SPATIAL ANALYSIS OF CROSS MEDIAN CRASHES

The use of traditional statistics (like regression) has been the mainstay of transportation research especially in traffic safety. Research studies often develop statistical models under the assumption that data occurring in a space are independent of each other. However, the presence of spatial dependencies or spatial autocorrelation, where values at one location are influenced by the presence of other values in its geographic proximity, often violates the assumption of independence that is implicit in many statistical analyses. The failure to account for spatial autocorrelation can lead to serious errors in statistical analysis for geographic data (4, 5). Crashes are a spatial phenomenon and the spatial independence assumption is often violated; therefore, it is important to analyze the spatial heterogeneity/homogeneity of such data spread in space. The measurements of spatial dependencies through spatial pattern analysis can help analyze crash data to identify hotspots and contributory factors.

A Spatial Analytical Framework consisting of new and modified spatial statistical methods in network space was defined to extend the analysis of crash data into the realm of spatial data analysis. The Spatial Analytical Framework consists of two parts; namely, the Theoretical and Computational Frameworks. The Theoretical Framework consists of new and modified methods based on different variants of K-Function (distance-based statistic) adapted to network space to resolve issues identified in the literature pertaining to network vs. planar space, the uniform and non-uniform network problem, anisotropy in transportation data analysis, and the need for variable distance-based statistic. The Computational Framework consists of specific programs and tools developed to facilitate the practical implementation of Spatial Analytical Framework while addressing issues relating to the analysis of multiple point patterns and computation of local network statistic.

The essential principle behind K-functions is to compare the observed number of events (crashes in our case) within a certain distance of an event (can be a crash or bridge or ramp or curve) to an expected number of events if the underlying process is completely spatially random (CSR). If the observed number of events is greater than the upper bound for CSR, the events are said to be clustered. Conversely if the observed number of events is smaller than the lower bound for CSR, the events are said to be dispersed. In this research Cross median crashes (CMC) on interstate highways from 2001-07 were analyzed to identify hotspots and factors affecting the crashes in spatial context identifying not only the magnitude but also the extent of spatial relationships.

SPATIAL CLUSTER ANALYSIS

The first and foremost fundamental question that traffic safety engineers face is the question of where crashes occur. The global network K-Function and incremental K-Function are two methods for the analysis of hotspots of crashes which provide a holistic view of the spatial patterns in point data in the entire study area at various scales.

Global Network K-Function and Incremental K-Function Results

The results of the observed Global Network and Incremental K-Functions were plotted on a graph showing clustering (hotspots) or dispersion (coldspots) of crash points at various distance scales. The expected values under Complete Spatial Randomness (CSR) were also plotted on the same graph to determine the statistical significance of the observed values using the upper and lower bounds for two confidence levels (2.5 and 5 percent). The upper and lower bounds based on Monte Carlo simulations were calculated for two confidence levels to analyze the sensitivity of the results at different confidence levels. If the observed values are above the upper bound or below the lower bound, the pattern is said to be statistically significant clustered or statistically significant dispersed respectively. If the observed values are within the upper and lower bounds, there is no statistically significant spatial pattern and the points are randomly distributed. The total extent of the distance on which the analysis was conducted was plotted on the x-axis, and the values of the observed and expected K-Functions were plotted on the y-axis.

The Global Network and Incremental K-Function analysis for CMC on Interstate Highways in Wisconsin between 2001 and 2007 was conducted up to a total distance of 20 km (in each direction) and in distance increments of 1 km. The selection of the 20 km and 1 km values was based on a number of factors keeping in view the computational intensity of using smaller distance increments and the need for capturing the spatial patterns in crash data.

Figure 1 and Figure 2 present the results of the Global Network and Incremental K-Function for CMC, respectively. The results in Figure 1 and Figure 2 show that on average there is clustering of CMC. Global network K-Function analysis results in Figure 1 displays clustering throughout the range of analysis (20 km) masking variations within distance increments. Figure 2 displays the spatial patterns within each distance increment showing the magnitude of clustering, which is highest at 1 km distance and becomes random at 5 km distance. After 5 km, the patterns fluctuate between clustering and random. The spatial pattern in Figure 2 strongly indicates the presence of local clusters especially at smaller distance scales (1 km to 5 km). However, Global Network and Incremental K-Functions can only provide global overview of the spatial patterns in the entire region and cannot identify variations at local scale or the locations of hotspots. The limitations of the global methods were overcome by developing local methods, namely the local network K-Function and LSPP Graphs.
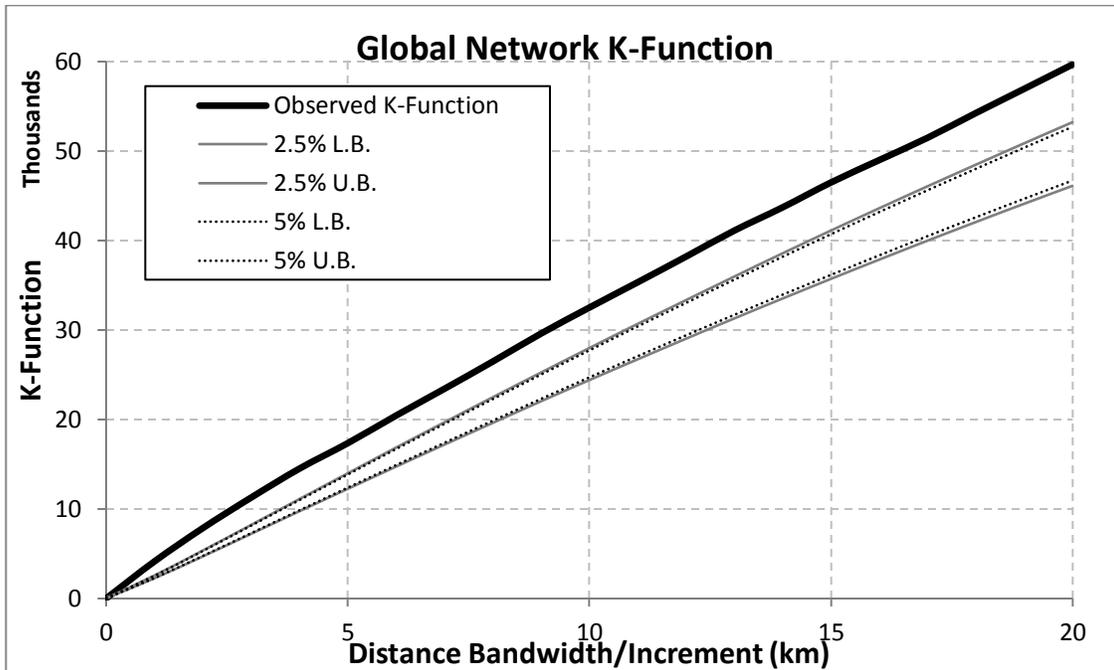
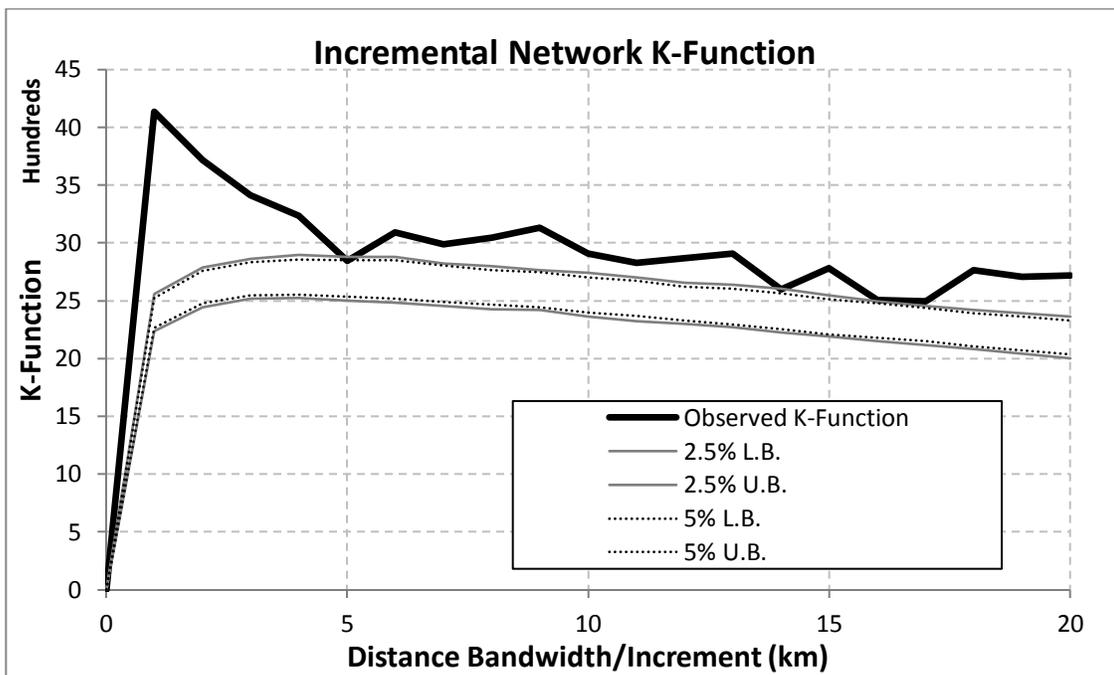**FIGURE 1 GLOBAL NETWORK K-FUNCTION ANALYSIS OF CMC**

FIGURE 2 INCREMENTAL K-FUNCTION ANALYSIS OF CMC

Local Network K-Function

The local network K-Function method describes the location of spatial patterns on a local level to identify detailed patterns and variations masked by global methods. The local network K-Function was calculated for every individual CMC location because they were the smallest possible scale at which local statistic could be calculated; hence, providing the greatest detail and no loss of information due to aggregation. Using individual crash locations meant that 699 local network K-Functions have to be calculated which presented a significant computational challenge. The tools developed as part of the Spatial Analytical Framework provided an efficient implementation solution for the calculation of local network K-Function.

