10	0.09%	
Snowy Road	735	6.98%	
Icy Road	1,942	18.44%	
Slushy Road	382	3.63%	
Foreign Material Road	29	0.28%	
With Road Treatment	78	0.74%	
Dry with Icy Road Treatment	8	0.08%	
Wet with Icy Road Treatment	1	0.01%	
Snowy with Icy Road Treatment	3	0.03%	
Icy with Icy Road Treatment	13	0.12%	
Slushy with Icy Road Treatment	5	0.05%	
Unknown Road Condition	382	3.63%	100.00%
Total Accidents:	**10,533**		
Total Number of Locations:	**2,463**		

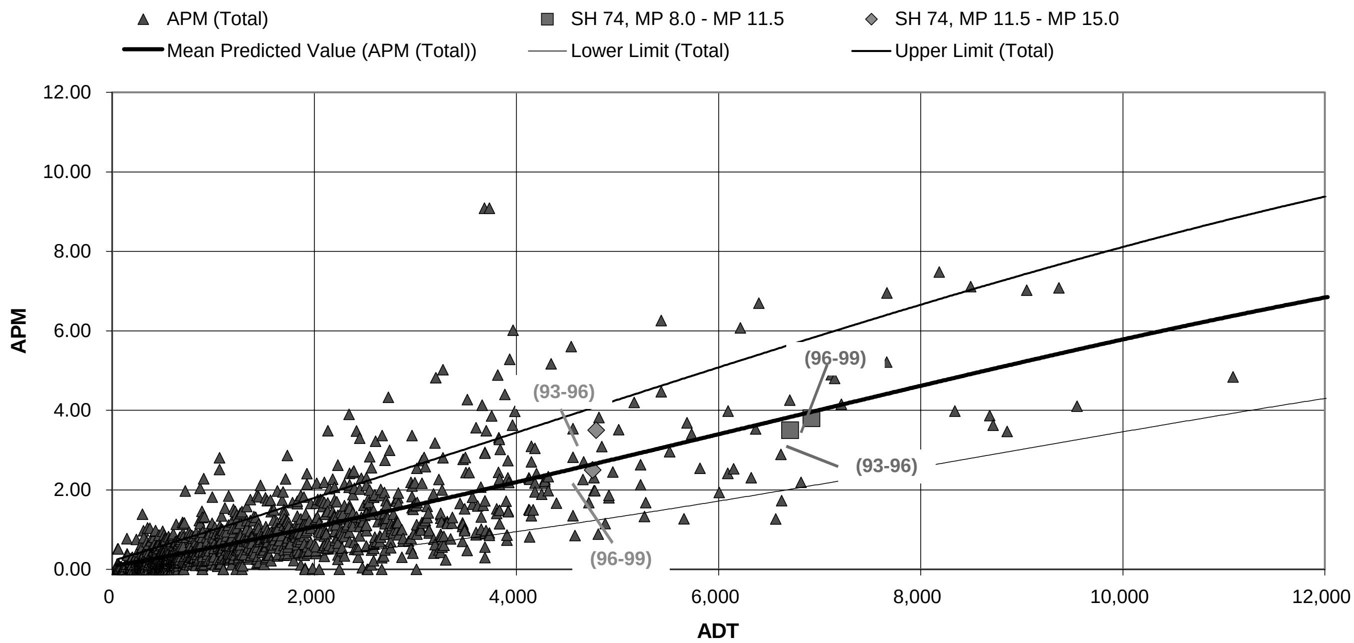

FIGURE 2 SPF graph for two-lane mountainous highways containing project locations (APM = accidents per mile, MP = milepost).

pattern in random accident occurrence, one needs to devise some form of continuous statistical testing for use along the length of roadway segments. This continuous testing is expected to delineate the boundaries of "abnormal" accident occurrence within project limits. In other words, it is expected to reveal locations with potential for accident reduction. To achieve this goal, the following procedure can be implemented.

