Building a Spatial Crash Information System:  
Examples from the Houston-Galveston Metropolitan Safety Planning Program  
Ned Levine, PhD  

Abstract  

This paper discusses the development of a spatial crash information system, which includes the attribute characteristics of crashes that are usually taken from crash reports, but combines it with a GIS interface and a series of spatial analysis tools. The advantage of this system is that it can conduct the usual types of analysis of crashes typically conducted by State and local departments of transportation, such as the analysis of drivers or vehicle characteristics and the enumeration of the number crashes at particular locations or road segments, but can also conduct a variety of spatial analyses that are difficult, if not impossible, with a tabular system. These include spatial selection, hot spot identification and analysis, and crash risk interpolation. These types of analysis can be conducted on selected road segments or on an entire metropolitan area, something which is impossible with a tabular data base. Examples from the Houston-Galveston Metropolitan Safety Planning Program are taken to illustrate such a system. Using their system, the Houston-Galveston Area Council identifies hot spots that form the basis for safety engineering studies as well as conducts corridor analysis to identify high risk crash hot spots. Finally, the paper discusses the need to improve spatial accuracy of information. Eventually, it will be possible to conduct the preliminary engineering analysis of hot spots throughout a region, simultaneously while being identified, once the spatial accuracy of crash locations is improved.
In this paper, I want to discuss the use of a spatial crash information system and illustrate how it can be used to facilitate safety planning. In doing so, I will illustrate its use with examples from the Houston Metropolitan Safety Planning Program (http://www.h-gac.com/safety).

The Need for Crash Information

Safety planning is a data-driven activity in that any public agency wishing to address safety needs to have information about safety. Whether the agency is a State or local department of transportation, a local police department, or a county health agency, having information about motor vehicle crashes is essential to making informed decisions.

The usual basis for such information is the crash report that is collected normally by local police departments. In virtually all States, crashes have to be reported to the local police department, usually using a standardized reporting form. These forms are then entered into a local data base. Again, in virtually all States, copies of many of those forms are sent to a State agency, which compiles data from every jurisdiction in the State. These records form what is known as a Crash Records System, a collection of information about motor vehicle crashes that can be used for a variety of purposes, such as for planning, implementation or legal cases.

Whether all or some of the forms are sent to the State agency depends on the unique reporting criteria. For example, in Texas, all crashes involving fatalities, injuries, or property damage in excess of $1,000 is required to be reported to the local police departments. For the State agency that compiles the data across all jurisdictions, the Crash Records Bureau of the Texas Department of Public Safety, there is a slightly different reporting requirement: all crashes involving fatalities or injuries must be reported, but property damage only crashes in which one or more vehicles was towed need be reported. Thus, non-injury crashes where all vehicles are driven away from the scene (the so-called ‘fender bender’) are not reported to the State agency.

The Federal Highway Safety Improvement Process

Irrespective of which criteria are used to report crashes, these data can be used for planning purposes. They can be used to identify the characteristics of drivers or victims or used to identify the locations where crashes occur. For example, for improvements to the roadway, most States follow the Federal Highway Safety Improvement Process (known as HSIP; FHWA, 1981a; 1981b). This is a four-step process that involves 1) collecting crash data on road locations in the State, including those at railroad-highway grade crossings; 2) analyzing available data to determine where the hazardous locations or road segments are found; 3) conducting an engineering study to determine crash patterns at these locations; and 4) establishing priorities for implementing highway improvements based on a benefit-to-cost analysis. The legislation for this came from a series of acts that were enacted by Congress from 1966 onward.
However, whether the crash data that have been collected for the purpose of identifying characteristics of drivers or victims, or whether the data have been collected to identify hazardous locations, it is important that the data be accurate and relatively comprehensive.

**Advantages of a GIS-Based Crash Information System**

It is my contention that a crash information system based on a Geographic Information System (GIS) represents a tool that can advance safety planning (AECOM et al., 2002). GIS-based crash information systems have been around since the 1990s. We developed one of the first systems in Honolulu in 1994 and 1995 with a metropolitan-wide system (Kim, Levine & Nitz, 1995; Levine, Kim, & Nitz, 1995a; 1995b; Kim and Levine, 1996). Since then, many jurisdictions around the country have developed similar systems, such as Detroit (through the Southeast Michigan Council of Governments), New York City, New Jersey, and Greenville (SC). In Texas, the cities of Austin and Sugar Land have built such systems.