The results of the local network K-Function are presented in Figure 3. The local network K-Function was calculated for distance bandwidths ranging from 1 to 20 km in increments of 1 km, at 2.5 percent confidence level. The results show locations of local clusters of CMC at various scales. The results also indicate that as the distance increment increases, the small scale local clusters disappear; whereas, some clusters increase in size due to aggregation of data at larger distance increments. The varying nature of CMC clusters at various distance bandwidths indicate the significance of analyzing data at various scales because it is impossible to select a single result from Figure 3 as true representation of clusters of CMC.

The actual extent of the CMC hotspots in terms of a segment length of the road network cannot be determined from the results in Figure 3, which shows statistically significant clustered individual CMC points. Therefore, the results from local network K-Function were further processed to determine hotspot segments of CMC. The hotspot segment was defined as the length of the road segment, upstream and downstream of the cluster point, equivalent to the distance bandwidth at which the cluster point was identified as a statistically significant cluster. In cases where there were hotspot segments for two cluster points overlapped for a specific distance bandwidth, the hotspot segments were merged together to create one hotspot segment. The results of the hotspot segments for the CMC data are presented in Figure 4 for distance bandwidths ranging from 1 km to 9 km. The calculations for distance bandwidths ranging from 10 km to 20 km were not included because the hotspots merged together at larger scales and there was no discernable change detected. The hotspot segments results as shown in Figure 4 are easier to interpret as the mile markers for start and end points of the hotspot segments can be easily identified to provide a practical solution to identifying the hotspots in the field.
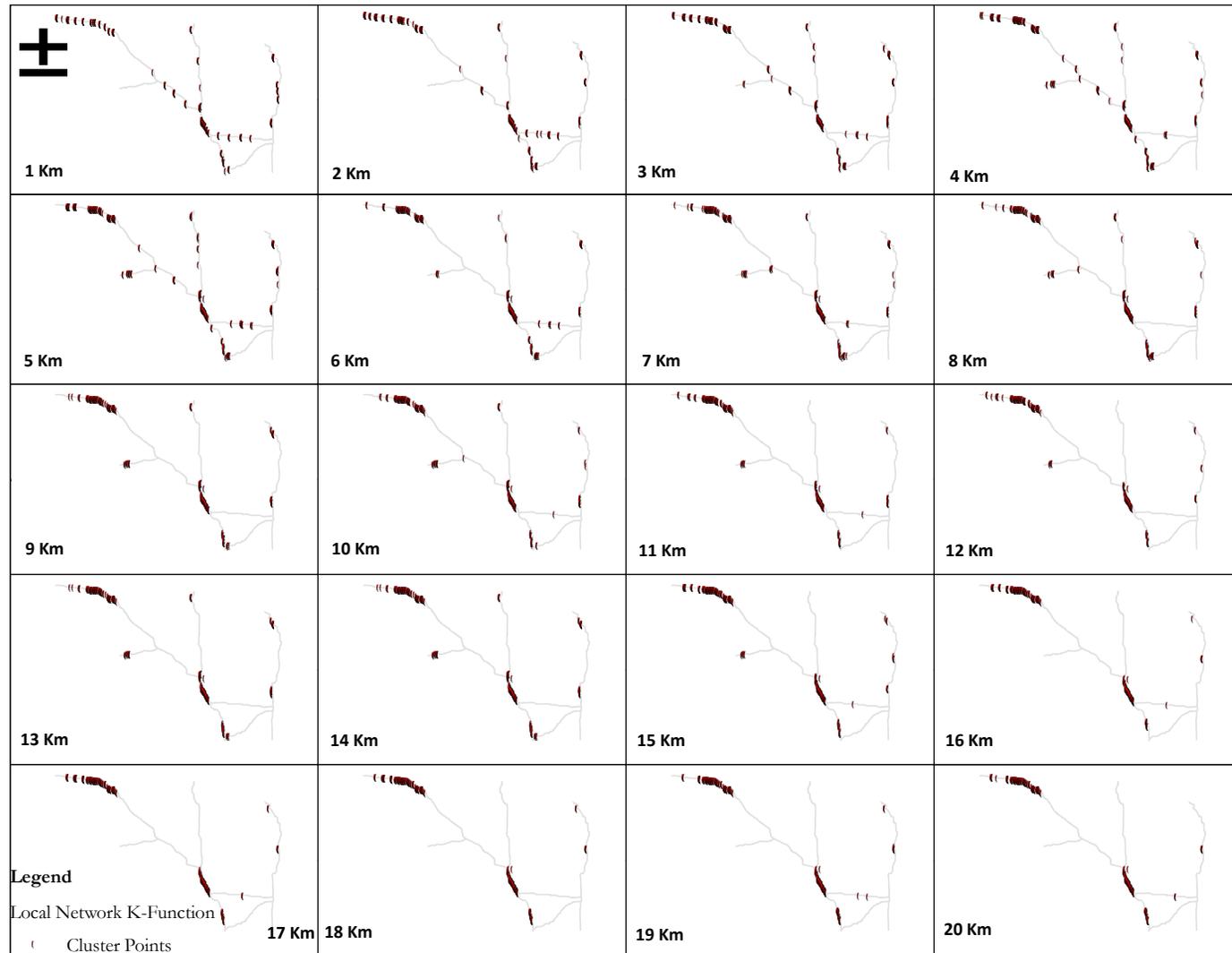
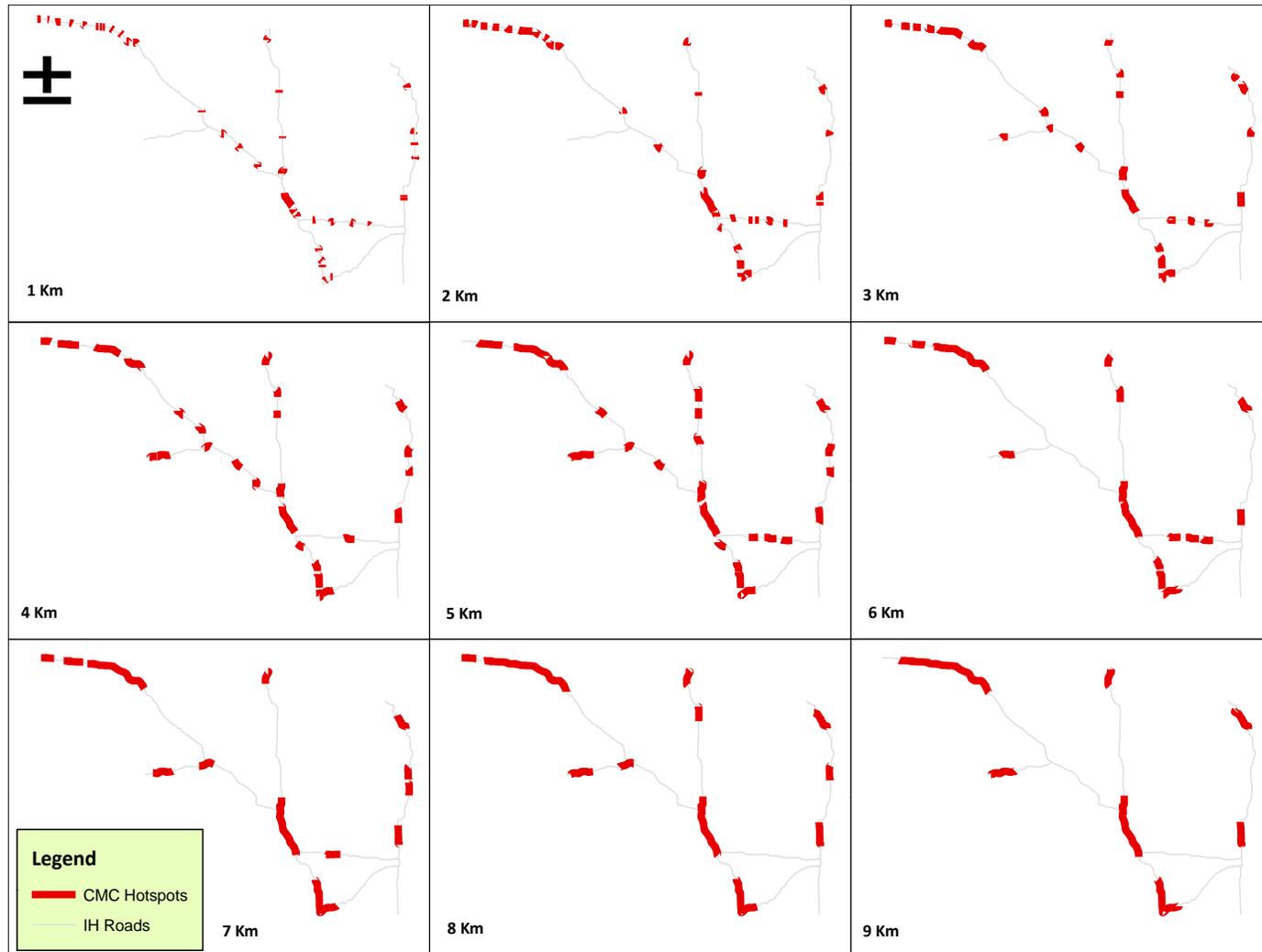**FIGURE 3 LOCAL NETWORK K-FUNCTION CLUSTER POINTS AT VARIOUS DISTANCE BANDWIDTHS FOR CMC**

FIGURE 4 LOCAL NETWORK K-FUNCTION HOTSPOT SEGMENTS OF CMC

Threshold crash rates were calculated for identifying CMC hotspot segments using the upper bound of expected number of CMC under Complete Spatial Randomness (CSR) as a function of distance and AADT. The crash rates were computed using equation 1.

$$R_{st} = \frac{\left(\dfrac{C}{2t}\right)}{Y} \tag{1}$$

Where:

$C = 2.5\%$ Upper Bound$\left[LK_{is}(t)\right]$;

$LK_{is}(t)$ = local network K-Function expected value for segment class "s" and distance bandwidth t;

Y = number of years of CMC data used in the analysis; and

R_{st} = Threshold Crash Rate for CMC Hotspot Segment for segment class "s" and distance bandwidth t.

