Spatial Analysis of Weather Crash Patterns

Ghazan Khan[1]; Xiao Qin, Ph.D., P.E.[2]; and David A. Noyce, Ph.D., P.E.[3]

Abstract: Spatial statistical techniques can be an effective tool for analyzing patterns and autocorrelation in crash data, especially weather-related crashes. Since weather is a geographic phenomenon, it tends to show distinct geographic patterns affecting certain locations more than others. Accordingly, "weather-related" crashes may also display similar distinct patterns or clustering. The objective of this research was to use spatial statistical techniques to identify significant patterns of weather-related crashes. Weather-related crashes, defined as those crashes which occurred in adverse weather conditions, were analyzed using the Getis-Ord $G_i^*(d)$ statistic. The statistic reveals spatial patterns for weather-related crashes which are clustered at different locations depending upon weather conditions (snow, rain, and fog). The results also show geographic areas (counties) of statistically significant high and low relative crash rates for each weather condition. Furthermore, the resulting patterns of crashes were validated by comparing counties of high and low crash rates with areas of varying weather data. The establishment of this relationship between weather and crashes is imperative in identifying the variables contributing to these crash types and the implementation of effective countermeasures for road weather safety audit purposes.

Introduction

One of the most critical questions that traffic safety engineers face is where to implement safety countermeasures such that the most significant impact on safety can be achieved. The accurate identification of safety deficient areas on a broader scale for planning purposes and on a local scale for highway safety treatments is the key to a successful and comprehensive safety program.

Safety professionals realize that an important aspect of road safety is weather and its contribution to crashes. Weather inevitably affects road safety throughout the year in all climates through snow, rain, fog, ice, sleet, or wind. The effects of these adverse weather conditions on road safety are obvious in terms of reduced pavement friction, reduced driver visibility, and deteriorated traffic control device functions. For a traffic safety engineer or road safety auditor, the first task is identifying a potential or existing problematic area prior to addressing the root causes of crashes in relation to weather interaction with roadway geometric features, and traffic conditions.

In 2004, there were 6,181,000 reported vehicle crashes in the United States (NHTSA 2004). About 16% of these crashes—approximately 991,000—were determined to be "weather-related" crashes as rain, sleet, snow, fog, ice, or some combination of these were present at the time of crash. Weather-related crashes in 2004 resulted in nearly 5,000 fatalities and approximately 276,000 injuries. Crash statistics were fairly similar in 2003 with just over 16% of all the crashes being weather related (NHTSA 2003). Between 1995 and 2001, the percentage of injury and fatal crashes related to adverse weather conditions was approximately 22% for both crash types (Goodwin 2002). These figures translate into a significant economic and social loss.

The transportation system in Wisconsin experiences all weather types and is a prime example of the effects of adverse weather on traffic safety. According to the Wisconsin Department of Transportation (WisDOT), approximately 1,430 persons were killed on Wisconsin roadways from 1999 to 2002 and 116,790 were injured in crashes that took place during adverse weather conditions. In the wake of the deadliest crash in Wisconsin history on October 11, 2002, where dense fog led to a 50 vehicle crash that left ten people dead and 39 injured, Wisconsin has implemented a more aggressive and proactive approach to improving weather-related traffic safety on state highways. One of the actions considered was to introduce a formal road weather safety audits procedure into the WisDOT facilities development process, symbolizing a pronounced step towards incorporating traffic safety seamlessly into its project and program development.

Ideally, road weather safety audits should be performed at all locations to assess the safety of the roadway system. Clearly, the magnitude of this task makes it impossible. Therefore, weather prone locations must be identified and prioritized so that audit efforts can be focused on those locations in which weather is a significant contributor to safety issues.

In this research, weather prone locations or clusters in the state of Wisconsin were identified by considering historical weather-related crashes. Normal techniques, such as plotting weather-related crash rates or numbers cannot provide a clear picture of the most weather prone locations (crash rate or frequency map),

[1]Research Assistant, Dept. of Civil and Environmental Engineering, Univ. of Wisconsin-Madison, 2205 Engineering Hall, 1415 Engineering Dr., Madison, WI 53706. E-mail: gkhan@wisc.edu

[2]Associate Researcher, Traffic Operations and Safety (TOPS) Laboratory, Univ. of Wisconsin-Madison, 1214 Engineering Hall, 1415 Engineering Dr., Madison, WI 53706. E-mail: xqin@engr.wisc.edu

[3]Associate Professor, Department of Civil and Environmental Engineering, University of Wisconsin-Madison, 1204 Engineering Hall, 1415 Engineering Drive, Madison, WI 53706 (corresponding author). E-mail: noyce@engr.wisc.edu

nor does it provide reliable safety information because crash occurrence is a stochastic process with latent variables in error. An innovative way to investigate patterns of high and low weather-related crash locations in a quantifiable manner is through spatial statistical techniques. Through spatial analysis, statistically significant patterns can be recognized in the form of clustered or nonrandom crash events. Moreover, the use of these techniques can provide a measure to quantify the safety worthiness of weather-prone areas for road safety audit focus and renewed safety emphasis.

This research used spatial statistical techniques to identify weather-related crash patterns. Spatial patterns were revealed through spatial autocorrelation by observing clusters of areas with similar attribute values—high or low crash rates given that weather is a geographic phenomenon. Weather demonstrates patterns which are influenced by geographic conditions and limitations; hence, crashes related to weather should also display similar patterns. The identification and prioritization of these locations is the first step towards closely analyzing the prevailing conditions on roadways for the underlying causes of crashes.

Literature Review

Traditional road safety studies evaluate the impacts of various parameters on safety performance, either qualitatively or quantitatively, such as the influence of various geometric features on crash occurrences (Shankar et al. 1995; Khattak and Knapp 2001a). These studies often develop statistical models under the assumption that crash events occurring in a specific space are independent of each other. This assumption may hold true for crashes happening in a small area where geographical conditions are assumed to remain similar; the same cannot be said for county—or statewide analysis areas. 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 because features lying in space influenced by geographical factors are bound to display some sort of spatial dependencies (Getis and Ord 1992 Cressie 1993).

Lack of spatial independence in geographic data has given rise to statistical techniques that measures spatial autocorrelation in data, which can be incorporated into modeling procedures to eliminate errors and account for spatial dependencies. For crashes that are influenced by weather or some other spatial phenomenon, the spatial independence assumption is often violated. It is important to analyze the spatial heterogeneity/homogeneity of these data spread in space, especially when analyzing them from a geographical context to make correct assumptions about the nature of the data and the analysis conducted (Cressie 1993). Road crashes can be analyzed from different spatial contexts to establish spatial associations. The measurements of these spatial dependencies or spatial autocorrelation integrated with geographic information system (GIS) can help analyze spatial patterns and clusters in crash data as well as help improve modeling procedures and error estimates.

Several studies have been conducted to establish spatial patterns in vehicle or pedestrian crashes for identification of critical locations (Jones et al. 1996). Kim and Yamashita analyzed spatial patterns of pedestrian crashes in Honolulu, Hawaii using K-means clustering techniques (Kim and Yamashita 2004). These spatial patterns identified areas of high pedestrian crashes that were present in light of various demographic features such as population or land use. Similarly, Levine conducted a spatial analysis of Honolulu crashes in the context of varying conditions and noted the limitations of "blackspot" analysis in terms of describing the location and causation of different types of crashes (Levine et al. 1995). Flahaut carried out a study for black zones and found several advantages in defining black zones (i.e., high crash frequency locations) using spatial autocorrelation and kernel methods on road segments (Flahaut et al. 2003).

