UNIVERSITY OF PETROLEUM AND ENERGY STUDIES

Project topic:

CRASH SAFETY OF HYDERABAD CITY USING ArcGIS TECHNIQUES

Mentor: Dr. SABYASACHI MAITI

Student Information

P.VISWATEJA

B. TECH GEOINFORMATICS

VIII SEMESTER

R350308024

N. GEETA MOUNIKA

B. TECH GEOINFORMATICS

VIII SEMESTER

R350308018

ABSTRACT

This report is based on the study done by the students of UNIVERSITY OF PETROLEUM AND ENERGY STUDIES, BIDOLI. The report is an attempt to enhance the Crime Mapping and Analysis using ArcGIS techniques and also brief overview of the geographic information system (GIS).

The goal of the study was to determine if there are benefits to using a GIS-based crash referencing system over the traditional data bases that contain crash, roadway feature, and operations data that can be linked by location. Specific objectives for the project included:

1) Identifying high-truck-crash locations using the GIS-based system, and

2) Exploring the applicability of non-traditional data bases to this type of analysis

Computerized crash analysis systems in which crash data, roadway inventory data, and traffic operations data can be merged are used in many States and municipalities to identify problem locations and assess the effectiveness of implemented countermeasures.

By integrating this traditional system with a GIS which offers spatial referencing capabilities and graphical displays, a more effective crash analysis program can be realized. In a recent FHWA sponsored study, a crash referencing and analysis system within a GIS was developed within the Highway Safety Information System (HSIS) State of North Carolina for the area of Wake County and provides the functions needed to edit both tabular and spatial crash and roadway data as well as perform crash analyses related to intersections, roadway segments, and special highway features such as bridges. A complete description of the GIS-based crash referencing and analysis system is provided in a separate HSIS summary report.

Within Wake County, specific roadways are designated as part of the truck route network. All trucks greater than 8 ft in width or 45 ft in trailer length must travel on the designated routes with two exceptions. First, all trucks are allowed to travel up to 3 drivable miles off the network to reach their destination. For example, a truck could turn onto a roadway that intersects with the designated truck route, travel two miles and then turn onto another connecting roadway and travel another mile.

The second exception refers to trucking terminals that are outside the 3-mi buffer; trucks are allowed to reach their terminals, but the route must be approved by the NCDOT. Of the 1,821 mi of roads in the county, 135 mi (7% of the total) are designated truck routes. An additional 692 mi (38%) are roads that are within 3 driveable miles of a designated truck route. Combining these values, more than 45% of the roadways in the county are considered to be part of the designated truck system.

Acknowledgements

An activity can be termed as an accomplishment only when the purpose it fulfilled. The accomplishment of any activity involves a continuous unflinching effort, motivation and support from its mentor. We would like to extend our heartfelt gratitude to our instructor and mentor Dr. Sabyasachi Maiti for having constant faith in us throughout this project and directing and supporting us in every possible way at each step and for constantly guiding, monitoring and helping us improve at each stage. We extend our thanks to all the students and people for being very supportive. Above all, we thank each and every one of those who have been instrumental in the successful compilation and presentation of this report.

Thank you!

P. Viswateja
N. G. Mounika

Signature(s) of Student(s)
Date: 26/4/12

Maiti

Signature of mentor
Date:

TABLE OF CONTENTS

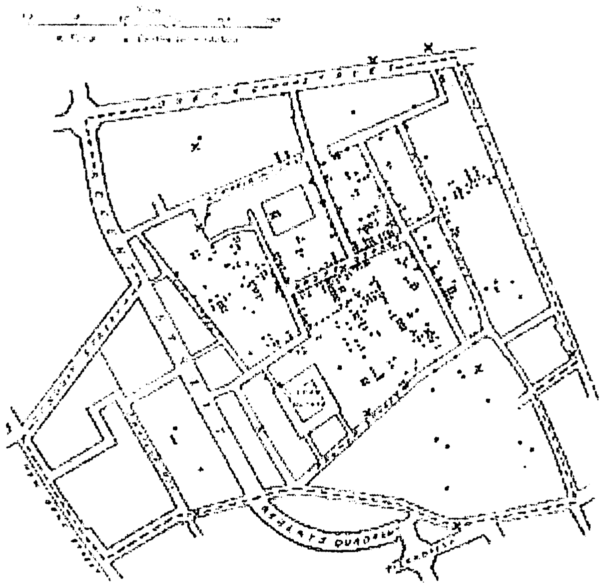
Table of Figures

LIST OF TABLES

LIST OF FLOWCHARTS

1. Geographical Information System:

In 1854, JohnSnow depicted a cholera outbreak in London using points to represent the locations of some individual cases, possibly the earliest use of the geographic method. His study of the distribution of cholera led to the source of the disease, a contaminated water pump (the Broad Street Pump, whose handle he had disconnected, thus terminating the outbreak) within the heart of the cholera outbreak.



