Proactive detection of high collision concentration locations on highways

Koohong Chung[a]1*, Kitae Jang[b], Samer Madanat[b], Simon Washington[c]

[a]*California Department of Transportation, United States*
[b]*Institute of Transportation Studies, University of California, Berkeley, United States*
[c]*Queensland University of Technology, Brisbane, Australia*

Abstract

In previous research (Chung et al, 2009) the potential of the Continuous Risk Profile (CRP) to proactively detect the systematic deterioration of freeway safety levels was presented. In this paper this potential is investigated further, and an algorithm is proposed for proactively detecting sites where the collision rate is not sufficiently high to be classified as a high collision concentration location but where a systematic deterioration of safety level is observed. The approach proposed compares the weighted CRP across different years and uses the cumulative sum (CUSUM) algorithm to detect the sites where changes in collision rate are observed. The CRPs of the detected sites are then compared for reproducibility. When high reproducibility is observed, a growth factor is used for sequential hypothesis testing to determine if the collision profiles are increasing over time. Findings from applying the proposed method using empirical data are documented in the paper together with a detailed description of the method.

Keywords: hot spot identification; proactive detection; continuous risk profile

* Corresponding author. Tel.: +1-510-622-5429; fax: +1-510-643-9922.
E-mail address: koohong_chung@dot.ca.gov.

1. Introduction

The vast majority of existing approaches for detecting high collision concentration locations on highways are reactive – they detect "hot spots" after observed collision rates exceed a predetermined threshold. Such approaches cannot proactively detect sites where safety is gradually deteriorating due to adverse changes that occur over time, such as worsening of pavement skid resistance, growing of vegetation that restricts sight distance, capacity constraints that lead to safety problems at locations such as freeway off-ramps and intersection turn bays, and negative influences from changes in nearby land uses.

This paper proposes a method for proactively detecting sites where the collision rate is systematically increasing over time. The proposed approach uses the weighted continuous risk profile (CRP) (Chung et al, 2009) to detect sites where changes in collision rates are observed. The CRPs of the detected sites are normalized and compared for reproducibility. When high reproducibility is observed, the growth factor (the factor used to normalize the CRP plots of detected sites) is then used in a sequential hypothesis testing (Wald, 1945) framework to evaluate the systematic changes in the collision profiles.

The description of the proposed method is provided in section 2. Findings from applying the method using empirical data are documented in section 3. This paper ends with future research plans and concluding remarks in section 4.

2. Description of the proactive detection approach

The objective of the proposed approach is to detect a site that displays systematic increase in collision rate, not necessarily a site with significantly high collision rate; the proactive detection method is different from traditional approaches in this aspect. The proposed approach compares the weighted continuous risk profile (CRP) from different years and uses an extension of the cumulative sum (CUSUM) algorithm (Basseville and Nikiforov, 1993) to detect sites where changes in collision rates are observed. The CUSUM algorithm is designed to detect abrupt changes. When changes in the weighted CRP are gradual, the CUSUM algorithm may not be effective. For this reason, the sites detected by the CUSUM algorithm are further evaluated using sequential hypothesis testing. A detailed description of the proposed method is presented in this section.

Continuous Risk Profile

The CRP is fitted to the underlying true risk, and reflects a measure of risk interpretable as collision risk per unit distance of roadway. Empirical analysis of traffic collision data from previous research (Chung et al, 2009) found that gradual deterioration of safety levels of a facility are revealed through the peaks in CRP plots that grow over the years. Remarkably, the locations of these growing peaks did not deviate from year to year; only the area under the curve and the size of the peak varied. Systematic detection of these gradually growing peaks can be achieved by monitoring the normalized sum of squared CRP plots from different years.

The proactive detection method starts by detecting peaks identified using the CRP whose location and shape are reproducible as the size of the peaks varies over the years. Let $A(d)$ denote the number of collisions per unit distance observed in the vicinity of location d (see Figure-1) and $M(d)$ denote the weighted average number of collisions over the window $[d-L, d+L]$ (see equation (1)).

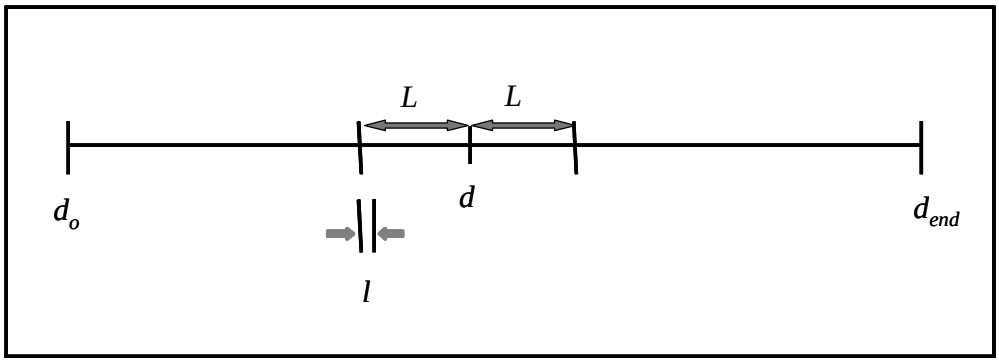

Figure-1 Hypothetical Highway Segment showing the window used for averaging

$$M(d) = \frac{\sum_{i=-\min(L/l,(d-d_0)/l)}^{\min(L/l,(d_{end}-d)/l)} (L/l - |i| + 1) \times A(d + i \times l)}{\sum_{i=-\min(L/l,(d_{end}-d)/l)}^{0} -i \quad + \sum_{j=0}^{\min(L/l,(d-d_0)/l)+1} j \quad + (L/l+1)} \tag{1}$$