These studies were conducted at segment, corridor, or intersection level at preselected locations. There were no specific studies identified on a broader regional scale. The objective of the research described in this paper was to identify focus areas in the state as a first step and the second step would be to perform an in depth and detailed level analysis in light of various contributing factors. Additional information on the second step can be found in the research work completed by the authors (Qin et al. 2006).

As demonstrated in these microscopic-based research efforts, it is logical to extend the use of spatial pattern recognition to identify patterns in weather-related crashes.

There are two ways of assessing spatial patterns in geographic data using spatial association:
- Global measures; and
- Local measures.

Global measures of spatial association analyze patterns on a large scale to show whether data are clustered, dispersed, or randomly distributed in space. Specifically, global measures show the overall patterning of the data in the region under study. Some examples of global measures methods are the Moran's I index and Getis–Ord General G statistic (Anselin 1995). As the global measures of spatial association can be used to test general patterns in data, the identification of statistically significant patterns of high (hot spots) or low (cold spots) attribute values within the study area is also interesting and necessary.

The local measures of spatial association can quantify spatial autocorrelation at a small scale which may be masked by global measures. Both distance statistics, i.e., Getis–Ord Gi statistics and local Moran's I or local indicators of spatial association (LISA) proposed by Anselin are well known types of local measures of spatial association (Getis and Ord 1992; Anselin 1995; Ord and Getis 1995).

The calculation of spatial autocorrelation requires the identification of an extent of the neighborhood surrounding an individual location. The extent can be defined in terms of contiguity (boundary based) or distance (band width) based. Additionally the nature of the spatial relationships between that location and its neighboring localities based on which spatial associations are calculated is required. The nature of the interaction can be defined as Euclidian (shortest path), network constrained (distances measured on road surfaces), cost, or inverse weighted distance.

This conceptualization of the spatial relationships is represented in a spatial weight matrix. In the most general form, a first-order contiguity based weight assigns a value of 1 to all neighbors bordering a particular location and assigns a value of 0 to nonbordering locations. Similarly for distance, all locations within a specified distance band are assigned a value of 1 while areas outside the cutoff distance are assigned a value of 0. According to Flahaut, the choice of weighting matrix is never objective because of the limitations in defining clear geographic boundaries regarding the subject matter under study (Flahaut et al. 2003).

The choice of a spatial weight matrix is difficult due to the

variable and unpredictable nature of weather events. Nevertheless, effort was made to ensure objectivity in this choice. Several studies had considered spatial relationships described by spatial weight matrices depending upon the nature and objective of the study (Getis and Ord 1992; Anselin 1995; Ord and Getis 1995; Flahaut et al. 2003). For instance, the choice of a spatial weight matrix can be dependent on some prior knowledge about the area under study or a general idea of the interaction between neighboring locations with regards to the subject matter. The choice of spatial weight matrix is explained in detail in the "Methodology" section.

In order to analyze the crash data, definitions had to be established to properly define weather and nonweather-related crashes. Furthermore, since different weather phenomenon show different patterns throughout a year, it was also decided to further analyze the weather crashes according to different weather types. Several authors have defined weather-related crashes; however, these studies used varying definitions (Andrey et al. 2003; Andrey and Yagar 1993; Satterthwaite 1976; Khattak and Knapp 2001a,b; Brodsky and Hakkert 1988). According to the Federal Highway Administration (FHWA), weather-related crashes are defined as crashes that occur in adverse weather (presence of rain, snow, sleet, fog) or slick pavement conditions (wet, snowy, slushy, or icy pavement) (Goodwin 2002). The FHWA definition of weather-related crashes was applied in this research.

Data Collection and Assembly

Three years of Wisconsin crash data (2000–2002) were obtained to complete this research. Wisconsin crash data contain two sections regarding weather conditions at the time of the crash, namely "weather condition" and "road condition." Weather-related crashes selected for this analysis included one or more of the following three types of weather and/or road conditions:
1. Fog (crashes reported in foggy weather conditions);
2. Snow (crashes reported in snowy weather conditions and/or roads covered with snow); and
3. Rain (crashes reported in rainy weather conditions and/or wet pavement conditions).

Assembling the data required an aggregation of crash data within a geographic boundary. It was decided to aggregate the data set of all crashes on a county level because it provided a well-defined jurisdictional-based picture of the state in terms of planning and programming of safety implementations. County level analysis was the first step by which the over 19,000 segments of varying length throughout the state could be prioritized for further microscopic analysis. Relative crash rates were defined for each county as the percentage of individual weather-type crashes divided by the total number of crashes. Relative crash rates were used because crashes aggregated on a county level had to be normalized by some measure of exposure to facilitate the comparisons between counties. Traditional exposure values such as volume and vehicles miles of travel (VMT) data were not available for local roads.

It would be ideal to clearly identify and exclude all those crashes where human factors (i.e., alcohol, age, etc.) are the only cause of a weather-related crash. There are several studies in literature suggesting that between 70% and 90% of all crashes are human error related, for example, a study sponsored by the Australian Transport Safety Bureau Report in January 2006 suggests that 75% of all crashes are human error related (Salmon et al. 2006). However, there are certain locations where driver-error related crashes could become more prevalent under adverse weather conditions as compared to nonadverse weather conditions. In that case, ideally the problem should be addressed both from the driver as well as weather standpoint. If the crashes are driver-error related, then no clusters should be detected given the random nature of these crashes. On the other hand, weather-related crashes should cluster in weather-prone areas. It is much easier to identify the commonalities in the environment and infrastructure of weather-related crashes and take corrective measures that make roads more forgiving rather than completely eliminate human error.

Wisconsin is divided into 72 counties and covers a total land area of approximately 150,000 km^2 with almost its entire eastern border alongside Lake Michigan. Wisconsin extends approximately 500 km from north to south and 400 km from east to west. The weather patterns vary throughout the state in all directions depending on the geographic proximity. Comprehensive weather data were available and obtained from the National Weather Service (NWS) Cooperative Observer Program (NWS COOP 2006).

Figs. 1 and 2 present total annual snowfall and rainfall trends generated from NWS COOP weather stations data. These continuous weather maps were generated from point weather station data using geostatistical interpolation techniques (Universal Kriging) to cover all areas of the state (Qin et al. 2006). As one would expect, areas close to Lake Michigan experience more rainfall than other regions of the state, whereas the northern regions of the state close to Lake Superior experience greater snowfall throughout the year.

To understand the magnitude of the weather-related crash problem, Table 1 presents a summary of weather and nonweather-related crashes for Wisconsin from 2000 to 2002. In Table 1, it is evident that the frequency of rain-related crashes is the largest among all weather-related crashes. The number of fog-related crashes decreased in 2002 after being consistent for the previous 2 years.

Methodology

The objective of this research was to identify locations that experienced a significantly higher percentage of weather-related crashes through pattern detection techniques and that the occurrence of these crashes was not a random or chance event. In order to analyze patterns of spatially distributed features, overall regional pattern measures can be visually interpreted through frequency, mean, or proportion measurements on GIS-based maps. Although the weather-related crashes could easily be plotted statewide, creating a visualization-based map, the ability to visually discern spatial patterns and to identify hot spots was limited. There was a need to identify statistical processes to provide a quantifiable measure of spatial patterns rather than predefined ranking or number based classifications because visual interpretations alone cannot provide conclusive results. Spatial patterns supported by statistically significant quantities, which describe those patterns accurately, can resolve the issue by providing a quantifiable method of analysis.

