공학석사학위논문

계층적 군집분석 기반의 Continuous Risk Profile을 이용한 고속도로 사고취약구간 선정

Identifying Hotspots on Freeways using the Continuous Risk Profile with Hierarchical Clustering Analysis

2013년 2월

서울대학교 대학원

건설환경공학부

이 서 영

계층적 군집분석 기반의 Continuous Risk Profile을 이용한 고속도로 사고취약구간 선정

Identifying Hotspots on Freeways using the Continuous Risk Profile with Hierarchical Clustering Analysis

지도교수 전 경 수

이 논문을 공학석사 학위논문으로 제출함.

2012년 10월

서울대학교 대학원

건설환경공학부

이 서 영

이서영의 공학석사 학위논문을 인준함.

2012년 12월

위 원 장 ____________________(인)

부위원장 ____________________(인)

위　　원 ____________________(인)

ABSTRACT

Identifying Hotspots on Freeways using Continuous Risk Profile with Hierarchical Clustering Analysis

Lee, Seoyoung
Department of Civil and Environmental Engineering
The Graduate School
Seoul National University

Crashes that occur on freeways generally cause extensive damage and injuries. Therefore, there is a need for the development of techniques for managing and reducing the number of crashes that occur by identifying hotspots efficiently within a limited budget.

Among existing network screening methods, the Continuous Risk Profile(CRP) model well known to have performance that is superior to competing methodologies. However, to identify hotspots, the CRP model requires the use of safety performance functions which are used as a rescaling factor.

In this study, I utilized hierarchical clustering analysis to use the Continuous Risk Profile, which had great results for identifying hotspots in nations and regions in which no safety performance functions have been established.

I identified hotspots by replacing safety functions that are used as a rescaling factor in the CRP model with expected average crash frequency following groups that were obtained by hierarchical clustering analysis.

I compared the hotspots identified by the existing CRP model and the hotspots identified by the CRP model using hierarchical clustering analysis. Also, I compared the hotspots identified by the CRP model using hierarchical clustering analysis and the Sliding Moving Window method and the Peak Searching method. These comparisons indicated that the CRP model using hierarchical clustering analysis, just like the existing CRP model, was more effective at identifying hotspots on freeways than other network screening methods.

Keywords : Continuous Risk Profile, hierarchical clustering analysis, hotspots, rescaling factor, safety performance functions

Student Number : 2011-20993

〈List of Contents〉

<List of Tables>

<List of Figures>

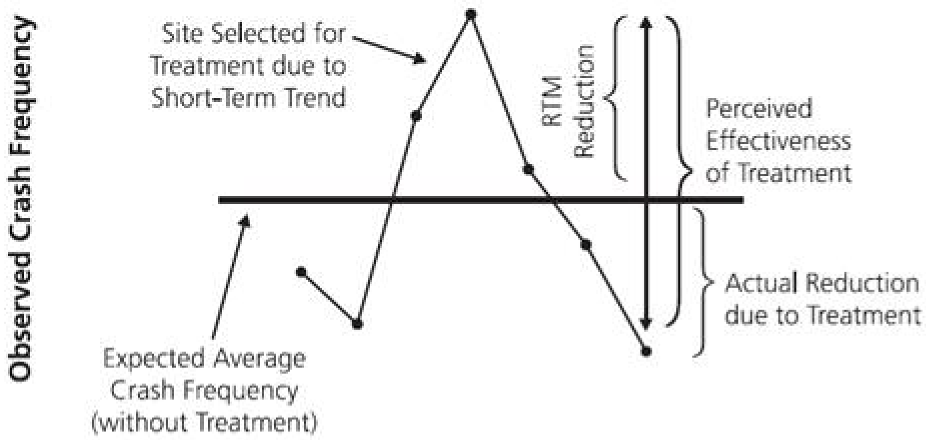
Chapter 1. Introduction

1.1. Background and Purpose of the Study

Most crashes that occur on freeways usually involve multiple vehicles, and there is a higher probability of casualties than in crashes on general roadways. Therefore, we need to enhance the safety of freeways by developing and implementing countermeasures for hotspots identified by network screening methods. In the identification process, more effective network screening methods can reduce the experts' time and labor for investigating hotspots on freeways. Also, the identification process will increase the effectiveness of local governments in that their limited budgets result in their preferentially investigating sites that are expected to have a huge potential for improving safety.

In domestic case, the Korea Expressway Corporation is identifying hotspots of freeway and preparing countermeasures of hotspots. There are many network screening methods for identifying hotspots, but the Korea Expressway Corporation uses a method that focuses on the frequency of crashes because it is generally applicable. However, the use of crash frequency has several disadvantages.

Most importantly, its effectiveness in reducing crashes is limited because hotspots are identified under inflexible standards that do not consider traffic volumes and regional characteristics. Also, it has possibility of identifying hotspots incorrectly due to the regression-to-the-mean (RTM)[1].



〈Figure 1-1〉 The Regression-to-the-Mean

(source : AASHTO 2010, page 3-12)

Three network screening methods, i.e., the Sliding Moving Window, Peak Searching, and Continuous Risk Profile method, are used extensively in the identification of hotspots. The Continuous Risk Profile (CRP), which was developed by the California Department of Transportation (Caltrans) in 2007, has several advantages over other network screening methods.

First, the Continuous Risk Profile can observe the variation and passage of peaks by year, which is an indicator of hotspots. the CRP is different from other network screening methods with respect to the way it observes the continuous passage of changing peaks. Second, CRP' s false positive rat e[2] is less than those of other network screening methods. The Sliding

---

[1] regression-to-the-mean (RTM): When a period with a comparatively high crash frequency is observed, it is statistically probable that the following period will be followed by a comparatively low crash frequency. (AASHTO 2010, page 3-11)

[2] false positives: sites generated by network screening methods, but where no specific highway deficiency is identified through further follow-up (Chung and Ragland 2007, page 3)

Moving Window and Peak Searching methods identify hotspots on freeways based on aggregated data for traffic collisions irrespective of the years in which they occurred, but the CRP determines hotspots whether or not reproducibility by year exists. Therefore, CRP can reduce the rates of false positives in contrast with other network screening methods. Finally, CRP can proactively detect the systematic deterioration of sites that have the possibility of becoming hotspots. In this case, peaks on a graph produced using CRP may not appear in the early analysis period, but, with time, some meaningful peaks are generated. Proactively detectable sites that are not reproducible over the entire analysis period have high probability of being hotspots, so careful monitoring of these sites in advance contributes greatly to the goal of preventing collisions.

Because of these strengths, CRP identifies hotspots more effectively than other network screening methods. However, CRP requires safety performance functions(SPFs)[3] that are used as rescaling factors. Few nations and regions have established SPFs along freeways, so most of them cannot use CRP because effective performance cannot be guaranteed if CRP is used without SPFs.

Therefore, the goal of this study was to identify hotspots through CRP that nations and regions without SPFs can utilize. Thus, I wanted to prove that the false positive rates of CRP can be less than those of other network screening methods if I used expected average crash frequency by clustering

---

[3] safety performance functions (SPFs): Statistical base models are used to estimate the average crash frequency for a facility type with specified base conditions (AASHTO 2010, page 3-16)

based on traffic volumes and the number of lanes as a rescaling factor in the CRP method.

The specific approach used to accomplish the goal of the study is described below.

First, hotspots were identified by the existing CRP method and by the CRP method using hierarchical clustering analysis, and the results of the two methods were compared. Second, the results of the CRP method were compared with those other network screening methods, i.e., the Sliding Moving Window and Peak Searching method.

The above comparisons indicated that the CRP method, just like the existing CRP, is more effectively identified hotspots on freeways than the other network screening methods.

1.2. Scope of the Study

In chapter two, three network screening methods (Sliding Moving Window, Peak Searching, and Continuous Risk Profile) are explained, and the literature related to the CRP method is reviewed. In chapter three, the results are presented and discussed when I compared the existing CRP method developed by Caltrans and the modified CRP method that I proposed using hierarchical clustering analysis. In chapter four, I present comparisons of the hotspots identified by the various network screening methods, i.e., the Continuous Risk Profile, Sliding Moving Window, and Peak Searching methods. Finally, chapter five presents my conclusions and the contribution this study makes.

Chapter 2. Network Screening Methods and Literature Review

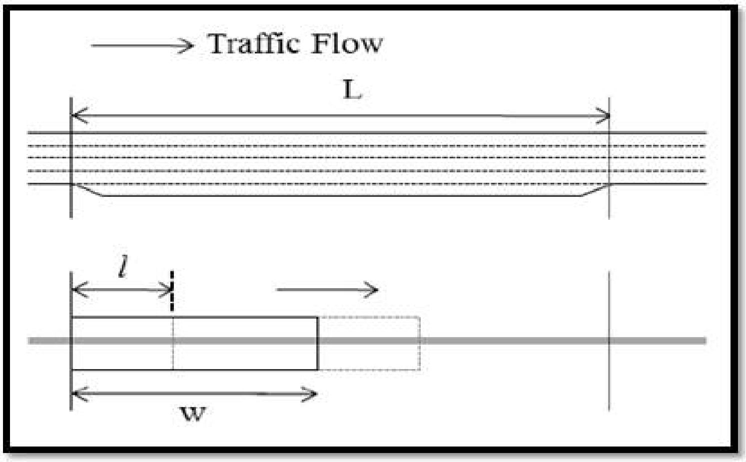
2.1. Network Screening Methods

2.1.1. Sliding Moving Window Method

The basis for the Sliding Moving Window method is stated as follows: "A window of a specified length, is conceptually moved along the road segment from beginning to end in increments of a specified size. The performance measure chosen to screen the segment is applied to each position of the window, and the results of the analysis are recorded for each window. From all the windows that pertain to a given segment, the window that shows the most potential for reduction in crash frequency out of the whole segment is identified and is used to represent the potential for reduction in crash frequency of the whole segment. The potential safety improvements from all the windows are compared, and the maximum value is used to represent the potential for collision reduction for the whole segment" [4]

The Sliding Moving Window method has two characteristics. First, it divides the entire roadway into segments with a homogeneous property. Second, each window overlaps the next window.

[4] AASHTO 2010, page 4-15



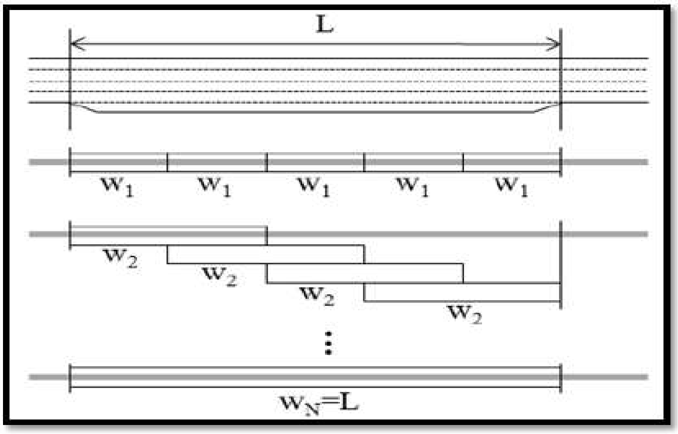
〈Figure 2-1〉 The Sliding Moving Window Method

(source : Kwon et al. 2012, page 9)

2.1.2. Peak Searching Method

The basis for the Peak Searching method is stated as follows: "Each individual roadway segment is subdivided into windows of similar length, potentially growing incrementally in length until the length of the window equals the length of the entire roadway segment. The windows do not span multiple roadway segments. For each window, the chosen performance measure is calculated, and the maximum value is used to represent the potential for collision reduction for the whole segment, like the Sliding Moving Window Method. Then, the results are subjected to precision testing. The precision of the performance measures from Peak Searching Method can be assessed by the coefficient of variation. If none of the performance measures for the initial windows is found to have the desired precision, the length of each window is incrementally moved forward." [5]

The Peak Searching method uses the process of segmentation in the same way as the Sliding Moving Window method, but the windows in the Peak Searching method do not overlap, with the possible exception that the last window may overlap the previous window.



