System for Digitizing Information on Wisconsin's Crash Locations

Arup Dutta, Steven Parker, Xiao Qin, Zhijun Qiu, and David A. Noyce

One of the key hurdles in identifying unsafe intersections and roadways in Wisconsin is the lack of a complete crash location map, especially for crashes that occurred on local streets. Crash locations are reported in terms of relative offset from an intersection on the basis of on- and at-street name information, which identifies the intersection, and direction and distance information, which identifies the offset. For intersection crashes, the offset distance is typically set to zero. As described in this paper, the Traffic Operations and Safety Laboratory at the University of Wisconsin, Madison, has developed a system to automate the mapping of Wisconsin local road crash locations. The location mapping algorithm involves the integration of two separate Wisconsin Department of Transportation databases: the Wisconsin crash database of police traffic accident reports and the Wisconsin Information System of Local Roads (WISLR). The application of WISLR, which is an inventory of local roads with details such as traffic information, pavement condition, and roadway geometry, provides invaluable access to more comprehensive safety analysis. Although the methodology introduced is specific to these two databases, the general ideas can be applied to any similar sets of crash and geographic information system databases. The final result is a pinpoint map of all the intersection and segment crashes that occurred on local roads in Wisconsin, along with the complete crash information associated with each mapped crash. The algorithm developed with this methodology is able to map approximately 79% of the intended pool of available crashes. Quality evaluations indicate that the mapping is almost 98% accurate.

During the 10-year period from 1996 to 2005, there were on average more than 130,000 police-reported crashes annually on public roadways in the state of Wisconsin (*1*). Of those, nearly 5,500 crashes per year resulted in death or serious injury. The need to mitigate crash hazards continues to be an important goal for Wisconsin transportation planners and engineers. Knowledge of both where an incident occurs and how the chain of events took place is necessary to improve the design and operation of intersections and minimize the consequences of a traffic accident. In general, however, there is no reliable system in place at this time to map all Wisconsin reported crashes onto a geographic information system (GIS) digital map for crash safety analysis.

In 1999, the Wisconsin Department of Transportation (WisDOT) Bureau of Traffic Safety (BOTS) secured federal Section 411 incentive grant funding to perform a traffic records assessment that supported the development and activities of a State Traffic Records Coordinating Committee (TRCC). In 2002 and 2003, TRCC gave top priority to automating crash records and improving location data. The Department of Motor Vehicles (DMV) during the same time period successfully implemented the National Model Traffic and Criminal Software (TraCS) for internal entry of Wisconsin motor vehicle accident reports (MV4000). The DMV began pilot-testing the new crash data collection system with four law enforcement agencies in September 2004.

DMV and BOTS have taken important steps to accelerate the deployment of TraCS; however, the automation of crash data and location collection cannot take place in the short term because not every police vehicle is installed with a Global Positioning System (GPS) unit. Moreover, there is no plan in the near future to equip every vehicle in Wisconsin. As a result, it is clear that the process of automating crash location data collection and follow-up analysis is a long-term prospect that will take several years to implement fully. Of additional concern is the fact that current deployment of the new technologies, software, and recommendations to the MV4000 does not affect historical crash data, information that may be extremely valuable in the understanding of crash locations and patterns.

As a reaction to these needs, WisDOT invested in the process of hand-coding all crashes that occur on state-managed highways to a crash reference point and relative distance system for GIS mapping. However, there is still no ready way to map crashes that occur on local roads. One solution—which is the one adopted by this project—is to develop an automated system.

Over the years, researchers have attempted to develop, analyze, and disseminate crash-related geocoding procedures and digital maps to suit the needs of their applications (*2–4*). The current study adds to the existing literature for developing methods to digitize historical crash data into geospatial maps. The purpose of this study is to describe an algorithm and software tool developed by the Traffic Operations and Safety (TOPS) Laboratory at the University of Wisconsin, Madison, to automate the process of digitizing Wisconsin local road crash information on a GIS map. In particular, the TOPS project's objectives were to

- Develop an algorithm for automating the crash location data mapping process for crashes occurring on local roads with respect to existing WisDOT base maps and crash forms; the algorithm would translate location information from a database of police crash reports to a geospatial map and create a pinpoint map from the crash information;
- Conduct a quality check for accuracy of the mapped crash locations for a sample subset of crash data; and
- Determine the drawbacks, potential improvements, and recommendations based on the results obtained.

Department of Civil and Environmental Engineering, University of Wisconsin, 2205 Engineering Hall, 1415 Engineering Drive, Madison, WI 53706. Corresponding author: A. Dutta, arup@cae.wisc.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2019, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 256–264.
DOI: 10.3141/2019-30

DATA SOURCES

There are two primary data sources for this project: the WisDOT crash database of police-reported crashes and the Wisconsin Information System for Local Roads (WISLR) GIS of geospatial information for all local roads in Wisconsin.