There are several statistical techniques available for analysis of spatial patterns in an effort to identify clusters of high or low attribute values. Some of these techniques have been mentioned in previous sections. The primary reason behind using spatial statistical techniques is the fact that classical statistical procedures (aspatial techniques) do not consider geographic proximity. Spatial statistical techniques go one step further by incorporating the
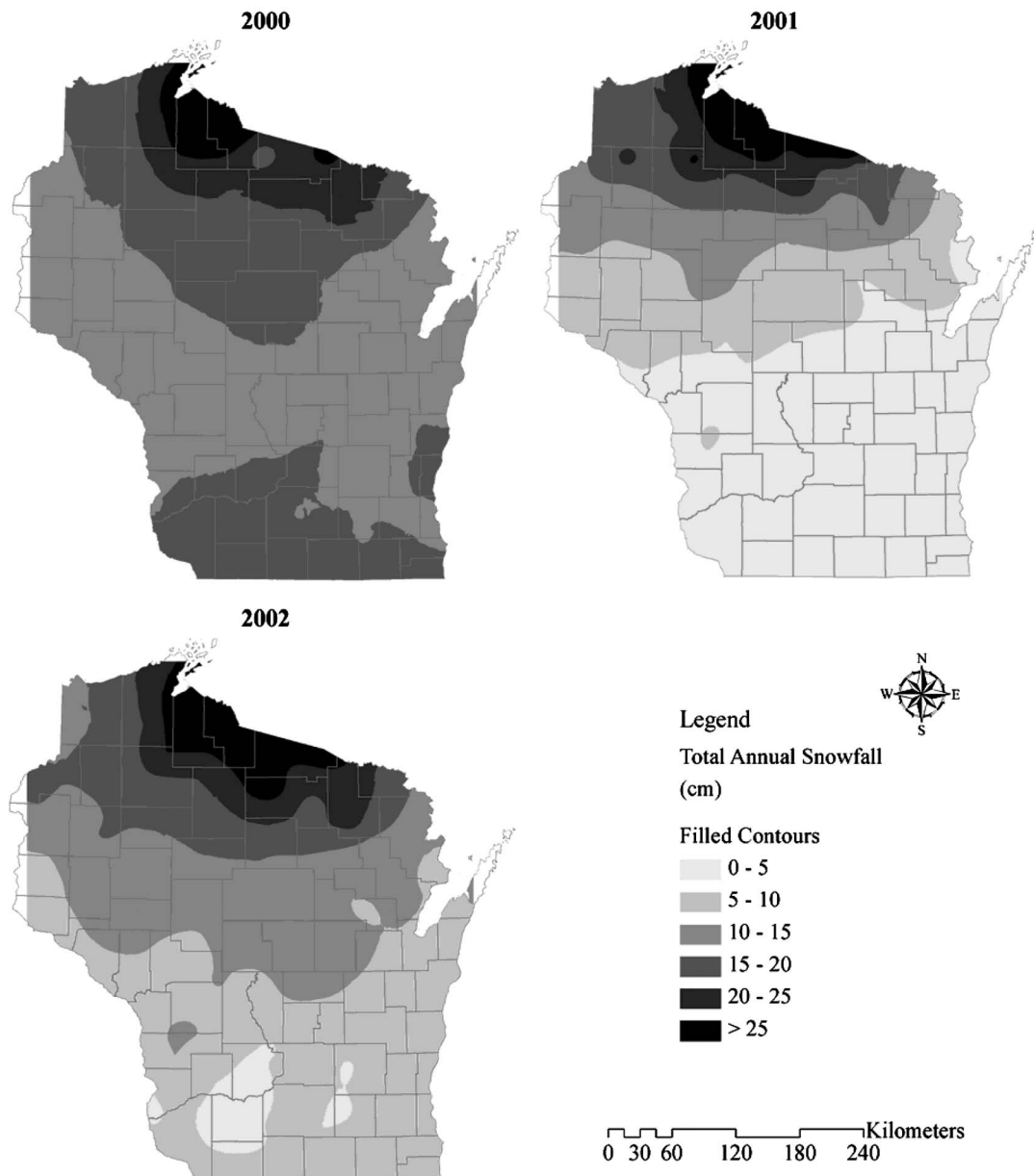
Fig. 1. Continuous surface for total snowfall amount in Wisconsin 2000–2002

relationships between adjacent areas, or spatial autocorrelation, which is inherent in geographically distributed data such as weather-related crashes. Since adverse weather events are inevitably focused in particular areas due to their dependence on geographic conditions and atmospheric proximities, crashes affected by such weather events should therefore show similar patterns. The use of spatial statistical techniques adds depth and significance to the results since data are assumed to be independent. With these requirements in mind, the Getis–Ord $G_i^*(d)$ statistic was used to analyze patterns of weather-related crashes on a county level for Wisconsin (Ord and Getis 1995).

Given the Getis–Ord $G_i^*(d)$ statistic, the following set of hypothesis was defined for this research: $\mathbf{H_0}$=weather-related crashes display no clustered or dispersed patterns and are randomly distributed across space; and $\mathbf{H_a}$=weather-related crashes show clustered patterns signifying that they are affected by weather.

The above hypothesis is based on the premise that weather patterns concentrate in certain areas; therefore, crashes primarily or partially caused by weather will also concentrate in those areas. If crashes happening in adverse weather are caused by some other reasons, their patterns would not correlate with weather patterns. Instead of identifying the underlying causes or circumstances such as traffic conditions, roadway geometrics, roadside hazards, or human factors involved in weather-related crashes, the goal is to identify areas that should be prioritized for efforts related to improving traffic safety.