〈Figure 2-2〉 The Peak Searching Method

(Source : Kwon et al. 2012, page 9)

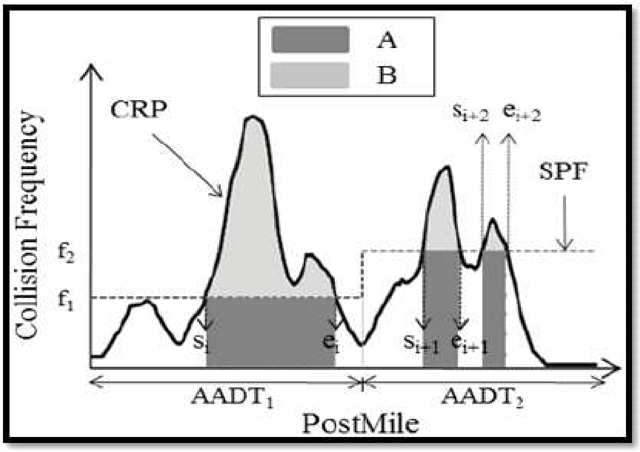
2.1.3. Continuous Risk Profile

The basis for the Continuous Risk Profile method is stated as follows: The CRP method is "a new method for assessing collision risk along a roadway that addresses the limitations of the fixed window approach. This method is fitted to the underlying true risk and reflects a measure of risk interpretable as collision risk per unit distance of roadway. This method can both proactively and reactively monitor the changes in risk over the years.

5 AASHTO 2010, page 4-16

The resulting risk profile produces variations in risk over freeway segments that are highly reproducible over time" [6]. This reproducibility identifies hotspots on freeways.

The CRP method has two characteristics. First, it does not use the process of segmentation that is used in the Sliding Moving Window and Peak Searching methods. Second, it requires SPFs that are used as rescaling factors.



〈Figure 2-3〉 The Continuous Risk Profile Method

(Source : Kwon et al. 2012, page 9)

[6] Chung and Ragland 2007, page 4; Chung and Ragland 2009, page 468

2.2. Literature Review

〈Table 2-1〉 Literature Review about the CRP Method

Paper	Performance Measures	Removing Random noise	Rescaling Factor	Note
Chung (2007)	Average Crash Frequency	Moving Average	Average Crash Frequency (during the analysis periods)	the first stage of the development
Yu (2008)	Average Crash Frequency	Moving Average	Average Crash Frequency (during the analysis periods)	the first stage of the development→ the application of korean collision data
Chung (2009)	Average Crash Frequency	Moving Average	SPFs	the identification of false positive rate's reduction
Oh (2009)	Average Crash Frequency	Moving Average	SPFs	the application of CRP under wet Condition
Chung (2010)	Average Crash Frequency	Weighted Moving Average	SPFs	the identification of proactive detection about the systematic deterioration
Kim (2011)	Average Crash Frequency & EPDO Average Crash Frequency	Moving Average & Weighted Moving Average	-	the estimation of crash expenses for hotspots
Kwon (2012)	Average Crash Frequency	Weighted Moving Average	SPFs	the application of SPFs by Caltrans and SPFs by the study

The American Association of State Highway and Transportation Officials (AASHTO) (2010) published a Highway Safety Manual that contained the analysis procedure related to the management of roadway safety. In the chapter related to network screening, the Manual discussed the Sliding Moving Window, Peak Searching, and Simple Ranking methods and presented some examples about these methods.

Chung and Ragland (2007) developed the CRP method to supplement the Sliding Moving Window method. In their study, an average crash frequency per unit distance of roadway was used as a performance measure. Sites that presented continuous peaks during the analysis periods using the CRP method were identified as hotspots on the roadways. The strengths of the CRP method are reproducibility and the identification of variations in continuous risks. Their study was the first stage of the development of the CRP method, so it had a very simple form.

Yu (2008) adjusted some variables in the CRP method that was developed in 2007 to make it better suited for circumstances in Korea. The collision data occurred on a Korean four freeway sections (207 km) over a 10-year period was applied to the CRP method. His study identified the continuous risks associated with the Korean freeway and enhanced the precision associated with the identification of hotspots. Also, the influence of safety improvements was expanded. The contribution of his study was that it applied Korean collision data to the first stage of the development of the CRP method.

Chung and Ragland (2009) studied the proactive detection of hotspots and the benefit of safety improvements observed at the location of the project and at neighboring sites. And, in this study, the number of collisions used by the Caltrans was used as a rescaling factor to meet the requirement of a significance level of 99.5%. Through these procedures, the hotspots identified by the CRP method and the hotspots identified by the Sliding Moving Window method were compared, and it was verified that the CRP method has a lower rate of false positives.

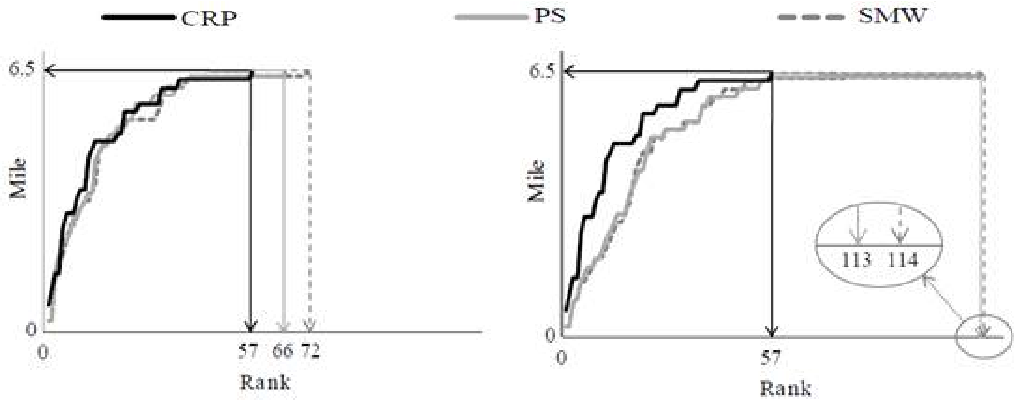
Oh et al. (2009) used the CRP method to identify the common features of sites that may contribute to high collision rates when the pavement is wet. The results reported in this paper indicated that speeding was the primary factor in collisions, irrespective of the condition of the pavement, and it became an even more dominant factor when the pavement was wet at all observed locations. Other collision factors are rapid spatial changes, narrow lanes, the absence of a median, and increased width of the freeway, and diminished drivers' visibility and so on.

Chung et al. (2010) researched an algorithm for proactively detecting sites that have undergone systematic deterioration but do not yet have collision rates that are high enough for the sites to be classified as high-collision locations. In this paper, the authors identified sites that had undergone systematic deterioration by the weighted moving average and the cumulative sum (CUSUM) algorithm. Also, the length of the 2L, the window was used in the CRP method, was changed from 0.1 mi to 0.4 mi in an attempt to determine the optimal length. The optimal length was determined to be 0.2 mi to remove a random noise to the maximum ewtent possible and not to act on peaks of sites.

Kim (2011) used collision data from the Kyeongbu expressway and the

Seohaean expressway collected over a period of three years in the CRP method and compared the results with the differences followed by the Equivalent Property Damage Only (EPDO), the length of the window, and Average Annual Daily Traffic (AADT). The results indicated that it is relatively more stable to use the average crash frequency provided by EPDO as the performance measure. Also, the author estimated the expenses of crashes containing AADT for hotspots.

Finally, Kwon et al. (2012) fully demonstrated the sensibility of analysis caused by segmentation and the difference of performance among three network screening methods. In their paper, they evaluated the performance of three network screening methods. Identical input data were used in the analyses that were conducted in all of the methods, but the methods differed is the ways that the roadways were segmented. The SPFs used in Caltrans and the SPFs developed by this study were used as rescaling factors. The conclusion of this paper was expressed in terms of segmentation, i.e., long segments and short segments. In this paper, the number of ranks means the number of sites required to detect true hotspots. Also, a large number of sites implies that the analysis will take a longer time to complete, which decreases the effectiveness of the approach and increases the danger that drivers face. The Sliding Moving Window and Peak Searching methods have different numbers of sites, depending on how the segmentation was done. However, the CRP method has a constant number of sites irrespective of segmentation, because it does not include the process of segmentation. Also, the CRP method had the smallest ranks, irrespective of segmentation. Thus, in this paper, the authors concluded that the performance of the CRP method was the best among the three network screening methods.



Analysis by Long Segments　　　Analysis by Short Segments

〈Figure 2-4〉 Comparison of the Performance among the Three Network Screening Methods

(Source : Kwon et al. 2012, page 13)

The CRP method is in the early stages of development, and the paper written by Kwon at el. (2012) determined that its use of SPFs as a rescaling factor was effective in identifying hotspots, compared with other network screening methods. However, verifying the effective performance of the CRP method depends on the establishment of SPFs in nations and regions. Until that is done, we cannot conclusively state that the CRP method without SPFs can identify hotspots better than other network screening methods. Therefore, the goal of the present study was to prove that the CRP method can effectively identify hotspots using hierarchical clustering analysis instead of SPFs as a rescaling factor.

Chapter 3. The Process of Continuous Risk Profile

3.1. Raw Data

The objective of this study was to compare the performance of the existing CRP method with the performance of the CRP method using hierarchical clustering analysis. Korean collision data are of limited use because SPFs have not been established for Korean expressways. Thus, in this study, I attempted to conduct analyses by using the collision data and the SPFs for the I-880 freeway in Alameda County, California, from 2004 through 2008.

These are the scope for analyses and requiring data as presented below.

〈Table 3-1〉 The Temporal · Spatial Scope of Analysis

Division	Contents
Temporal Scope	2004~2008 (for 5 years)
Spatial Scope	I-880 freeway in Alameda County, California, United States
requiring data	the collision data, AADT, the number of lanes

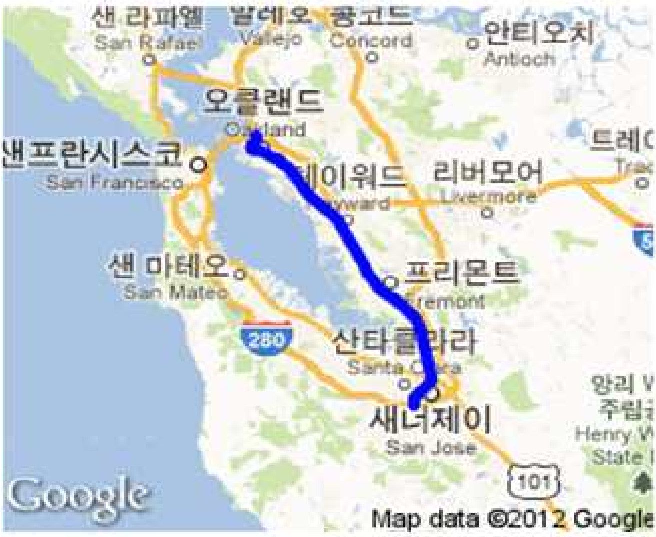
Source : Statewide Integrated Traffic Records System (SWITRS)

〈Table 3-2〉 The Number of Collisions for the I-880 Freeway

(Unit : number)

Year	I-880 Northbound	I-880 Southbound
2004	1,589	1,744
2005	1,634	1,675
2006	1,479	1,560
2007	1,460	1,507
2008	1,376	1,399

Source : Statewide Integrated Traffic Records System (SWITRS)



〈Figure 3-1〉 Location for the I-880 Freeway

(source : http://pems.dot.ca.gov/)

The collision data for the I-880 freeway in Alameda County, California, was collected by the Statewide Integrated Traffic Records System (SWITRS) and is expressed in relative postmiles from adjacent cities, so I had to convert relative postmiles to absolute postmiles. The conversion process was accomplished using an Excel spreadsheet and the guidelines in the Performance Measurement System (PeMS) of Caltrans[7], as shown in Figure 3-2.