WisDOT Crash Database

Wisconsin traffic crashes are, by statutory definition, reportable if someone is killed or injured or if the property damage exceeds a certain threshold ($1,000 for property-related crashes or damage to government-owned vehicles and $200 for all other government-owned property, such as traffic control devices). Crash information is generally reported by a dispatched police officer via the Wisconsin MV4000 police form and is eventually archived in the WisDOT DMV crash database. The WisTransPortal data management system at the TOPS Laboratory contains a local copy of all crashes in the DMV database for the years 1994–2005. For the purposes of algorithm development, analysis, and quality evaluation, this project focused on crash records for the city of Madison for 2003, downloaded from the WisTransPortal system. Figure 1 shows a subset of the records (rows) and fields (columns) from the WisTransPortal crash database. The definition for each field is as follows:

- Accident-number: computer-system-generated number to uniquely identify a crash (this column is not shown in Figure 1 because of privacy concerns);
- On-street: name of the local street on which a crash took place;
- At-street: name of the street that intersects with the street on which a crash took place;
- At-highway: name of the intersecting or nearest highway on which a crash took place;
- Refpoint-number: state trunk highway reference point number where a crash occurred;
- Accident-location: type of location at which a crash occurs (public road intersection, public road nonintersection, parking lot, private property);
- Intersection-direction: direction from the listed intersection;
- Intersection-distance: distance from listed intersection location in hundredths of a mile;
- Municipality-code: Wisconsin municipality in which a crash occurred; and
- On-highway: name of the highway on which a crash occurred.

WISLR Software

The WISLR is a GIS-based software package developed and maintained by WisDOT that combines data for local roads in Wisconsin with interactive mapping functionality. The database associated with WISLR contains roadway- and intersection-specific information that is used to build the geospatial map. According to WisDOT (*5*),

> With WISLR, users can produce maps that show the location of road-related data and see trends that might otherwise go unnoticed. For this reason alone, WISLR aids with organized and logical assessments about local road data. This is just one example of what WISLR can do—and there are many other benefits.

WISLR was chosen for this project, as it is the official local roads' GIS database in WisDOT, and in addition it provides an opportunity to link important physical roadway characteristics information to the crash reports for safety engineering analysis.

Figure 2 shows a portion of the city of Madison local roads map clipped from WISLR. It shows the nodes (intersections) and links

On-Street	At-Street	At-Highway	Refpoint-Number	Accident-Location	Intersection-Direction	Intersection-Distance	Municipality-Code	On-Highway
PARKING LOT	POST RD			PL		0	1373	
UNIVERSITY AVE	N RANDALL AVE			I		0	1373	
S PARK ST	W BADGER RD			I		0	1373	
ACEWOOD BLVD2005	GOLDFINCH			N	S	1	1373	
N MILLS	UNIVERSITY AVE			N	S	1	1373	
W JOHNSON ST	WISCONSIN AVE			N		0	1373	
GAMMON PL				N		0	1373	
SUDBURY WAY	LAMPLIGHTER WAY			N		0	1373	
MARQUETTE ST	HAUK ST			I	E	0	1373	
		12	244	N		0	1373	39
FREEPORT RD	VERONA FRONTAGE RD			N	W	1	1373	
PARKING LOT	JOHN NOLEN DR			PL		0	1373	
MINERAL POINT RD	S HIGH POINT RD			I		0	1373	
PARKING LOT	JOHN NOLEN DR			PL		0	1373	
GORHAM ST	WISCONSIN AVE			I	W	0	1373	
WHITNEY WAY		12	340	I		0	1373	
N BROOM ST	W MIFFLIN ST			N		0	1373	
E WASHINGTON AVE	BALDWIN ST		73	I	E	1	1373	151
S STOUGHTON RD	E BUCKEYE RD			N	S	2	1373	
PARKING LOT	EAST TOWNE MALL			PL		0	1373	
OHMEDA DR	FEMRITE DR			I		0	1373	
E WASHINGTON AVE	N STOUGHTON RD	51	77	I	E	2	1373	151
ODANA RD	S WHITNEY WAY			I	W	2	1373	
	WESTBOUND	12	244	N	S	0	1373	39
GREENWAY TRL	CARNWOOD RD			I	S	1	1373	

FIGURE 1 Sample crash information from WisTransPortal crash database.

FIGURE 2 WISLR links and nodes for city of Madison.

(segments connecting two intersections) that are used as the base map for locating crashes. Understanding the relational information behind this map is critical to the development of the mapping algorithm. Many tables exist within the WISLR database. Detailed information about relations, tables, and fields relevant to the mapping algorithm will be provided in subsequent sections.

OVERVIEW OF CRASH MAPPING ALGORITHM

Intersection crashes are those crashes that occur at, or very close to, a roadway intersection. Segment crashes are those crashes that occur within the roadway link, which can be determined by two adjacent roadway nodes.

Specific implementation details are described more fully in the final project report, which is available on the TOPS Laboratory website (*6*).

Intersection Crashes

Intersection Crash Mapping Data Sources

Three fields from the crash database as described earlier are required to map a crash record at the intersection level, namely, municipality-code, on-street, and at-street. The location information in these fields originates from the MV4000 form, which is hand-coded by a police officer at the crash scene and subsequently manually entered into the DMV database.

The important WISLR tables and their relationships relevant to the determination of an intersection at or near which a crash occurs are shown in Figure 3. Each table is described as follows:

- Roadway route: unique list of road names for each municipality;
- Alternate roadway route prefix: standard and alternate prefixes for road names;
- Alternate roadway route name: standard and alternate spelling for common road names;

