Identification of Hazardous Highway Locations Using Knowledge-Based GIS: A Case Study

GARY S. SPRING AND JOSEPH HUMMER

The work described in this paper was conducted at North Carolina Agricultural and Technical State University and North Carolina State University. The study used the increased capabilities offered by geographic information systems (GISs), along with the detailed mapping (which contains highway features and geometrics) available for Guilford County, North Carolina, to demonstrate the use of engineering knowledge regarding accident causation to identify hazardous locations. The general approach taken was that of a pilot study, in which a subset of information is used to demonstrate how a new technology (in this case GIS) may be used to solve a particular problem or problems. The mapping data available from Guilford County, along with various other data files, the MapInfo GIS, and North Carolina's Accident Records System (ARS) were used to conduct the study. The project provided valuable information regarding the limitations and advantages of using engineering knowledge about accident causation to identify hazardous highway locations, and demonstrated the utility and difficulties of applying the GIS to ARS. The overall approach is focused on, and difficulties associated with implementation are discussed.

Accident Records Systems (ARSs) represent the first component of the Highway Safety Improvement Program (HSIP), which was established in 1979 by the FHWA (*1*) as part of its mandate (as set forth in the 1966 Highway Safety Act) to help states to develop safety programs.

ARSs are data bases that contain accident information, as well as traffic and physical information such as road inventories, traffic counts, pavement condition, railroad grade crossings, bridge locations, traffic signal and sign inventories, and traffic permit files. ARSs can help provide fast, safe, and high-quality service to the motoring public on both state and local facilities if they are used efficiently and effectively. This requires accurate and complete data input, as well as consistent and highly accessible data files.

This paper describes work completed at North Carolina Agricultural & Technical (A&T) State University and North Carolina State University. This project used the increased capabilities offered by geographic information systems (GISs), along with the detailed mapping (which contains highway features and geometrics) available for Guilford County, North Carolina, to demonstrate the use of engineering knowledge about accident causation to identify hazardous locations. The paper's focus is on issues associated with implementing knowledge-based GISs for identifying hazardous highway locations. Details on knowledge base development and evaluation of the system can be found in the Project Final Report (*2*).

MOTIVATION

Accidents are random and rare events. High accident experience has been since the inception of the HSIP the most common way to determine hazardous locations, probably because the statistics are easily generated and are readily available. The use of accident data alone, as is the practice in the majority of states (*3,4*), to focus resources on locations that may be hazardous has several problems and limitations. The problems and limitations are discussed in detail by Zegeer (*5*). Most are related to the poor quality, incompleteness, or inaccessibility of the data, such as data errors and inconsistencies in accident records; inaccurate location of accidents, particularly in rural areas; outdated accident data; and inconsistent referencing systems.

North Carolina's MERGE system (*6*) has two primary deficiencies that were of interest for this project and that hinder the effectiveness of North Carolina's HSIP. These are inaccurate location of accidents and incomplete information on highway features and geometrics. The latter deficiency, along with budgetary constraints, is an important reason why state agencies such as North Carolina's opt for the use of statistics to identify hazardous locations, which essentially equates high accident experience to hazardousness. The sole use of statistics for this purpose presents several problems, one of which is the "regression to the mean" phenomenon. High accident levels may be due to this statistical anomaly, as discussed by Pendleton and Morris (*7*), and not to a roadway problem. Various techniques have been discussed that attempt to overcome this and other problems associated with using statistics (*8–10*).

High-accident locations often represent problems on the roadway, but other locations may have equally high potential for a catastrophic event, even though accident experience is not yet abnormally high. Moreover, on a systemwide basis a particular element may have a high accident experience; thus, it may be more cost-effective to make a systemwide correction of a common element than to correct a high accident location. The need for a comprehensive program to address hazardous roadway elements is discussed by Zegeer (*5*). A commonly identified example of a potentially hazardous element is a roadway section with a low friction number.

> This condition can be identified in two ways: by searching for sites with high wet-weather accident experience and then checking skid resistance properties of those sites, or by friction testing sites throughout the highway system and listing sections with low friction numbers. If only the "high-accident" sections with low friction numbers are selected and improved, the problem has only been partly corrected. The likely result is that other sections with low friction numbers will develop high-accident experience in the future. Thus, the ideal solution is to systematically identify all of the problem sections and improve those with the greatest need (*5*).