	A	B	C	D	F	G	H	O	P
1	accident_number	district	county	route_name	pm_prefix	postmile	AbsPM	accident_date	accident_time
2	2988350	4	ALA	880	R	0.03	10.532	2008-02-11	1905
3	3099627	4	ALA	880	R	0.03	10.532	2008-11-19	1800
4	2972807	4	ALA	880	R	0.04	10.542	2008-01-24	2030
5	2986485	4	ALA	880	R	0.05	10.552	2008-02-29	1930
6	3019617	4	ALA	880	R	0.06	10.562	2008-04-13	1555
7	3048420	4	ALA	880	R	0.07	10.572	2008-07-03	1755
8	3051132	4	ALA	880	R	0.08	10.582	2008-08-24	1310
9	3088275	4	ALA	880	R	0.1	10.602	2008-10-02	1625
10	3012298	4	ALA	880	R	0.19	10.692	2008-04-13	1220
11	2978886	4	ALA	880	R	0.24	10.742	2008-02-15	1740
12	3089929	4	ALA	880	R	0.28	10.782	2008-10-23	10
13	3052786	4	ALA	880	R	0.3	10.802	2008-07-26	210
14	3103386	4	ALA	880	R	0.3	10.802	2008-11-26	404
15	3067661	4	ALA	880	R	0.36	10.862	2008-09-05	1720
16	3026517	4	ALA	880	R	0.4	10.902	2008-06-21	120
17	3028679	4	ALA	880	R	0.61	11.112	2008-06-01	100
18	3056350	4	ALA	880	R	0.8	11.302	2008-08-18	600
19	3056351	4	ALA	880	R	0.8	11.302	2008-08-18	602

〈Figure 3-2〉 The Collision Data for the I-880 Freeway in Alameda County, California

3.2. Calculation of a Performance Measure per Unit Distance

This process is to calculate a performance measure per unit distance. Possible performance measures for network screening include average crash frequency, crash rate, EPDO average crash frequency, excess predicted average crash frequency using SPFs, and expected average crash frequency with empirical Bayes (EB) adjustment.[8] In this study, I used average crash

[7] http://pems.dot.ca.gov/

frequency as a performance measure.

3.3. The Application of the Weighted Moving Average

The weighted moving average applied to the average crash frequency that chosen as a performance measure for network screening. The moving average is a process for removing random noise from a performance measure. Thus, the weighted moving average, which places emphasis on the center of the windows, can visually display peaks in the CRP method.

The variables used in the weighted moving average were 2L (Window) and l (Increment). The settings chosen for these variables affect the results of the CRP method. As the length chosen for a window increases, the random fluctuations of the CRP method decrease. This removes random noise, but the effects of peaks (hotspots) also are reduced. Therefore, 0.2 mi (2L) and 0.01 mi (l) were used as optimal lengths because they filtered random noise effectively and did not affect the peaks[9].

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

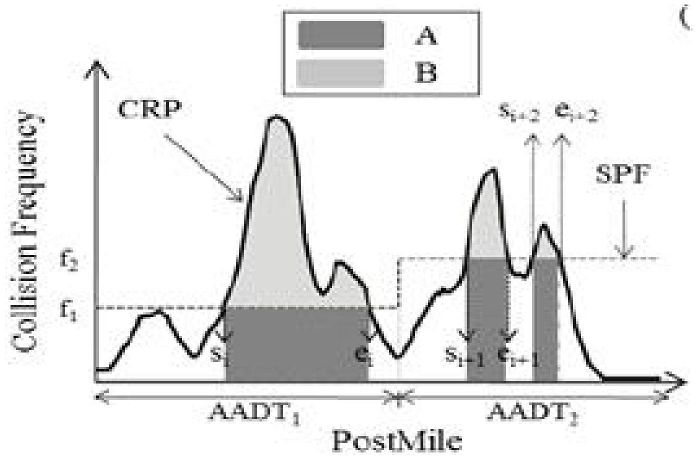
[8] AASHTO 2010, pages 4-10~14

[9] Chung et al. 2010, page 929

3.4. The Application of Rescaling Factors

3.4.1. The Existing Continuous Risk Profile

The existing CRP method uses SPFs as a rescaling factor in the procedure of identifying hotspots. M(d) means that the weighted moving average applies to a performance measure per unit distance, and B(d) [section A of Figure 3-3] means the expected average frequency of crashes generated by SPFs per unit distance. Therefore, the section of M(d) - B(d) [section B of Figure 3-3] is detected as excess crash frequency, i.e., such sections are hotspots on the freeway.



〈Figure 3-3〉 The Existing Continuous Risk Profile

(source : Kwon et al. 2012, page 9)

To get the SPFs that are used as a rescaling factor, the highway rate group, which is classified by, e.g., the number of lanes, AADT, and geometric structures, must be known. Caltrans classifies freeways and highways into 67 groups based on facility features. The I-880 freeway is classified by eight of the groups as follows in order to determine the expected average crash frequency by SPFs[10].

〈Table 3-3〉 The Highway Rate Groups for the I-880 Freeway

Division	Description	Base Rate	ADT Factor
H55	Rural Freeway 5-6 lanes	0.25	0.0050
H56	Rural Freeway 7 lanes or more	0.20	0.0035
H61	Suburban Freeway 5-6 lanes	0.20	0.0060
H62	Suburban Freeway 7 lanes or more	0.25	0.0035
H64	Urban Freeway 5-6 lanes	0.40	0.0055
H65	Urban Freeway 7-8 lanes	0.40	0.0035
H66	Urban Freeway 9-10 lanes	0.35	0.0030
H67	Urban Freeway 11 lanes or more	0.35	0.0025

Source : Caltrans 2002, page 18~19

In each segment of the I-880 freeway, the expected average crash frequency was calculated using the base rate, the Annual Daily Traffic (ADT) factor by the highway rate groups, and AADT. The formula used to determine the expected average crash frequency is expressed as:

[10] Caltrans 2002, page 17~19

$$R_E = Base\,Rate + ADT\ Factor \times Average\,ADT$$

$$N_E = \frac{R_E \times Travel}{10^6}$$

$$Travel = AADT \times 365 Days \times Length$$

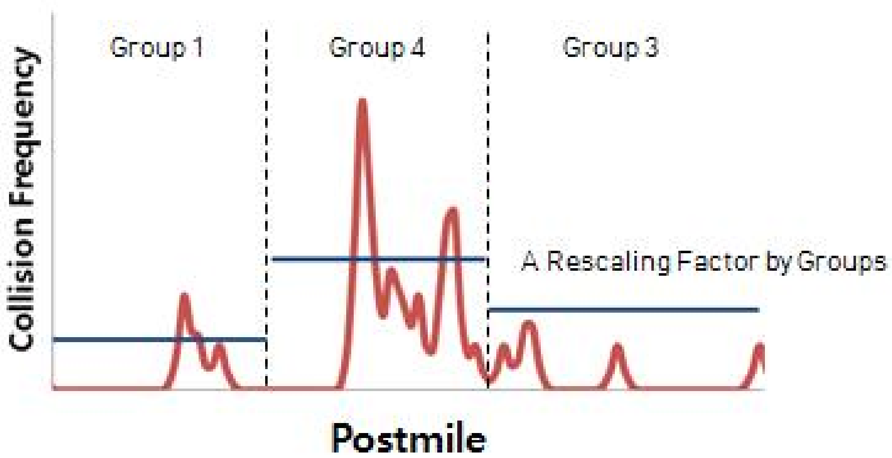
R_E = expected collision rate per Vehicle–Mile–Travel (VMT) determined for highway group E

N_E = expected number of collisions for highway group E

The process above is used to calculate N_E, the expected average crash frequency, from SPFs. It acts as a rescaling factor in the existing CRP method.

3.4.2. The Continuous Risk Profile using Hierarchical Clustering Analysis

In this study, I compared the results of the existing CRP method with the CRP method using hierarchical clustering analysis. The latter allows the use of other features of a freeway as a rescaling factor instead of requiring the use of SPFs. Useable characteristics of a freeway include traffic volumes, number of lanes, and geometric structures; in this study, I selected AADT, the number of lanes, and the two combined as variables, which were used to conduct hierarchical clustering analysis. The expected average crash frequencies were calculated by hierarchical clustering analysis according to three types of variables. These expected average crash frequencies became rescaling factors B(d), which was applied to M(d).



〈Figure 3-4〉 The Continuous Risk Profile using Hierarchical Clustering Analysis

The establishment of SPFs in the existing CRP method involves the subdivision of the roadway and regression analysis. Subdividing the roadway into homogenous segments is a process that has a great effect on the precision of the SPFs, because subdividing the roadway must consider traffic volumes as well as specific geometric structures. As a result, the process of establishing SPFs requires many related materials and is difficult to complete in a short period of time. Therefore, this study identified hotspots on the California freeway with the CRP method using hierarchical clustering analysis instead of SPFs as a rescaling factor.

In this study, hierarchical clustering analysis was conducted using AADT, the number of lanes, and the combination of the two as variables to examine the case of average crash frequency per unit distance. Using these variables and the specified case, the expected average crash frequency was calculated and used as a rescaling factor.

The purposes of clustering analysis are 1) to understand target groups through classification of observed objects and 2) to use them effectively. Hierarchical clustering analysis, in particular, gradationally categorizes observed objects to some groups that are internally homogeneous and forms clusters. Methods used to form clusters include minimum clustering, maximum clustering, mean linkage clustering, average linkage clustering between-group of intra-group, and Ward' s method[11]. In this study, hierarchical clustering analysis was conducted using Ward's method, decreasing in variance for the cluster using SPSS.

There are rescaling factors by groups of I-880 Northbound (2005) as

[11] Kim and Kim 2007, pages 173~174

follows:

〈Table 3-4〉 Rescaling Factors by Clustering (AADT + the Number of Lanes)

Group	Count	AADT		the Number of Lanes		Expected Average Crash Frequency	
		Sum	Average	Sum	Average	Sum	Average
1	1108	80590000	72734.66	3368	3.039711	239.7368	0.216369
2	986	91961000	93266.73	2958	3	240.7368	0.244155
3	2119	239532000	113040.1	9938	4.689948	1079.737	0.50955
4	340	18891000	55561.76	2380	7	69.57895	0.204644

〈Table 3-5〉 Rescaling Factors by Clustering (AADT)

Group	Count	AADT		Expected Average Crash Frequency	
		Sum	Average	Sum	Average
1	2042	159345500	78034.04	514.7895	0.252101
2	1991	215116500	108044.5	889.6316	0.446827
3	398	52555000	132047.7	211.8421	0.532267
4	122	3957000	32434.43	13.52632	0.110871

〈Table 3-6〉 Rescaling Factors by Clustering (the Number of Lanes)

Group	Count	the Number of Lanes		Expected Average Crash Frequency	
		Sum	Average	Sum	Average
1	2062	6186	3	472.1579	0.228981
2	677	2708	4	382.4737	0.564954
3	1474	7370	5	705.5789	0.478683
4	340	2380	7	69.57895	0.204644

The CRP method using 3-hierarchical clustering analysis (variables: AADT, the number of lanes, and the combination of the two) was generated by using the expected average crash frequency to M(d).