For $d = d_0 + k \times l$ and $k = 1,2,... \frac{d_{end} - d_0}{l}$

Where,

d_0 = beginning postmile of segment; d_{end} = ending postmile of segment; $d_0 < d_{end}$; l = increment

$2L$ = size of the moving average window

$k, \frac{L}{l}$ and $\frac{d_{end} - d_0}{l}$ are integers (they define the number of increments within the moving window)

The length of $2L$ in Figure-1 needs to be long enough to filter out the random noise in the data and should not be too long to affect the location of the pronounced peaks (i.e., hot spots). This statement can be further explained with the aid of CRP plots constructed using different size of $2L$ ranging from 0.1 to 0.4 miles in Figure-2. Note how most critical points (i.e., where the slope of CRP is zero) marked with white circles are smoothed out as $2L$ increased from 0.1 to 0.4 miles while the location of the spots marked with black circles does not deviate much with respect to the changes in $2L$. This figure graphically illustrates how the critical points created due to random fluctuations in the data (see the white circles in Figure-2) can be smoothed out as the size of $2L$ exceeds the domain of the random fluctuations. However, the size of $2L$ cannot be arbitrarily increased to eliminate the random fluctuations since large $2L$ can change the location of some of the pronounced peaks (see grey circles in Figure-2) by including the peaks from adjacent sites in estimating $M(d)$.

Figure-3 displays how the number of critical points changed with respect to the changes in $2L$ and shed light on the optimal length of $2L$. Note how the number of critical points rapidly decreases with the increase in $2L$ while $2L$ is less than 0.2 mile. The rate at which the number of critical points decreases with respect to increase in $2L$ slows down while $2L$ is between 0.2 and 0.5. The reduction in the number of critical points then becomes less affected by the increase in $2L$. Empirical analysis of the data revealed that the rapid reduction in the number of critical points is due to smoothing out of the excessive number of critical points in $M(d)$ created by random fluctuations. As $2L$ was increased beyond 0.2 mile, the location of the pronounced peaks started to deviate from their locations identified using smaller $2L$: the locations of the hot spots were affected. For the purpose of proactively detecting high collision concentration locations, one does not need to filter out all the random noise since there are additional steps in the procedure to identify sites that display progressive deterioration of the safety level and further remove the threat of random events. Therefore, to filter out the random noise as much as possible without affecting the locations of peaks, 0.2 mile was used as the optimal size for $2L$.