$G_i^*(d)$ statistic is described by Ord and Getis (Ord and Getis 1995). The statistic indicates the extent to which a location is surrounded by a cluster of high or low values (Ord and Getis 1995). $G_i^*(d)$ statistic shows areas where higher or lower than average values tend to be found near each other. The standardized G calculates a single z-score value for each location in the study area. A positive value indicates clustering of high attribute value locations and a negative value indicates clustering of low attribute
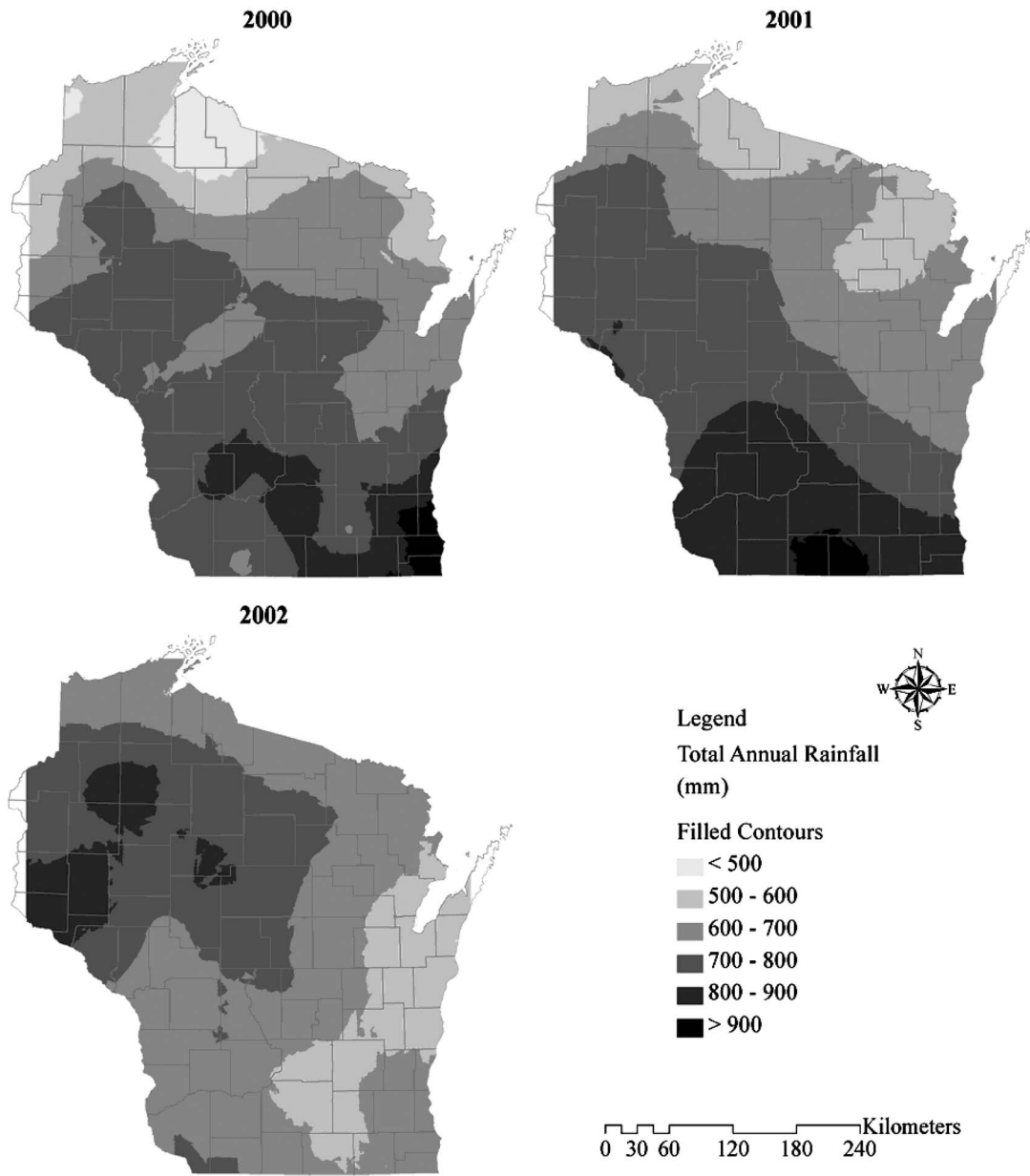
Fig. 2. Continuous surfaces for total rainfall amount in Wisconsin 2000–2002

value locations. The larger the absolute values, the more significant the results are.

The $G_i^*(d)$ statistic can be presented as follows

$$G_i^*(d) = \frac{\sum_{j=1}^N w_{ij}(d)x_j - \bar{x}\sum_{j=1}^N w_{ij}(d)}{S\sqrt{\frac{[N\sum_{j=1}^N w_{ij}^2(d) - (\sum_{j=1}^N w_{ij}(d))^2]}{N-1}}} \qquad (1)$$

where

Table 1. Total Crash Summary by Weather for Wisconsin 2000–2002

	Total weather crashes				
Year	Fog	Snow	Rain	Total nonweather crashes	Total
2000	1,061	24,687	19,066	103,018	152,649
2001	1,109	8,340	18,563	104,432	137,988
2002	610	12,227	15,535	108,868	141,445

$$\bar{x} = \frac{\sum_{j=1}^N x_j}{N} \quad \text{and} \quad S = \sqrt{\frac{\sum_{j=1}^N x_j^2}{N} - (\bar{x})^2}$$

$G_i^*(d)$=Getis–Ord $G_i^*(d)$ z-score value including the value at site i; x_i=relative crash rate of site i; x_j=relative crash rate of neighboring locations to site i; d=fixed band radius around site i; w_{ij}=spatial weight matrix for all sites j within distance d; and N=number of weighted points, each representing relative crash rate for each county.

In Eq. (1), i=site with an attribute value x_i where the $G_i^*(d)$ statistic is being calculated and x_j=neighboring locations up to a distance d with similar attribute values. For analysis, the attribute values were multiplied by the spatial weight matrix, w_{ij}, that defines which locations was included in the analysis and corresponding weight. The sum of these observed values was subtracted from the expected value, the sample mean. Then, this difference was divided by the standard deviation to obtain standardized z-score, values for each site i. The z-score value from the

result decided the statistical significance of the clustered pattern of high or low attribute value. If there was spatial autocorrelation and clustering of high or low values in and around site i, the resulting value would be positive or negative depicting spatial patterning. A positive z-score value represents clustering of high values and a negative z-score represents clustering of low values around site i.

The locations i and j in the above equation are depicted by the geometric centroid of individual counties since the data were aggregated at a county level. The attribute values used at these sites were relative crash rates for each weather-type crash, which has already been defined earlier.

In any type of analysis of spatial autocorrelation, one of the most important questions is that of the conceptualization of spatial association among the features, or the construction of the spatial weight matrix $W(d)$. The value d is defined as the band distance which decides the extent of the neighborhood during calculations. In this research, that choice created a particular challenge as weather patterns or interactions are hardly consistent in terms of their size or movements. Sometimes a weather condition might only occur inside one county and will have no impact on surrounding areas. Other times the situation could be exactly the opposite and would occur in many counties at the same time. Moreover, the movement of weather conditions from one county to the other was not predefined. The spatial extent of weather events from year to year is also not fixed nor were they consistent. Because of such variability, a single decision about how weather effects change over locations is impossible.

The proposed choice for this analysis was a fixed distance band of 75,000 m with row-standardized weights (sum of row quantities in spatial weight matrix equals one). This distance band was selected because the minimum distance at which all counties had at least one neighbor was calculated to be 71,910 m. Any distance greater than that would result in a larger number of neighbors. Since at times, the influence of weather conditions does not even stray outside the boundaries of a single county, considering its effects on such a vast area was deemed rare and unique. Therefore, it was decided that a distance of 75,000 m would serve the best possible option.

Results

The results of Getis–Ord $G_i^*(d)$ analysis of snow- rain- and fog-related crashes for the 3-year period from 2000 to 2002 can be seen in Figs. 3–5, along with a summary statistics of Wisconsin counties provided in Table 2. These maps present a spatial clustering of high and low attribute value (relative crash rates for respective weather-related type crash) locations in different regions of the state. The $G_i^*(d)$ statistic gives negative spatial correlation and positive spatial correlation as clustering of low and high attribute values, respectively, as opposed to the general idea of clustering of similar and dissimilar values (regardless of the magnitude of the attribute value). Each county is represented by a standardized z-score value in four categories as calculated by the $G_i^*(d)$ statistic. The z-score values above +2 represents counties that lie in a cluster of high attribute value (relative crash rate) area, statistically significant at approximately 95% confidence level. This means that if the attribute values of the counties are randomly distributed among the 72 counties, the counties with a z-score value above +2 would consistently (more than 95% of the time) show above average percentage of weather-related crashes.

These counties cluster together in particular regions of the state and their location is consistent throughout the analysis period. Moreover, their location also overlaps the weather patterns which suggest weather influences crashes at these locations. On the other hand, counties with a z-score between +2 and −2 represent locations that may have a high or low relative crash rate value, but are not part of a statistically significant spatial pattern or cluster. There is a greater than 5% chance that such counties having a high percentage of weather-related crashes could be the result of unusual weather patterns or conditions for that particular year. These counties should be isolated and should not overlap the location of general weather patterns, and hence could be omitted as chance locations.