3.5. Review of Reproducibility

K(d) means applying a rescaling factor to M(d). The process of reviewing reproducibility determines whether K(d) has exceeded the expected average crash frequency during the analysis period at a certain postmile. If continuous excess average crash frequencies were recorded at a certain postmile during the entire analysis period (2004~2008), the reproducibility of that postmile will be verified.

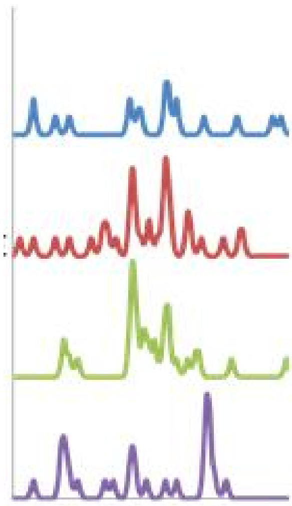
	A	B	C	D	E	F	G	H	I	J	K
1	기존	START	MID	END	Repro		mix	START	MID	END	Repro
77		23.74	23.84	23.94	5			2.57	2.67	2.77	5
78		23.75	23.85	23.95	5			2.58	2.68	2.78	5
79		23.76	23.86	23.96	5			2.59	2.69	2.79	5
80		23.77	23.87	23.97	5			2.6	2.7	2.8	5
81		23.78	23.88	23.98	5			2.61	2.71	2.81	5
82		23.79	23.89	23.99	5			2.62	2.72	2.82	5
83		23.8	23.9	24	5			3.35	3.45	3.55	5
84		23.81	23.91	24.01	5			3.36	3.46	3.56	5
85		23.82	23.92	24.02	5			3.37	3.47	3.57	5
86		23.83	23.93	24.03	5			3.38	3.48	3.58	5
87		23.84	23.94	24.04	5			3.39	3.49	3.59	5
88		23.85	23.95	24.05	5			3.4	3.5	3.6	5
89		23.86	23.96	24.06	5			3.41	3.51	3.61	5
90		23.87	23.97	24.07	5			3.42	3.52	3.62	5
91		23.88	23.98	24.08	5			3.43	3.53	3.63	5
92		23.89	23.99	24.09	5			3.44	3.54	3.64	5

〈Figure 3-5〉 Reviewing Reproducibility at a Certain Postmile

In addition to the reproducibility of excess average crash frequency during the entire analysis period, the tendency of crash frequency at the postmile can be identified by correlation.

$$S_{y,j}(d) = \frac{K_y(d)}{\int_{f_j}^{s_j} K_y(x)dx} \qquad r_{y,y-1}(j) = \frac{\int_{f_j}^{s_j}(S_{y,j}(d) - \overline{S_{y,j}})(S_{y-1,j}(d) - \overline{S_{y-1,j}})dx}{\sqrt{\int_{f_j}^{s_j}(S_{y,j}(d) - \overline{S_{y,j}})^2 dx}\sqrt{\int_{f_j}^{s_j}(S_{y-1,j}(d) - \overline{S_{y-1,j}})^2 dx}}$$

In the CRP method, the correlation changes CRP' s area per year of hotspots into the same unit area and identifies the relation before and after the year. CRP' s graphs of the identified hotspots are not affected by small changes in the distance of the postmile, due to the length of the freeway and the average speed on the freeway[12]. So, to facilitate the correlation between before-year data and after-year data, the CRP method' s peaks per year were fixed at the same postmile, and then, the analysis was completed.



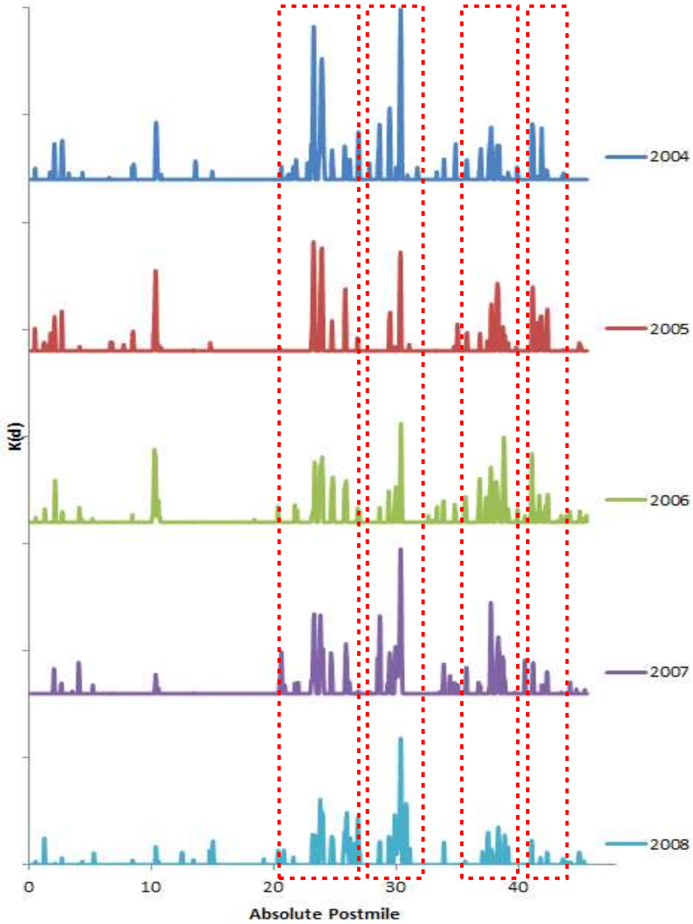
	2004	2005	2006	2007
2004	1			
2005	0.5925	1		
2006	0.4177	0.8283	1	
2007	0.3301	0.5749	0.7975	1

〈Figure 3-6〉 The Correlation between Before and After Year

[12] Chung and Ragland 2007, page 8

3.6. The Identification of Final Hotspots

Sites completing from the beginning of the CRP analysis to the review of reproducibility were selected as final hotspots by the CRP method. The CRP method identified major hotspots on the I−880 northbound freeway, as discussed below.



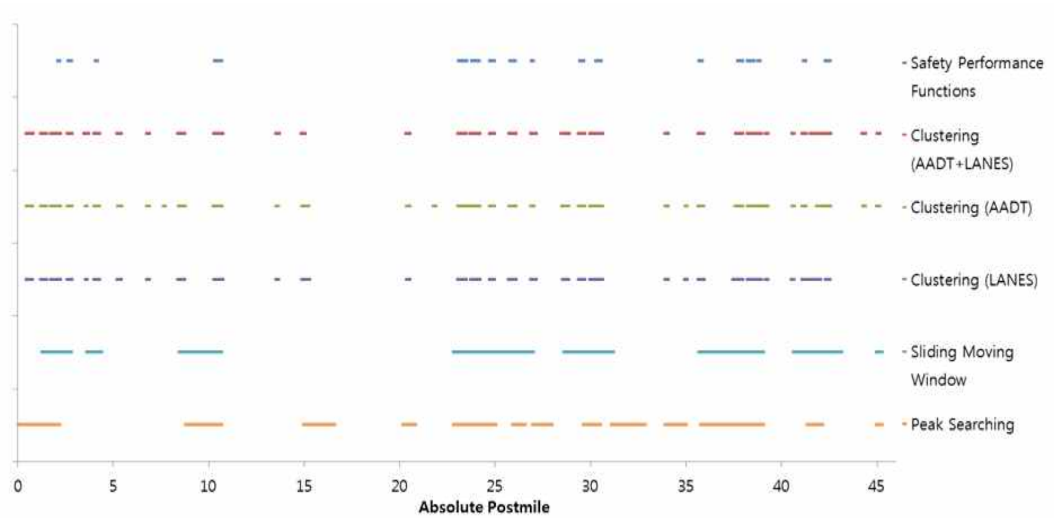
〈Figure 3−7〉 Final Hotspots in I−880 Northbound Freeway (the Existing CRP)

Chapter 4. Results

4.1. Hotspots on the I-880 Northbound Freeway

4.1.1. Hotspots Identified by Various Network Screening Methods

The various methods used to identify hotspots were the existing CRP method using SPFs, the CRP method using 3-hierarchical clustering analysis (variables: AADT, the number of lanes, and the combination of the two), the Sliding Moving Window method, the Peak Searching method, and their results were arranged in order of absolute postmile.



〈Figure 4-1〉 Comparison between Absolute Postmile of Hotspots on the I-880 Northbound Freeway

According to the above diagram, the existing CRP method using SPFs as a rescaling factor identified the smallest number of hotspots on the entire freeway. Also, the hotspots identified by the CRP method using 3-hierarchical clustering analysis included all of the hotspots identified by the existing CRP method. This indicates that the CRP method using hierarchical clustering analysis does not generate false negatives[13].

The Sliding Moving Window method and the Peak Searching method identified longer lengths of hotspots in comparison with the existing CRP method and the CRP method using hierarchical clustering analysis. This is because the window that shows the most potential for reduction in crash frequency out of the whole segment is identified and is used to represent the potential for reduction in crash frequency of the whole segment[14].

According to this result, the CRP method using hierarchical clustering analysis has higher false positive rates than the existing CRP method, but it had a lower false positive rate than the Sliding Moving Window method and the Peak Searching method.

The table below shows the lengths of the hotspots identified by each network screening method.

---

[13] false negatives : sites that require safety improvements but that not identified. Chung and Ragland 2007, page 3)