**Inventorying Crashes at Intersections**

Still, the slow diffusion of GIS to crash analysis has limited many States and local governments in planning for safety improvements. Traditionally, State DOT’s and local government have used a tabular approach to identify hazardous locations. Typically, a location will be pre-determined and the crash information system will be searched to locate all records that document crashes at that location. Sometimes this process can be quite tedious since many local governments catalog crashes by the block. For example, to locate the crash records for the intersection of Westheimer and Fondren (to use a Houston example), it is necessary to search for the block on Westheimer to the west of Fondren, the block on Westheimer to the east of Fondren, the block on Fondren to the north of Westheimer, and the block on Fondren to the south of Westheimer.

A GIS-based crash information system can circumvent such a procedure since it possible to select all crashes that occurred at that location with the click of a mouse. While this is an improvement in speed, the tabular database can still produce the results. However, a GIS-based crash information system can accomplish a lot more than a tabular database. For example, for any complex selection, a tabular can be queried with respect to the characteristics. For example, to select all crashes where speeding was a factor, the tabular database can find those records, sort then, and output them; the speed depends on the efficiency of the database.

**Attribute Selection of Crashes**

A GIS-based system can do the same thing since its attribute database is essentially tabular. But the linkages of the database to spatial location (the geographical display part of the program) can allow much more complex relationships to be identified. For example, to find all speeding crashes that occurred within a quarter mile of a particular intersection, in the GIS program one would need to specify a central location –
the intersection, and a search radius and select out all records that fell within that search radius. This operation will take, at most, a minute to be implemented. To do that operation in a tabular database would require identifying all blocks that fell within a quarter (measuring it manually on a map) and then find the records associated with the blocks. It is a very tedious and slow process.

**Hot Spot Analysis of Crashes**

Beyond complex selection, a GIS-based crash information system can identify relationships that are almost impossible to determine with a tabular database. This makes a GIS-based crash information system the basis of a spatial information system. A spatial information system combines several elements: 1) the tabular database traditionally associated with crash analysis; 2) a GIS which links the crash records to a specific location; and 3) a set of spatial analysis tools that extract information from the records based on their spatial location. By taking advantage of the spatial nature of the data, usually contained in a GIS program, the spatial information system provides insights into crashes that can aid planning.

One simple example is hot spot analysis. A hot spot is a concentration of crashes that occurs at or near a location. Sometimes, a distinction is made between a hot spot location and a hot spot area (Block and Block, 1999). A hot spot location is a specific location (e.g., an intersection) where many crashes occur while a hot spot area is a small area where there is a concentration of events. To conduct this analysis with a tabular database is almost impossible to do. With the tabular database, one can choose a location or set of segments and identify the number of crashes that occurred there, but one cannot do the opposite and search the database for locations where there are many crashes. With a spatial information, one uses software that can count the number of events at each location or examine clusters of incidents that occur near each others, for example SatScan or CrimeStat (Kulldorff, 1997; Levine, 2004). Hot spot identification is an important planning tool for safety since it identifies where priorities need to be established by the State DOT or local government (step 1 in the HSIP mentioned above).

**Crash Risk Analysis**

Another example illustrates the power of a spatial crash information system. Crash risk is an important property of safety planning. Usually crash risk is defined as the number of crashes relative to the baseline traffic (or sometimes population). A particular road may have a large number of motor vehicle crashes, but if it is also a high volume road, then the higher number of crashes would be expected. For example, freeways have typically much higher crash numbers. However, since they also carry very large volumes, their crash risk is frequently much lower than with major arterials. Conversely, some arterials have a limited number of crashes. However, compared to their traffic volume, they have a high crash risk. In short, crash risk is a useful measure in determining where priorities need to be placed.
Now let’s imagine trying to analyze crash risk on an arterial road with both a tabular crash database and with a spatial information system. In either case, two separate databases have to be used – a crash database and a database of traffic volume (e.g., vehicle miles traveled, VMT). Without a spatial crash information system, the process involves at least four steps. First, the segments have to be selected. Second, the number of crashes on each segment must be determined. Third, the VMT of each segment has to be obtained from the VMT database. Fourth, for each segment, the number of crashes are divided by the VMT, and then usually multiplied by a large number to get rid of decimal places (traditionally, 100 million).