G. S. Spring, Department of Civil Engineering, North Carolina A&T State University, Greensboro, N.C. 27411. J. Hummer, Department of Civil Engineering, North Carolina State University, Box 7908, Raleigh, N.C. 27695.

To summarize, North Carolina makes exclusive use of statistical analyses applied to its ARS to focus its limited resources on areas of the highway system identified as "hazardous." Primary reasons for this are the problems (typical of problems faced by other state departments of transportation) associated with its ARS data, primarily the lack of highway feature and geometric data. Although the incompleteness of accident location data is a problem in North Carolina, as it is in other states, it was beyond the scope of this project and therefore was not addressed further.

A GIS SOLUTION

Because of the spatial character of ARS data, GIS technology greatly simplifies their extraction and presentation, provides a higher degree of user friendliness, and provides better access to the data. GISs also provide a means to integrate data from many sources (e.g., U.S. Census data, U.S. Geological Survey [USGS] data, accident records, pavement conditions, etc.).

The GIS is a computerized data base management system that provides graphic access (capture, storage retrieval, analysis, and display) to spatial data. The most visually distinctive feature of GIS software is a map display that allows thematic mapping and graphic output data overlaid on a map image. The key element that distinguishes GISs from other data systems is the manner in which geographic data are stored and accessed. GISs store geographic data using topological data structures: objects' locations relative to other objects are explicitly stored and therefore are accessible. These data structures allow analyses to be performed that are impossible using traditional data structures. Standard GIS functions that are useful for this application include thematic mapping, statistics, charting, matrix manipulation, decision support systems, modeling algorithms, and simultaneous access to several data bases.

In order for a GIS to be useful, a set of detailed base maps, of an acceptable scale and precision for ARS applications, must be available. Guilford County, North Carolina, has mapped (planimetric information only) approximately 90 percent of the county at a 1 in 2,400 scale to USGS mapping standards, which is adequate for locating accidents.

The general approach taken for the project was that of a pilot study, in which a subset of information is used to demonstrate how a new technology (in this case GIS) may be used to solve a particular problem or problems. The mapping data available from Guilford County, the MapInfo GIS, and North Carolina's MERGE system (along with other data described later) were used to accomplish the project's objective. A workstation-based prototype knowledge-based GIS (KBGIS) was developed for the project's demonstration purposes. The system was anticipated to be accessible through microcomputers tied to the system. This was not in fact possible, due to the limitations of the software used. However, that software does allow direct transfer of data between various hardware platforms. Thus, a data base developed on a UNIX-based workstation platform is directly transferable to Windows or Macintosh platforms. Limitations are discussed later in the paper.

Hardware Setup

Given the diversity of ARS users' needs, goals, and current activities, the best hardware configuration to implement the small pilot study's ARS in GIS was determined to be a local area network (LAN) at North Carolina A&T, consisting of two workstations, two IBM PCs, and three Macintosh Quadras. The LAN, an ethernet network using X-windows and the TCP/IP communication protocol, allows evaluation of a decentralized system; that is, users may access the workstation network, which contains ARS data, from their stand-alone PCs or from a workstation. The LAN allows software to be run under the DOS, MacOS, and Windows environments, thus providing maximum flexibility in accessing the data base. In addition to providing flexibility, the LAN satisfies the special needs associated with geographic information systems: large mass storage capacity, portability, rapid and powerful computing abilities, precision digitizing and quality plotting capabilities, and high-quality graphics. The LAN also provides access to the Internet, which facilitated the sharing of large data files among the project team.

Software

Because of unforeseen problems with Ultimap, the GIS software originally chosen for use on the project, the Guilford County map data were converted to MapInfo format. Ultimap Corporation abandoned the version of software originally intended for use in this project and entered bankruptcy. MapInfo was chosen as the replacement for several reasons, the most important of which were as follows:

- It allows access to true object geometry, thus allowing queries about, for example, highway curves' radii.
- It provides a built-in full-featured structured programming language, MapBasic, which was used to interface the GIS and knowledge-based components (which were written in MapBasic).
- MapInfo (Version 2.1) allows access to a wide variety of data base formats.
- The same version of the software runs across multiple platforms, which makes it a strong choice for multi-user environments.
- MapInfo is widely used across North Carolina. Additionally, several large GIS packages provide "hot links" to MapInfo files.