E. W. Gilbert's version (1958) of John Snow's 1855 map of the Soho cholera outbreak showing the clusters of cholera cases in the London epidemic of 1854

While the basic elements of topography and theme existed previously in cartography, the John Snow map was unique, using cartographic methods not only to depict but also to analyze clusters of geographically dependent phenomena for the first time.

The early 20th century saw the development of photolithography, by which maps were separated into layers. Computer hardware development spurred binuclear research led to general-purpose computer "mapping" applications by the early 1960s.

The year 1962 saw the development of the world's first true operational GIS in Ottawa, Ontario, Canada by the federal Department of Forestry and Rural Development. Developed by Dr. Roger Tomlinson, it was called the "Canada Geographic Information System" (CGIS) and was used to store, analyze, and manipulate data collected for the Canada Land Inventory (CLI) – an effort to determine the land capability for rural Canada by mapping information about soils, agriculture, recreation, wildlife, waterfowl, forestry, and land use at a scale of 1:50,000. A rating classification factor was also added to permit analysis.

"Canada invented GIS to inventory its natural resources. As a large country with a small population, they had to devise means of gathering data, so turned to technology. ..."

CGIS was the world's first such system and an improvement over "mapping" applications as it provided capabilities for overlay, measurement, and digitizing/scanning. It supported a national coordinate system that spanned the continent, coded lines as "arcs" having a true embedded topology, and it stored the attribute and locational information in separate files. As a result of this, Tomlinson has become known as the "father of GIS," particularly for his use of overlays in promoting the spatial analysis of convergent geographic data. CGIS lasted into the 1990s and built a large digital land resource database in Canada. It was developed as a mainframe based system in support of federal and provincial resource planning and management. Its strength was continent-wide analysis of complex datasets. The CGIS was never available in a commercial form.

In 1964, Howard T Fisher formed the Laboratory for Computer Graphics and Spatial Analysis at the Harvard Graduate School of Design (LCGSA 1965-1991), where a number of important theoretical concepts in spatial data handling were developed, and which by the 1970s had distributed seminal software code and systems, such as 'SYMAP', 'GRID', and 'ODYSSEY' -- which served as literal and inspirational sources for subsequent commercial development—to universities, research centres, and corporations worldwide.

By the early 1980s, M&S Computing (later Intergraph), Environmental Systems Research Institute (ESRI), CARIS and ERDAS emerged as commercial vendors of GIS software, successfully incorporating many of the CGIS features, combining the first generation approach to separation of spatial and attribute information with a second generation approach to organizing attribute data into database structures. In parallel, the development of two public domain systems began in the late 1970s and early 1980s. MOSS, the Map Overlay and Statistical System project started in 1977 in Fort Collins, Colorado under the auspices of the Western Energy and Land Use Team (WELUT) and the U.S. Fish and Wildlife Service. GRASS GIS was begun in 1982 by the U.S. Army Corps of Engineering Research Laboratory (USA-CERL) in Champaign, Illinois, a branch of the U.S. Army Corps of Engineers to meet the need of the U.S. military for software for land management and environmental planning. The later 1980s and 1990s industry growth were spurred on by the growing use of GIS on UNIX workstations and the personal computer. By the end of the 20th century, the rapid growth in various systems had been consolidated and standardized on relatively few platforms, and users were beginning to export the concept of viewing GIS data over the Internet, requiring data format and transfer standards.