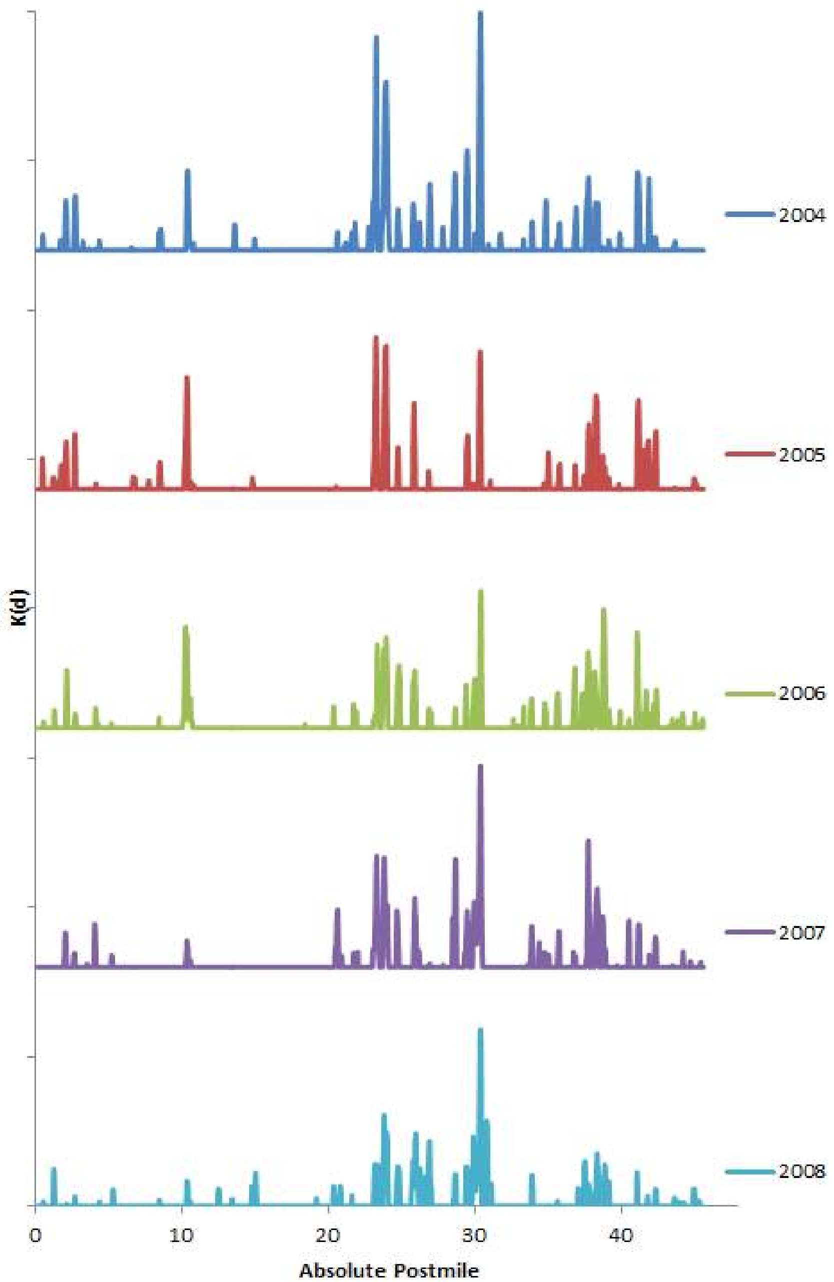
[14] AASHTO 2010, page 4-15

〈Table 4-1〉 The Lengths of Hotspots on the I-880 Northbound Freeway

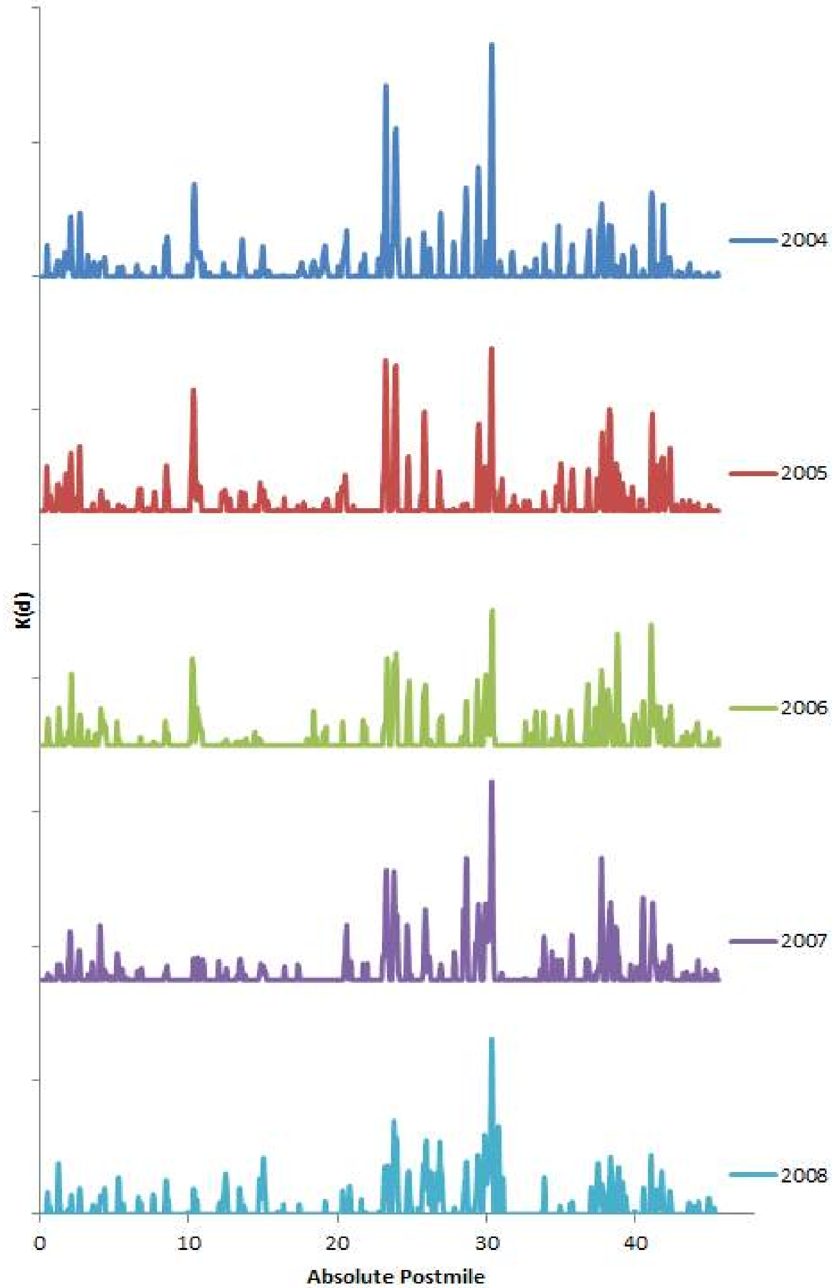
Division	The Existing CRP	The CRP using Clustering			The Sliding Moving Window	The Peak Searching
a Rescaling Factor	SPFs	Combination	AADT	# of lanes	-	-
the Lengths of hotspots (mile)	2.42	6.3	6.15	5.54	15.95	17.61
Rate (%)	5.29	13.78	13.45	12.12	34.89	38.52

The total length of the I-880 Northbound freeway is almost 46 miles. The methods, ranked in ascending order of their results for hotspot length, were the existing CRP method, the CRP method using clustering (variable: the number of lanes), the CRP method using clustering (variable: AADT), the CRP method using clustering (variable: the combination of the two), the Sliding Moving Window method, and the Peak Searching method. Subsequently, this order indicates descending order of false positive rate. According to this analysis, the CRP method using hierarchical clustering analysis had worse performance than the existing CRP method, but it had better performance than the other screening methods, i.e., the Sliding Moving Window and the Peak Searching method.

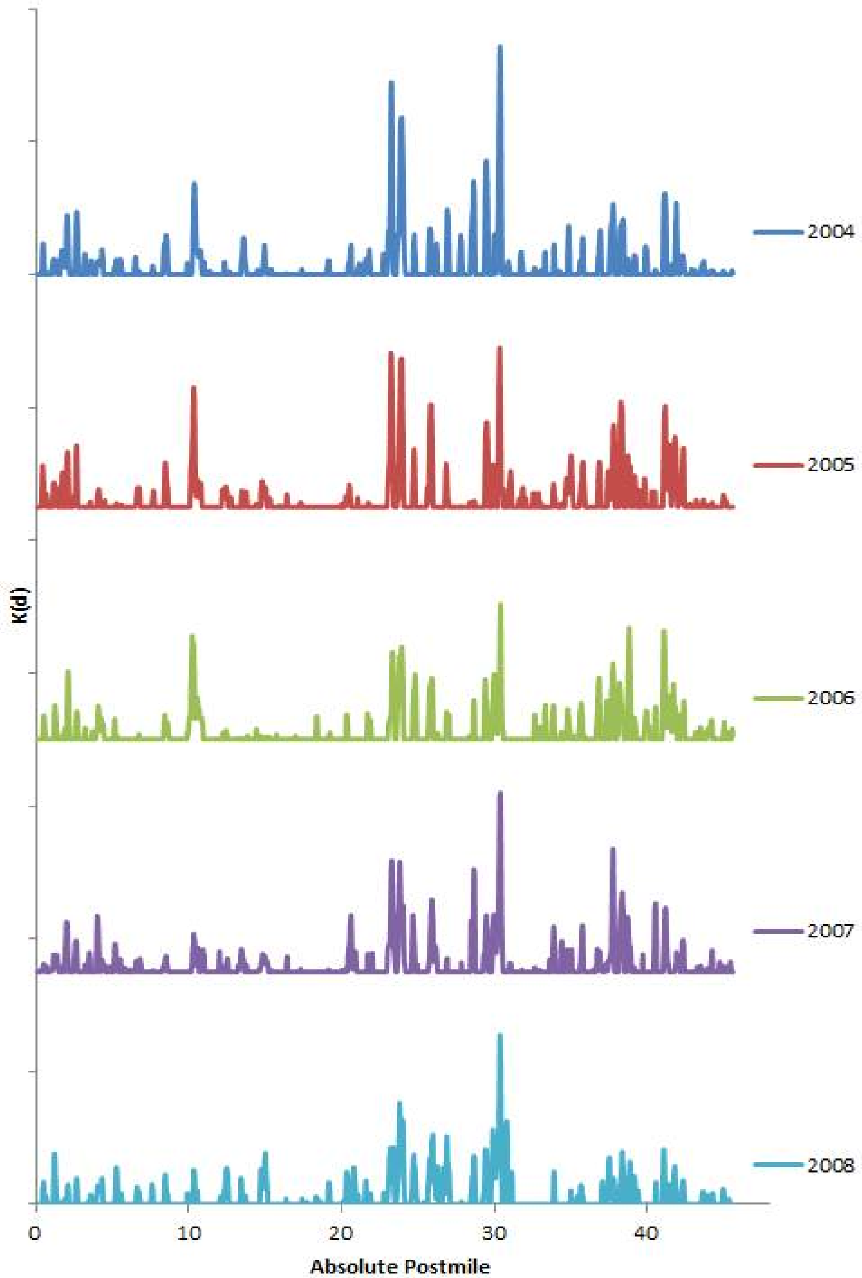
4.1.2. Hotspots Identified by 4-Continuous Risk Profile



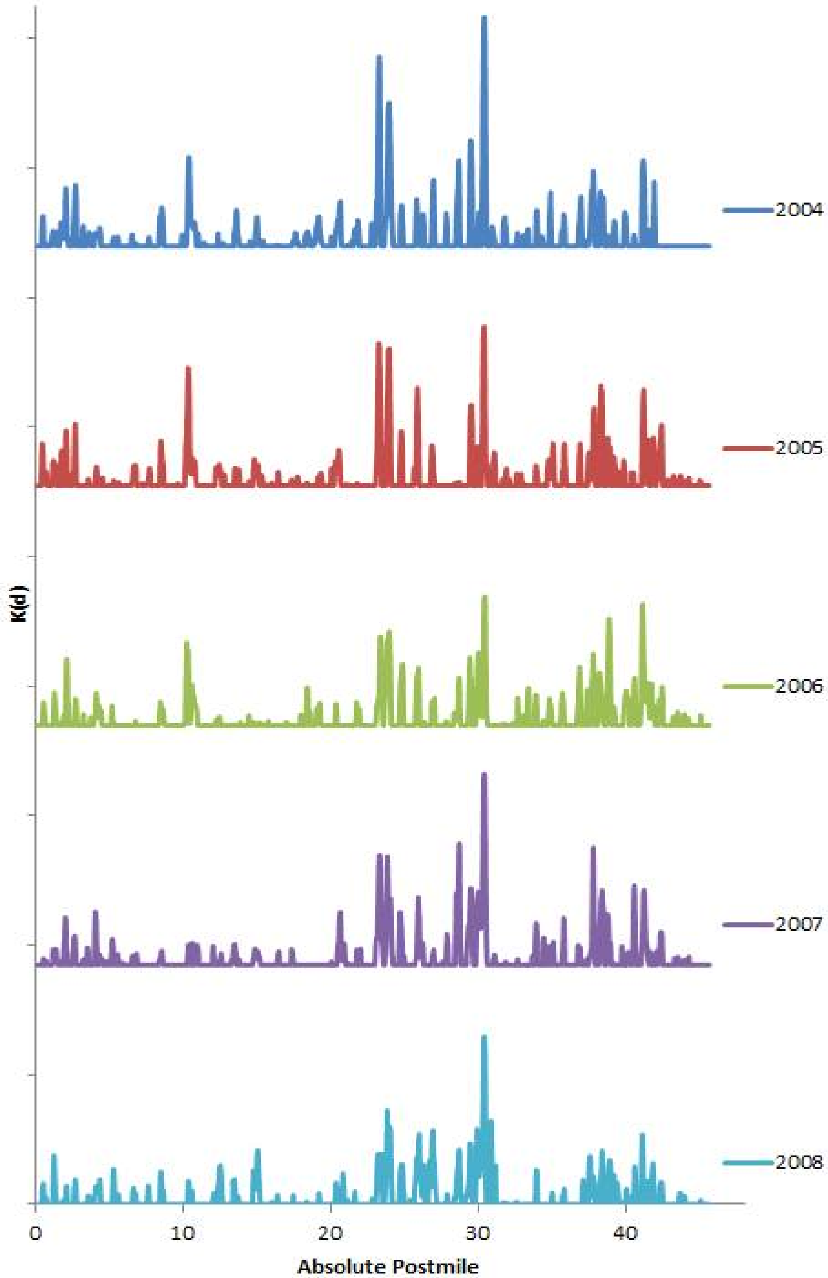
〈Figure 4-2〉 The Existing CRP on the I-880 Northbound Freeway