Data

The project team decided to take a feature-based approach to hazardous site identification. Features chosen were curves, bridges, and intersections. Therefore, data for each of these in addition to average daily traffic (ADT) and accident data were collected as part of this effort as summarized in Table 1. Deficiencies in and difficulties associated with the data are described later in this paper.

Accident data, obtained through the University of North Carolina's Highway Safety Research Center from the state of North Carolina's MERGE system, were coded to a 1:24,000 centerline map of Guilford County. The coding was performed by personnel at the University of North Carolina's Institute for Transportation Research and Education GIS laboratory. Both data sets (the accident data and the road centerline data) were transferred to the A&T network via the Internet. The data were then imported to the MapInfo software.

The North Carolina Department of Transportation (NC DOT) bridge maintenance unit provided project personnel with its federal bridge file for 751 bridges in Guilford County, dated 1992. The file was provided in ASCII format and was imported first into the FoxBase data base manager to provide data structure (the data were

TABLE 1 Project Data Summary

Description	Source	No. of Variables	No. of Records
Accident location data	MERGE	13	11,554
Accident attribute data	MERGE	77	63,899
Roadway centerline network (1:24K)	ITRE GIS laboratory	15	23,524
Bridges	NCDOT Bridge Maintenance Unit, Federal Bridge file for 751 bridges	21	448
Average daily traffic	NCGIA - derived from NCDOT HPMS	52	3,726
Planimetric data	Guilford County GIS Unit	1	500,000
Centerline data (1:150K TIGER files)	MapInfo StreetInfo+	6	36,201

space-delimited, which is not supported by the MapInfo software) and then into the MapInfo software. The bridge file contained 21 variables, including latitude and longitude. Bridges were geocoded using these values instead of milepost information because the latitude-longitude data were more complete.

Intersection data consisted solely of location and street names. They were created from enhanced TIGER files (called StreetInfo Plus) that were purchased from MapInfo Corporation, for Guilford County. No other data, such as turning movement counts, were available.

Curve data were in the map data obtained from Guilford's planimetric data base. The data base was digitized from digital orthophotos created by flying the county in 1991. The resulting maps are at a 1 in 2,400 scale and therefore provide a ±60-unit precision as mentioned earlier. Because of the change from the Ultimap software to MapInfo, it was necessary to transfer the data from the Apollo Domain workstation on which they were created in Ultimap to a UNIX-based workstation on MapInfo runs. The project team transferred more than 300 files via the DXF format. The resulting file contains more than 500,000 records and requires 4.5 megabytes of storage.

THE KBGIS MODEL

Figure 1 depicts the model's structure. The system's engine is the GIS. The user initiates sessions by choosing the type of site for which an analysis is to be performed. The GIS engine, through a series of graphic SQL queries (such as buffering) and traditional SQL queries of its data bases, provides the information listed in Table 2, in the form of "Danger Tables," to the knowledge-based component of the system. This information, along with user-provided information, is used for each site type (curve, intersection, or bridge) to identify and rank sites based on their hazardousness. Conceptually, the process consists of calculating an accident rate based on data and a rate based on models that have been developed by others: intersection models by Hauer et al. (*11*), bridge models by Turner (*12*), and curve models by Zegeer and Council (*18*). Adjustments to the rates and levels of confidence in the rates, extracted from the knowledge base, are used to calculate a combined accident rate and a level of confidence in that rate. Hazardousness level is then determined using a function depicted by Figure 2, derived from expert interviews.

In the Guilford County case study, only accident rates based on data for bridges were possible, because the bridge file had ADT data as one of its fields. The ADT file obtained from NC DOT for the other site types had several difficulties associated with it, which are described in a later section.