The spatial patterns of weather-related crashes were compared with annual snowfall and rainfall weather data in order to explain the occurrence of these patterns and to validate them. Weather data are presented in Figs. 1 and 2 in the form of continuous weather surfaces for comparison purposes. Fig. 6 shows the relative crash rates by county for rain-related crashes as an example to compare the patterns generated by simple plotting against statistically significant clusters. There are no clear or consistent patterns shown in Fig. 6. Moreover, it is difficult to decide what the cutoff number should be to identify rain-related crash prone areas. The use of cluster detection techniques (Getis–Ord $G_i^*(d)$) solves this problem by differentiating between statistically significant clusters and chance locations.

The results of the Getis–Ord $G_i^*(d)$ statistic present a very clear and distinct spatial pattern of clustering for each respective weather-related crash data analysis, thereby rejecting the null hypothesis ($\mathbf{H_0}$). Counties experiencing a higher percentage of snow-related crashes cluster in the northern regions of the state as shown in Fig. 3. The clusters tend to shift slightly between the years, but are generally consistent in their locations. Fig. 1, representing annual snowfall in centimeters, shows that the northern regions of the state experience more snowfall than the southern regions. Note that although the magnitude of snowfall changes considerably between the years as a result of varying intensity of winter for each year, the regions remain consistent.

Rain-related crashes also display spatial patterns of clustering as shown in Fig. 4, although these patterns are not as consistent as those for snow-related crashes. The patterns are mainly clustered around the southeastern region of the state moving along the coast of Lake Michigan. Over the years, the pattern extends northwards but is limited to the region close to the lake. The total annual rainfall surfaces in Fig. 2 show that between 2000 and 2001 the southeastern and southern regions of the state experience more rainfall than the northern regions. For 2002, the pattern of rainfall shifts completely in almost an east-west direction with northwestern and northeastern regions experiencing high and low rainfall, respectively.

The spatial patterns for fog-related crashes are also clustered predominantly in the southwest region of Wisconsin as shown in Fig. 5. Because of a lack of a sufficient number of weather stations collecting fog data, a continuous estimated surface through interpolation could not be generated. Moreover, since fog is a localized phenomenon, generalization over a large area would be incorrect. General knowledge indicates that fog tends to occur more frequently in valleys, rolling terrain, depressions, and near water bodies. Therefore, the patterns were overlaid with a hill shade map generated from a 1 km by 1 km digital elevation model (DEM) of Wisconsin. A hill shade map shows the nature of terrain and the variations in it. It can be observed from the hill
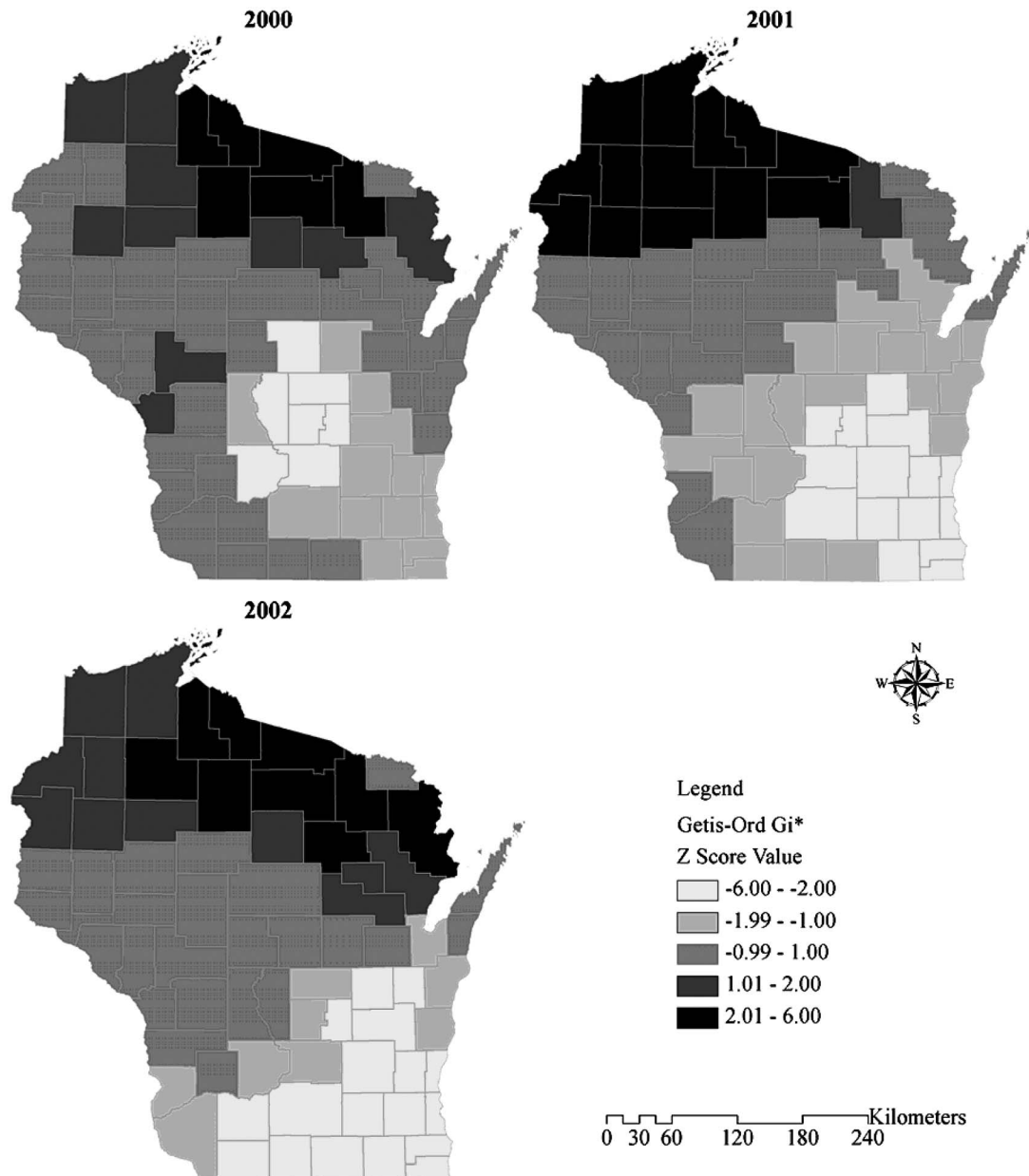
Fig. 3. Clusters of snow-related crashes in Wisconsin calculated by Getis–Ord $G_i^{\star}(d)$ statistic

shade map that the southwestern region shows greater variations and a rolling terrain as compared to the middle regions of the state which are more flat.

Discussion

A case has been presented for the use of spatial statistical techniques in analyzing crash data to examine spatial patterns in order to prioritize locations for road weather safety improvements. The results display statistically significant and consistent patterns of clusters for crashes of individual weather types over a 3-year period and justify the use of spatial statistical techniques. The observed clusters for an individual weather type are a strong indication that weather plays an important role in influencing crashes at these locations, since crashes are generally considered to be random events. The fact that a particular weather-type crash displays consistent patterns and lies in an area experiencing more

adverse weather establishes a strong connection between weather and crashes at these locations. A comparison with relative crash rates signifies that traditional crash rates or numbers cannot discern underlying spatial patterns detected by the use of spatial statistical techniques.