With a spatial crash information system, the same process can be followed. But a simple procedure can allow a quick approximation to be produced. The spatial information system can, first, interpolate crashes to a grid; second, interpolate vehicle miles traveled (VMT) to the grid; and, third, divide the number of crashes by the VMT for each grid cell. The advantage is that this can be applied to either the specific arterial or to the entire region. That is, one can do this for an entire region at the same time, irrespective of how many arterials are in the region. There are various software packages that can do such an approximation, for example Spatial Analyst and CrimeStat (ESRI, 1997; Levine, 2004). But, such a process is impossible in a tabular database.

There are many other operations that can be conducted with a spatial crash information system which are difficult, if not impossible, with a tabular system, such as examining spill-over effects of crashes along a road or in an area, identifying higher-level hot spots such as neighborhoods or sub-regions, or examining relationships between the number of crashes at a location and factors contributing to the analysis (spatial relationships).

The Houston-Galveston Metropolitan Safety Planning Program

To illustrate the value of a spatial crash information system, let’s look at the Houston-Galveston Metropolitan Safety Planning Program. The Houston-Galveston Area Council (H-GAC) is the Metropolitan Planning Organization for the Houston-Galveston metropolitan area, an eight county area that has currently more than 5 million persons. Figure 1 shows a map of Texas counties with the eight counties outlined while figure 2 shows a map of the Houston-Galveston metropolitan area with an outline of the City of Houston, the fourth largest city in the United States.
**Geo-coding crashes**

H-GAC started its safety program in 2001. The first step was to build a GIS-based crash information system. We followed a process that was similar to the one developed in the 1990s in the Honolulu study (Levine and Kim, 1999). We obtained data from the Crash Records Bureau of the Texas Department of Public Safety (DPS). Their data was kept in a very old information system that coded information using the old 80 column punch card format. Like most crash records systems, there were three types of records: 1) a crash record that gave information about the location and time of the crash as well as its characteristics; 2) a driver/vehicle record that gave information about the vehicles and drivers involved for each vehicle involved in the crash. For pedestrian and bicycle crashes, there were separate records also provided; and 3) a victim record that provided information about each non-driver victim of the crash.

The geographical location of the crashes was not very well defined. This information was listed in several ways. The county and city in which the crash occurred were listed as was the type of roadway (e.g., interstate, city road). For the actual road on which the crash occurred, there was a five digit field which identified the name of the road. For the cities of Houston and Galveston, fortunately, there was a five digit numerical code that each city had used to identify the specific road name. However, for the other non-State roadways, only the first five digits of the street name were provided. For the State roads, information was provided using a State roadway identifier system (roadway, control number, section number, mile marker). In addition, information was provided about whether the crash occurred at an intersection or not as well as information about the direction of the vehicle that caused the first impact.

To geocode the crashes, we broke the records into 13 separate files. The cities of Houston and Galveston were geocoded as separate files since there were numerical codes for street names. The cities of Pasadena (TX) and Texas City were also coded as separate files since there was additional information provided for crashes within these cities. A separate file was created for crashes occurring on the State road system that were identified by State roadway identifiers. Finally, separate files were created for each of the eight counties for those crashes not occurring in the above mentioned file; within each county, we coded each city separately.

The geocoding was done with two sets of files. The first were SAS files that translated five digit letters or numbers into proper street names. As mentioned, there were 13 separate SAS programs written. In all, more than 110,000 lines of code were written to translate the letters or names into proper street names. The second set of files was standardized GIS base maps from which the street names are linked to spatial locations. We used five different digital base maps (layers) that corresponded to the different SAS files – an accurate street map maintained by H-GAC and the surrounding cities and counties; Census Bureau TIGER files for two of the counties; a freeway layer for crashes occurring on freeways; and a State roadway identifier file. To do the geocoding, we used two different GIS programs: ArcView (ESRI, 1998a) and Atlas*GIS (ESRI, 1998b).
Overall, we were able to geocode 81% of the crashes. Much of the non-coverage came from non-intersection crashes in the smaller cities of the region (see endnote 2). We tested accuracy by drawing a sample of 500 geocoded crashes and then checking whether the geocoding algorithms had correctly placed the record in the correct location. Using these tests, we estimated that the geocoding had an overall accuracy of 90%. Both the matching percentage and the accuracy increased toward the center of the region. For example, with City of Houston crashes, we were able to geocode 90% of the crashes with 94% accuracy.
Initially, we prototyped the system on a single year’s worth of data (1999). Then, we extended the geocoding to cover all released information by DPS from 1998 till 2001, the latest year for which they have released information. As might be expected, the initial prototyping took a long time to accomplish (close to a year and half). However, subsequent geocoding occurred much faster due to the accumulated information in the conversion files. We were able to code 1998, 2000 and 2001 in about three months.