Accident Frequency Calculations

The procedures used to determine accident frequency are generic and are used for all three site types. This is essentially a buffering problem, the objective of which is to find and count all objects of Type 1 (e.g., accidents) that lie within some specified radius of

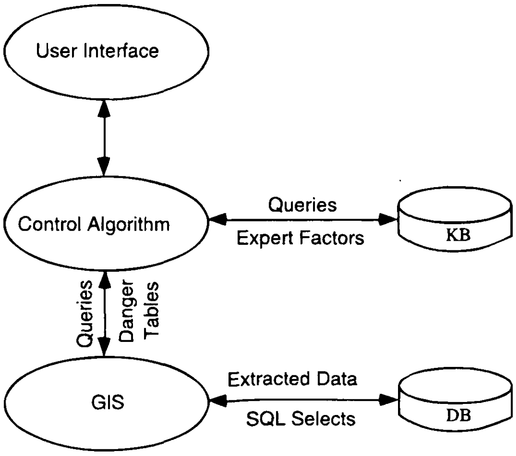


FIGURE 1 Model structure.

TABLE 2 Information Passed To Kb

Category	Derived Data Item
Curves	Length of curve Degree of curve Number of accidents occurring within specified distance
Bridges	Accident rate based upon data Accident rate based upon model
Intersections	Number of accidents occurring within specified distance

objects of Type 2 (e.g., intersections). The buffering function provided by MapInfo results in one count for this query, namely, the total number of accidents that lie within some radius of *all* intersections. The count of interest, however, is the number of objects of Type 1 that lie within a specified radius of *each* object of Type 2. Thus, it was necessary to write special code that would do the latter. Using the intersection example, the code has two parts:

1. Convert intersection point objects to circles with radius as specified.
2. Count the number of accident objects that fall, geographically, within the circle objects just created.

Figure 3 provides an example of high-accident locations selected based on a criterion of greater than four accidents within a specified radius from intersections. Figure 4 depicts a thematic map generated from this information.

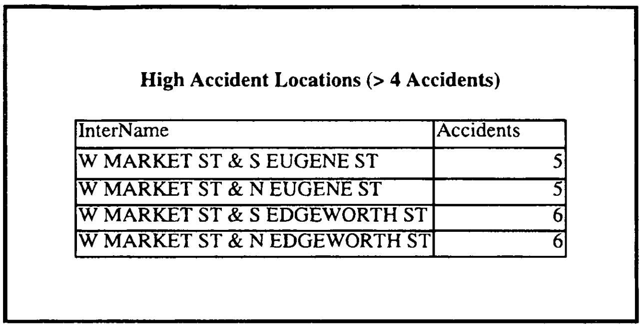
Accident Model Rate Calculations

As mentioned earlier, bridges were the only site type for which it was possible to calculate model rates, given the limitations of the data available for this project. To generate the bridge model rate tables shown in Figure 5, the following steps were used.

Overall Confidence \ Overall Accident Rate	0.1 Very Low	2	4	6	8	10 Very High
Very High 10	36 36, 36	22.9 9,35	13.1 7,25	6.5 3,11	2.6 2,4	1 1,1
8	33.2 24,35	23.7 14,34	14.5 7,26	8.2 6,12	4.8 4,5	2.7 2,4
6	32.5 25,34	24.1 15,31	17.6 13,27	10.9 9,13	8 6,11	5.4 3,10
4	31.9 28,33	25.1 20,27	19.5 15,25	15.4 12,22	11.8 8,17	9.9 5,16
2	31.8 29,34	25.8 21,29	22.8 20,27	19.2 15,23	16.7 10,22	14.5 7,21
Very Low 0.1	31.3 28,35	29.1 22,34	26.8 19,31	24.2 18,30	22.2 16,29	20.6 11,28

Key
Mean
Low,High

FIGURE 2 Hazardousness function.



High Accident Locations (> 4 Accidents)

InterName	Accidents
W MARKET ST & S EUGENE ST	5
W MARKET ST & N EUGENE ST	5
W MARKET ST & S EDGEWORTH ST	6
W MARKET ST & N EDGEWORTH ST	6

FIGURE 3 Sample table used to identify high-accident intersections.

1. Select a unique field and structure number, and create a derived field using the bridge rate model.
2. Normalize the model rates in Figure 5 to a scale of 10.

The curve rate model uses curve parameters that may be calculated from curve radius and delta. It also uses road width, which is not available in the data and therefore must be obtained directly from the user. Radius and delta values were calculated using the MapInfo-supplied access to true curve geometry along with a series of complex geographic selections.

Data used for these calculations are summarized in Table 2. The information depicted in Figures 3 through 5 are passed back to the knowledge-based components of the system. The system then prompts the user for more detailed information on a site-specific basis, for example, ADTs on curves, turning movements at intersections, and so on.