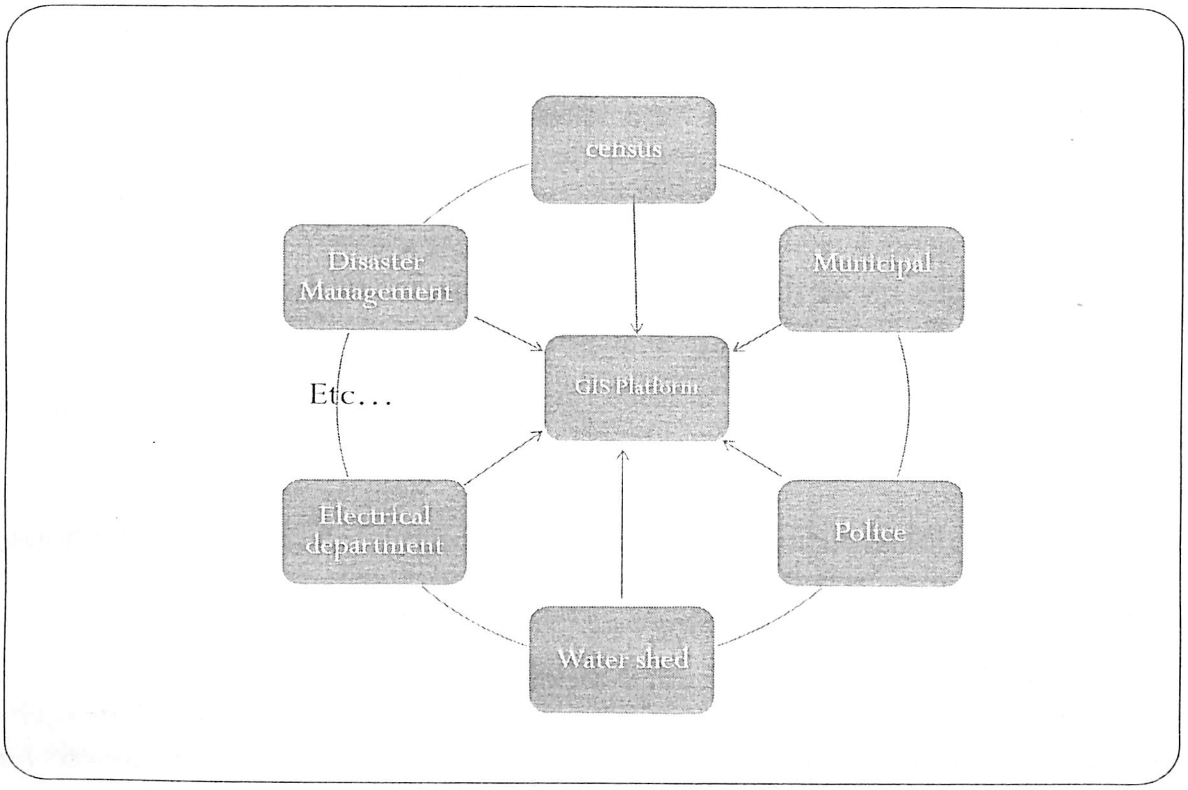
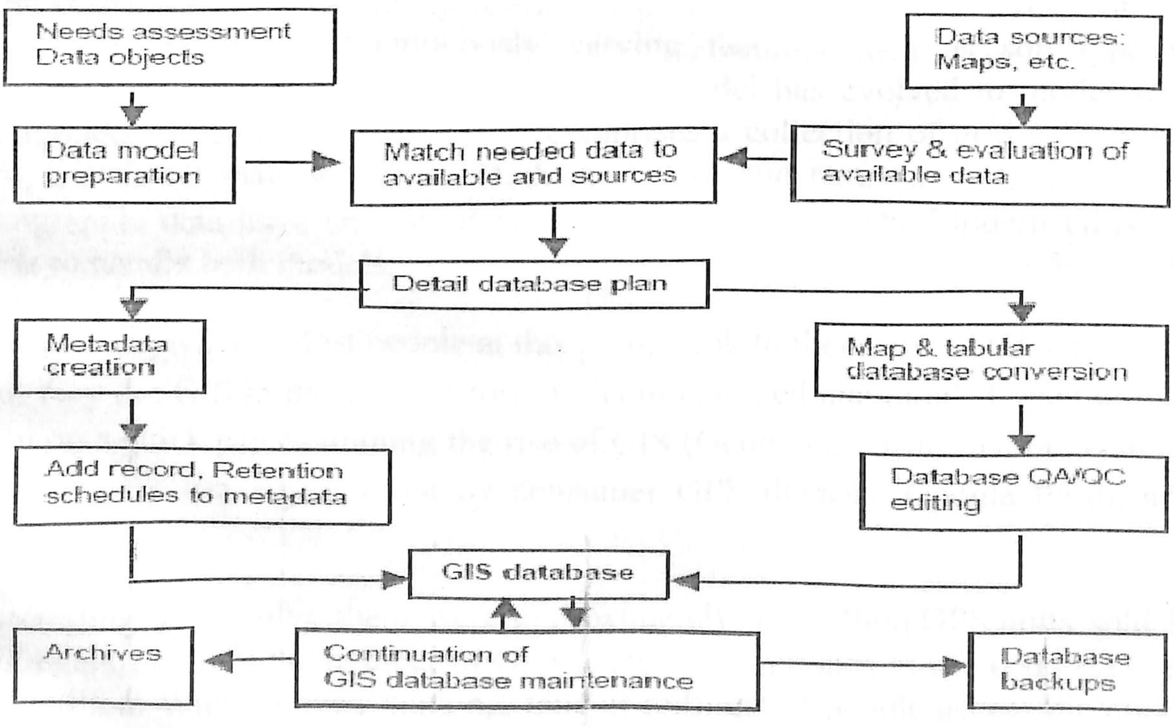