Spatial correlation has been used as a tool in this research to identify statistically significant locations. Although spatial correlation can be attributed to correlated but unobserved effects between counties, it is not an issue here because the objective is to identify weather-prone locations rather than the factors affecting weather-related crashes. Regardless of the factors, if there is positive spatial correlation (clustering of high attribute values), there are bound to be similarities between counties, which will help identify them as clusters of high crash locations. Negative spatial correlation would indicate a marked difference in characteristics between two counties (one having high and the other having low crash rates) suggesting they are different from each other. Hence those areas would not show up as clusters of high crash rates. It
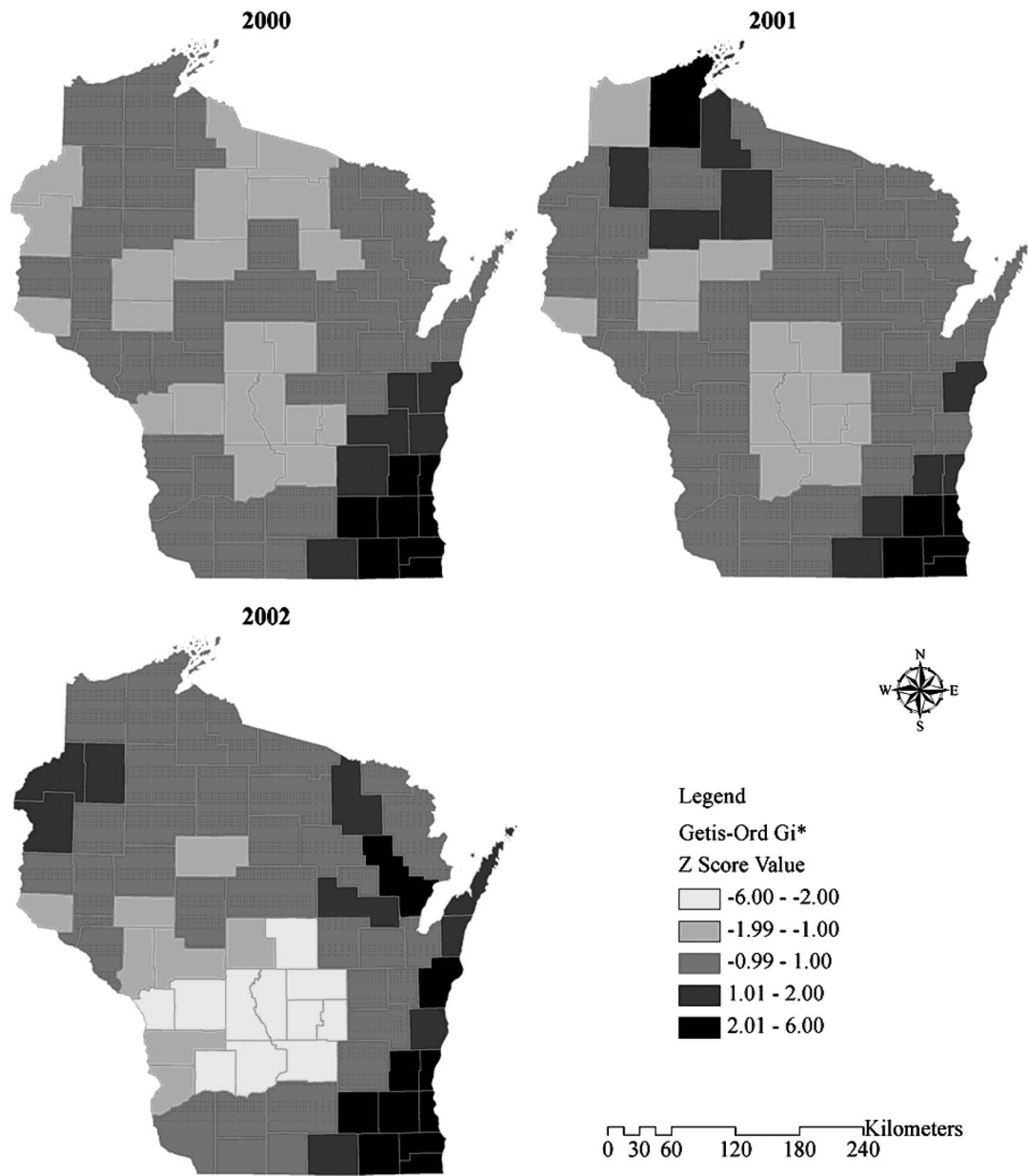
Fig. 4. Clusters of rain-related crashes in Wisconsin calculated by Getis–Ord $G_i^*(d)$ statistic

does not necessarily mean that weather does not affect crashes in the county with a higher percentage of weather-related crashes, but it could also be chance occurrence in the case where the road or other characteristics are similar between the two counties.

It is true that some drivers may change their driving behavior as weather conditions change. This may reduce the risk in adverse weather from a driver's standpoint. If that is the case, then theoretically those areas would show a smaller percentage of weather-related crashes and they would not be part of a cluster of high crash locations. On the other hand, if certain locations experience a higher percentage of weather-related crashes even with potential behavior changes, it could suggest that the interaction of roadway characteristics and adverse weather causes or contributes to the crash occurrence. These locations have been identified through this research for further in depth scrutiny into possible causal factors and mitigation efforts.

The reason for spatial clusters of snow-related crashes in the northern region is likely due to the fact that northern counties in

Wisconsin experience more snowfall and snowstorm events. Since the locations of these counties are also consistent over the years, this presents a clear relationship between snow-related crashes and snowfall.