Data Integration

Although the use of the MapInfo software was not initially anticipated, the choice proved to be serendipitous. MapInfo provides a fairly full-featured, structured programming language called MapBasic, which provides transparent access to MapInfo's functionality. Therefore, all programming for this effort was performed in MapBasic. Additionally, problems associated with integrating GIS, knowledge base rules, and inferencing mechanism were avoided in this way. This has been the case in a great many other KB GIS efforts described in the literature (*14–18*). As was described earlier, all data were imported to the MapInfo format for use.

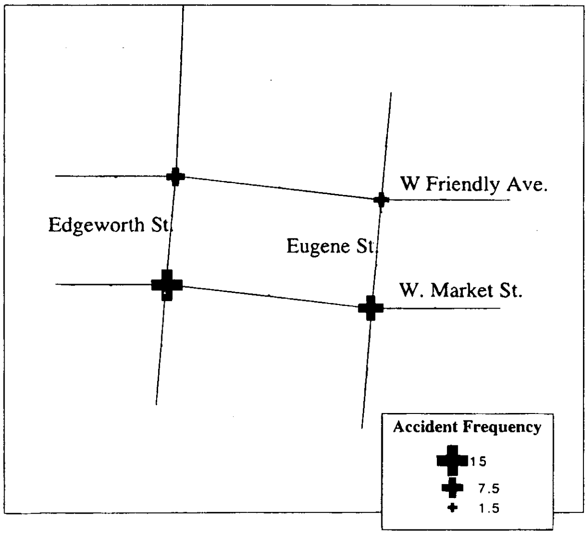


FIGURE 4 Sample thematic map of high-accident locations.

THE KNOWLEDGE-BASED SYSTEM ELEMENT

The key feature of knowledge-based systems (KBS) with regard to data base applications is their ability to use relations among objects stored in an existing data base to infer, with varying degrees of certainty, other objects. Since they infer new data using existing data, queries for which there are no data explicitly stored may be posed and answered. Consider, for example, a family tree. To store the tree in a conventional data base would require that *all* family relationships be explicitly defined as fields. That is, for each person in the family, the data base would have to contain explicit information regarding his or her relationship to all other family members. One way to do this would be to have each row represent a family line and each field represent a relationship. This method requires full specification of all relationships for which queries are to be made. For example, if one were to ask the data base "Who is Fred's cousin?" the answer could only be answered if the cousin relationship were stored in the data. Using a KBS in combination with the data base would require specification of only one relationship (such as parent), along with a set of rules representing all other relationships. This simplifies the data base and allows new relationships to be added easily (or existing ones to be changed or deleted) without disturbing the data. Relationships that have some degree of uncertainty associated with them may also be used in this process. For example, inheritance rules governing eye color or congenital defects could also be included if these types of information were of interest.

Information regarding accident causation derived from previous studies [for example Zegeer and Council (*18*), Harwood and Warren (*19*), and Spring (*20*)] were used in concert with GIS analysis strategies to identify hazardous highway locations using engineering knowledge rather than pure statistics. The project team conducted a series of interviews with state and local traffic engineers regarding the forensics of hazardous highway locations. Many of the questions used for the interviews were derived from a thorough review of literature such as that mentioned previously. The principles used for construction of the knowledge-based prototype were described by Mouradian (*21*). He advocated early prototype development using generally numerical (ratio) data. He also advocated using a simple scoring method to weight different pieces of information. This was the approach taken for this project. A set of models was taken from the literature, along with traffic engineers' confidence in those models, and were used to quantify hazardousness.