The results of equation 1 were calculated for various segment classes based on AADT categories as defined in Table 1. Furthermore, the results of Table 1 are presented in Figure 5 as a nomograph for hotspot identification based on AADT segment classes and distance bandwidths. The results of the nomograph can be used as a threshold to identify if a specific section of the Interstate Highway with a given CMC rate is a hotspot or not. The nomograph provides threshold values for segments with different AADT. The nomograph also provides different values for distance bandwidths, the selection of which will depend upon the nature of the road segment under analysis. A lower bandwidth value would be selected on segments where the nature of interaction between CMC with respect to segment characteristics is considered to be on a small scale. Similarly, a larger bandwidth value would be selected for long and uniform segment lengths. Interestingly, the results in Figure 5 show that the nomograph lines are indistinguishable beyond 5 km distance bandwidth.

TABLE 1 THRESHOLD CRASH RATES FOR CMC HOTSPOT SEGMENTS

| AADT Class	1 km	2 km	3 km	4 km	5 km	6 km	7 km	8 km	9 km	10 km	11 km	12 km	13 km	14 km	15 km	16 km	17 km	18 km	19 km	20 km
10000 - 20000	.21	.14	.12	.11	.09	.10	.09	.09	.10	.09	.10	.10	.10	.10	.10	.10	.11	.11	.11	.11
20000 - 30000	.29	.18	.14	.13	.11	.12	.11	.11	.10	.10	.10	.10	.10	.10	.10	.10	.10	.10	.09	.09
30000 - 40000	.29	.21	.19	.18	.16	.15	.14	.14	.13	.14	.14	.14	.13	.13	.13	.13	.13	.13	.14	.14
40000 - 50000	.36	.25	.21	.20	.19	.18	.17	.17	.17	.16	.16	.17	.16	.16	.17	.17	.16	.16	.16	.16
50000 - 60000	.43	.29	.26	.23	.21	.21	.20	.20	.20	.20	.19	.20	.19	.19	.19	.19	.19	.19	.19	.19
60000 - 70000	.43	.32	.29	.27	.26	.24	.23	.23	.23	.23	.23	.22	.23	.22	.22	.22	.22	.22	.21	.21
70000+	.50	.36	.31	.29	.27	.26	.26	.24	.24	.24	.23	.23	.22	.22	.21	.21	.21	.21	.21	.21
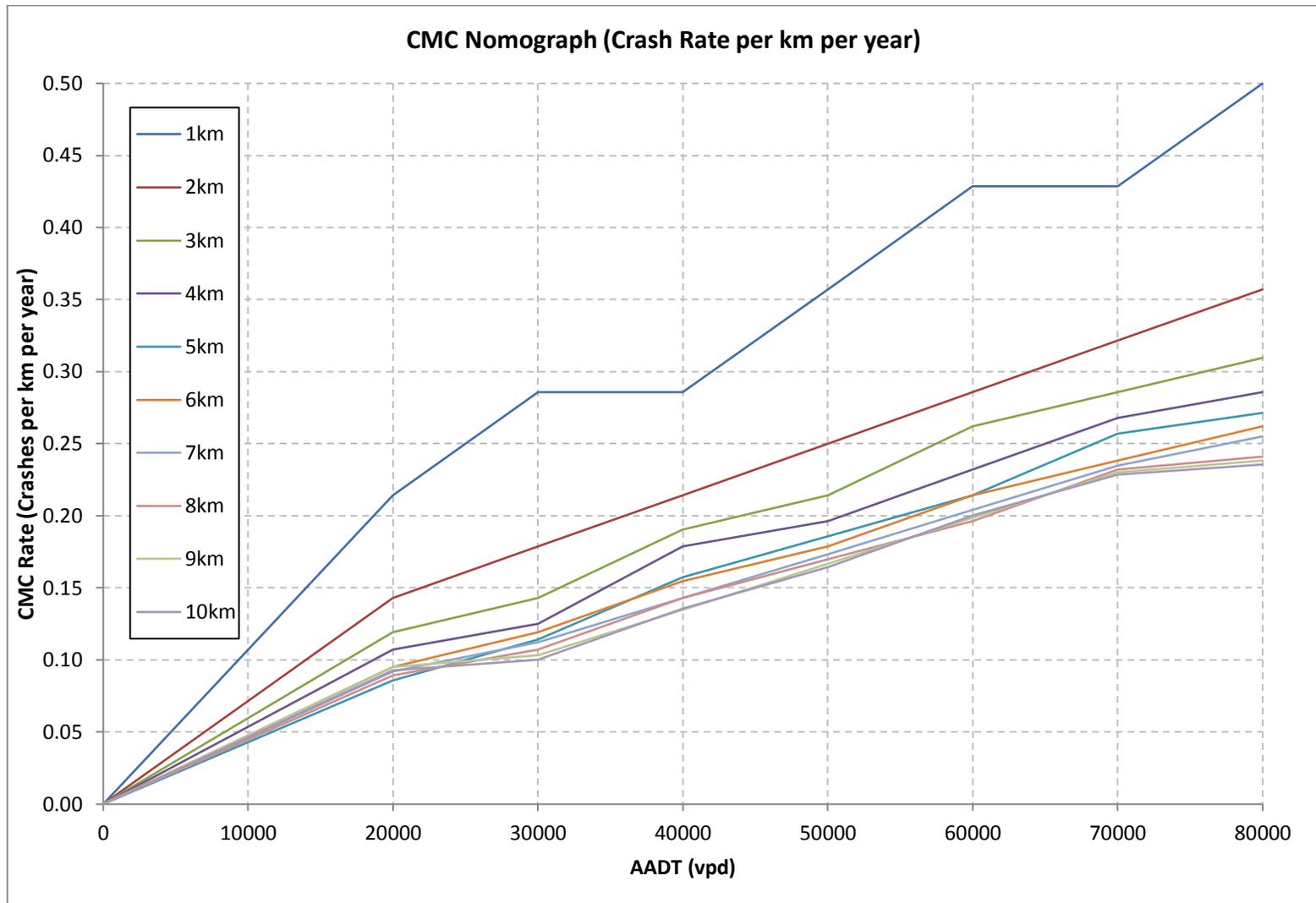
FIGURE 5 CMC HOTSPOT NOMOGRAPH BY SEGMENT CLASS AND DISTANCE BANDWIDTH

CONTRIBUTORY FACTORS ANALYSIS

The second objective of this research was to identify factors affecting CMC using the newly described global network, incremental, variable distance, and incremental variable distance cross K-Functions to identify the presence, magnitude, and extent of the relationship between crashes and various geometric features of interest. The cross K-functions are used to compare two different point patterns (crashes and geometric features). In this research the following geometric features were analyzed to determine their effects on the occurrence of CMC on Wisconsin Interstate Highways.

1. Bridges
2. Entrance Ramps
3. Exit Ramps
4. Left Curves
5. Right Curves

The results of the analysis of factors affecting CMC on Wisconsin Interstate Highways are presented in the following sections. The analyses were conducted using the anisotropic K-Function variant in order to model the effects of directionality in the relationship between CMC and aforementioned geometric features.

Analysis of CMC and Bridges

The premise behind the analysis of CMC downstream of bridges was to determine the effects of precipitation on bridge decks and other factors on the occurrence of CMC. Figure 6 and Figure 7 show the results of the global network and incremental cross K-Functions for CMC and bridges based on 2.5 and 5 percent confidence levels. The x-axis shows the increasing distance downstream from end point of the bridge up to 1600 m or 1.6 Km (one mile) and the y-axis shows the K-Function calculations. Although both Figure 6 and Figure 7 indicate that CMC cluster downstream of bridges, the results of incremental cross K-Function present a clearer picture and more detail of the variations within distance increments. Figure 7 shows that CMC cluster downstream of bridges up to a distance of 100 meters. The results provide conclusive evidence that bridge locations affect the occurrence of CMC. More importantly, the results suggest that the extent of effects that bridges have on CMC extend up to a distance of 100 meters on average.