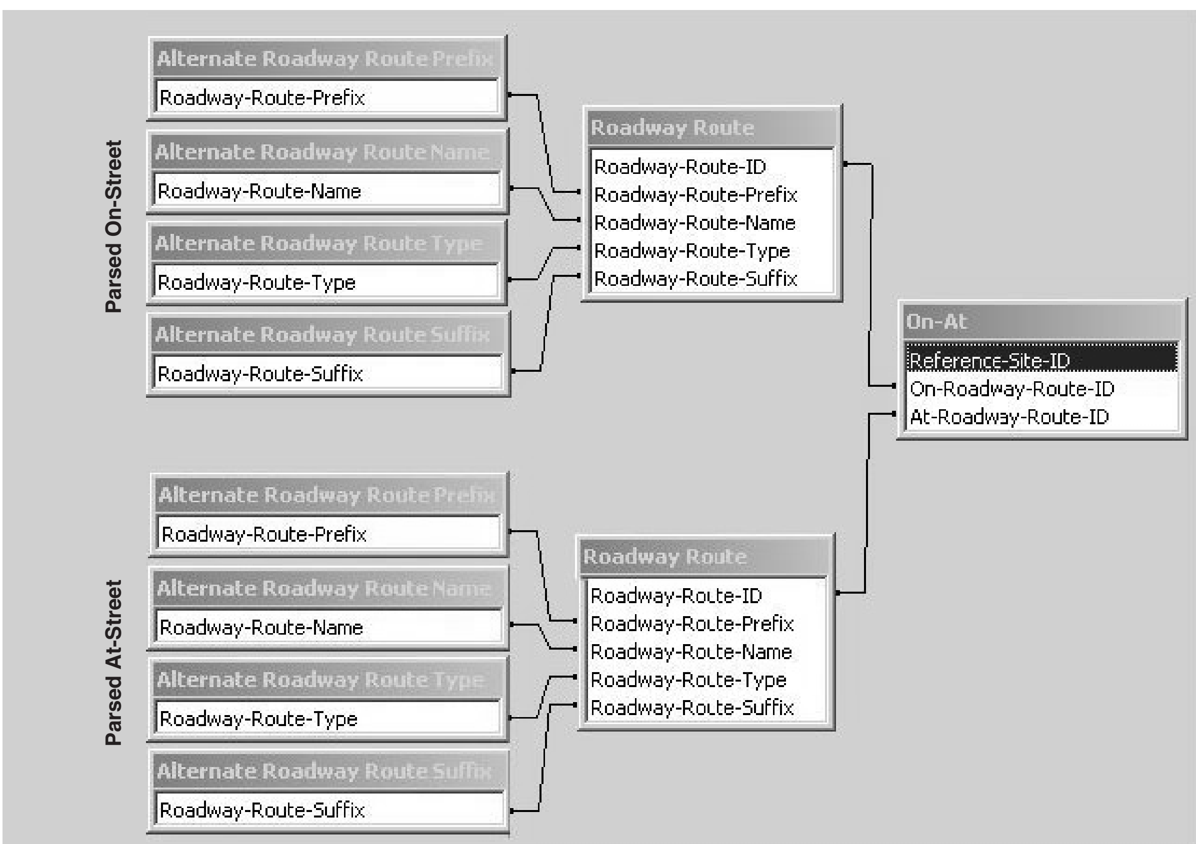

FIGURE 3 WISLR tables and relationships.

- Alternate roadway route type: standard and alternate types for road names;
- Alternate roadway route suffix: standard and alternate suffixes for road names;
- Standard roadway route prefix: standard prefixes for road names;
- Standard roadway route type: standard types for road names;
- Standard roadway route suffix: standard suffixes for road names; and
- On-at: combination of road names that intersect with each other.

The roadway route table contains a unique list of road names in WISLR separated into four parts: directional prefix, road name, road type, and directional suffix. For example, "E Washington Ave" would be separated into three fields and an empty fourth field: roadway-route-prefix = E, roadway-route-name = Washington, roadway-route-type = Ave, and roadway-route-suffix = ___. Each street name in this table is associated with a unique Roadway-Route-ID. The four alternate tables are used to take alternate spellings (aliases) that are part of a road name and standardize them. The three standard tables contain the entire list of standard prefixes, types, and suffixes used in WISLR. The role of the standard tables will be described in subsequent sections.

In WISLR, intersections are identified with nodes listed in the On-At table. Each intersection is identified with a unique Reference-Site-ID. Two street names represented by their roadway-route-ID (the on-roadway-route-ID and the at-roadway-route-ID) form a node. (It should be noted that the roadway route table is represented twice in Figure 3 to illustrate the fact that a combination of the on-roadway-route-ID and the at-roadway-route-ID is required to obtain the Reference-Site-ID.)

Intersection Crash Mapping Methodology

The key to determining the intersection location for a crash record is to associate a WISLR reference-site-ID value in the on-at table shown in Figure 3 with a pair of on-street and at-street street names from the crash database.

To create this association, the following steps are performed:

1. For each crash record, parse the on-street and at-street street names into prefix, name, type, and suffix components.
2. Match the parsed component information with the roadway-route-prefix, roadway-route-name, roadway-route-type, and roadway-route-suffix information in the WISLR roadway route table for the same municipality in which the crash occurred. The result is a WISLR roadway-route-ID for each on-street and at-street street name in the original crash record. These are represented by on-roadway-route-ID and at-roadway-route-ID, respectively.
3. Determine the reference-site-ID in the WISLR on-at table based on the on-roadway-route-ID and at-roadway-route-ID.

Once the reference-site-ID is obtained for a particular crash record, it is possible to generate an intersection-level mapping since the reference-site-ID represents the intersection in WISLR at which the crash occurred.

Implementation of Intersection Crash Mapping Algorithm

The intersection crash mapping algorithm was implemented as a Java program to automate the process of assigning WISLR reference-site-IDs to crash database accident-numbers. Specifically, the intersection mapping program performs the following steps:

1. Imports crash information from a crash data file downloaded from the WisTransPortal.
2. Removes certain crash records from processing (this step is described further in subsequent sections).
3. Parses each crash record on-street and at-street field into its four parts: prefix, name, type, and suffix. Parsing is performed by splitting the information given in the on-street (or at-street) field into multiple words and then utilizing the WISLR tables to analyze each word in order to determine if it should be a prefix, name, type, or suffix. The following assumptions are used by the parsing mechanism:
   – At least one word in the parsed street name will be mapped onto the name field,
   – Only the first word can be tested to see if it is a prefix,
   – The last word can be tested to see if it is a suffix. If the last word is a suffix, the immediately preceding word can be tested to see if it is a type (if that does not violate the first assumption). If the last word is a not a suffix, it can be tested to see if it is a type.
   – If a word is not a prefix, type, or suffix, it has to be added to the name field.