EVALUATION

The objective of the evaluation method chosen was to assess the validity of the underlying system model of the real world, which requires that evaluation criteria and acceptable levels of performance be established. A common evaluation criterion, which was used for this study, consists of a simple comparison of system conclusions with human conclusions. How closely they agree may be used as a measure of performance. There are problems intrinsic to this approach, described by Spring (*22*). However, it provides a gross but, for the purposes of this preliminary system, sufficiently

STRUCTURE_NO	MR (Accidents/yr)	STRUCTURE_NO	MRAdjusted
40185	2.00295	40185	7.65E-05
40187	1.4051	40187	5.37E-05
40188	0.96556	40188	3.69E-05
40189	79.1197	40189	0.0030237
40193	0.420459	40193	1.61E-05
40194	9.463	40194	0.000361645
40195	364.573	40195	0.0139328
40197	12404.01	40197	0.474041
40198	821.483	40198	0.0313944
40202	204022.53	40202	7.79708
40203	39.1525	40203	0.00149628
40204	0.478173	40204	1.83E-05
40205	48.2972	40205	0.00184576
40206	92.1084	40206	0.00352009
40207	72.6338	40207	0.00277583

FIGURE 5 Sample bridge model rate tables.

accurate assessment of system performance. The approach is common for preliminary evaluations such as this one (23). A list of locations that were programmed for improvement in the period after 1992 by NC DOT was obtained and compared to the list of sites, ranked by level of hazardousness, output by the KBGIS. Table 3 presents a summary of these results, which were subjected to a simple χ^2 test, which tested whether or not the two sets of conclusions are related. It was concluded that, at the 95 percent level of confidence, they are related. Given several limitations associated with this study (and perhaps in light of the limitations), overall agreement was deemed adequate. A complete description of the evaluation process and its results may be obtained from Spring and Hummer (*2*).

DIFFICULTIES

Difficulties encountered during project implementation fall essentially into two categories, software and data. As is explained in the following paragraphs, most of the major difficulties faced were due to data problems or limitations. Although these may seem daunting, they are the same difficulties that any GIS development effort faces, and so are not specific to this particular application. They would be addressed as part of an agency's GIS implementation program.

Software

In the process of using the Ultimap and MapInfo (including MapBasic) software packages several difficulties were encountered, some of which were bugs in the software and some of which were simply missing or poorly designed features. In the case of Ultimap, difficulties may be attributed to its lack of robustness; perhaps this is one reason why the company was bankrupted. MapInfo, although adequately robust for most applications, was not originally designed as a full-featured GIS package and therefore lacked robustness for the purposes of this project. Another function that would be important if agencies implemented the system developed for this project is networkability, which MapInfo does not support. Three bugs in MapInfo were discovered that caused delays and had a negative impact on the Project's final product:

- For certain graphic objects, the MapBasic object geography function, which provides access to object attributes, does not consistently return correct values for those attributes.
- Result tables from multiple selections sometimes do not retain derived data columns, within MapBasic. (Using MapInfo for this yields acceptable results.)
- The MapInfo version that runs on the workstation (Version 2.0.2) does not allow the user to choose a subset of layers from AutoCAD DXF translation files.

Additionally, MapBasic does not provide access to many of the excellent built-in functions provided by MapInfo. For example, MapInfo has a powerful address matching capability to which only minimal access is given through MapBasic. Even when some built-in function is available in MapBasic, it may be difficult to find. Many tricks and processes that would smooth the way are not given anywhere in the written documentation. Additionally, MapInfo does not support a networked environment, eliminating one of the advantages of the LAN described earlier.

TABLE 3 Evaluation Format

	NCDOT	
KBGIS	Not Hazardous	Hazardous
Not Hazardous	7	16
Hazardous	14	4

Data

Table 4 summarizes the difficulties encountered with the data used for the project. Of the 23,524 road segments contained in the 1:24,000 mileposted centerline file used to geocode accidents, only 3,126 had names. Essentially all of the named segments were mileposted. Additionally, only about 31 percent of the accidents obtained from the MERGE data base had milepost information (that is, were locatable). This resulted in only 11,554 of the 38,157 acci-

TABLE 4 Data Limitations

Data File	No. of Elements	No. Usable	Reason
Roadway centerline network (1:24K)	23,524	3,126	Unnamed segments
Accidents	38,157	11,554	Not mileposted
Bridges	753	448	Missing lat/long information
ADT	52	0	Inconsistent road name conventions

dents that occurred in Guilford County in the last 3 years being geocoded. This is a major limitation to the validity of project results and is certainly one of the biggest obstacles facing implementation of a GIS-based ARS. However, given that most of the mileposted accidents are in rural areas and that the majority of named roads happened to be in rural areas, the validity of project results was enhanced. Locating accidents and other features (ADT and so on) on a map requires a match between referenced road names. The apparent absence of a consistent convention for naming rural roads, among the various data sources used for this study, therefore also contributed greatly to the difficulties in tying accidents and other feature data to the map. This was especially critical for the ADT data file. Road names used in the file were inconsistent with any other naming system that could be found. This prevented the project team from locating the ADTs and using them for rate calculations.