〈Figure 4-3〉 The CRP using Clustering (AADT + the Number of Lanes) on the I-880 Northbound Freeway



〈Figure 4-4〉 The CRP using Clustering (AADT) on the I-880 Northbound Freeway

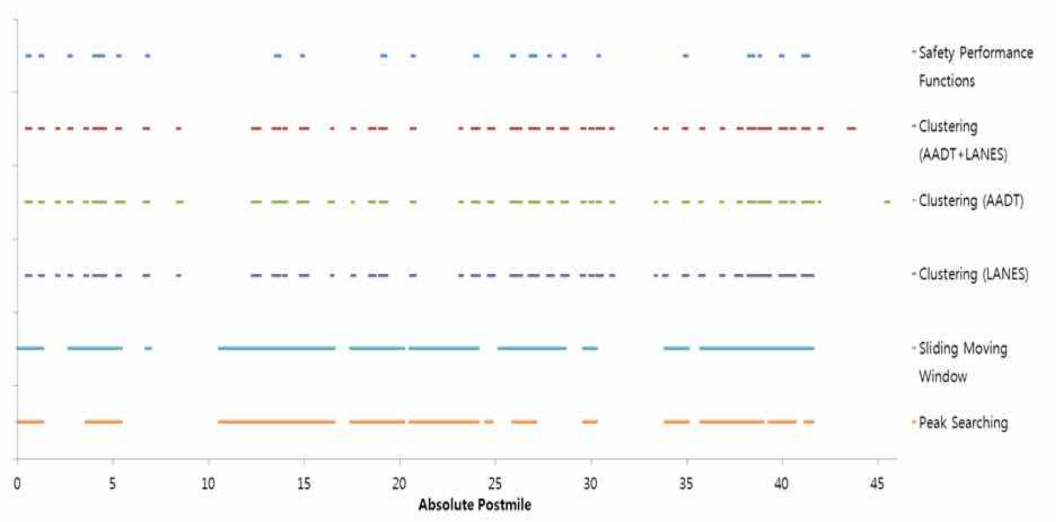


〈Figure 4-5〉 The CRP using Clustering (the Number of Lanes) on the I-880 Northbound Freeway

4.2. Hotspots on the I-880 Southbound Freeway

4.2.1. Hotspots Identified by Various Network Screening Methods

As on the I-880 Northbound freeway, hotspots were identified by the CRP method using SPFs, the CRP method using 3-hierarchical clustering analysis (variables: AADT, the number of lanes, and the combination of the two), the Sliding Moving Window method, and the Peak Searching method. The methods were arranged in order based on their absolute postmile results.



〈Figure 4-6〉 Comparison between Absolute Postmile of Hotspots on the I-880 Southbound Freeway

Also, on the I-880 Southbound freeway, the existing CRP method using SPFs as a rescaling factor identified the smallest number of hotspots on the total freeway. Also, hotspots identified by the CRP method using 3-hierarchical clustering analysis included all of the hotspots identified by the existing CRP method. As on the I-880 Northbound freeway, the CRP method using hierarchical clustering analysis did not generate any false negatives.

The Sliding Moving Window method and the Peak Searching method identified longer lengths of hotspots than the existing CRP method and the CRP method using hierarchical clustering analysis. Therefore, the latter had higher false positive rates than the existing CRP method, but it had lower false positive rates than the Sliding Moving Window method and the Peak Searching method.

Hotspots identified by the Sliding Moving Window method on the I-880 bi-directional freeway included hotspots identified by the existing CRP method, the Peak Searching method failed to identify several hotspots that were identified by the existing CRP method. This occurred because the Peak Searching method is affected by the initial value for the coefficient of variation(CV) (0.3 in this study) used in precision testing. So, adjusting the initial value for the coefficient of variation can reduce the number of errors that are generated. For example, the Peak Searching method identified all of the hotspots identified by the existing CRP when the initial value for the coefficient of variation was set at 0.5, but it identified a very long length of hotspots.

The table below indicates the lengths of the hotspots identified by each network screening method.

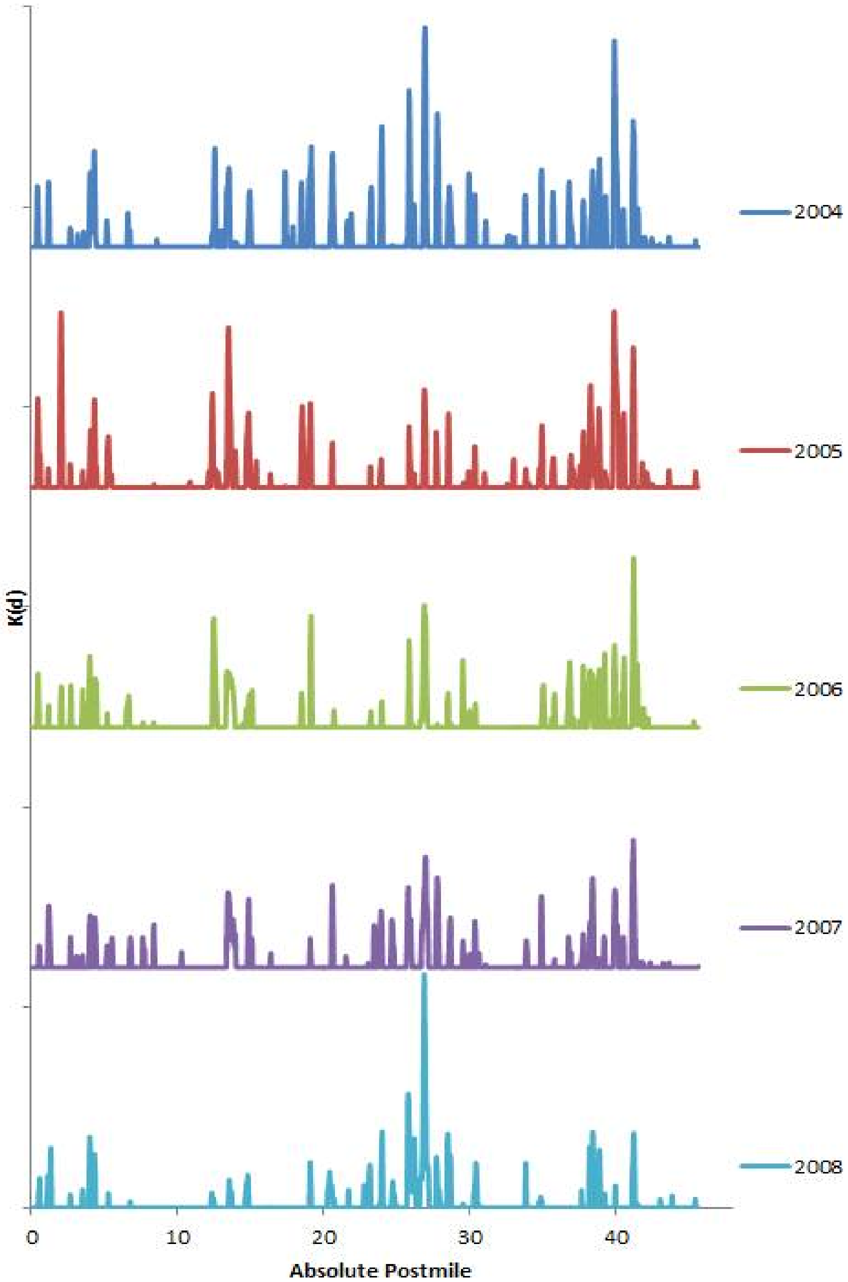
〈Table 4-2〉 The Lengths of Hotspots on the I-880 Southbound Freeway

Division	The Existing CRP	The CRP using Clustering			The Sliding Moving Window	The Peak Searching
a Rescaling Factor	SPFs	Combination	AADT	# of lanes	-	-
the Lengths of hotspots (mile)	2.15	7.23	7.41	8.03	26.37	23.12
Rate (%)	4.70	15.81	16.21	17.56	57.68	50.59

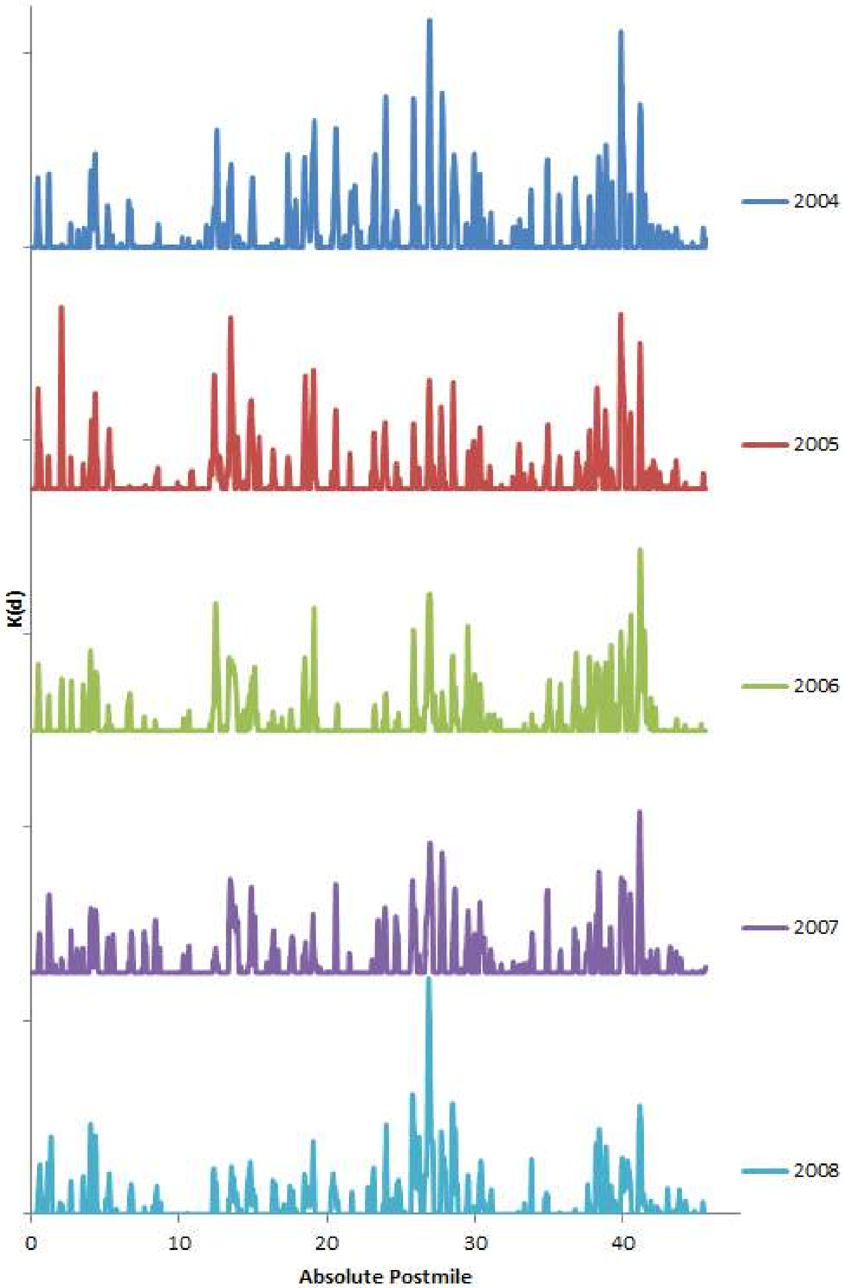
The total length of the I-880 Southbound freeway is almost 46 miles. The methods, ranked in ascending order of their results for hotspot length, were the existing CRP method, the CRP method using clustering (variable: the number of lanes), the CRP method using clustering (variable: AADT), the CRP method using clustering (variable: the combination of the two), the Peak Searching method, the Sliding Moving Window method. Subsequently, this order indicates descending order of false positive rate.