Select a scanning interval Δs of fixed length containing a feature vector $\boldsymbol{X_i}$, consisting of n_t total accidents, n_i of which are of the type *i*. The scanning interval, in its first position, is tested for the presence of accident pattern *i* using a direct diagnostics method, then the scanning interval is moved a fixed distance δi, known as a scanning increment. After sliding the scanning interval a distance of one scanning increment, a new feature vector $\boldsymbol{X_i} + \mathbf{1}$ is obtained and tested again for the presence of pattern *i*. Continuous sliding and testing of the scanning interval will reveal the limits of the accident patterns if they are present. Figure 4 depicts the reference framework for continuous testing for the presence of accident patterns by moving a scanning interval within study limits. A pattern recognition algorithm can be described as follows:

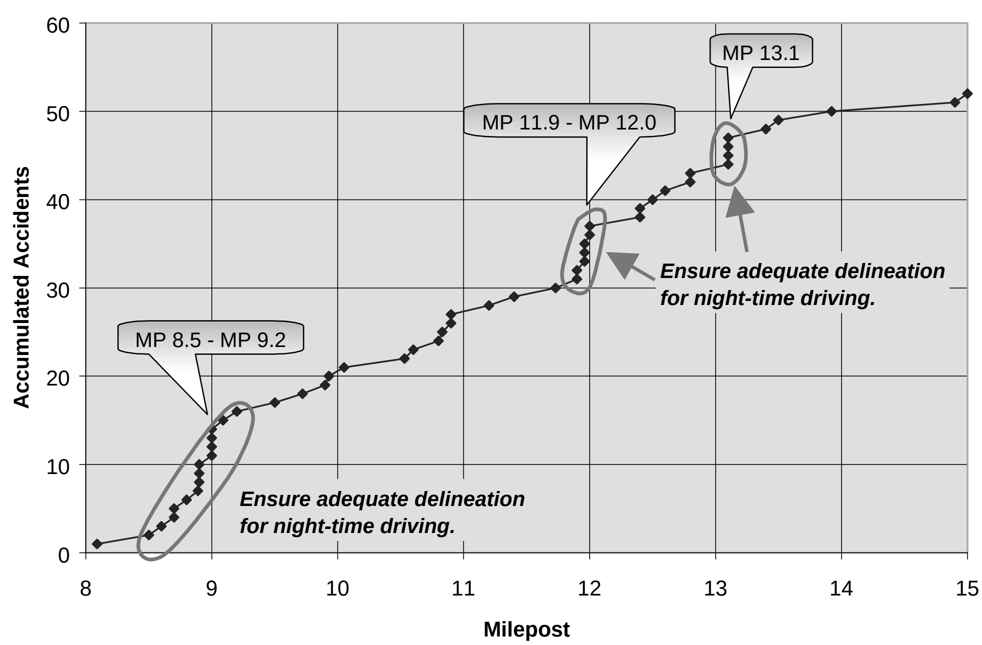

FIGURE 3 Cumulative graph of nighttime accidents [State Highway 74, MP (milepost) 8.0–15.0].

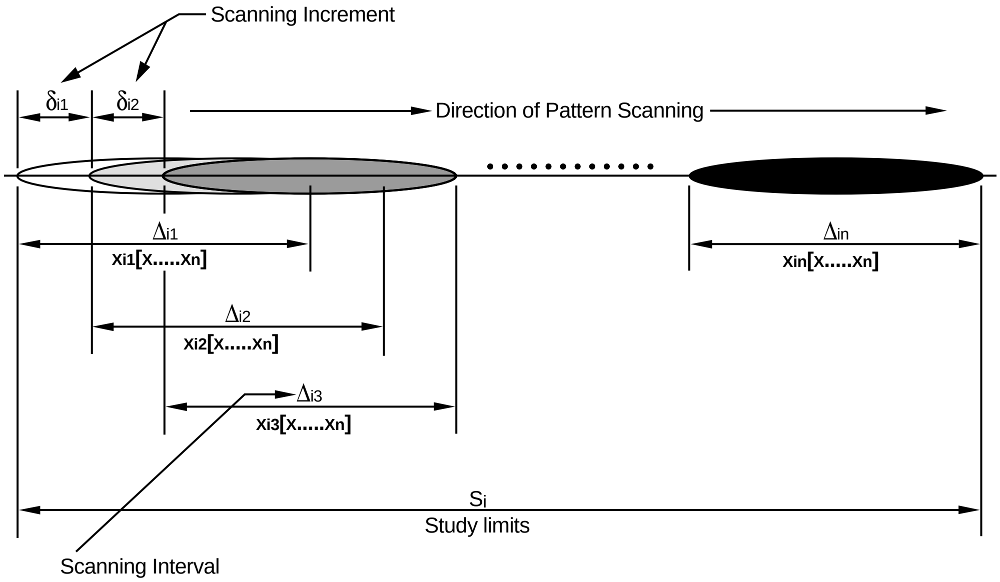

FIGURE 4 Reference framework diagram for continuous pattern recognition.