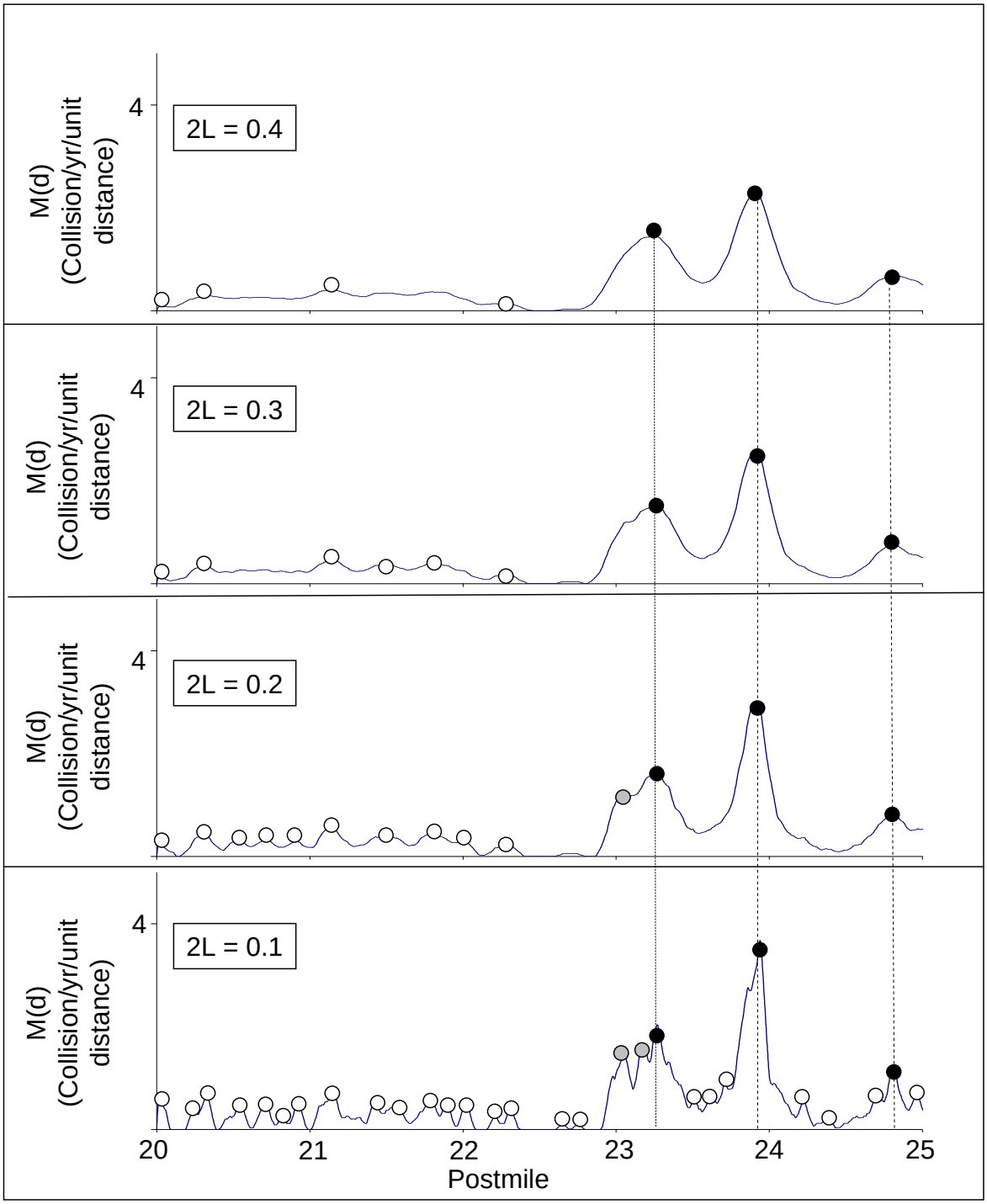

Figure-2 M(d) with different size of 2L (traffic collision data along I-880N in between postmile 20 to 25).

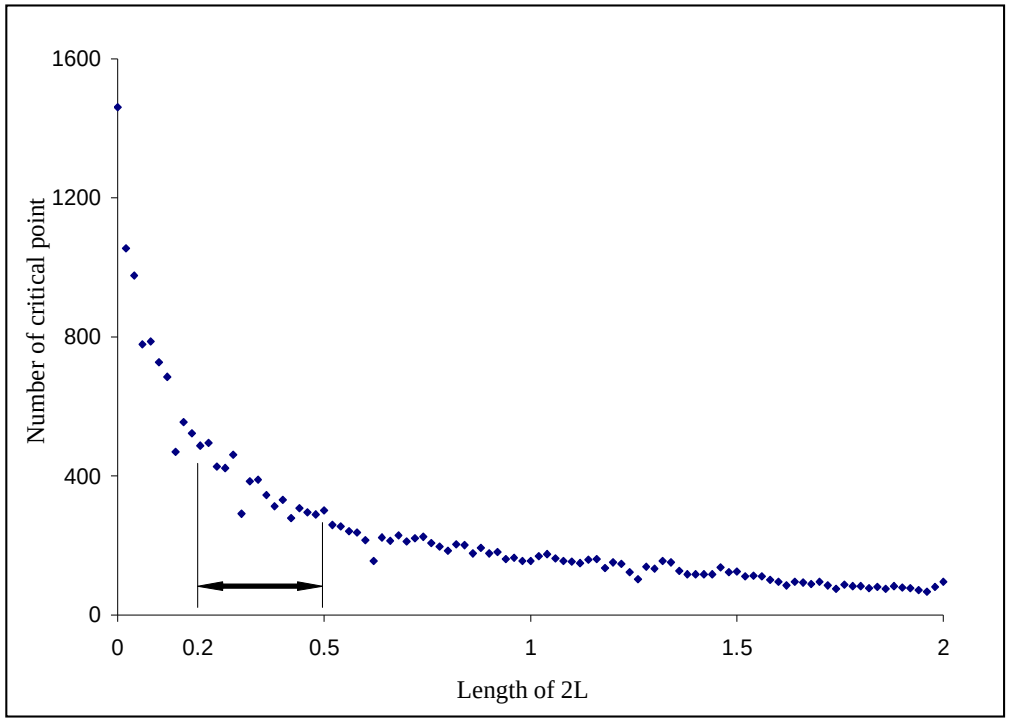

Figure-3 Changes in critical points with respect to changes in 2L

Proactive Detection

$M_y(d)$ in equation (2) denotes the profile in year y at postmile d and $M_{y-1}(d)$ is the profile in the previous year. The normalized sum of squared $M(d)$ between year y and y-1 is shown in equation (2). Figure-4(a) shows $SM_{y,y-1}(d)$ of each year along I-880N between postmile 20 and 32. The locations marked by sudden surges in slope are sites whose risk profile significantly increased or decreased with respect to other sites along the corridor. In the example shown in Figure-4, postmile 20 and 32 were used as the start point, *S*, and end point, *E*, respectively.

$$SM_{y,y-1}(d) = \frac{\int_S^d [M_y(x) - M_{y-1}(x)]^2 dx}{\int_S^E [M_y(x) - M_{y-1}(x)]^2 dx} \quad (2)$$

The logic behind normalizing the sum of squared $M_y(d)$ is to prevent changes in exogenous factors (traffic volume, rainfall intensity, etc.) from obscuring the detection of progressive deteriorations of a site. When the changes in exogenous factors have a similar effect at all sites along the route, monitoring only the changes in $M_y(d)$ could produce high false positive rates since overall increases or decreases in collision rate in year y-1 compared to y cause the changes seen in year y to appear more significant than other years with no change in exogenous factors.

The figure in the box (see Figure-4(b)) shows the value of ln(h) with respect to its percentile, where h is the slope of $SM_{y,y-1}(d)$. The enlarged view of the portion of the graph enclosed in the dotted box labeled A are shown below in the same figure. Note how the distribution ln(h) is reproducible from year to year. Such reproducibility enables us to apply the cumulative sum algorithm (CUSUM) (Basseville and Nikiforov, 1993) using the same threshold value, h^*, across years to detect sites where pronounced changes in collision risk occur as shown in equation (3):

$$H_y(d) = 1 \text{ if } SM_{y,y-1}(d) - SM_{y,y-1}(d-l) > h^* \\ = 0 \text{ otherwise} \quad (3)$$

h* is the threshold for detecting a site; the 97th percentile was used in this study.