Figure 7 also shows clustering at distance of 800 and 1100 meters which may be because many bridges are located at interchanges; hence they are followed by entrance ramps downstream of the bridge where CMC could be clustering. Figure 7 clearly demonstrates the advantages of distance based statistics by showing the spatial scales at which clusters of CMC happen.
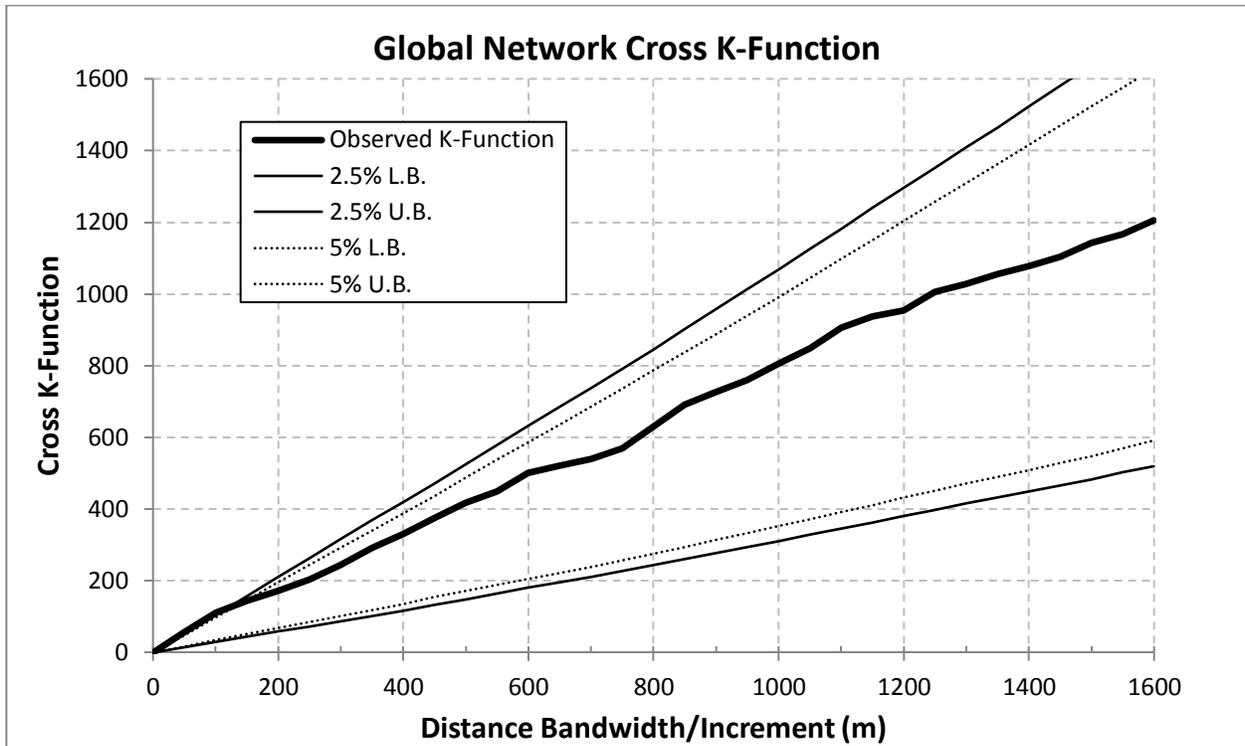
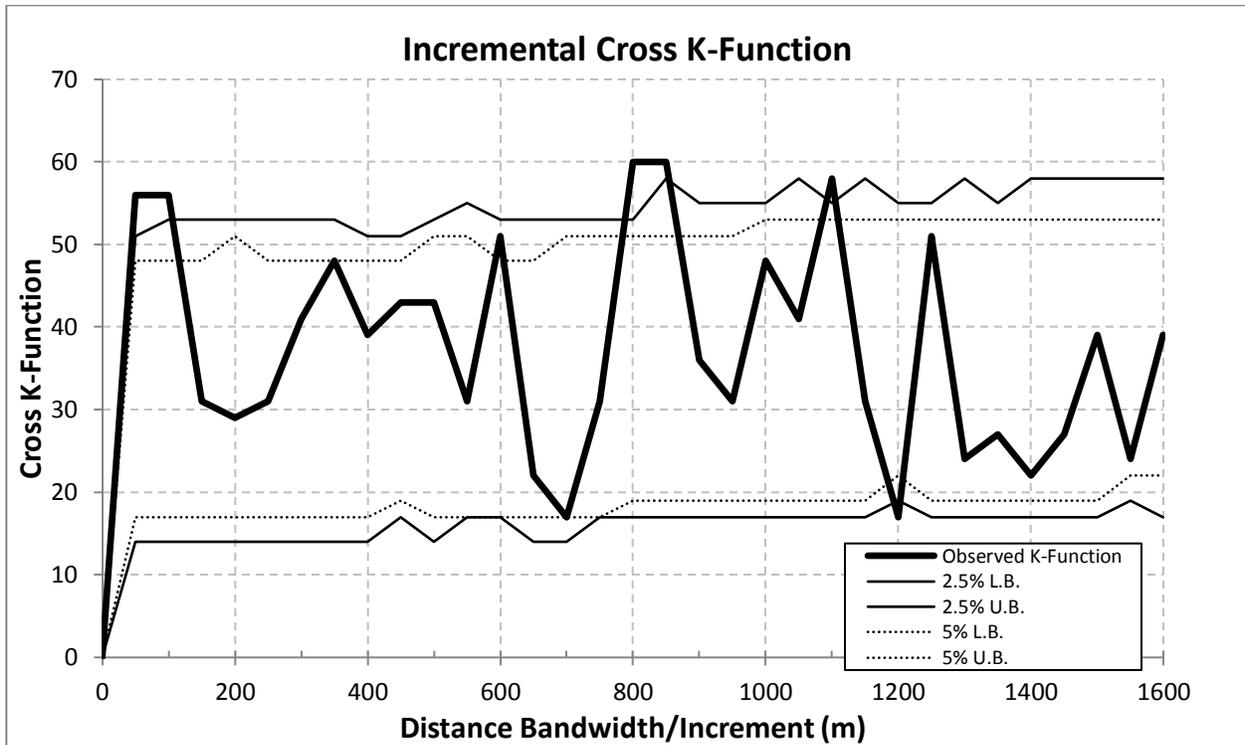**FIGURE 6 GLOBAL NETWORK CROSS K-FUNCTION FOR BRIDGES AND CMC**

FIGURE 7 INCREMENTAL CROSS K-FUNCTION FOR BRIDGES AND CMC

Analysis of CMC and Entrance Ramps

The analysis of CMC in relation to the entrance ramp locations was based on the premise that merging vehicles and speed variations could increase the probability of CMC, e.g., in cases of severe sideswipe crashes forcing a vehicle to go into the opposing lanes. Figure 8 and Figure 9 show the results of the global network and incremental cross K-Functions based on 2.5 and 5 percent confidence levels. The x-axis shows the increasing distance downstream from start point of the entrance ramp, which was defined as gore location represented by the pavement marking, up to a distance of 1.6 Km (1 mile). Although both Figure 8 and Figure 9 indicate that CMC cluster near entrance ramps, the results of incremental cross K-Function present greater details especially within distance increments. Figure 9 shows that CMC cluster around the 100, 300, and 700 meters mark downstream of the start of the entrance ramps, which means that entrance ramp locations affect CMC. The results also suggest that the influence area of entrance ramps with respect to CMC extends up to a distance of approx. 700 meters on average.
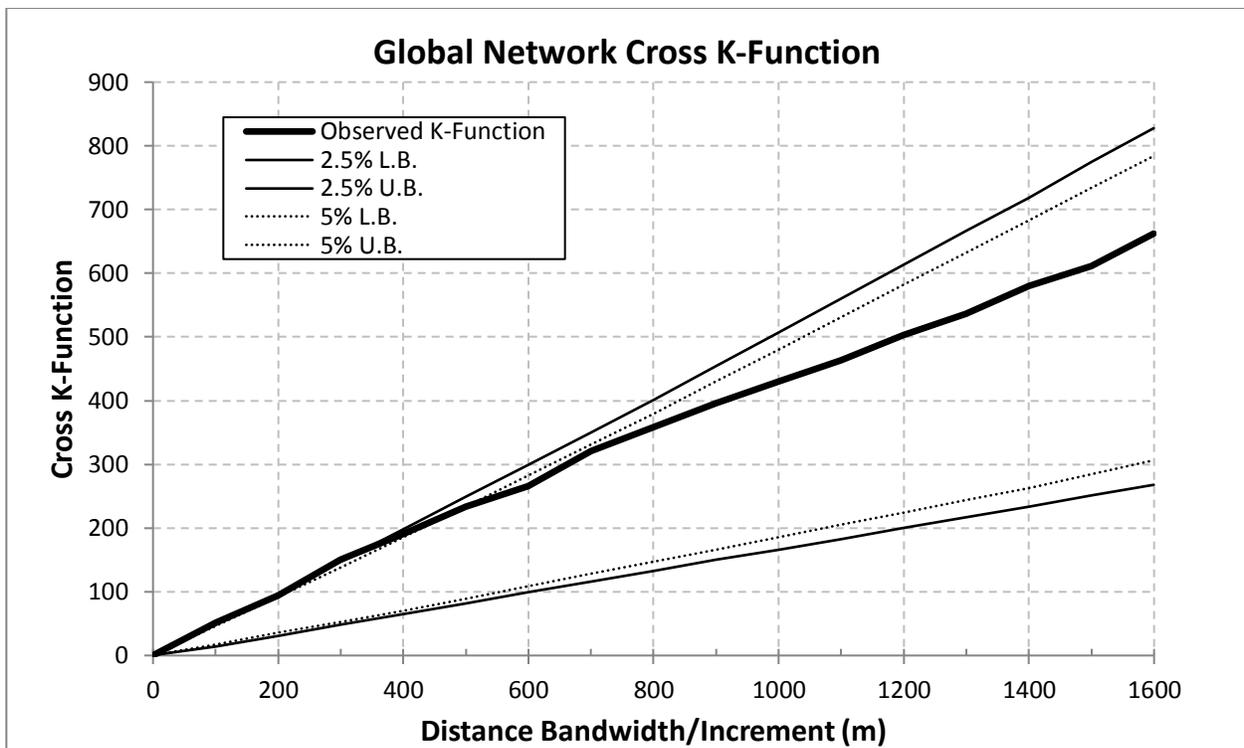