According to this analysis, the CRP method using hierarchical clustering analysis on the I-880 bi-directional freeway had worse performance than the existing CRP method, but it had better performance than the other screening methods, i.e., the Sliding Moving Window and the Peak Searching method.

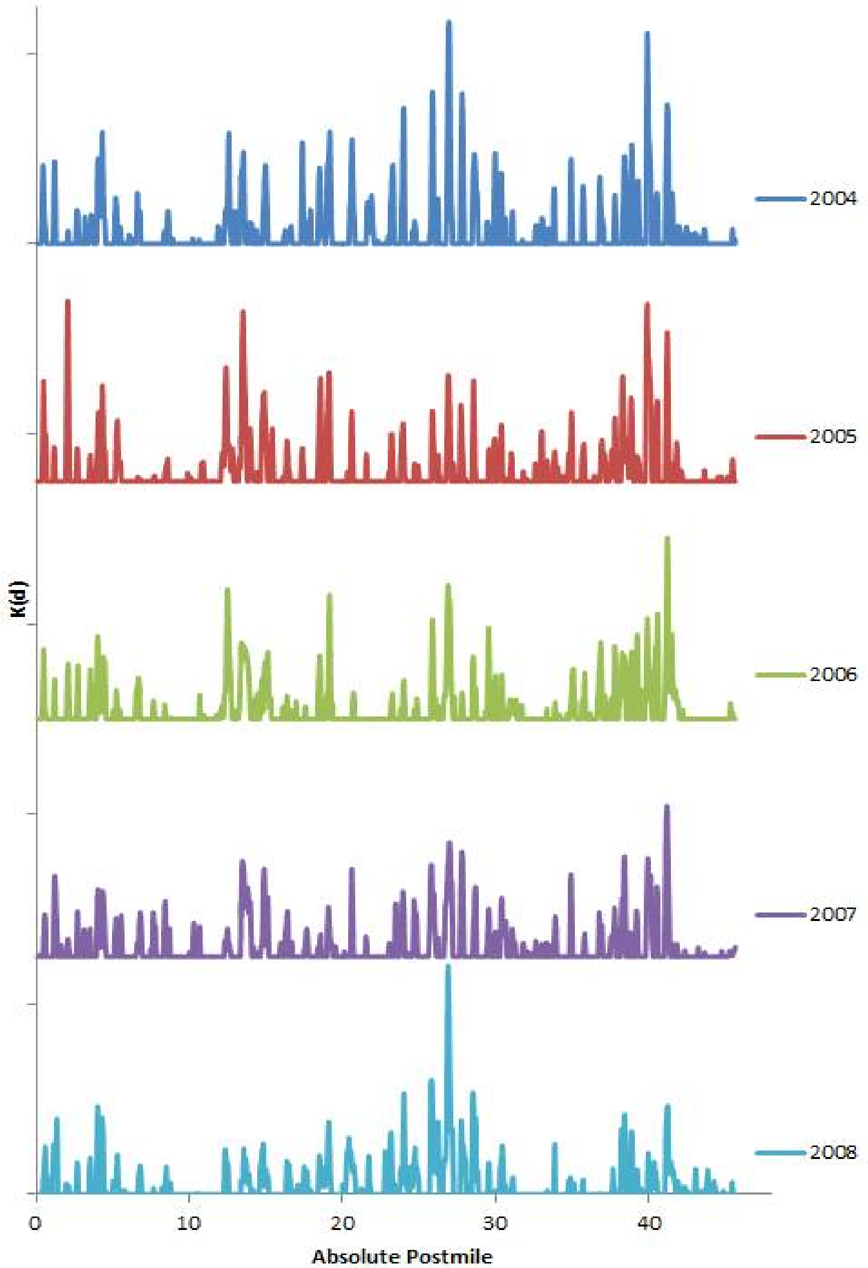
4.2.2. Hotspots Identified by 4-Continuous Risk Profile



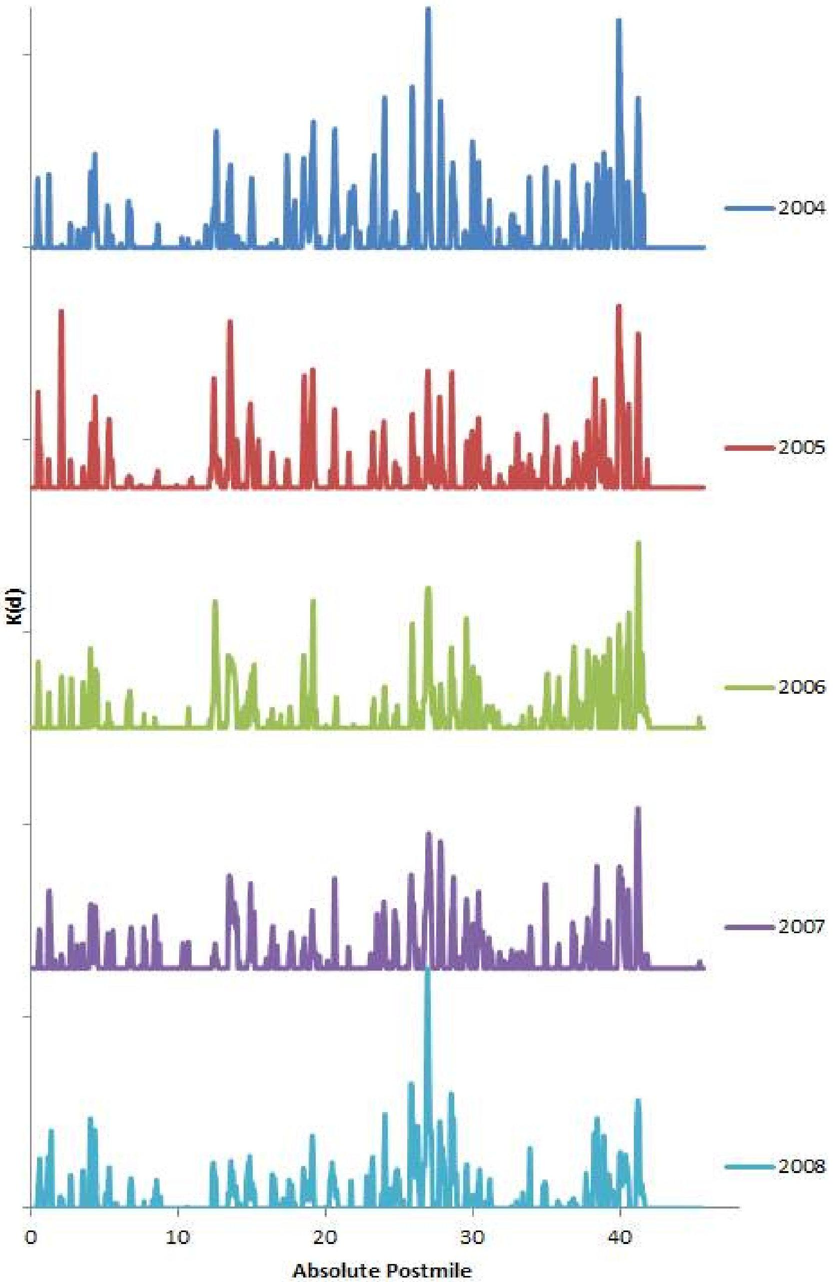
〈Figure 4-7〉 The Existing CRP on the I-880 Southbound Freeway



〈Figure 4-8〉 The CRP using Clustering (AADT + # of lanes) on the I-880 Southbound Freeway



〈Figure 4-9〉 The CRP using Clustering (AADT) on the I-880 Southbound Freeway



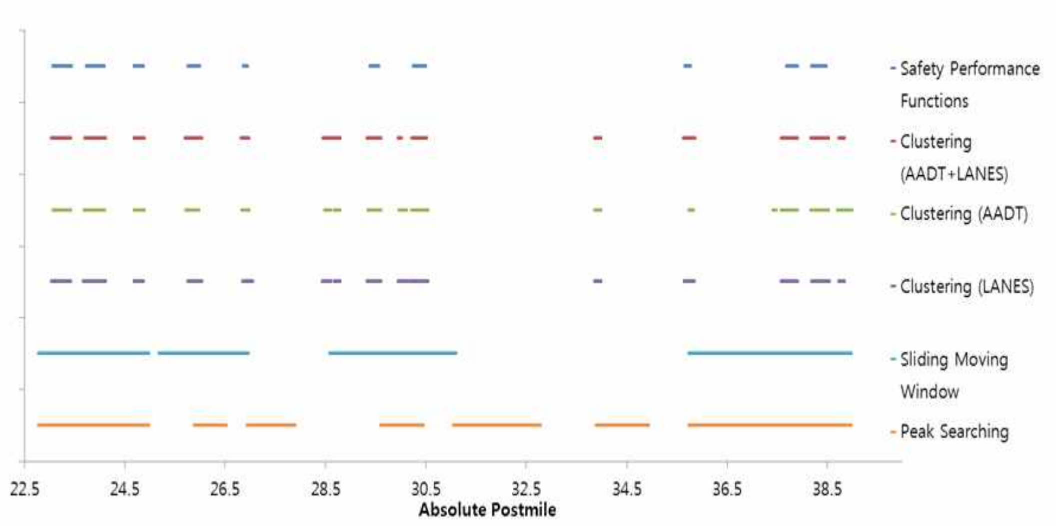
〈Figure 4-10〉 The CRP using Clustering (the number of lanes) on the I-880 Southbound Freeway

4.3. Reanalysis of Bi-directional Collision Concentration Locations

The results obtained from the effort to identify hotspots on the I−880 bi−directional freeway indicated that collision concentration locations existed at Northbound absolute postmile 22.77 ~ 39.98 miles and Southbound absolute postmile 10.54 ~ 30.91 miles were collision concentration locations.

These collision concentration locations were segregated from the whole freeway, which did not affect the segmentation of the Sliding Moving Window method and the Peak Searching method. Then, reanalysis of these collision concentration locations was conducted. According to this reanalysis, it was confirmed that conclusion derived from the whole freeway was meaningful. There are figures arranging hotspots in order of absolute postmile and tables representing the lengths of hotspots identified by each network screening method.