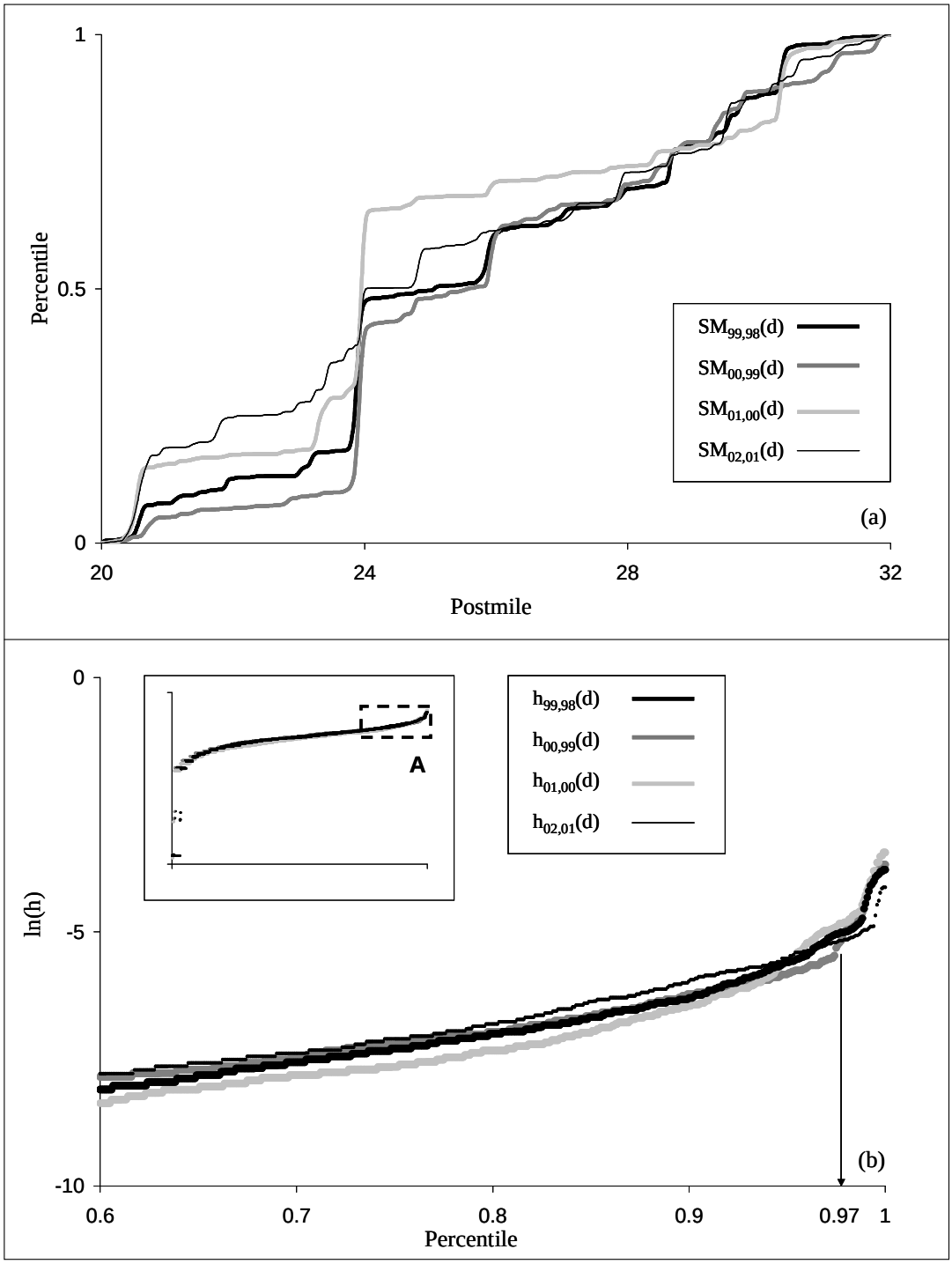

Figure-4 (a) $SM_{y,y-1}(d)$ along I-880 northbound; (b) ln(h) percentile.

The shortcoming of applying only equations (2) and (3) is that it can also detect sites where the collision rate is minimal but display large fluctuations in collision rate compared to their magnitude. These sites can be filtered out by applying equation (4).

$$EC_y(d) = Max(M_y(d) - r^*, 0) \tag{4}$$

r^* is the parameter used to prevent sites with minimal collision rates but display large fluctuations compared to their rate, from being detected. $EC_y(d)$ is the positive difference between $M_y(d)$ and r^*. In practice, corridors are comprised of different roadway sections that are classified into different roadway groups depending on their attributes (number of lanes, traffic volume and observed speed distribution) (Chung et al, 2009). Different roadway groups have different threshold collision rates, which is more appropriate than using one r* value for the entire corridor. r^* needs to be replaced with $r^*(d)$ in practice depending on the roadway group at location d. For illustration purposes, the 75th percentile of $M_y(d)$ was used as r* in the subsequent analysis.