- Select scanning interval $\Delta_i = \Delta_1 = \Delta_2 = \ldots \Delta_n$. Scanning interval is generally 1 mi for the analysis of roadway segments in rural areas. Selection of the scanning interval is data driven and can be extended beyond 1 mi in areas with low ADT and reduced in the areas of high ADT.
- Select a scanning increment $\delta_i = \delta_1 = \delta_2 = \ldots \delta_n$. The scanning increment is also data driven and is generally in the range between 0.01 mi and 0.1 mi. A scanning increment of 0.10 mi was found adequate for most roadway environments.
- Obtain feature vector $\boldsymbol{X_{i1}\{x \ldots x_n\}}$ reflecting accident history within segment Δ_1.
- Identify the number of total accidents nt_1 and the number of accidents of a specific type n_{i1} contained within feature vector $\boldsymbol{X_{i1}\{x \ldots x_n\}}$.
- Select appropriate p_i reflecting the Bernoulli probability of success for each trial related to pattern *i*. In this context SPF$_i$ represents Safety Performance Function *i*.
- Select critical value P_α for making classification decision, P_α = .01 for most cases; when identifying patterns in the areas with low ADT, P_α = .05 can be used.

Compute

$$P(SF_i/n_{1i}) = 1 - \sum_{i=0}^{n_{1t}-1} \frac{n_{t1}!}{(n_{t1}-i)!\,i!} p_i (1-p_i)^{n_{t1}-i}$$

If $P(SPF_i / n_i) \le P_\alpha$, then the classification decision is made that

$$\Delta_1 \begin{bmatrix} x\{x \ldots x_n\} \\ n_{i1}/n_{t1} \end{bmatrix} \notin SPF_i$$

If $P(SPF_i / n_{i1}) \le P_\alpha$, then the classification decision is made that

$$\Delta_1 \begin{bmatrix} x\{x \ldots x_n\} \\ n_{i1}/n_{i1} \end{bmatrix} \in SPF_i$$

- Following the completion of the first pattern recognition cycle, the scanning interval Δ is shifted an incremental distance δ. At the second position of Δ, a feature vector $\boldsymbol{X_{i2}\{x \ldots x_n\}}$ reflecting accident history of the roadway segment delineated by Δ_2 is obtained and $P(SPF_i / n_{i2})$ is computed, at which point a second classification decision is made.

In this incremental fashion the entire distance *S* representing project limits is tested for the presence of accident pattern *i*. If no indication of pattern development is observed at the ends of the study area, the number of direct diagnostics tests can be computed as follows:

$$N_{\text{test}} = \frac{S - \Delta}{\delta}$$

If some evidence of pattern development is observed at either end of the study, the author recommends extending pattern recognition analysis a distance of one Δ outside the project limits. It will ensure detection of accident patterns that are partially observed within project limits and are extending outside of the study area.

In order to quantify the strength of the pattern, the concept of pattern intensity index (PI) is introduced.

$$\mathrm{PI} = \frac{P_\alpha}{P(SPF_i/n_i)}$$

$$= \frac{P_\alpha}{1 - \sum_{i=0}^{n_i-1} \frac{n_t}{(n_t-i)!\,i!} p_i(1-p_i)^{n_t-i}}$$

$$\lim_{P(SF_i/n_i)\to 0} \mathrm{PI} = \infty$$

Pattern intensity monotonically increases as conditional probability of observing *n* accidents of type *i* monotonically decreases. PI is only computed for the conditions when

$$P(SPF_i/n_i) \le P_\alpha$$

Assigning the PI a value of 0 when $P(SPF_i / n_i) > P_\alpha$ allows for easy delineation of the accident pattern boundaries on the PI graph.

$$\mathrm{PI} = \begin{cases} \dfrac{P_\alpha}{P(SPF/n_i)} & \forall P(SPF/n_i) \le P_\alpha \\ 0 & \forall P(SPF/n_i) > P_\alpha \end{cases}$$