Spatial patterns of rain related crashes indicate clustering in the southeastern region of the state. This can be attributed to the fact that weather data indicate greater rainfall in this region and hence more frequent events. For 2000 and 2001, the statistically significant counties showing a higher percentage of rain-related crashes are near the region experiencing the most rainfall. However, for 2002, this is not the case. A possible explanation for this may be the fact that the NWS COOP stations record precipitation accumulated daily. Precipitation includes the summed accumulation of rainfall and water equivalent of snowfall. Since it was impossible to determine the amount of rainfall on days experiencing rain mixed with snow, researchers included data for days on which only rain occurred. Nevertheless, the resulting similarities between areas showing a higher percentage of rain-related crashes
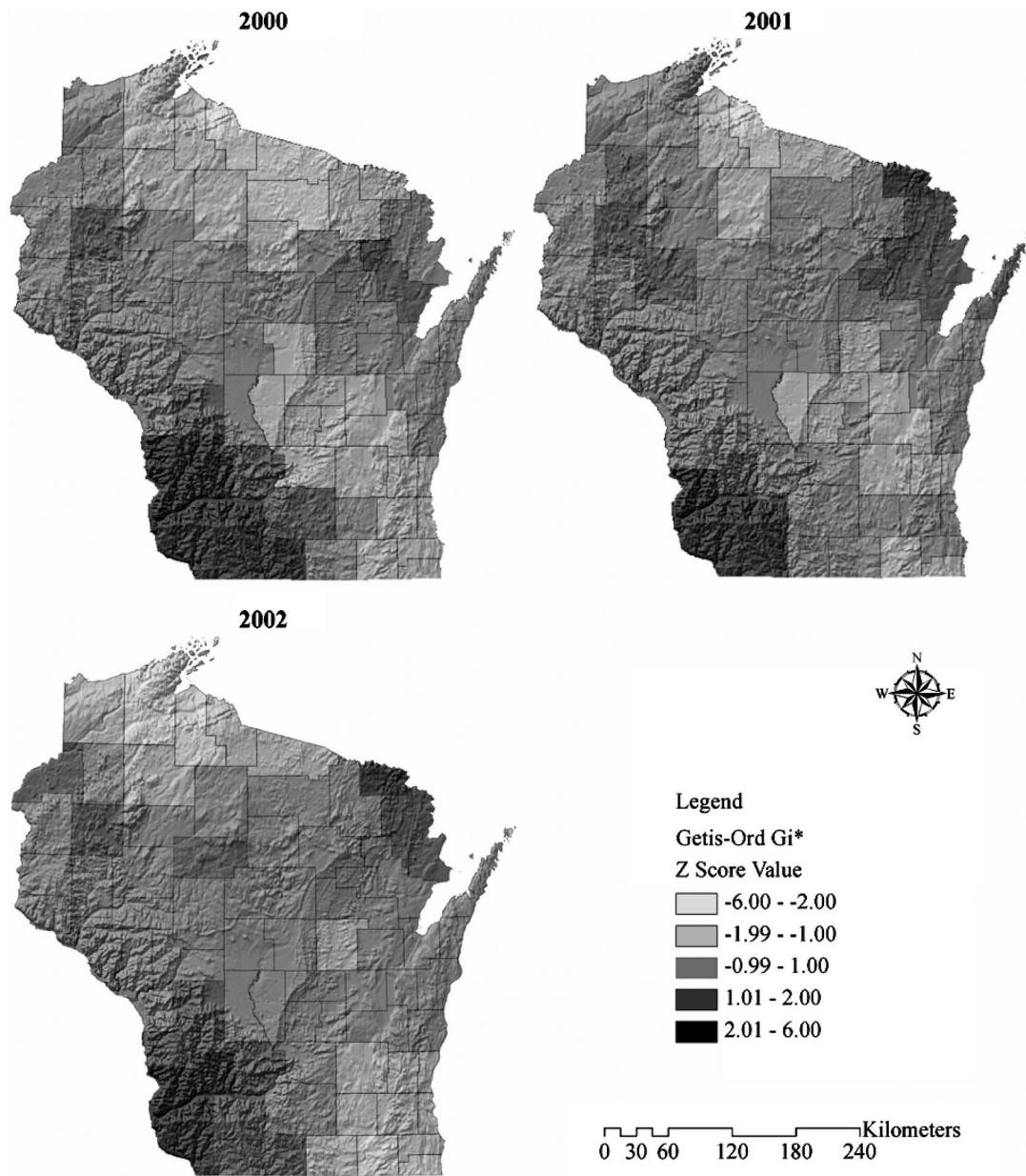
Fig. 5. Clusters of fog-related crashes in Wisconsin calculated by Getis–Ord $G_i^\star(d)$ statistic

and greater rainfall in 2000 and 2001 gives a strong indication of rainfall affecting these crashes.

The spatial patterns for fog-related crashes clustered in the southwest region because of the terrain with more valleys and depressions. These frequent variations in the topology are ideal locations for fog occurrences in the presence of moisture and optimum temperature conditions. The overall results of the comparison between spatial patterns of weather-related crashes and weather data present strong evidence to validate the spatial patterns.

Conclusions

Weather-related crashes aggregated on a county level have been analyzed to identify spatial patterning for different types of weather-related crashes. Fig. 7 summarizes the counties which

were calculated to be part of statistically significant clusters of high attribute values, for all 3 years, for each type of weather-related crash.

Based on the results and discussions presented, the alternative hypothesis of weather-related crashes showing clustered patterns is validated. This suggests that weather impacts the occurrence of crashes at those particular locations. The fact that clusters of areas experiencing a higher percentage of weather-related crashes overlap areas experiencing more rainfall or snowfall indicates that weather has some role to play in the occurrence of these crashes. If weather had no effect on these crashes, the observed patterns would be of complete spatial randomness (no patterns or clusters of similar attribute value locations) with no spatial autocorrelation. Moreover, if these patterns were a result of some other cause, they most likely would not be located in the same region showing a greater amount of adverse weather. The use of spatial statistical techniques as opposed to the use of crash rates and

Table 2. Summary Weather-Related Crash Statistics of Wisconsin Counties

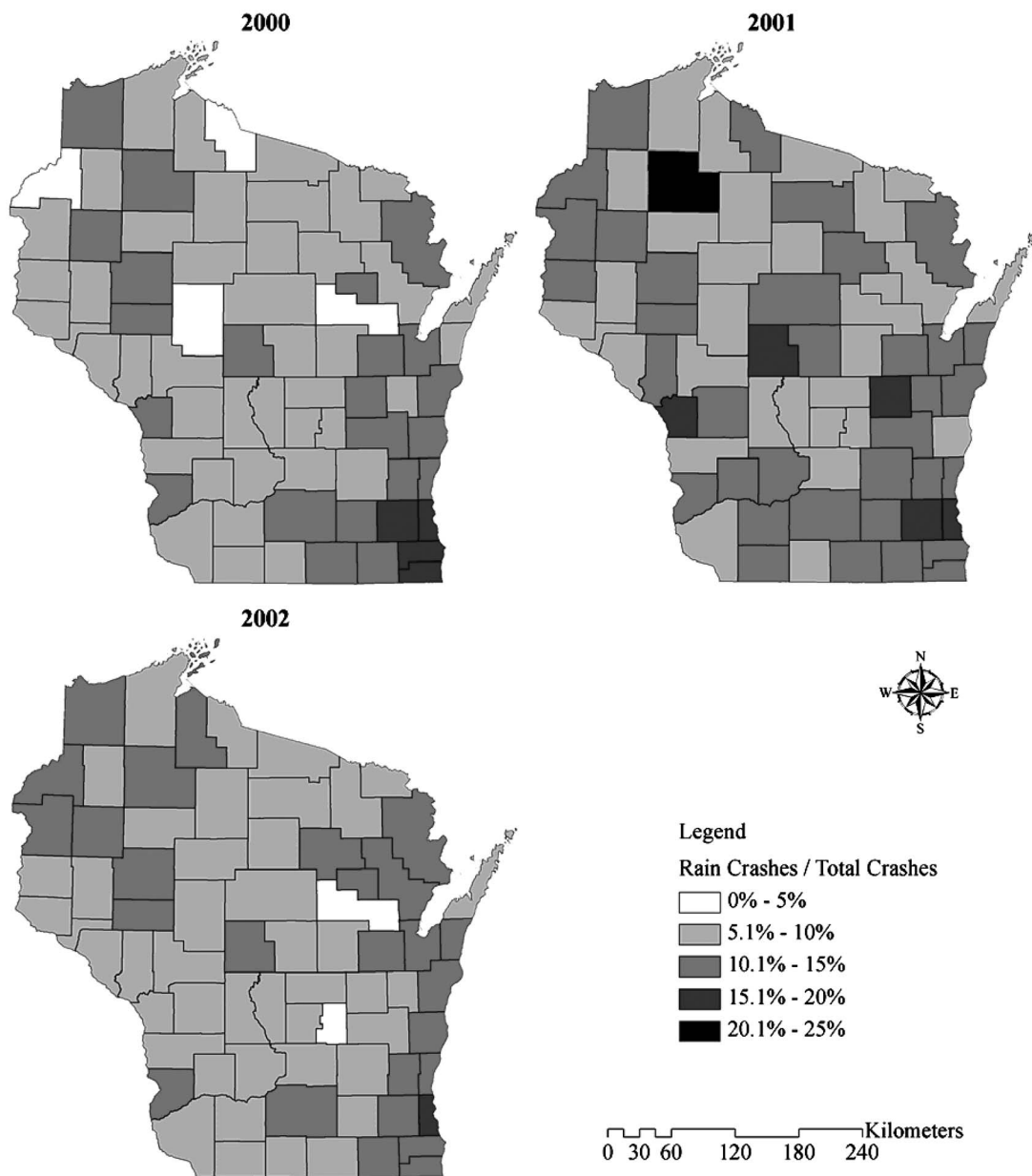|	Relative crash rate				$G_i^*(d)$ z-score			
Year	Mean	Std. error	Min	Max	Mean	Std. error	Min	Max
2000								
Snow	0.174	0.004	0.112	0.279	−0.146	0.179	−2.999	3.603
Rain	0.048	0.002	0.007	0.104	0.112	0.217	−1.727	5.209
Fog	0.009	0.001	0.000	0.027	−0.083	0.181	−2.304	5.561
2001								
Snow	0.050	0.003	0.015	0.146	−0.221	0.247	−2.453	5.296
Rain	0.061	0.002	0.028	0.103	0.010	0.133	−1.948	2.408
Fog	0.011	0.001	0.000	0.029	−0.101	0.138	−2.351	3.199
2002								
Snow	0.077	0.003	0.033	0.213	−0.138	0.219	−2.937	3.901
Rain	0.048	0.002	0.024	0.091	−0.020	0.205	−3.408	3.691
Fog	0.006	0.000	0.000	0.018	−0.051	0.152	−2.564	3.398