FIGURE 8 GLOBAL NETWORK CROSS K-FUNCTION FOR ENTRANCE RAMPS AND CMC
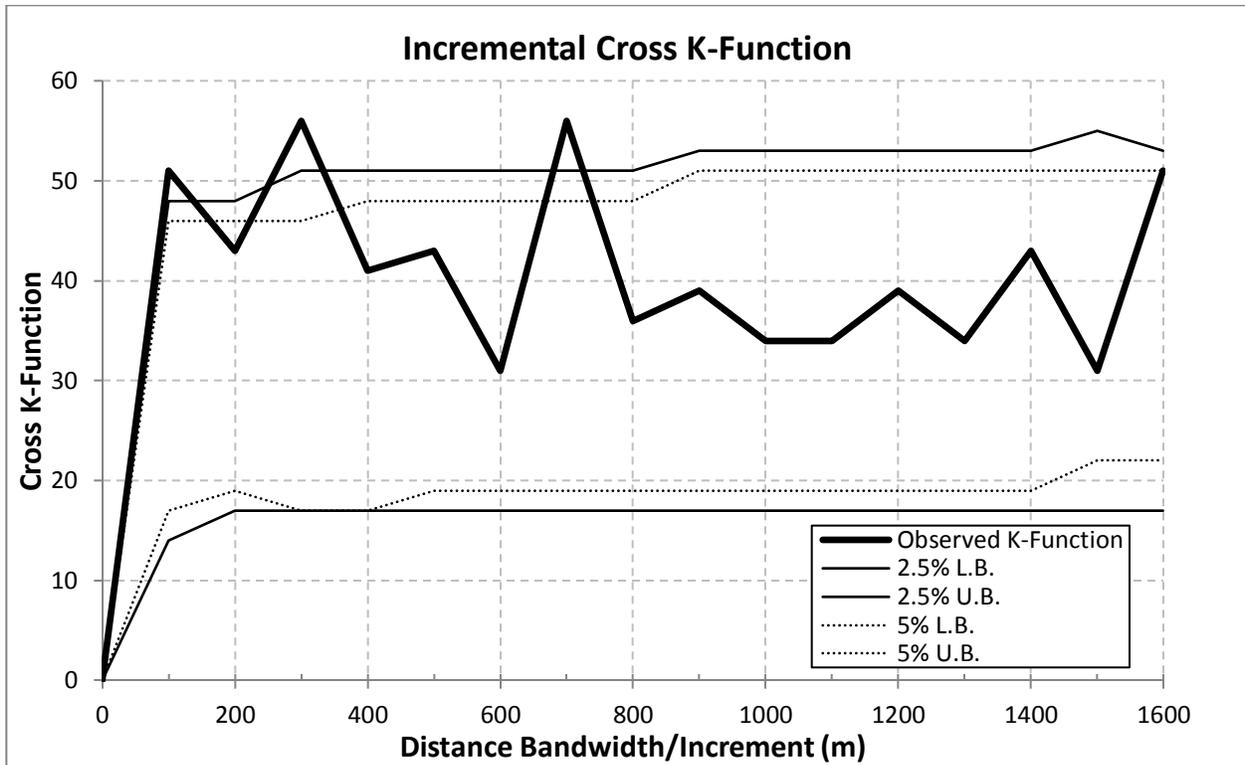
FIGURE 9 INCREMENTAL CROSS K-FUNCTION FOR ENTRANCE RAMPS AND CMC

Analysis of CMC and Exit Ramps

Similar to the analysis of CMC and entrance ramp locations, the exit ramp locations were also analyzed to determine their effects on CMC. Figure 10 and Figure 11 show the results of the global network and incremental cross K-Functions based on 2.5 and 5 percent confidence levels. Both Figure 10 and Figure 11 indicate that CMC do not cluster near exit ramps. The results of incremental cross K-Function in Figure 11 shows some dispersion in CMC but in general, the pattern is random. The results suggest that exit ramp locations do not affect the occurrence of CMC. The results of the analysis of CMC and exit ramps as presented in Figure 10 and Figure 11 contradict the findings based on traditional Negative Binomial regression, where the effect of exit ramps on CMC was statistically significant positive. The reason for the contradiction may be the fact that in the dataset used for NB regression, crashes and presence of features (such as bridges, ramps, curves) were aggregated by segment lengths of half and one mile. Since most exit ramps at interchange locations are immediately followed by bridges and/or entrance ramps, crashes downstream of exit ramp end points but clustering near bridges/entrance ramps could have been falsely attributed to exit ramps. Hence, the advantages of using distance-based statistic are clearly demonstrated by the results of CMC and exit ramp analysis illustrated in Figure 10 and Figure 11.
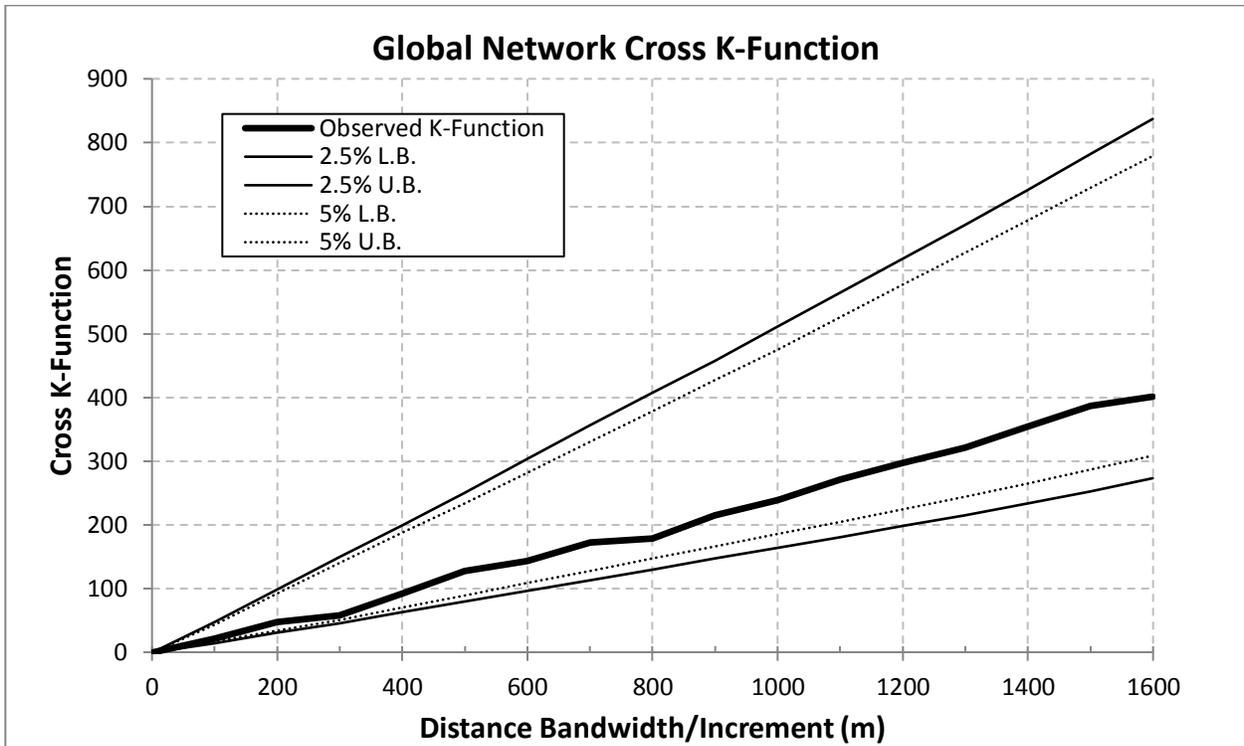
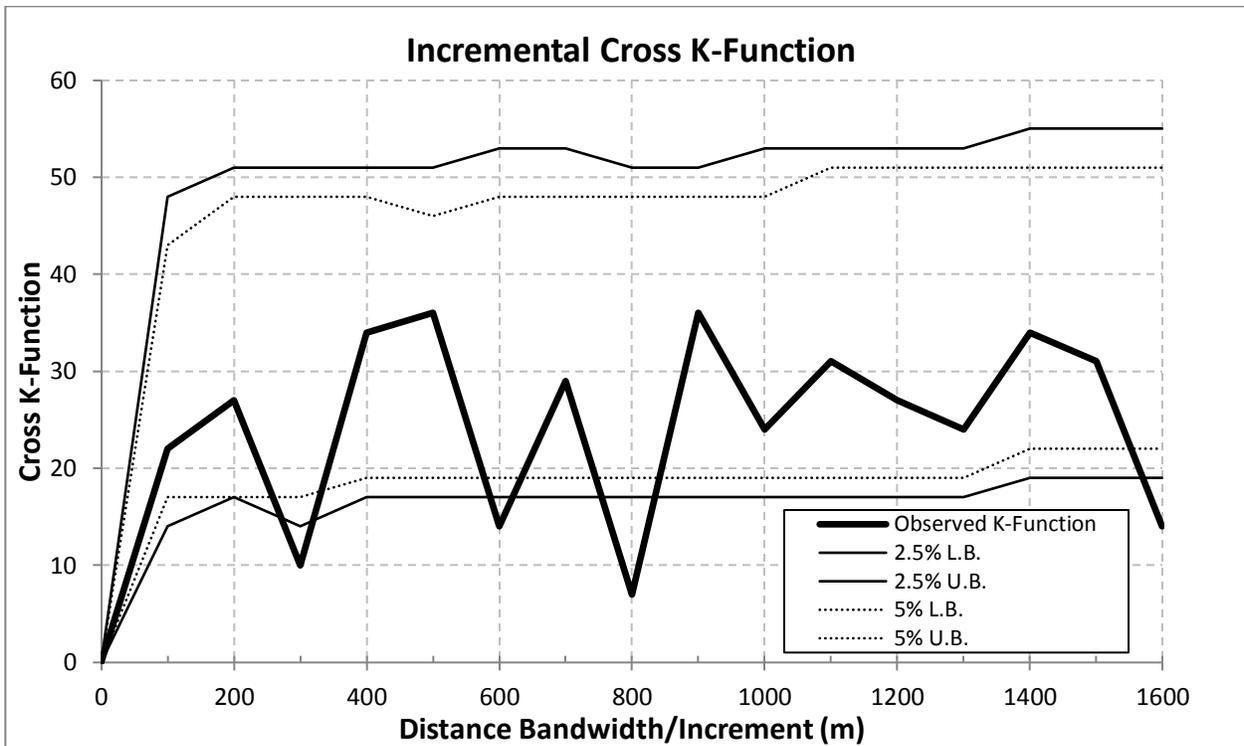**FIGURE 10 GLOBAL NETWORK CROSS K-FUNCTION FOR EXIT RAMPS & CMC**