Figure 5(a) shows the result of applying equations (2) to (4). The grey boxes in Figure 5 (b) show sub-segments whose start and end locations are defined by non-zero $EC_y(x)$. Each sub-segment in year y is denoted as $B_{y,1}(s_1, f_1)$ where s_1 and f_1 are the start and end postmiles of the sub-segment which will be referred to as a bin from here on.

$$B_{y,j}(s_j, f_j) = 1 \text{ if } H_y(d) = 1 \text{ anywhere between } s_j \text{ and } f_j$$
$$= 0 \text{ otherwise} \tag{5}$$

Where, j is index for the bin, j = 1,2..n

$$K_y(d) = EC_y(d) \times B_{y,j}(s_j, f_j) \tag{6}$$

The resulting graph is shown in Figure-5 (c).

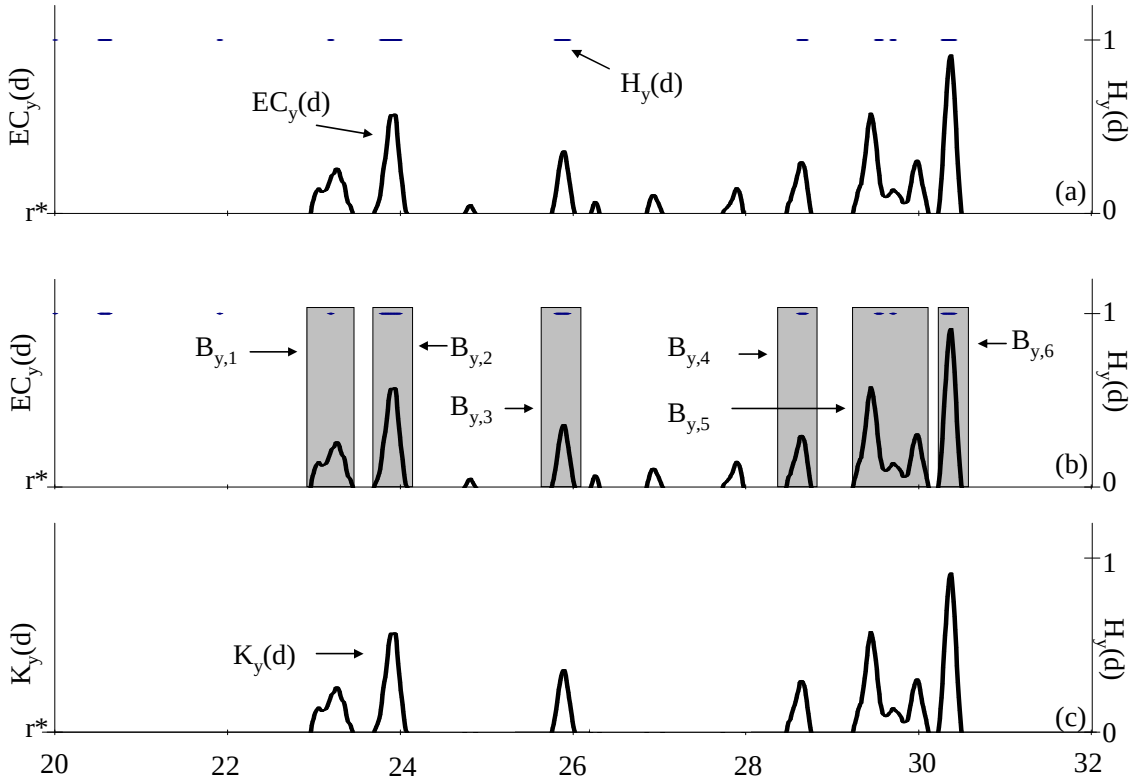

Figure-5 Graphical illustration of the process for selecting peaks where changes in collision profile are detected (data source: I-880 northbound, year 1999)

3. Findings

Attributes of the bins

The value of $K_y(d)$ within each bin marks the peaks in $M(d)$ whose value exceeds r^* and displays significant changes in CRP compared to the previous year. The terms $K_y(d)$ and $K_{y-1}(d)$ between s_j and f_j of $B_{y,j}$ are normalized as shown in equation (7) to estimate the cross-correlation between $S_{y,j}(d)$ and $S_{y-1,j}(d)$. Equation (8) shows the corresponding growth factor, $G_{y,y-1}$,for the j^{th} bin and the change in area under $K_y(d)$ in a bin compared to the previous year. Since the growth factor is a ratio, monitoring only the growth factor can be misleading when the net change in collision number is not considered simultaneously: even a small net change in collision number could cause a significant change in growth factor of a bin when the area under $K(d)$ is small. The net change (see equation (9)) in the number of excess collisions needs to be monitored simultaneously. The distribution of the growth factor and the net change are shown in Figure 6. The growth factor and net change followed log-normal and normal distributions, respectively.

$$S_{y,j}(d) = \frac{K_y(d)}{\int_{f_j}^{s_j} K_y(x)dx} \tag{7a}$$

$$r_{y,y-1}(j) = \frac{\int_{f_j}^{s_j} (S_{y,j}(d) - \bar{S}_{y,j})(S_{y-1,j}(d) - \bar{S}_{y-1,j})}{\sqrt{\int_{f_j}^{s_j} S_{y,j}(d) - \bar{S}_{y,j})^2} \sqrt{\int_{f_j}^{s_j} S_{y-1,j}(d) - \bar{S}_{y-1,j})^2}} \tag{7b}$$

where

$r_{y,y-1}(j)$ = cross-correlation between two successive values of $S_{y,j}(d)$ and $S_{y-1,j}(d)$.