Analyzing Safety with the Spatial Information System

The Houston-Galveston area has a severe safety problem, the worst in Texas and among the worst in the nation. Figure 3 shows a map of 252,240 serious crashes that occurred in the Houston-Galveston area between 1999 and 2001, an average of 84,080 serious crashes a year. Over the three year period, there was an average of 627 fatalities and 93,971 injuries each year from these crashes. We estimated that the likelihood of a driver being involved in a fatal or injury crashes was 36% higher than the State of Texas average and 149% higher than the U.S. average for the three year period. As might be expected the region has more crashes of virtually every type than any other region in the
state – fatal crashes, injury crashes, pedestrian crashes, bicycle crashes, and commercial motor vehicle crashes.

Using these data, we can conduct various types of analysis, both traditional and spatial. For example, figure 4 shows a map of fatal crashes in the region for the three year period while figure 5 shows a map of alcohol-related (Driving-Under-the-Influence – DUI) crashes and hot spots. The hot spots were identified by the CrimeStat program using a nearest neighbor hierarchical clustering routine (Levine, 2004).

**Crash Analysis of Roadways**

One of our major outputs from the spatial crash information system is the analysis of safety along various roadways. We are continually conducting analysis of roadways, usually in response to a request from a city, county or the Texas Department of Transportation. For example, figure 6 shows a map of the 1998 to 2000 crashes along Kirby Drive, a major arterial in the City of Houston. The arterial has a very high crash risk (433 serious crashes per 100 million VMT compared to 247 for Harris County – the central county in the region, or 233 for arterials in the State of Texas). Figure 7 shows an interpolation of crash risk along Kirby Drive. The interpolation was again done with the CrimeStat program. As seen, even though crash risk is high for the arterial, there are certain locations where it is particularly high. These are high risk hot spots.

Figure 3
Figure 4

Safety in the Houston Metropolitan Region: 1999-2001
Location of Fatal Crashes

Legend
- Freeway
- State road
- Water
- Region
- Counties

84.5 Miles
Figure 4
Figure 5

DUI Crashes and Crash Hot Spots: 1999-2001
Location of Crashes and Hot Spots
Figure 6

Safety on Kirby Drive: 1998 to 2000
Location of Crashes
High Risk Locations on Kirby Drive: 1998 to 2000
Crashes Per 100 Million Vehicle Miles Traveled
**Safety Engineering Studies**

One very useful application of this methodology is the identification of hazardous locations as a basis for conducting a detailed safety engineering study, following the HSIP methodology discussed above. Using these data, we have conducted five studies of hazardous locations in four cities within the region (two in the City of Houston, and one each in the cities of Pasadena, Galveston, and Sugar Land). In each study, we formed a partnership with the City and identify which locations will be studied. We then put out a Request for Proposal and contract with a consulting firm to do the analysis. The firm then examined in detail the crashes using both our data and the most recent data from the City. For each intersection, a pattern of crashes was identified and several mitigation measures were proposed. Using a standard benefit-to-cost (B/C) methodology specific to the State of Texas, the consultant prioritized improvements and ranked them based on the B/C ratio. Figure 8 below the location of five intersections in the City of Pasadena that were studied. Every study that we conduct is posted on our safety web page in addition to being given to the City (see http://www.h-gac.com/safety).

**Conclusion**

There are many other applications with which we use our spatial information system, both traditional and spatial, for example the analysis of behavioral factors associated with crashes and the spatial modeling of crashes to identify congestion bottlenecks. However, these are the main ones.

I’ll finish with two points. First, a spatial crash information system, which relies on a GIS program, can provide insights and geographical specificity which are very difficult to achieve with a traditional, tabular crash records systems. The spatial component of the data base allows the explicit analysis of spatial relationships whereas the tabular data base can only do this indirectly. In a sense, a GIS-based crash information system involves both the traditional tabular data base as well as information about the spatial location of the incidents. It is this latter type of information which opens up many applications. The use of spatial analysis tools make explicit these spatial relationships and covert the records system into a spatial information system.

Second, on the other hand, the accuracy and, therefore, utility of the spatial crash information system does depend on the accuracy of the original data. Unfortunately, the accuracy of the existing State records system is fairly primitive. As mentioned above, we have been only able to match about 80% of the total records with about 90% overall accuracy. Further, the spatial accuracy of the matches is only to the nearest intersection (for intersection crashes) or to mid-block.