The parsing mechanism performs two levels of analysis with respect to the WISLR tables. Level 1 analysis attempts to parse on-street and at-street fields into the prefix, name, type, and suffix fields based on the contents of WISLR standard tables. Level 2 uses the alternate tables in WISLR in order to convert nonstandard formats into standard ones during the parsing procedure. For instance, the alternate prefix "North" could be standardized into N in Level 2 parsing. Use of the alternate tables is necessary, as all street names are standardized in WISLR. The algorithm only parses using alternate tables if the street is not found in WISLR after parsing with standard tables.

4. Matches the parsed crash record street information with the contents of the roadway route table to obtain WISLR roadway-route-IDs. The match is made with a rigorous algorithm that considers spelling errors, roadway name aliases, and incomplete crash report information. The primary challenge in developing the matching algorithm was the existence of incomplete street name information in the crash records. To handle this situation, five levels of matching were established based on the amount of street name information used in the matching step:
   – Name matching. The Name field of the parsed crash field is matched to the roadway-route-name field in WISLR roadway route table. The additional prefix, type, and suffix information is ignored.
   – Prefix-name matching. Both prefix and name fields of the parsed crash record are matched to WISLR. Suffix and type information is ignored.
   – Name-type matching. Both name and type fields of the parsed crash record are matched to WISLR. Prefix and suffix information is ignored.
   – Prefix-name-type matching. Prefix, name, and type fields of the parsed crash record are matched to WISLR. Suffix information is ignored.
   – Prefix-name-type-suffix matching. Prefix, name, type, and suffix information of the parsed crash record is matched to WISLR. This level takes into account all available street name information to find the WISLR roadway-route-ID.

Spelling errors are most critical in the name field of the parsed crash record. If the name field cannot be matched, all match levels will be unsuccessful. Therefore the name field is assigned a spelling sensitivity if all match levels are unsuccessful. A spelling sensitivity of *n* causes the matching process to compare only the first *n* percent of the characters in the crash record street name fields to the roadway-route-name field in the WISLR roadway route table. Spelling sensitivity can be implemented on the name field while checking any of the five aforementioned match levels.

5. Match the roadway-route-IDs to intersection reference-site-IDs. Given the roadway-route-IDs, find the reference-site-ID that, as mentioned earlier, represents the intersection in WISLR.

Now that the five steps have been described, it is possible to describe the intersection crash mapping algorithm that implements these steps. First the various parameters and their levels used in the algorithm are defined:

- parse_level = 1 or 2: parse_level = 1 checks the parsed on-street or at-street against the standard tables and parse_level = 2 checks against the alternate tables in WISLR.
- match_level = 1, 2, 3, 4, or 5. Match Level 5 performs a full prefix, name, type, and suffix match, whereas Match Level 1 performs only a name match.
- num_remove = 0 or 1. This parameter is introduced, as it was found that some crash records had additional address information, for example, street numbers included in the street name field for the crash record. Retaining this information would cause a mismatch in the name field. At the same time there may be streets present that have numeric names that are not street numbers (such as “1st St”). Therefore a parameter num_remove is introduced—num_remove = 0 implies that no numeric values were removed from street names; num_remove = 1 implies that all numeric values were removed from street names.
- spell_match = 0 or 1: spell_match = 0 implies that no spelling match is done for the name field; spell_match = 1 implies that spelling match is performed for the name field.

The logic of the algorithm is to start with the most rigorous matching process and gradually relax the conditions until a successful match is found. In particular, the algorithm attempts to minimize any modifications of a street name given in the on-street or at-street field of the crash record (such as replacing possible aliases by using alternate roadway route tables, removing street address information, and performing a spelling match) and at the same time attempts to find a match at the highest match_level value. If a match is not found, the match_level value is reduced. If no matches are found even at the lowest match level, the least possible amount of modification of street name is introduced, and again there is an attempt to find matches from highest match_level to lowest match_level. This process is repeated until one or more Roadway-Route-IDs are found or all match level and modification options are exhausted.

Segment Crashes

Segment crashes are those crashes that occur within the roadway link, which can be determined by two adjacent roadway nodes.

Segment Crash Mapping Data Source

Four fields from the crash database are required to map a crash record at the segment level, namely, on-street, at-street, intersection-direction, and intersection-distance. The intersection-direction and intersection-distance fields provide information to locate roadway segment crashes relative to given intersections.

In order to map segment crashes, first the nearby intersection given by the on-street and at-street fields needs to be determined as described in the previous section. Next, two additional WISLR tables are required for mapping into a segment:

1. The roadway link table contains information for all roadway links. A roadway route can consist of several links. The field relevant to this discussion from this table is roadway-link-ID. The Roadway-Link-ID gives the identification in the WISLR database for each individual roadway link.

2. The roadway route link table gives the relation between roadway links and roadway routes. At least one roadway-route-ID can be found in this table for each roadway-link-ID.