TABLE 2 Pattern Intensity Index Along Study Segment

Loc. MP	Mid-point	Acc. Dark-Unlighted	Total Number of Accs.	Chance of Dark-Unlighted Acc.	P(X>Ni,Nt,p)	Score	Modif. Score (PI)
8.09-9.09	8.59	14	27	0.3421	0.0446	0.22	0.00
8.19-9.19	8.69	13	17	0.3421	0.0005	21.90	21.90
8.29-9.29	8.79	14	17	0.3421	0.0001	154.59	154.59
8.39-9.39	8.89	14	17	0.3421	0.0001	154.59	154.59
8.49-9.49	8.99	13	17	0.3421	0.0005	21.90	21.90
8.59-9.59	9.09	14	20	0.3421	0.0012	8.49	8.49
8.69-9.69	9.19	13	23	0.3421	0.0233	0.43	0.00
8.79-9.79	9.29	12	24	0.3421	0.0809	0.12	0.00
8.89-9.89	9.39	11	30	0.3421	0.4557	0.02	0.00
8.99-9.99	9.49	9	23	0.3421	0.3827	0.03	0.00
9.09-10.09	9.59	7	34	0.3421	0.9727	0.01	0.00
9.19-10.19	9.69	6	33	0.3421	0.9871	0.01	0.00
9.29-10.29	9.79	5	37	0.3421	0.9989	0.01	0.00
9.39-10.39	9.89	5	37	0.3421	0.9989	0.01	0.00

NOTE: Acc(s). = accident(s).

Table 2 shows a tabulation of PI along a segment of road discussed in the previous example and Figure 5 contains a graph reflecting a pattern of nighttime accidents. It is important to note that performing hundreds of tests such as these carries an inherent statistical danger of false positives, but it is more than offset by a significant benefit of identifying hidden accident patterns susceptible to correction.

SUMMARY

Detection of an accident pattern suggests a presence of an element or elements in the roadway environment, which triggered a deviation from a random statistical process in the direction of reduced safety. Identification of such an element through engineering investigation, which typically includes a site visit and plans review, always provides a critical clue to accident causality. Development of the diagnostic knowledge base and a pattern recognition algorithm led to the following finding: Existence of accident patterns susceptible to correction may or may not be a accompanied by the overrepresentation in accident frequency reflected by the safety performance functions or high accident rates. In fact it can be said that detection of accident patterns provides a more direct link to the development of the countermeasure strategy than a mere increase in accident frequency. Furthermore, most of the frequency clusters are merely reflections of specific patterns. In many cases, the expected or normal proportion of accidents is counterintuitive, which further emphasizes the need for the creation of a framework of diagnostic norms for various types of roadways. These diagnostic norms should be calibrated locally for various road classes in various environments. Accident type, severity, road condition, spatial distribution of accidents, and lighting conditions are only a few of the many important symptoms of the accident problem, which can be detected using pattern recognition methodology.

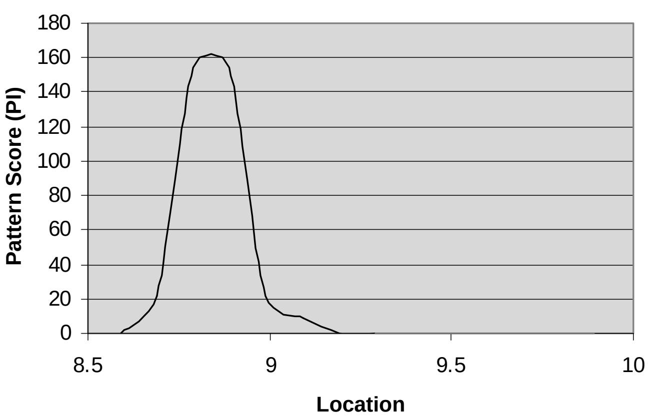

FIGURE 5 Pattern recognition graph for nighttime accidents.

REFERENCES

1. Heisenberg, W. *Physics and Philosophy: The Revolution in Modern Science.* Harper, New York, 1958.
2. Hauer, E., and B. Persaud. *Safety Analysis of Roadway Geometric and Ancillary Features.* Transportation Association of Canada, 1997.
3. Hauer, E. Identification of Sites with Promise. In *Transportation Research Record 1542,* TRB, National Research Council, Washington, D.C., 1996, pp. 54–60.

Publication of this paper sponsored by Committee on Statistical Methodology and Statistical Computer Software in Transportation Research.