FIGURE 11 INCREMENTAL CROSS K-FUNCTION FOR EXIT RAMPS AND CMC

Analysis of CMC and Left Curves

The idea behind the analysis of CMC in relation to horizontal curves was to determine if the probability of CMC increased in the vicinity of horizontal curves. The analysis was conducted for left and right curves separately due to the difference in the behavior of vehicles traversing the curves. Figure 12 and Figure 13 show the results of the global network and incremental cross K-Functions for CMC and left curves, based on 2.5 and 5 percent confidence levels Although both Figure 12 and Figure 13 indicate that CMC do not cluster near left curves, the results of incremental cross K-Function provides more details especially within distance increments. Figure 13 shows that CMC are dispersed around the start of the curve and at 800 meters distance; but are generally distributed randomly indicating that left curves do not affect the occurrence of CMC. However, the results were not considered reliable due to varying lengths of the curves in the study area.
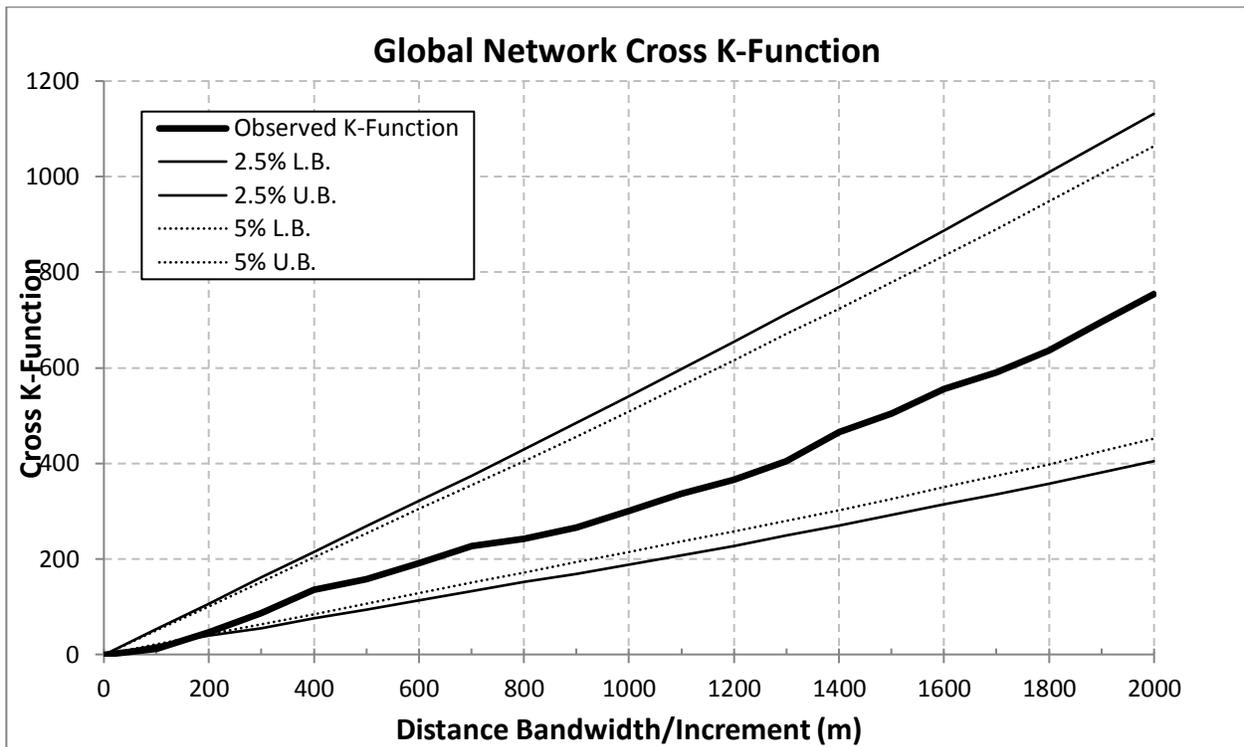