$\bar{S}_y$ & $\bar{S}_{y-1}$ = means of the corresponding series in the jth bin.

$$G_{y,y-1} = \frac{\int_{f_j}^{s_j} K_{y,j}(x)dx}{\int_{f_j}^{s_j} K_{y-1,j}(x)dx} \tag{8}$$

$$N_{y,y-1} = \int_{f_j}^{s_j} K_{y,j}(x)dx - \int_{f_j}^{s_j} K_{y-1,j}(x)dx \tag{9}$$

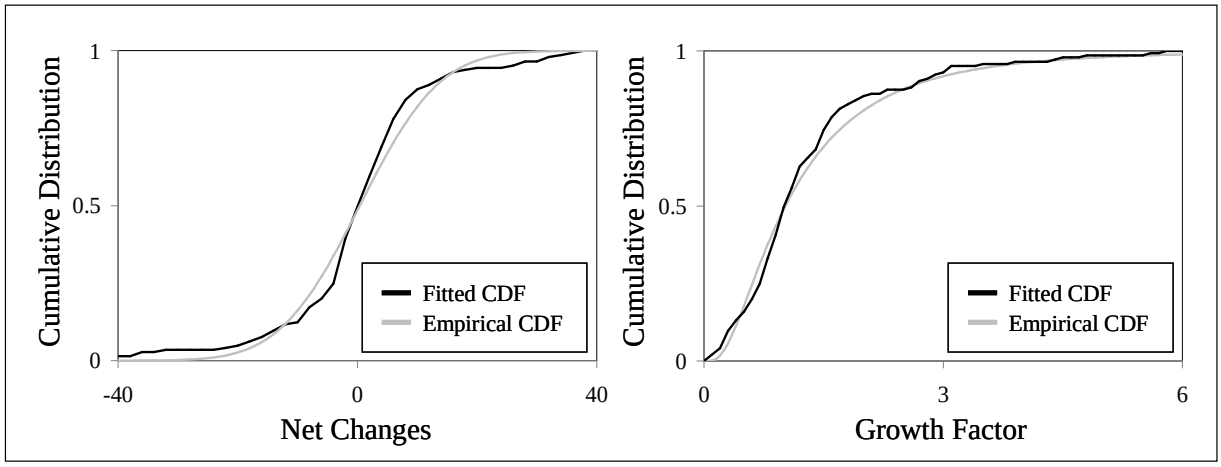

Figure-6 Distribution of growth factor and net change

The cross-correlation between $S_{y,j}(d)$ and $S_{y-1,j}(d)$, the growth factor, and the net change together define the attribute of a bin and they can be plotted on three-dimensional graph as shown in Figure-7(a). The points are obtained from analyzing traffic collision data along 82 miles of freeways in the San Francisco Bay Area. Note that all the points in the figure have cross correlations greater than 0.75 since the proposed approach only detects sites that display reproducible patterns in $M(d)$.

Figure-7(b) shows the points whose growth factor and net change are greater than the 70th percentile projected onto the cross correlation and growth factor plane. These points are subsequently used to detect the candidate sites for proactive detection.