Segment Crash Mapping Methodology

Segment crash mapping is essentially an extension of the intersection crash mapping algorithm. Segment crashes are coded in the crash database in terms of direction (intersection-direction) and distance (intersection-distance) from an intersection. Hence, the first step is to determine the unique intersection reference-site-ID related to each segment crash as described under intersection mapping. The key to determining the segment location for a crash record is then to determine the roadway-link-ID in the roadway link table in WISLR and shift the point from the intersection mapping result into a new location along the intersection direction based on the intersection-direction and intersection-distance information. Several technical details, assumptions, and procedures associated with segment crash mapping are not covered in this discussion. Further details can be obtained from the final project report on the TOPS Laboratory website (*6*).

Interface Development for Intersection and Segment Crash Mapping

The mapping procedure for intersection and segment crashes was implemented as a Visual Basic interface integrated with ESRI Map Objects 2.3. Specifically, the crash mapping program performs the following steps:

1. Imports the list of reference-site-IDs from the intersection crash mapping algorithm. The reference-site-IDs are used to find all possible roadway links surrounding the intersection for each individual segment crash record.

2. Classifies crashes as intersection- or segment-related by using a customized algorithm that weighs several factors (accident-location, intersection-direction, and intersection-distance) from the crash record. The first step to identify whether a crash is intersection- or segment-related is based on the information given in the accident-location field (*I* representing intersection crashes and *N* representing nonintersection, or segment, crashes in this field). It is not possible, however, in general to reliably determine this identification directly from the accident-location field in the crash database because of conflicting or insufficient information. For instance, some crash records are coded as segment-related crashes in the accident-location field, but the values of the distance to intersection are null. Other discrepancies and inconsistencies exist in the crash record, such as a value’s being assigned in the intersection-direction and intersection-distance field

TABLE 1 Refined Definition of Intersection-Related and Segment-Related Crashes

Accident–Location	Intersection–Direction	Intersection–Distance	I/S
I	Null	(0, 2)	I
I	Null	(2, +inf)	I
I	Not null	(0, 2)	I
I	Not null	(2, L)	S
I	Not null	(L, +inf)	S
S	Null	Null	I
S	Null	Any	I
S	Not null	(0, 2)	I
S	Not null	(2, L)	S
S	Not null	(L, +inf)	S

NOTE: I/S is the refined identification of intersection-related crash or segment-related crash.

for crashes that are specified as intersection crashes. From discussions with WisDOT officials, a refined definition was developed for intersection and segment crashes (Table 1). The segment crash mapping algorithm only handles crash records that are categorized as segment-related crashes in this refined definition.

3. For segment crashes, uses the reference-site-ID found in the intersection level mapping and intersection-direction from the crash record to identify the WISLR roadway link (roadway-link-ID) associated with the crash.

4. If a unique roadway-link-ID can be found, maps to the segment by using the intersection distance (intersection-distance) given in the crash report.

5. Generates a final crash map on the WISLR.

Figure 4 shows the Visual Basic interface that was developed for the automatic crash mapping program, named the crash mapping automation tool. The output of this application program is a shape (SHP) file, which can be viewed by most of the popular GIS map software programs.

CRASH MAPPING ALGORITHM RESULTS

The results of implementing the algorithms discussed in the previous section are described here. The automated crash mapping process was tested in the developed software for crashes that occurred in the city of Madison in 2003.

Intersection- and Segment-Level Mapping Results

As described earlier, both intersection and nonintersection (segment) crashes are first mapped to the intersection level. Segment-level mapping requires further processing once the intersection is found. The results obtained through the intersection-level crash mapping program are summarized next.