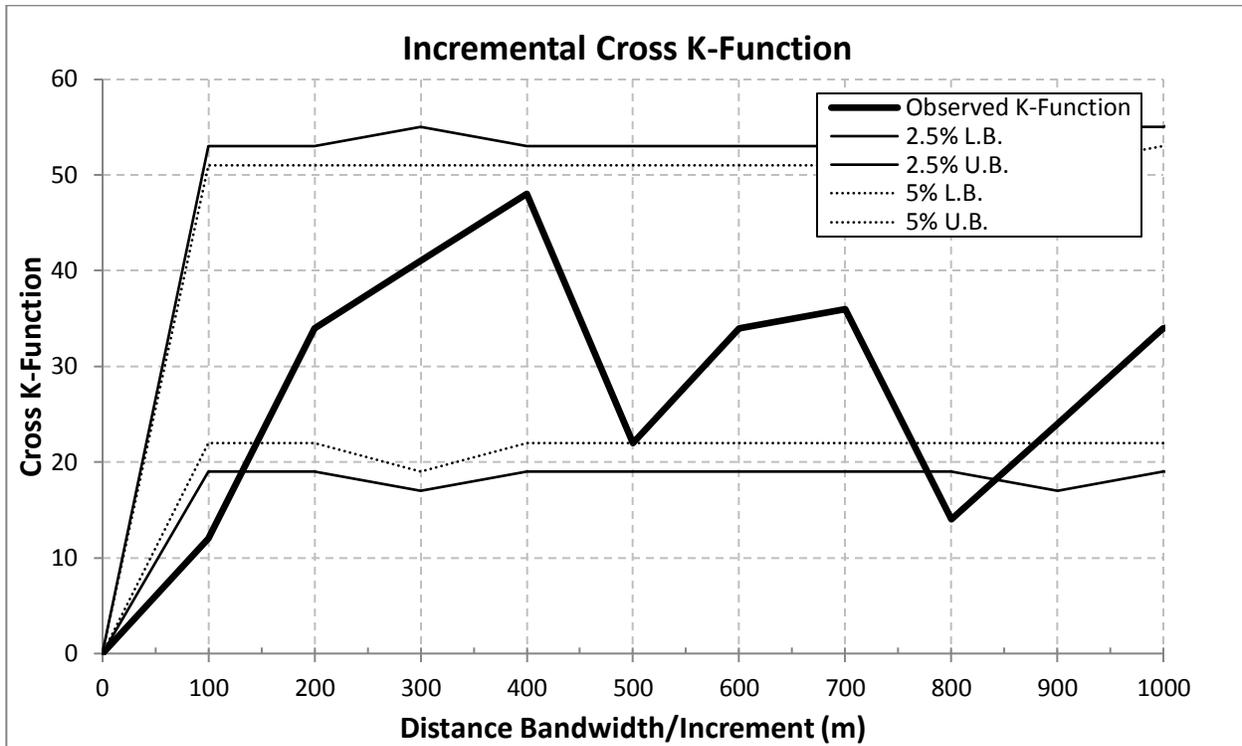**FIGURE 12 GLOBAL NETWORK CROSS K-FUNCTION FOR LEFT CURVES & CMC**

FIGURE 13 INCREMENTAL CROSS K-FUNCTION FOR LEFT CURVES AND CMC

The analysis of CMC near horizontal curves presented a unique challenge due to varying lengths of the curves and the nature of interaction between crashes and horizontal curves. Hence, new methods, namely variable distance and incremental variable distance cross K-Functions were developed to counter the effects of varying lengths of geometric features under special circumstances. The results of the variable distance and incremental variable distance cross K-Functions are presented in Figure 14 and Figure 15 based on 2.5 and 5 percent confidence levels. The x-axis shows the increasing distance downstream from the start point of left curves in terms of the proportion of length of individual left curves. For example, 0.25L would represent a distance increment of 25 and 50 meters for left curves of lengths 100 and 200 meters respectively; as compared to the case of regular K-Functions where the values at 25 and 50 meters would be used separate from each other. The results of the analysis of left curves and CMC using incremental variable distance cross K-Function in Figure 14 and Figure 15 present greater detail and different results from the ones shown in Figure 12 and Figure 13.
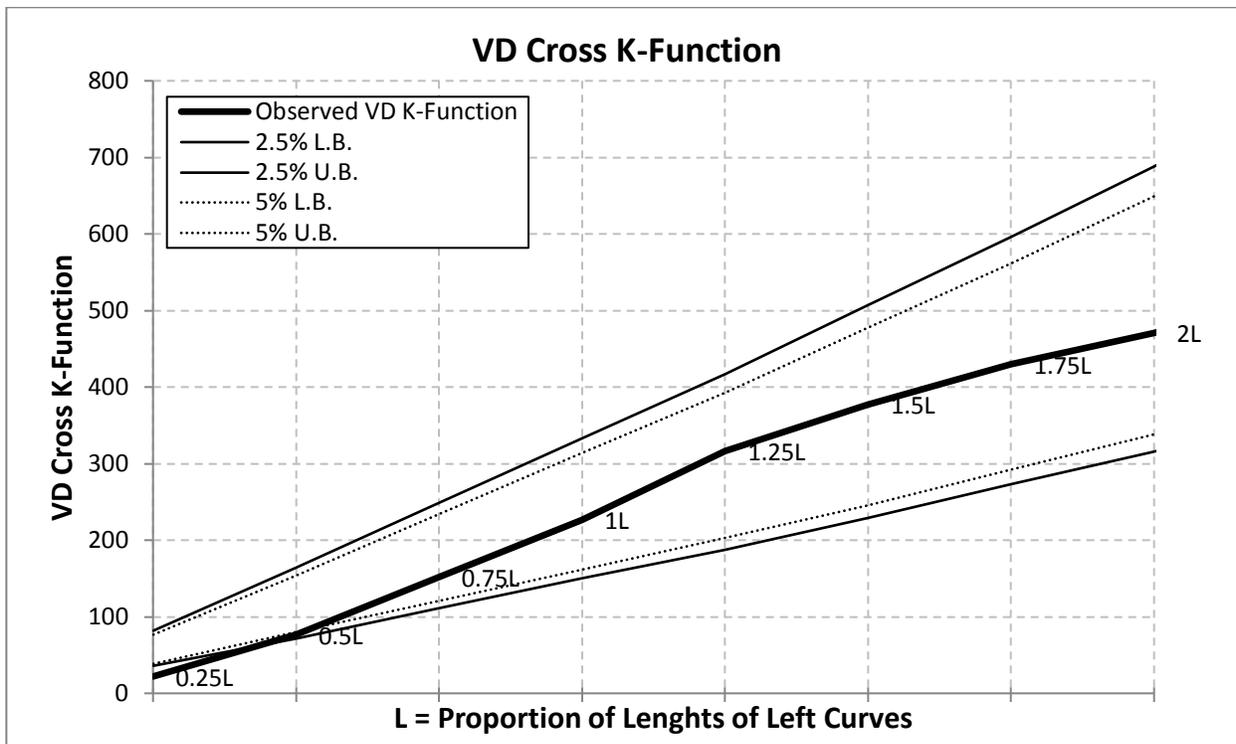
FIGURE 14 VARIABLE DISTANCE CROSS K-FUNCTION FOR LEFT CURVES AND CMC
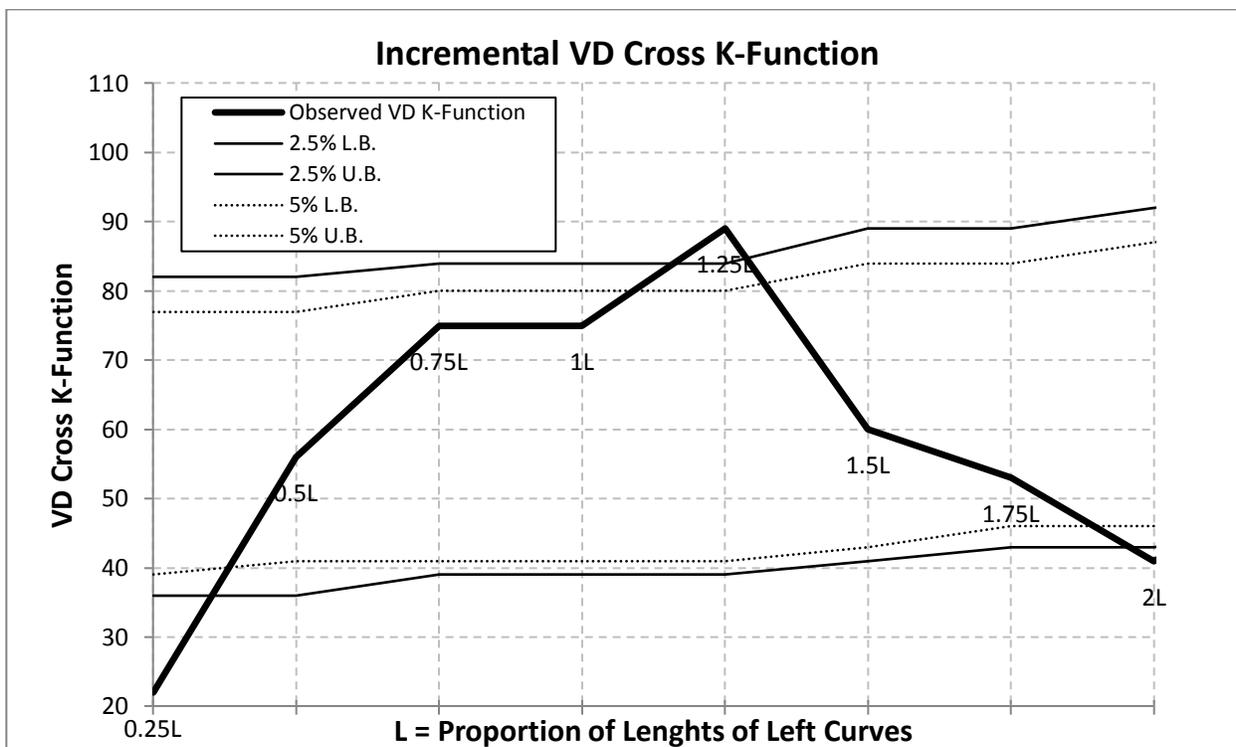

FIGURE 15 INCREMENTAL VARIABLE DISTANCE CROSS K-FUNCTION FOR LEFT CURVES AND CMC

Although CMC are dispersed near the start of the left curves, they become clustered at a distance of 1.25L which is downstream of end point of the left curves. Clustering of CMC shows that the probability of CMC increases near left curves. The fact that there was dispersion at the start of the curve and clustering at a distance of 1.25L further reinforces the results. If a CMC occurring on a left curve was influenced by the nature of the left curve, the vehicle would have to traverse some portion of the length of the curve before ending up in a crash at some point downstream of the start point of the curve. Hence the chance of observing a CMC at the start of the left curve would be far less and any clustering would be observed downstream of the start point.

The analysis of CMC near left curves using the same dataset and negative binomial regression showed that CMC are reduced by the presence of left curves which is consistent with the results of incremental cross K-Function showing dispersed pattern in Figure 13. However, the results in Figure 14 and Figure 15 contradict the findings from NB regression and clearly demonstrate the advantages of variable distance based K-Function analysis by showing clustering of CMC downstream of left curves.

Analysis of CMC and Right Curves

CMC were also analyzed with respect to right curve locations. Figure 16 and Figure 17 show the results of the global network and incremental cross K-Functions based on 2.5 and 5 percent confidence levels. The x-axis shows the increasing distance downstream from start point of the right curves. Figure 17 shows that CMC are dispersed with respect to right curves only at a distance of 1100 meters from the start of the curve. For the rest of distance bandwidths, both Figure 16 and Figure 17 indicate that CMC are random with respect to right curves. The results show that right curves do not influence the occurrence of CMC.
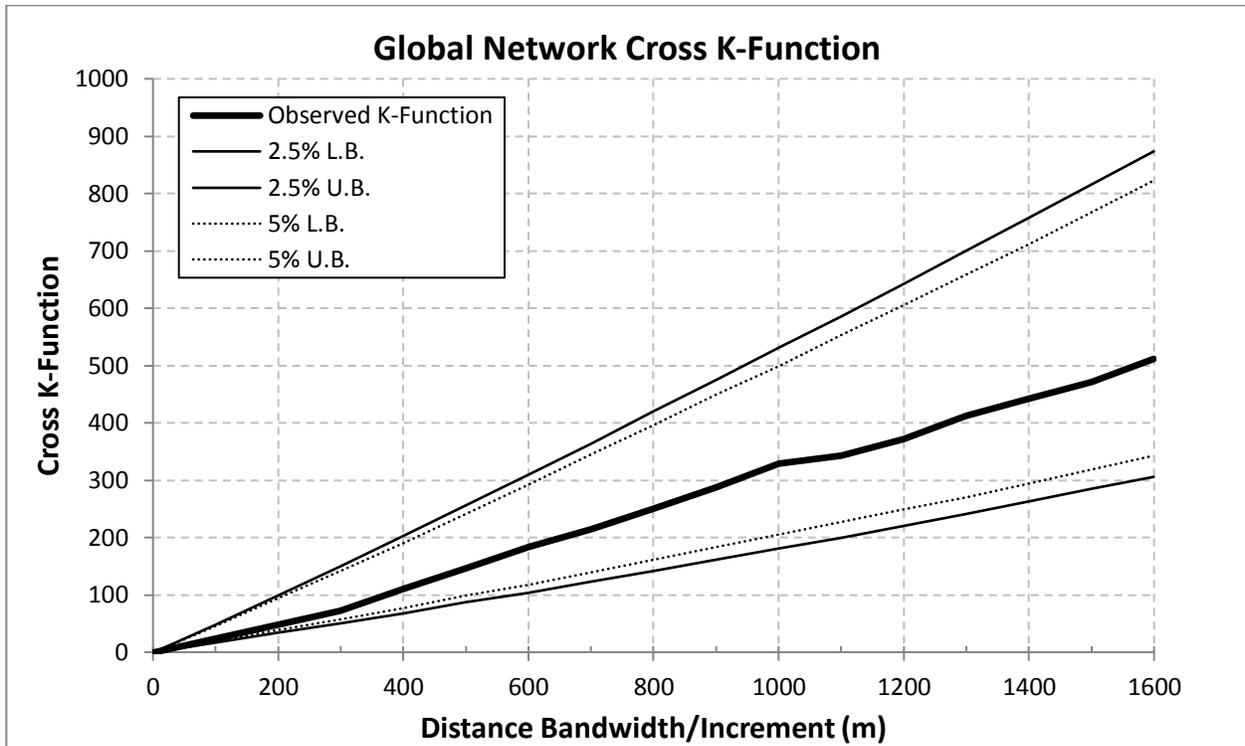
FIGURE 16 GLOBAL NETWORK CROSS K-FUNCTION FOR RIGHT CURVES & CMC
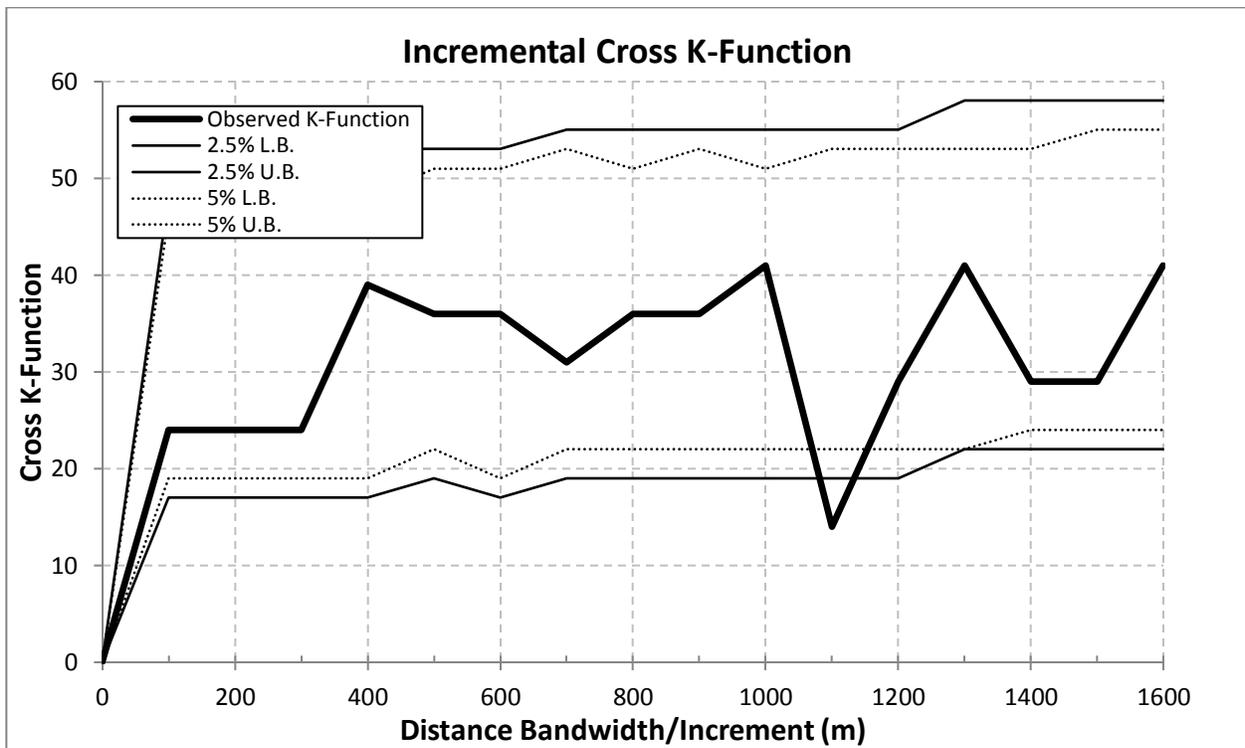