There were additional problems encountered in translating planimetrics from the original format to the MapInfo format via DXF. These all were related to bugs in the Ultimap GIS software in which the map data were originally stored. The platform on which it was anticipated programming would be done, Apollo/Domain, was not used due to the abandonment of Ultimap, nor was the A&T network completely functional. It was necessary for a second workstation to be purchased, as well as a PC-based computer, and for all components to be connected for a successful transfer to take place. Both software packages have bugs that contributed to the lower quality of these graphic data. Line and curve segments, after conversion to DXF and transfer to UNIX-based workstation, are discontinuous and some curves came through as full circles. This is due to a bug in the Ultimap software which creates these anomalies in the data during the conversion to DXF.

Another difficulty faced was due to the third MapInfo bug described earlier. The original data files had approximately 25 to 30 layers of data. Ultimap handled these as features that could be turned on and off and so all data were contained in the original map files. The MapInfo system is a layering system which requires separation of those features during the importation process. If all layers were included in the same data file, there would be no way to differentiate among the various data items such as bridges, roads, census tracts, and so on. The size of the resulting data file would be unwieldy as well. Unfortunately, the MapInfo Corporation, at the time of this project, had issued an upgrade to Version 2.1 only for its IBM PC Windows product, due to the bug described earlier. This required that all 300 files be transferred to an IBM PC that had MapInfo Version 2.1, which allows this selectivity. After being imported to MapInfo, the files were assembled into one large file and transferred back to the workstation. The resulting map file, even with only road edge information, contained over 595,000 elements for Guilford County alone.

CONCLUSIONS

Given the spatial nature of accident data, the use of GIS for ARS makes good sense. Within GIS software, different types of data are easily related, either graphically or in report form, thus making the data more easily accessible and providing a friendlier and more flexible user interface. For example, pinpointing problem spots on the highway network by displaying sites whose signs have reflectivity values below a certain level may be done with ease. These qualities also help to provide better quality data. The use of accident causation information will enhance hazard elimination programs by avoiding, or in some cases eliminating, problems associated with the use of accident records alone in identifying problem locations. The GIS also provides a link between the various ARS data files. Presently, inconsistent data files—that is, data files that use inconsistent referencing systems—make it extremely difficult to fully utilize the ARS as an accident analysis tool (i.e., all available information cannot be used together in one analysis). Often, data files, when created, were intended for purposes other than accident analysis and therefore often have different referencing systems (e.g., mile marker versus link node). The fact that in GIS all locations are referenced to the same map eliminates this problem.

The knowledge-based approach described in this paper creates a synergy by providing consistent access to a common pool of engineering knowledge. This also provides an excellent means of computer-based training for novice traffic engineers.

With these features, a KBGIS can provide more cost-effective, safer, and more efficient highway systems for the user community. This project demonstrates the process of integrating GIS with ARS and, it is hoped, will demonstrate that, with the dramatic advances in small computer hardware technology over just the last few years, the limits to what can be done are imposed by what users are willing to do, rather than by technology.

ACKNOWLEDGMENT

The authors thank the traffic engineers who participated in the interview process for this project, and members of the Project Advisory Group whose valuable input made the happy outcome of the project possible. This work was sponsored by the U.S. Department of Transportation's Southeastern Transportation Center.