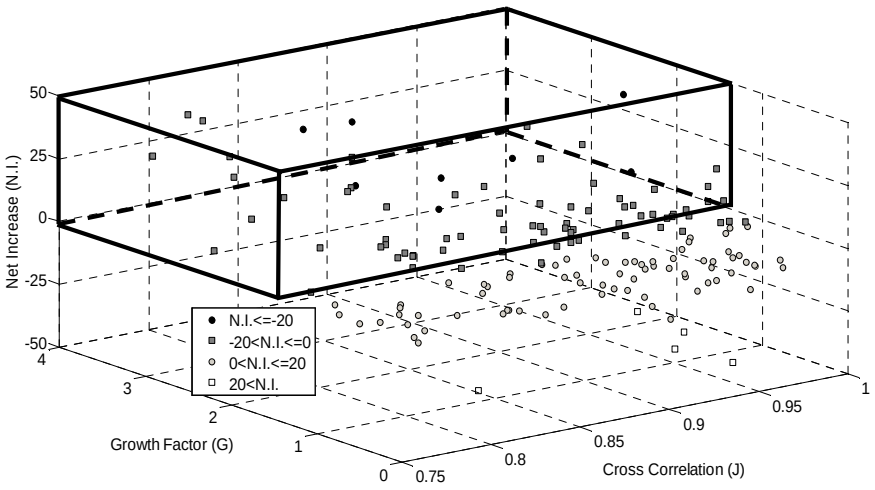

Figure 7(a) Attributes of bins detected along I-880 and I-580 in 1999

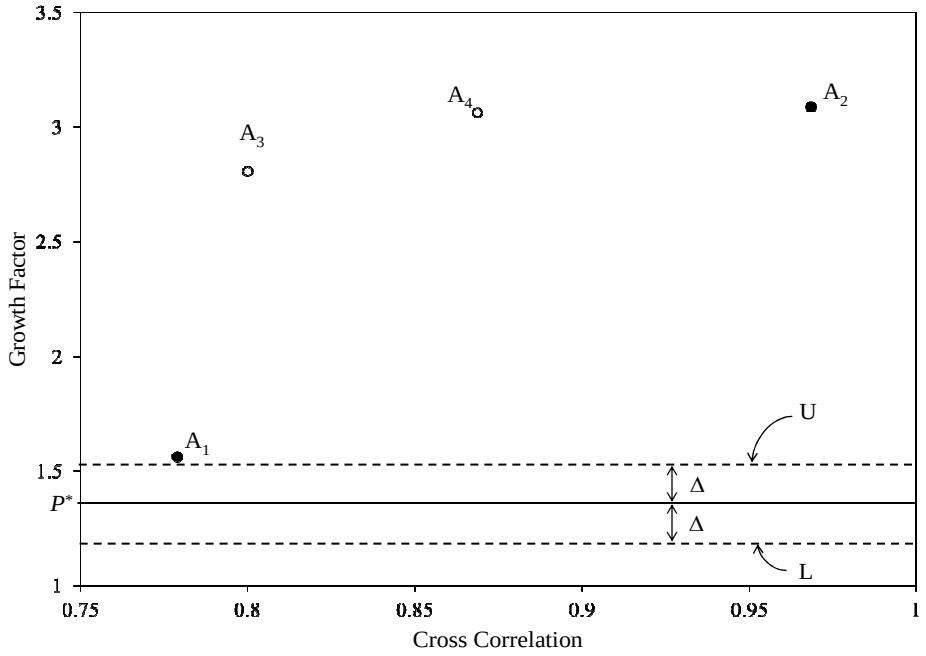

Figure 7(b) Attributes of bins that have cross-correlations between $S_{y,j}(d)$ and $S_{y-1,j}(d)$ greater than 0.75; the growth factor and the net change are greater than the 70th percentile

Sequential Hypothesis Testing

Typical hypothesis testing evaluates two alternatives; the result of the test is subject to both false positives (i.e., accepting the hypothesis when it is not valid) and false negatives (i.e., rejecting the hypothesis when it is valid). When the cost associated with the false positive or negative is much greater than the cost of making another observation, sequential hypothesis testing becomes a useful approach.

In sequential hypothesis testing, the decision line is divided into three contiguous segments. The outer two segments correspond to accepting or rejecting the hypothesis; the middle segment corresponds to the decision to take an additional observation. If the hypothesis is accepted (or rejected), no subsequent observation is taken. If an observation is taken, the decision-maker faces the same choices in the following period. The thresholds between regions are determined on the basis of the relative costs of accepting the hypothesis when it is false, rejecting it when it is true, and taking an additional observation.

In the context of proactive detection of freeway hot spots, these costs are unknown and cannot be estimated from field data. Instead of making assumptions about the costs, our approach was to make reasonable judgements about the size of the middle segment. In Figure 7(b), P* represents the point at which the decision-maker is completely indifferent between accepting and rejecting the hypothesis. The middle segment is thus assumed to have a length of 2 Δ, and to be centered around P*.

The solid black line in Figure-7(b) represents the 65th percentile and the two dotted lines labeled U and L represent the 70th and 60th percentiles, respectively. The proposed method rejects a site from further consideration once it falls below L. When a site falls between U and L, the site will be neither accepted nor rejected but will be re-evaluated in the subsequent time period. When a site lies above U, the site will become a candidate site to be finalized in the following analysis period. The grey horizontal line in Figure-8 indicates the $P^* \pm \Delta$ region shown in Figure-7(b).

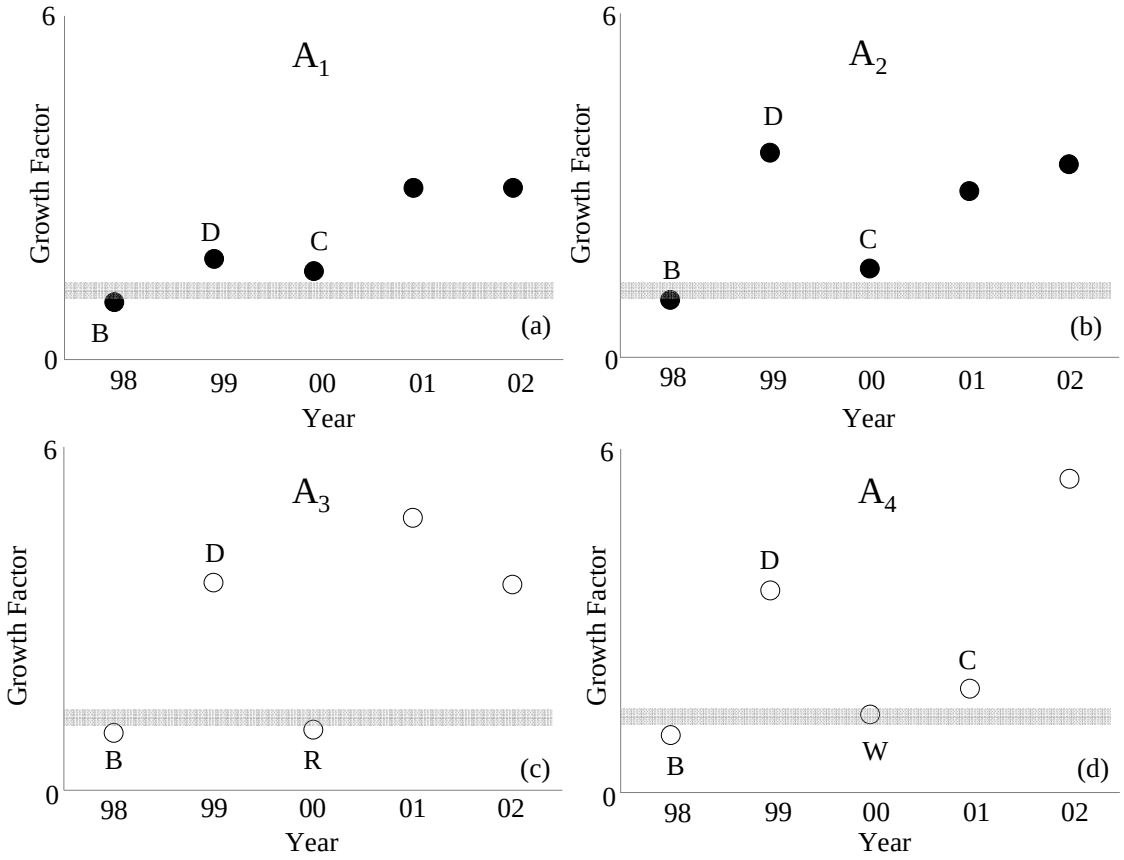

Figure-8 Changes in growth factor of the four sites shown in Figure-7(b)