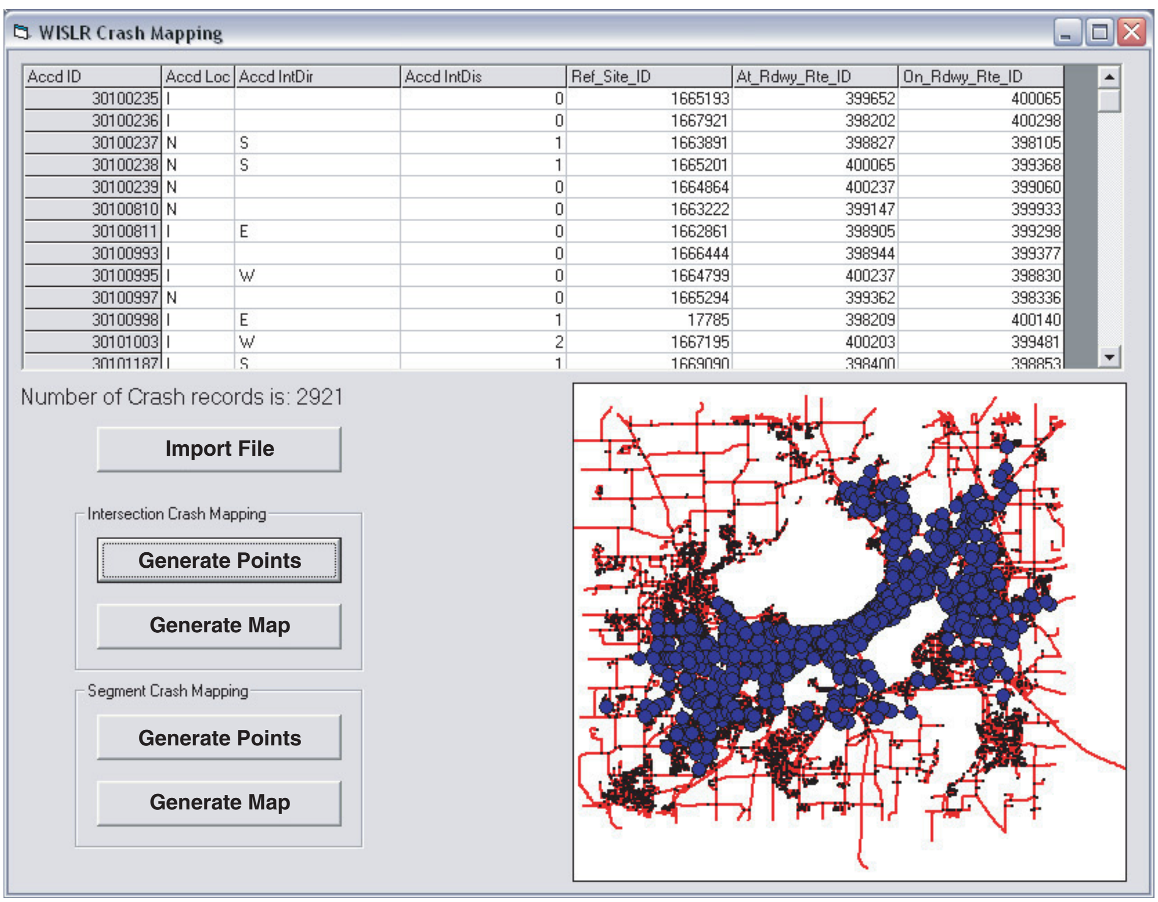

FIGURE 4 Interface for WISLR crash mapping program.

Filter Results

At the start of the process, the crash mapping algorithm applies three filters to the raw crash data to exclude certain crash records from processing:

1. Accident-location filter. Only public roadway crashes are processed. In particular, only crash records with intersection (I) and nonintersection (N) accident-location types are retained by the mapping algorithm. Crashes marked as parking lot (PL) and private property (PP) crashes are excluded.
2. On-street filter. An on-street field in the crash record may be null (i.e., does not contain any information) for two reasons. First, the on-street field information was not entered because of human error, so the data are incomplete. Second, it is possible that the street on which the accident took place was a highway, and it is therefore not a local road accident. In either case, the records where on-street field is null are removed from further processing.
3. At-street and at-highway. If both at-street and at-highway are null, the crash record has insufficient information to be mapped and is excluded from further processing. If at-street is null but at-highway is not null, the record will be a candidate for mapping provided that On-Street is also not null.

A total of 5,504 crashes occurred in Madison in 2003. A total of 4,351 crashes are available for mapping after applying Filters 1, 2, and 3.

Mapping Results

The first step is to map both intersection and segment crashes to the intersection level. A total of 3,273 records were mapped to unique intersections, that is, 75% of the 4,351 candidate records available for mapping after filter application and 59.5% of the 5,504 police-reported crashes that occurred in Madison in 2003. A total of 143 records were mapped to multiple locations (in such situations, the correct intersection is likely to be one of the multiple intersections found). Of these, 141 records were mapped to two intersections, and two records were mapped to three intersections. Multiple mapped records generally occurred for two reasons. The first case is related to "horseshoe" structures on the local roadway system, that is, where one road curves around and intersects a second road in two locations. In such a case, there is an ambiguity in the on-street and at-street information. The second case is from incomplete street information in the crash table. Since the multiple records were mapped to only a few candidate locations (two or three locations), it is believed that with some human intervention and judgment, they can be mapped quite easily to their correct location. Therefore, a total of 3,416 records were mapped to intersections of the combined single and multiple mapped records, that is, 78.5% of the 4,351 candidate records available for mapping after filter application and 62% of total 5,504 police-reported crashes that occurred in Madison in 2003. These 3,415 crashes consisted of both intersection and segment crashes, mapped to the level of the intersection. The remaining 936 crashes could not be mapped for several reasons, including spelling errors, incorrect or insufficient location information in the crash record, and missing WISLR links or nodes, or both.

Segment crashes were identified from these 3,416 crashes on the basis of the refined definition given in Table 1. There were 594 crashes that were segment-related, and all these crashes were found to belong to the subgroup of 3,273 uniquely mapped crashes. In all, 590 of 594 segment-related crashes can be mapped into a unique roadway link. Four of the segment crashes could not be adequately mapped in WISLR for reasons such as the inability to get the correct directional information of the link and link errors in WISLR. Because such situations were rare occurrences, the details are not discussed here and can be obtained from the final project report on the TOPS Laboratory website (*6*).

In summary, of the 3,416 crashes that were first mapped at the intersection level, 2,679 crashes were identified and mapped as intersection crashes, 590 crashes were identified and mapped as segment-level crashes, 4 crashes could not be mapped to segment level although they were identified as segment crashes, and 143 crashes were identified as intersection crashes but were mapped to multiple locations. For the four segment crashes that could not be mapped and the 143 multiple-mapped crashes, these were all mapped to the intersection level. These 147 crashes would require some human intervention to be mapped more accurately.

Figures 5 and 6 present the mapped results for the 3,273 uniquely mapped crashes up to the intersection level. Figure 5 shows the frequency of crashes mapped up to the intersection level for all intersection and segment crashes, and Figure 6 shows a sample of mapped segment crashes. The segment crashes are offset from the intersection, as expected.

QUALITY CHECK OF MAPPING RESULTS

The crashes mapped by using the algorithm were compared with manual mapping from two sources:

1. A digitized map of Madison 2003 crashes provided by the city of Madison. These crashes were digitized by hand directly from the MV4000 police reports.
2. Google Maps (*7*) online mapping service.

Quality evaluations were performed on the basis of the comparisons for the crashes mapped through the developed algorithm. In addition, the reasons for crashes that could not be mapped also were investigated.