4.3.1. I-880 Northbound Absolute Postmile (22.77 ~ 39.98 miles)

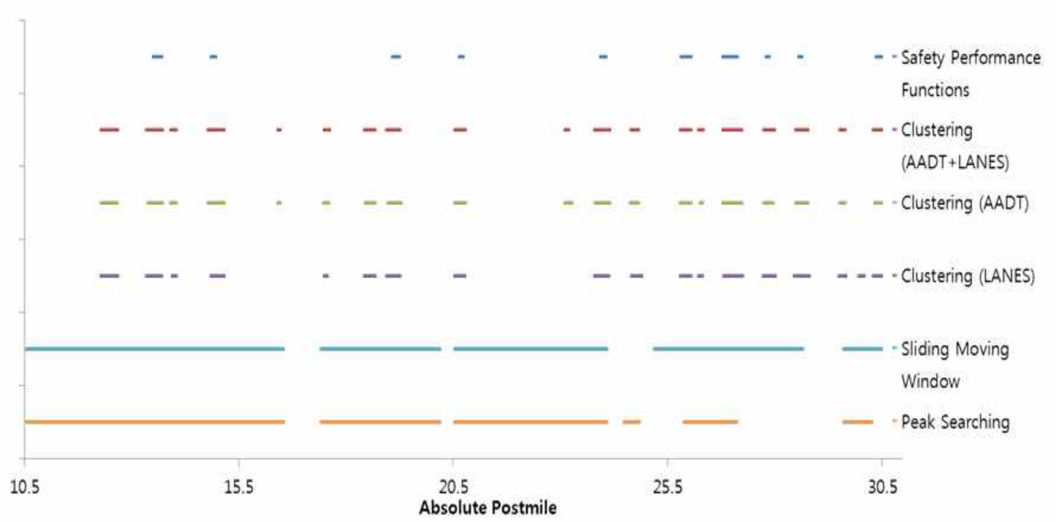


〈Figure 4-11〉 Comparison between Absolute Postmile of Hotspots on the I-880 Northbound (22.77 ~ 39.98 miles)

〈Table 4-3〉 Comparing the Lengths of Hotspots on the I-880 Northbound (22.77 ~ 39.98 miles)

Division	The Existing CRP	The CRP using Clustering			The Sliding Moving Window	The Peak Searching
a Rescaling Factor	SPFs	Combination	AADT	# of lanes	-	-
the Lengths of hotspots (mile)	1.9	3	2.94	3.07	9.08	10.72
Rate (%)	11.03	17.42	17.07	17.83	52.79	62.25

4.3.2. I-880 Southbound Absolute Postmile (10.54 ~ 30.91 miles)



〈Figure 4-12〉 Comparison between Absolute Postmile of Hotspots on the I-880 Southbound (10.54 ~ 30.91 miles)

〈Table 4-4〉 Comparing the Lengths of Hotspots on the I-880 Southbound (10.54 ~ 30.91 miles)

Division	The Existing CRP	The CRP using Clustering			The Sliding Moving Window	The Peak Searching
a Rescaling Factor	SPFs	Combination	AADT	# of lanes	-	-
the Lengths of hotspots (mile)	1.11	3.57	3.44	3.6	15.51	14.16
Rate (%)	5.45	17.53	16.90	17.68	76.23	69.55

4.3.3. Results of Reanalysis

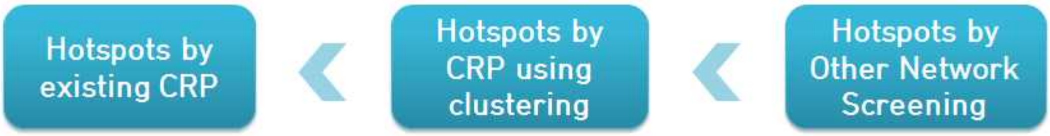
The various methods, ranked in ascending order of their indicated lengths of hotspots on the I-880 Northbound freeway (22.77 ~ 39.98 miles), were the existing CRP method, the CRP method using clustering (variable: AADT), the CRP method using clustering (variable: combination of AADT and the number of lanes), the CRP method using clustering (variable: the number of lanes), the Sliding Moving Window method, and the Peak Searching method. And the results of hotspots on the I-880 Southbound freeway (10.54 ~ 30.91 miles), ranked in ascending order of length were the existing CRP method, the CRP method using clustering (variable: AADT), the CRP method using clustering (variable: combination of AADT and the number of lanes), the CRP method using clustering (variable: the number of lanes), the Peak Searching method, and the Sliding Moving Window method. Subsequently, this order indicates descending order of false positive rate.

In the two cases above, there was a difference in the sequence of the Peak Searching method and the Sliding Moving Window method, but the results of the reanalysis were verified as follows.

The CRP method using hierarchical clustering analysis had worse performance than the existing CRP method, but it had better performance than the Sliding Moving Window method and the Peak Searching method.

Chapter 5. Conclusions and Further Advancement of the Study

5.1. Conclusions and Contribution of the Study



〈Figure 5-1〉 Comparison with Length of Hotspots by Each Network Screening Method

In previous studies, the existing CRP method had lower false positive rates than the Sliding Moving Window method and the Peak Searching method, which indicates that the existing CRP method is better than other network screening methods in terms of its performance in the identification of hotspots.

To identify the performance of the CRP method using hierarchical clustering analysis as a rescaling factor, in this study, I compared the performance of the CRP method using clustering with other network screening methods (the existing CRP method, the Sliding Moving Window method, and the Peak Searching method) based on the I-880 freeway in Alameda Country, California.

Hotspots identified by the CRP method using expected average crash frequency from hierarchical clustering analysis (variables: AADT, the number of lanes, and the combination of the two) included hotspots identified by the existing CRP method and had a lower false positive rate than the Sliding Moving Window method and the Peak Searching method. Accordingly, the CRP method using hierarchical clustering analysis is worse than the existing CRP method, but it is better than the Sliding Moving Window method and the Peak Searching method.

These results appeared in collision concentration locations on the I-880 bi-directional freeway as well as the entire I-880 freeway. Therefore, nations and regions without SPFs can utilize the CRP method using hierarchical clustering analysis, which was verified to have better performance than the Sliding Moving Window method and the Peak Searching method.

Applying the CRP method using hierarchical clustering analysis to identifying hotspots can reduce the false positive rate, so hotspots are identified effectively. Also, it can be used effectively with a limited budget to effectively identify areas that offer a huge potential for safety improvement, while reducing the time required by experts and the associated costs for investigating true hotspots of many freeways.

5.2. Further Advancement of the Study

Based on the results of this study, I concluded that the CRP method using hierarchical clustering analysis has better performance than the Sliding Moving Window method and the Peak Searching method. Recommended future advancements in the study are provided below:

(1) Difference of Hierarchical Clustering Analysis Depending on Variables

In this study, AADT, the number of lanes, and a combination of the two were used as variables in the hierarchical clustering analysis. Irrespective of the variable used, the CRP method using hierarchical clustering analysis had a more effective performance than other network screening methods. Future studies should use case studies to determine the effects of setting different values for the variables in hierarchical clustering analysis acts on the identification of hotspots.

(2) Setting of Default Value for the Coefficient of Variation in Peak Searching

Peak Searching contains the precision of the performance measures for verifying meaningful performance measures of segments. If the coefficient of variation for a given segment is greater than the initial value for the CV, the window gradually becomes wider than it was in the previous analysis. However, there is no specific initial value for the CV, so this affects the results of the Peak Searching method. Therefore, future studies should determine the extent to which the initial value for the CV influences the ability to identify hotspots.

References

California Department of Transportation (2002). *Table C task force: Summary Report of Task Force's Findings and Recommendations*. Caltrans.

Chung, K., and D. R. Ragland (2007). A Method for Generating a Continuous Risk Profile for Highway Collisions. *Transportation Research Board Annual Meeting 2007,* Paper #07-2935.

Chung, K., and D. R. Ragland (2009). The Continuous Risk Profile Approach for the Identification of High Collision Concentration Locations on Congested Highways. *Transportation and Traffic Theory 2009.*

Chung, K., K. Jang, S. Madanat, and S. Washington (2010). Proactive detection of high collision concentration locations on highways. *Transportation Research Part A: Policy and Practice,* Volume 45, Issue 9, November 2011.

Federal Highway Administration (2010). *Highway Safety Manual, 1st Edition.* American Association of State Highway and Transportation Officials.

Hauer, E. (1996). Identification of sites with promise. *Transportation Research Record,* Vol. 1542.

Hauer, E., J. Kononov, B. Allery, and M. S. Griffith (2002). Screening the Road Network for Sites with Promise. *Transportation Research Record: Journal of the Transportation Research Board,* Vol. 1784.

Hwang, T., K. Chung, D. R. Ragland, C. Chan (2008). Identification of High Collision Concentration Locations Under Wet Weather Conditions. *Safe Transportation Research & Education Center, Institute of Transportation Studies (UCB),* UC Berkeley.

Jang, K., K. Chung, D. R. Ragland, C. Chan (2008). Comparison of collisions on HOV facilities with limited and continuous access during peak hours. *Institute of Transportation Studies,* UC Berkeley.

Kononov, J. (2002). Identifying location with potential for accident reductions: use of direct diagnostics and pattern recognition methodologies. *Transportation Research Record,* Vol. 1784.

Kwon, O., M. Park, K. Chung, and H. Yeo (2012). Comparing the Performance of Sliding Moving Window, Peak Searching, and Continuous Risk Profile Methods for Identifying High Collision Concentration Locations. *Transportation Research Board 91st Annual Meeting,* Paper #12-2293.

Ragland, David R., and Ching-Yao Chan (2007). *High Collision Concentration Location : Table C Evaluation and Recommendations. California PATH Program,* Institute of Transportation Studies, University of California at Berkeley.

Oh, S., K. Chung, D. R. Ragland, and C. Chan (2009). Analysis of Wet Weather Related Collision Concentration Locations: Empirical Assessment of Continuous Risk Profile. *Recent Work, Safe Transportation Research & Education Center,* Institute of Transportation Studies (UCB), UC Berkeley.

Sahaji, R. K., P. Y. Park, G. Eguakun, and A. Gardiner (2012). Identifying High Collision Locations without Traffic Volume Data. *Transportation Research Board 91st Annual Meeting,* Paper #12-1424.

Qin, X., and A. Wellner (2012). Segment Length Impact on Highway Safety Screening Analysis. *Transportation Research Board 91st Annual Meeting,* Paper #12-0644.

김영주, 김희갑 (2007). **환경통계학**. 동화기술.

김용일 (2011). 사고 심각도를 반영한 고속도로 사고취약구간 판별 기법 연구 : Continuous Risk Profile 기법을 중심으로. 석사학위논문, 서울대학교.

유준석 (2008). CRP(Continuous Risk Profile) 기법을 활용한 고속도로 사고 잦은 지점 분석방법 연구. 석사학위논문, 한양대학교.

국문초록

고속도로에서 발생하는 사고는 상대적으로 그 규모가 크며 인명피해 가능성이 높으므로 고속도로의 사고취약구간을 선정하여 관리할 필요가 있으며, 이를 통해 효과적인 사고취약구간을 선정함으로써 한정된 예산 투자의 효율성을 증대시킬 수 있다.

현존하는 사고취약구간 선정 방법론 중, 2007년 개발된 Continuous Risk Profile은 다른 방법론에 비해 그 성능이 뛰어난 것으로 알려져 있다. 하지만 Continuous Risk Profile을 이용하여 사고취약구간을 선정하기 위해서는 규모조정계수로 이용되는 안전성능함수가 존재해야 한다.

본 연구에서는 사고취약구간을 선정하는 효과가 뛰어난 Continuous Risk Profile을 안전성능함수가 구축되어 있지 않은 국가나 지역에서도 사용할 수 있도록 계층적 군집분석을 이용하고자 한다.

기존의 Continuous Risk Profile에서 규모조정계수로 사용하는 안전성능함수를 계층적 군집분석을 통해 도출한 Group별 예측사고건수로 대체하여 사고취약구간을 선정한다. 기존의 Continuous Risk Profile과 계층적 군집분석을 이용한 Continuous Risk Profile을 통해 선정된 사고취약구간의 비교, 계층적 군집분석을 이용한 Continuous Risk Profile과 다른 사고취약구간 선정 방법론을 통해 선정된 사고취약구간의 비교를 통해 계층적 군집분석을 이용한 Continuous Risk Profile도 다른 사고취약구간 선정 방법론보다 성능이 좋음을 입증할 수 있다.

주요어: Continuous Risk Profile, 계층적 군집분석, 사고취약구간, 규모조정계수, 안전성능함수

학 번: 2011-20993