A_1 and A_2 in Figure-7(b) are the sites that were detected in 1999 and confirmed in 2000 by the proposed proactive detection, while A_3 and A_4 are sites that were detected in 1999 but not confirmed in 2000. Figure-8 shows how the growth factor of these four sites changed over the years since 1998. The data displayed in Figure 8(a) and (b) are from sites located on I-880 N. The data points labeled D in both figures denote the year that the site was detected and the data points labeled C denote the year that the site was confirmed. The growth factor in 1999 and 2000 is based on excess collision observed in 1998 (i.e., the area under $K_y(d)$): the growth factor of the bin in 1998 is 1 (see data points labeled B in Figure-8) The growth factor in 2001 and 2002 shows continued increase.

A_3 was detected in 1999 due to a sudden increase in growth factor, but, was rejected in 2001 (see the white circle labeled R in Figure-8(c)). Although detailed site conditions that resulted in marked increase in the growth factor at A_3 were not confirmed, investigation of empirical data from other routes revealed that sudden surges in growth factor were often caused by construction or maintenance activities that require long term (several months) lane closures. The ideal situation would be keeping track of lane closure activities – not planned closure, but implemented closures - as part of the hot spot identification procedure. However, these data are not yet linked together in most states' department of transportation. To reduce falsely detected sites due to construction or maintenance activities, the procedure only reports sites that were not rejected by the sequential hypothesis testing (Wald, 1945) in two or more consecutive years. Figure-8(c) shows that this site would have been detected again in 2001 and rejected again in 2002. What was observed at A_3 could be the due to the regression-to-the-mean

phenomenon. A_4 was detected in 1999 and its growth factor lied between $P^* \pm \Delta$ such that it was neither rejected nor accepted in 2000 (see data point labeled W in Figure-8(d)). The site was re-evaluated in 2001 and was confirmed.

Notice how the changes in growth pattern displayed in Figure-8(b) and (d) are similar. If P^* was set at a lower value or a smaller Δ was used, A_4 could have been confirmed in 2000 instead of 2001. There are two parameters in the proposed method that will affect the number of bins detected. Lower values of r^* or $r^*(d)$ will generate more bins and increase in the size of bins. The value of h^* used in the current study is based on empirical data observed along 82 miles of freeways. Lower values of h will increase the number of non-zero $H_y(d)$ resulting in more bins to be evaluated in the subsequent part of the analysis. The criteria for detecting a site are based on the attributes of bins which indicate how reproducible the collision pattern was and the significance of the growth factor and net changes.

4. Concluding Remarks

In an effort to remedy the shortcomings of existing retroactive hot spot identification procedures, this paper has presented a method for proactively detecting a hot spot where its safety level slowly deteriorates over time. The proposed method first compares normalized CRP plots from previous years to detect sites with significant changes in collision profile and uses sequential hypothesis testing to identify the target sites.

To ensure that small peaks in CRP are not detected due to their large variance compared to their rate, the procedure only compares the peaks that have values greater than r^*. As explained, $r^*(d)$ needs to be used in practice to account for differences in threshold values used for different roadway groups. Using growth factors and net changes together, the sites where significant growth was observed were detected; many of those sites were detected as a result of construction activities that took place along the roadway. To minimize both false positive and false negative rates, the proposed method employed sequential hypothesis testing to detect sites whose growth factor remained above the 70^{th} percentile for more than two years. However, whether waiting two years was an adequate time period to properly address the issues that arises from the regression-to–the-mean phenomenon (Hauer, 1997) was not investigated in the current study. This task will remain the subject of future research.

The practical advantages of being able to proactively detect a hot spot cannot be overstated. Proactive detection allows safety engineers to stay “one step ahead” of safety issues as they arise and before they become worse and claim lives and injure motorists. It also helps compensate for the long periods of time needed to identify, procure, and implement safety countermeasures once a hot spot is identified in the present tense. The value of such a robust hot spot forecasting approach is significant and has the potential to shape hot spot prediction methodologies moving forward.

Reference

1. Chung, K., Ragland, D.R., Madanat, S and Oh, S. (2009), The Continuous Risk Profile Approach for the Identification of High Collision Concentration Locations on Congested Highways, *Proceeding of 19th ISTTT*, pp 463~480.
2. Basseville, M. and Nikiforov, I.V. (1993) Detection of Abrupt Changes: Theory and Application, Prent.Hall, www.irisa.fr/sigma2/kniga.
3. Wald, A. (1945) "Sequential Tests of Statistical Hypotheses". *The Annals of Mathematical Statistics* 16 (2), pp. 117–186
4. Hauer, E (1997) Observational before-after studies in road safety, Pergamon, Elsevier Science Ltd.