Intersection-Level Crashes

The city of Madison has developed a geospatial map in ArcGIS for all crashes that occurred in Madison in 2003 by manually locating each of the crash intersections. The geospatial map developed by the city was used to determine the accuracy of the crash locations given by the TOPS intersection-level mapping algorithm. Because of inherent differences in the structure of the city map and the map generated by the algorithm, which is based on WISLR, only a limited number of crash records could be extracted for comparison purposes and only for the crashes marked as intersection crashes. The locations for a pool of 1,958 records for intersection crashes were extracted from this map and compared with the corresponding locations given by the intersection-level mapping algorithm. Approximately 86% of the crash records matched in their locations, and 2.6% were verified to be mapped incorrectly. For the remaining 11.5% of the records, the accuracy of the street locations was verified in WISLR for both the On-Street and At-Street fields, which means that it is quite likely that these were mapped to the correct intersection. However, a full verification for the 11.5% of the records could not be performed because of difficulties in extracting their exact locations from the city-developed map, and some shortcomings in the city-developed map such as incorrectly located crashes.

One of the primary challenges in performing a quality check on the crash mapping results is that address information in the official

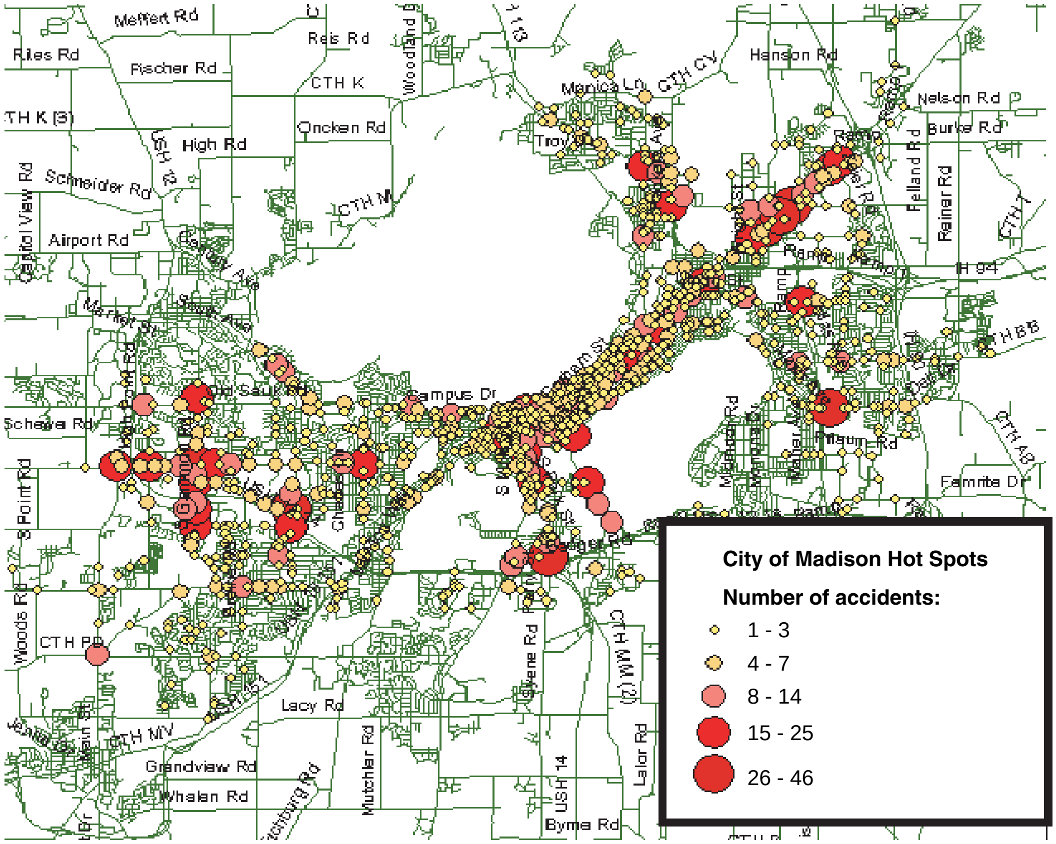

FIGURE 5 Hot spots in Madison, 2003.

police reports is not always clear. For a certain number of crashes, any manual mapping procedure will require a level of guesswork. As such, a second quality evaluation was performed by TOPS staff by comparing the results of the intersection-level mapping algorithm with manual mapping by using Google Maps.

The 3,273 uniquely mapped records were manually mapped in Google Maps based on the on-street and at-street information in each record. The intersection identified by Google Maps for each record was then compared with the intersection to which the same record was mapped in WISLR. In particular, for each of the 3,273 records the location in the WISLR map was first identified. Then the on-street and at-street information for the records was entered in the Google Maps website (*7*) as: "<On-Street> & <At-Street>, Madison, WI."

FIGURE 6 Sample of mapped segment crashes.

When Google Maps was unable to locate exact matches for <On-Street> and/or <At-Street>, it provided alternate candidates for <On-Street> and <At-Street> names that could be interactively selected by the user. Under such a situation, the best candidate was selected on the basis of the opinion of the evaluator. Then the crash location in Google Maps was visually compared with the crash location in the WISLR map given by the intersection-level mapping algorithm. If the two locations were found to match, the probability that the crash was mapped to the correct location was considered high.

From comparison of the intersection-level mapping algorithm with manual mapping through Google, it was observed that manually mapped crashes on Google agreed with 98% of the 3,273 records that were compared.

Segment Crashes

The segment crash mapping algorithm was implemented for the same data source as the intersection crash mapping, and results from intersection mapping are regarded as the input of segment crash mapping. In particular, the accuracy of segment mapping depends primarily on the accuracy of intersection crash mapping, since it is the precondition for finding the exact roadway link.