To improve beyond this accuracy will require improvements in both the geographical information provided on the State records as well as improvements in the identification of the crash location. For the former, the Texas Department of Public Safety is currently updating their records system to produce a more modern data base. We expect that many of the coverage issues will disappear once that system is operating.
(hopefully by early 2007). However, even that system will not provide detailed spatial accuracy for the exact location of the crash. For that to happen would require the use of Global Positioning System (GPS) receivers by investigating police office to accurately record the location of impact of the crashes.

Thus, the new frontier for crash information systems needs to be the establishment of very accurate information about the location of crashes (most likely using GPS receivers). Once that type of information is recorded, it will be possible to conduct preliminary safety engineering studies at the same time that the safety planning analysis is being conducted. There is software available for creating collision diagrams and pattern analysis based on the crash records (e.g., see CTRE, 1997; JMW, 2004; Arc, 2006). It will then be possible to identify hot spots and conduct a preliminary engineering assessment in a comprehensive and efficient manner, compared to the piecemeal way we do it now.

Figure 8
REFERENCES


http://www.accidentreconstruction.com/newsletter/nov02/cadzone.asp


http://www.ctre.iastate.edu/research/detail.cfm?projectId=112


http://www.jmwengineering.com/


Author Information

Ned Levine, PhD
Transportation Safety Coordinator
Houston-Galveston Area Council
3555 Timmons Lane, Suite 120
Houston, TX 77227
(713) 993-2457
(713) 993-4508 fax
Ned.Levine@h-gac.com
ENDNOTES

\(^1\) While Federal aid for highway safety goes back to the 1916 Federal-aid highway program, the initial piece of modern legislation was the Highway Safety Act of 1966 (23 U.S.C. Chapter 4). This act requires each State to have a highway safety program, approved by the Secretary of Transportation, designed to reduce traffic accidents and deaths, injuries, and property damage (NHTSA, 1998). The program is part of Title 23 (23 U.S.C. 402) and is referred to as the Section 402 Highway Safety Program. In its current form, this program sets forth the minimum requirements with which each State’s highway safety program must comply. It is administered jointly by NHTSA and FHWA. The program funds non-construction projects aimed at the reduction of injuries, deaths, and property damage from motor vehicle crashes. These include the development or upgrading of traffic record systems, the collection of crash data, traffic engineering studies, the development of technical guides for States and local highway agencies, work zone safety projects, the encouragement of the use of safety belts and child safety seats, roadway safety public outreach campaigns, enforcement to reduce impaired drivers, enforcement programs to combat drivers who speed or drive impaired, and the enforcement to reduce aggressive driving (FHWA, 1999).

Another Federal mandate for the safety of the roadways came with the Highway Safety Act of 1973. This law requires that each State conduct and systematically maintain a survey of all highways to identify high-hazard locations that may constitute a danger to vehicles and to pedestrians. The law establishes a benefit-cost methodology for identifying safety project locations and establishes a means for assigning priorities. Again, like the Highway Act of 1966, the 1973 Act provided mandates for States, a systematic approach towards safety improvements, and an ear-marked funding source for safety improvements. States could not use the excuse of lack of funding to avoid having to improve the safety of the highways.

The initial legislation was to make available construction funds for roads on the Federal-aid system other than a highway on the Interstate System, which was covered under other legislation. This program came to be known as the Section 152 Hazard Elimination Program (or HEP) and is part of Title 23 (23 U.S.C. 152). Also, the 1973 Act created a separate railway/highway crossings program, Section 130 (23 U.S.C. 130). Ear-marked funds were created for each of these programs separately.

In 1982, an amendment to this legislation was enacted that extended the Section 152 authorization to all public roads, except for the Interstate Highway System which had its own authorization (Highway Improvement Act of 1982). In 1991, the re-authorization extended the program to the Interstate Highway System and in 1998 to bicycle paths.

With the current re-authorization of the Federal transportation bill that occurred in 2005 (SAFETEA-LU), the 402 program remained but the 152 and rail-highway grade crossing programs were combined into a single funding category. As part of the 2005 legislation, each State is required to develop a Strategic Highway Safety Plan as a framework for determining investment priorities in safety.
For each record, if the crash occurred at an intersection, the record system provided information about both intersecting streets. For mid-block crashes, however, information about the nearest intersection was only provided for four cities: Houston, Galveston, Pasadena, and Texas City. Consequently, for mid-block crashes on other cities, we were unable to geocode the crashes since we had no information that could identify where the crash occurred. This is the major source on non-matching records.