FIGURE 17 INCREMENTAL CROSS K-FUNCTION FOR RIGHT CURVES AND CMC

As in the case of left curves, CMC near right curves were also analyzed using variable distance and incremental variable distance cross K-Functions. The results are presented in Figure 18 and Figure 19 based on 2.5 and 5 percent confidence levels. The x-axis shows the increasing distance downstream from the start point of right curves in terms of the proportion of length of individual curves. Figure 19 presents the results of incremental variable distance K-Function which shows clustering at a distance of 1L from the start of the right curve. The results of the variable distance based methods show that there is a relationship between CMC and right curves, which is similar to the case of left curve analysis in the sense that the use of regular K-Functions do not reveal any clustering but by using variable distance methods, the clusters were identified.
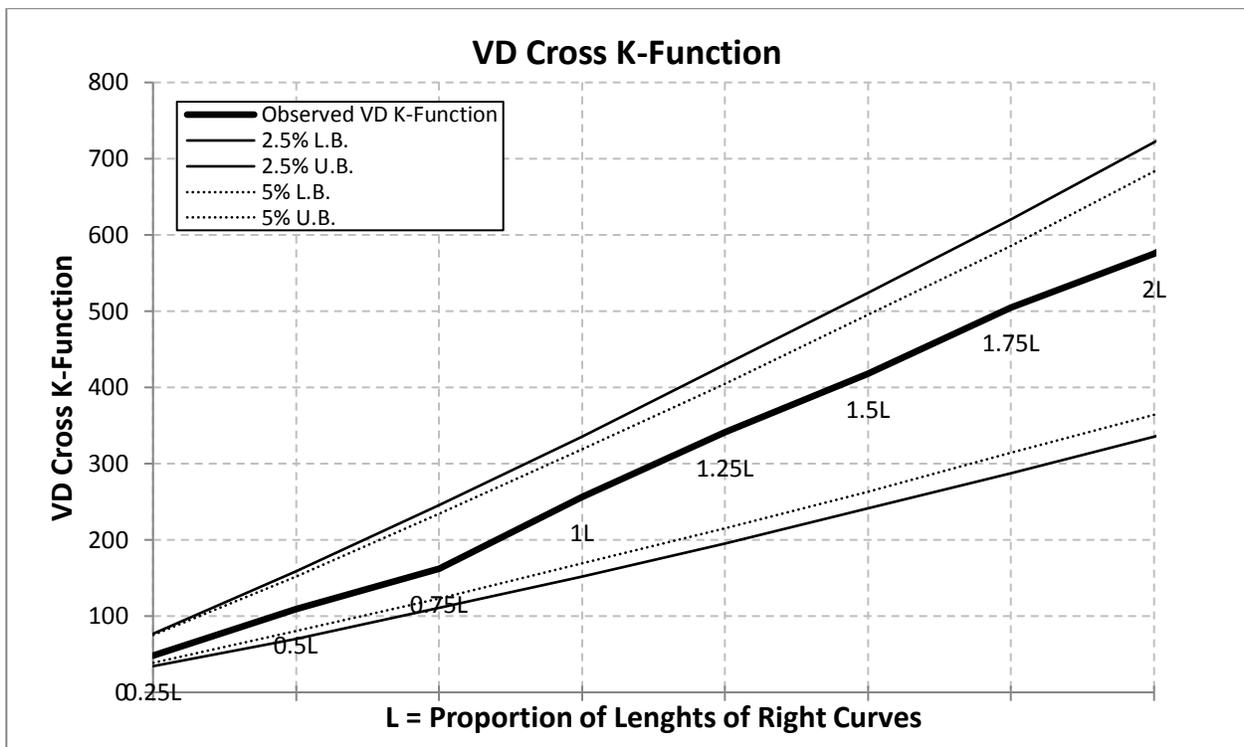

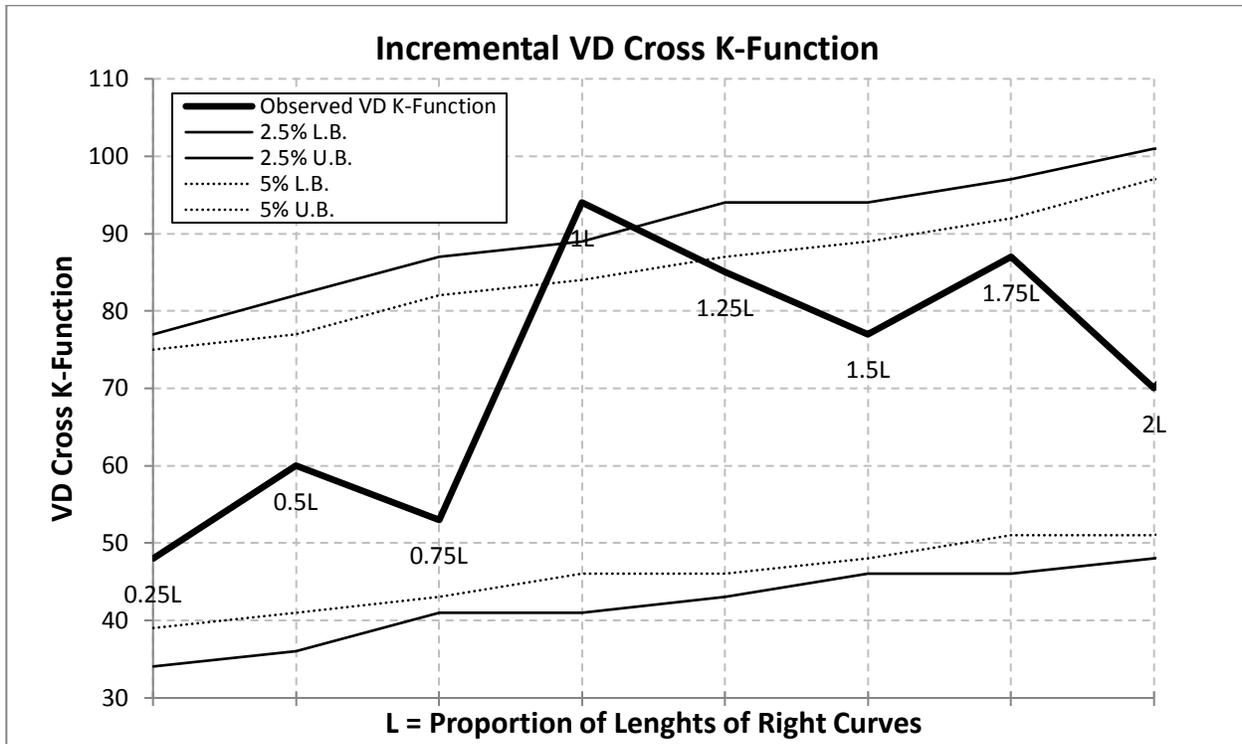**FIGURE 18 VARIABLE DISTANCE CROSS K-FUNCTION FOR RIGHT CURVES AND CMC**

FIGURE 19 INCREMENTAL VARIABLE DISTANCE CROSS K-FUNCTION FOR RIGHT CURVES AND CMC

CONCLUSIONS

Spatial statistical methods and tools were developed for analyzing CMC hotspots and contributing factors. Global Network K function was used to ascertain that CMC in Wisconsin are clustering and not completely spatially random. Local Network K function was used to identify CMC clusters and hotspots. In addition a nomograph was created for CMC crash rate thresholds that can be used for identifying CMC hotspots. The thresholds are a function of AADT and distance bandwidth.

The contributory factors analysis for CMC was conducted using the global network, incremental, and variable distance cross K-Functions; which were all part of the anisotropic K-Function. The results of the contributory factors analysis not only shows the nature of relationship between crashes and geometric features but also for the first time the extent of the relationship in terms of distance from the features either in one or both directions. CMC cluster downstream of bridges up to a distance of 100 meters. CMC cluster around 100, 300, and 700 meters mark downstream from start of entrance ramps. There is no relationship between CMC and exit ramps. Global network cross K-Function analysis of CMC show no relationship with left curves whereas variable distance cross K-Function analysis shows that CMC cluster downstream of left curves.

Global network cross K-Function analysis shows that CMC are dispersed near right curves whereas variable distance cross K-Function shows clustering at end of right curves.