A minor factor of segment mapping errors is that the generated directional information associated with the roadway links was incorrect, which occurred in four of the segment crashes that could not be mapped. The primary reason for the incorrect directional information was that the roadway link in WISLR had too much curvature to match the directional information given by the crash record.

In summary, on the basis of the tested results, more than 99% of segment crashes obtained after intersection-level mapping were accurately mapped to the correct roadway link.

Unmapped Records

There were four primary reasons why records were not mapped successfully: spelling errors, missing WISLR intersections, alternate road names, and insufficient information in the crash records. Possible solutions to each of these causes are as follows:

1. Spelling errors. The current spelling error-handling module was able to map 329 records that otherwise would not have been mapped. In addition, there are 115 remaining crash records that could not be mapped because of spelling errors. Those records potentially could be mapped with sufficient improvements to the algorithm, which would represent a 2.5% improvement.
2. Missing WISLR intersections. For various reasons, some intersections are not included in WISLR. TOPS is working with WisDOT to track missing intersections by municipality to recommend for inclusion in WISLR. Based on the analysis of the mapped data set, a further 362 records could be mapped if such intersections were added, which represents an 8.3% improvement.
3. Alternate road names. A few known alternate names were identified manually, which resulted in the mapping of 29 additional records (0.6% improvement for the 4,351 available records). However, an additional 205 records were mapped in Google Maps with manually entered alternate names known to the quality check evaluator but did not have an intersection associated with them in WISLR. These 205 records did have the correct on-roadway-route-ID and at-roadway-route-ID for the alternate names in WISLR. If an alternate road table for each municipality could be created and the missing WISLR intersections added, the combined improvement could map 596 additional records, a 13.7% improvement for the 4,351 available records.
4. Unintelligible records. A small percentage of records have either unintelligible or unmappable on-street or at-street fields (around 2% of the 4,351 records). There is no simple solution to map these records. Any possible mapping would have to be manual based on guesswork.

The automated mapping algorithm provided an encouraging breakthrough in the ability to analyze local road crash information in Wisconsin. Further improvements are possible through a more advanced spelling error-handling module, the development of alternate name roadway tables, and the inclusion of missing intersections in WISLR.

NEXT STEPS

Quality assessment demonstrates that automatic crash location mapping yields a high matching percentage and a reliable outcome without sacrificing accuracy. The successful implementation of the mapping algorithm for 2003 Madison crashes provides great promise for safety analysis on local highways or streets. To maximize its benefits, future expansion of the tool to a statewide application is of immediate interest. One improvement is to make the tool portable so that the algorithm can be conveniently transferred to other municipalities or agencies that are in great need of the crash location information.

Knowing the crash location is the first step to identifying safety-problem locations. Providing versatility and flexibility is the key to success in meeting users' needs. Consequently, the tool will be enhanced with more sophisticated query functions that allow users to conduct safety analysis on various bases such as hot-spot identification, corridor analysis, or network screening.

Safety data are far more than just crash information. Conducting meaningful crash analysis and developing feasible safety improvements demand a variety of data sources. Using WISLR as the base map provides a seamless integration of crash information to the highway inventory stored in the WISLR database. A user-friendly interface is required to facilitate more systematic and comprehensive safety analysis with valuable WISLR information such as roadway geometric characteristics, local traffic information, pavement conditions, intersection configuration, and more.

Ultimately, the tool can be upgraded to an Internet interactive map by using ESRI ArcIMS within the WisTransPortal framework. The upgrade not only provides maximum access to the highway safety community but also takes advantage of spatial query capability in ArcIMS such as proximity analysis and network analysis.

ACKNOWLEDGMENTS

This research project was sponsored through the FHWA state opportunity fund. The authors thank the project Technical Advisory Committee members for their valuable comments: Jonathan DuChateau (WisDOT), Martha Florey (WisDOT), Susie Forde (WisDOT), and William Bremer (FHWA). Special thanks also go to Andrew Graettinger for offering technical advice and to Jacob Notbohm for assisting in the quality assessment.

REFERENCES

1. Wisconsin Traffic Operations and Safety Laboratory, WisTransportal, University of Wisconsin, Madison. http://transportal.cee.wisc.edu.
2. Miaou, S., R. Tandon, and J. J. Song. *Providing Personalized Traffic Safety Information to the Public: Using Web-Based Geographical Information System Technologies.* Report SWUTC/05/167424-1. Texas Transportation Institute, College Station, 2005.
3. Steiner, R., I. Bejleri, X. Yang, and D. Kim. Improving Geocoding of Traffic Crashes Using a Custom ArcGIS Address Matching Application. *Proc., 22nd Environmental Systems Research Institute International User Conference,* San Diego Calif., 2003. http://gis.esri.com/library/userconf/proc03/p0698.pdf.
4. Zhan, C. Geocoding and Analysis of Freeway Service Patrol Data. *ITE Journal,* Nov. 2005.
5. WISLR website. Wisconsin Department of Transportation, Madison, http://www.dot.wisconsin.gov/localgov/wislr/index.htm.
6. Qin, X., D. A. Noyce, S. T. Parker, A. Dutta, and Z. Qiu. *Demonstration of Automatic Mapping of Wisconsin Local Crash Locations.* Final Report. Traffic Operations and Safety Laboratory, University of Wisconsin, Madison, Oct. 2006. http://www.topslab.wisc.edu/projects/documents/Final_Report.pdf.
7. Google Online Mapping Services. www.maps.google.com.

The Safety Data, Analysis, and Evaluation Committee sponsored publication of this paper.