Fig. 6. Relative rain crash rates by Wisconsin county 2000–2002

Fig. 7. Counties with Getis–Ord $G_i^\star(d)$ z-score greater than 2 throughout 3 years

frequencies has also been justified through comparison with crash rate maps.

The fact that positive spatial autocorrelation has resulted in these observed patterns suggests that weather was a contributor to the higher number of crashes in these areas. What further strengthens this fact is the consistency in the location of these spatial patterns over the 3-year analysis period. Since the statistically significant locations of high relative crash rate values are consistent over the analysis period, this shows that weather impacts crashes at these locations. Furthermore, the identity also suggests that these counties should be the focal point of future analysis and weather-based safety countermeasures. The statistically significant locations of high relative crash rate values can be further analyzed to identify the underlying crash causal factors.

The results presented in this research provide a unique and effective methodology to assess road weather safety on a broader scale for planning and decision making purposes. This method provides the first step towards localized analysis to identify specific locations in terms of road segments or corridors for weather specific improvements and countermeasure implementation as part of a comprehensive road weather safety audit. The Wisconsin Department of Transportation is already working towards conducting road safety audits in areas identified through this and other subsequent research completed by the writers.

Notation

The following symbols are used in this paper:

d = fixed band radius distance measured in meters or kilometers;

$G_i(d)$ = Getis–Ord $G_i(d)$ statistic z-score value excluding value at site i;

$G_i^*(d)$ = Getis–Ord $G_i(d)$ statistic z-score value including value at site i;

N = number of weighted points, each representing relative crash rate for each county;

w_{ij} = spatial weight matrix for all sites j within distance d of site i;

x_i = relative crash rate of site i; and

x_j = relative crash rate of neighboring location j to site i.

References

Andrey, J., Mills, B., Leahy, M., and Suggett, J. (2003). "Weather as a chronic hazard for road transportation in Canada." *Natural Hazards*, 28, 319–243.

Andrey, J., and Yagar, S. (1993). "A temporal analysis of rain-related crash risk." *Accid. Anal Prev.*, 25(4), 465–472.

Anselin, L. (1995). "Local indicators of spatial association—LISA." *Geogr. Anal.*, 27, 93–115.

Brodsky, H., and Hakkert, A. S. (1988). "Risk of a road accident in rainy weather." *Accid. Anal Prev.*, 20(3), 161–176.

Cressie, N. A. C. (1993). *Statistics for spatial data*, Wiley, New York.

Flahaut, B., Mouchart, M., Martin, E. S., and Thomas, I. (2003). "The local spatial autocorrelation and the kernel method for identifying black zones: A comparative approach." *Accid. Anal Prev.*, 35(6), 991–1004.

Getis, A., and Ord, J. K. (1992). "The analysis of spatial association by use of distance statistics." *Geogr. Anal.*, 24, 189–206.

Goodwin, C. L. (2002). "Analysis of weather-related crashes on US Highways." Mitretek Systems, Inc., ⟨http://ops.fhwa.dot.gov/weather/best_practices/CrashAnalysis2001.pdf⟩ (Mar. 14, 2004).

Jones, A. P., Langford, I. H., and Bentham, G. (1996). "The application of K-function analysis to the geographical distribution of road traffic accident outcomes in Norfolk, England." *Schizophr Bull.*, 42(6), 879–885.

Khattak, A. J., and Knapp, K. K. (2001a). "Snow event effects on interstate highway crashes." *J. Cold Reg. Eng.*, 15(4), 219–229.

Khattak, A. J., and Knapp, K. K. (2001b). "Interstate highway crash injuries during winter snow and non-snow events." *Transportation Research Record. 1746*, Transportation Research Board, Washington, D.C., 30–36.

Kim, K. E., and Yamashita, E. Y. (2004). "Using K-means clustering algorithm to examine patterns of pedestrian involved crashes in Honolulu, Hawaii." *J. Adv. Transp.*, 41(1), 69–89.

Levine, N., Kim, K. E., and Nitz, L. H. (1995). "Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns." *Accid. Anal Prev.*, 27(5), 663–674.

National Highway Traffic Safety Association (NHTSA). (2003). "Traffic safety facts 2003." ⟨http://www-nrd.nhtsa.dot.gov/pdf/nrd-30/NCSA/TSFAnn/TSF2003.pdf⟩ (June 23, 2005).

National Highway Traffic Safety Association (NHTSA). (2004). "Traffic safety facts 2004." ⟨http://www-nrd.nhtsa.dot.gov/pdf/nrd-30/NCSA/TSFAnn/TSF2004.pdf⟩ (June 23, 2005).

National Weather Service Cooperative Observer Program (NWS COOP). (2006). "National weather service cooperative observer program." ⟨http://www.nws.noaa.gov/om/coop/⟩ (June 23, 2005).

Ord, J. K., and Getis, A. (1995). "Local spatial autocorrelation statistics: Distributional issues and an application." *Geogr. Anal.*, 27, 286–306.

Qin, X., Khan, G., and Noyce, D. A. (2006). "Spatial statistical approach

to identifying snow crash-prone locations." *86th Annual Meeting Transportation Research Board*, TRB, Washington, D.C.

Salmon, P. M., Regan, M. A., and Johnson, I. (2006). "Human error and road transport: Phase two—A framework for an error tolerant road transport system." *Rep. No. 257*, Monash Univ. Accident Research Center, Clayton, Australia.

Satterthwaite, S. P. (1976). "An assessment of seasonal and weather effects on the frequency of road accidents in California." *Accid. Anal Prev.*, 8(2), 87–96.

Shankar, V., Mannering, F., and Woodrow, B. (1995). "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies." *Accid. Anal Prev.*, 27(3), 371–389.