REFERENCES

1. Zeegeer, C. V. *Highway Safety Improvement Program.* Report FHWA-TS-81-218. FHWA, U.S. Department of Transportation, 1981.
2. Spring, G. S., and J. Hummer. *Identifying Hazardous Locations Using Geographic Information Systems.* Final Report to Southeastern Transportation Center, U.S. Department of Transportation, Raleigh, N.C., 1994.
3. Spring, G. S., J. Collura, and P. W. Shuldiner. Analysis of High Hazard Locations: Is An Expert Systems Approach Feasible? In *Transportation Research Record 1111,* TRB, National Research Council, Washington, D.C., 1987, pp. 7–17.
4. Zegeer, C. V. *Highway Accident Analysis Systems.* National Cooperative Highway Research Program Synthesis 91, NCHRP, 1982.
5. Zegeer, C. V. *Methods for Identifying Hazardous Highway Elements.* National Cooperative Highway Research Program Synthesis 128, NCHRP, 1986.
6. Fischell, T., and E. Hamilton. Local MERGE—Bringing the Data Back Home. *Proceedings of the 14th International Forum on Traffic Records Systems,* National Safety Council, San Diego, Calif., 1988.
7. Pendleton, O. J., and C. N. Morris. Introducing the BEAST—A New Method for Accident Analysis. Presented at the 69th Annual Meeting of the Transportation Research Board, Washington, D.C., 1990.
8. Lau, M. Y., and A. D. May. Applications of Accident Prediction Models. Presented at the 68th Annual Meeting of the Transportation Research Board, Washington, D.C., 1989.
9. Higle, J. L., and M. B. Hecht. Comparison of Techniques for the Identification of Hazardous Locations. Presented at the 68th Annual Meeting of the Transportation Research Board, Washington, D.C., 1989.
10. Hauer, E., and B. N. Persaud. Problem of Identifying Hazardous Locations Using Accident Data. In *Transportation Research Record 975,* TRB, National Research Council, Washington, D.C., 1984, pp. 36–43.
11. Hauer, E., J. C. N. Ng, and J. Lovell. Estimation of Safety at Signalized Intersections. In *Transportation Research Record 1185,* TRB, National Research Council, Washington, D.C., 1988, pp. 48–61.
12. Turner, D. S. Prediction of Bridge Accident Rates. *Journal of Transportation Engineering,* Volume 110, Number 1, ASCE, New York, N.Y., 1984, pp. 45–54.
13. Leug, Y. Fuzzy Logic And Knowledge-Based GIS—A Prospectus. *IGARSS '89—Twelfth Canadian Symposium on Remote Sensing,* Vancouver, B. C., Canada, Part 1 (of 5), July 10–14, 1989.
14. McKinney, D. C., D. R. Maidment, and M. Tanriverdi. Water Planning Using an Expert GIS. Water Resources Planning and Management: Saving a Threatened Resource—In Search of Solutions. *Proceedings of the Water Resources Sessions at Water Forum.* ASCE, New York, N.Y., 1992, pp. 219–224.
15. McKinney, D. C., D. R. Maidment, and M. Tanriverdi. Expert Geographic Information System for Texas Water Planning. *Journal of Water Resources Planning and Management,* Vol. 119, No. 2, March–April 1993, pp. 170–183.
16. Stonebraker, M. The Integration of Rule Systems and Database Systems. *IEEE Transaction on Knowledge and Data Engineering,* Vol. 4, No. 5, Oct. 1992 p 415–423.
17. Gronlund, A. G. *A Knowledge-Based GIS Methodology for Forest Fire Management.* Master's thesis, University of North Carolina at Charlotte, 1993.
18. Zegeer, C., and F. Council. *Cost-Effective Geometric Improvements for Safety Upgrading of Horizontal Curves.* FHWA-RD-90-021. FHWA, McLean, Va., 1991.
19. Harwood, D. W., and D. L. Warren. Operational and Safety Effectiveness of Passing Lanes on Two Lane Highways. Presented at the 84th Annual Meeting of the Transportation Research Board, Washington, D.C., 1985.
20. Spring, G. S. An Expert System to Analyze Hazardous Intersections. *Microcomputers in Civil Engineering,* Vol. 3, No. 3, 1988, pp. 299–309.
21. Mouradian, W. H. Knowledge Acquisition in a Medical Domain. *AI Expert,* Vol. 5, No. 7, July 1990, pp. 34–38.
22. Spring, G. S. Validating Expert System Prototypes Using the Turing Test. *Transportation Research, Part C,* Vol. 1, No. 4, Pergamon Press, Ltd., Great Britain, 1993, pp. 293–301.
23. Spring, G. S., J. Collura, J. Shuldiner, and P. W. Watson. Testing, Verification and Validation of Expert Systems: A Case Study. *ASCE Journal of Transportation Engineering,* Vol. 117, No. 3, 1991, pp. 350–360.

Publication of this paper sponsored by Committee on Transportation Data and Information Systems.