Figure 1: LIFE CYCLE OF GIS DATABASE



Overview of GIS platform

Geographic References

Geographic information contains either an explicit geographic reference, such as a latitude and longitude or national grid coordinate, or an implicit reference such as an address, postal code, census tract name, forest stand identifier, or road name. An automated process called geocoding is used to create explicit geographic references (multiple locations) from implicit references (descriptions such as addresses). These geographic references allow you to locate features, such as a business or forest stand, and events, such as an earthquake, on the earth's surface for analysis.

Vector and Raster Models

Geographic information systems work with two fundamentally different types of geographic models--the "vector" model and the "raster" model. In the vector model, information about points, lines, and polygons is encoded and stored as a collection of *x,y* coordinates. The location of a point feature, such as a bore hole, can be described by a single *x,y* coordinate. Linear features, such as roads and rivers, can be stored as a collection of point coordinates. Polygonal features, such as sales territories and river catchments, can be stored as a closed loop of coordinates.

The vector model is extremely useful for describing discrete features, but less useful for describing continuously varying features such as soil type or accessibility costs for hospitals. The raster model has evolved to model such continuous features. A raster image comprises a collection of grid cells rather like a scanned map or picture. Both the vector and raster models for storing geographic data have unique advantages and disadvantages. Modern GISs are able to handle both models.

GIS is everywhere. Most people at this point think to themselves "I don't use it", but they do; GIS in its simplest form is "computerized mapping". I want to take you on a quick trip examining the rise of GIS (Geographic Information System) in everyday life, exemplified by consumer GPS devices, Google Earth, and geotagging.

According to Canalys there were approximately 41 million GPS units sold in 2008, and in 2009 the number of GPS enabled cell phones in use had exceeded 27 million. Without even thinking, tens of millions of people access directions and look-up local businesses from these hand-held devices every day. Let's tie this back to our big picture here, GIS. The 24 GPS satellites orbiting earth are constantly broadcasting data about their location and exact time. Your GPS device or phone receives and process the signals from three to four of these satellites to figure out where it is located. Points of interest, addresses (lines or points), and aerial or road data is all stored in a database that is accessed by your device. When you submit data, such as posting a geo-Tweet (a location-based Tweet on Twitter), checking in on Foursquare, or rating a restaurant you are adding data to one or more GIS data sources.

Popular GIS Applications

Before consumer GPS devices were so prevalent we used to have to go to a computer and look- up directions, such as with Bing Maps. (Bing Maps is a relatively new service, which grew out of Microsoft Virtual Earth.) Bing Maps has some great features such as oblique imagery (Bird's Eye View), Streaming Video, and Photosynth. Many websites incorporate data from Bing or other GIS sources to provide a limited mapping experience on their own websites (such as seeing all their physical storefronts).

Traditionally desktop GIS has dominated the GIS mindset. People think of ArcMap, MicroStation, or other enterprise-level GIS applications when they think desktop GIS. But the most prevalent desktop GIS application is free, and quiet powerful. With over 400 million total downloads (according to GeoWeb 2008 keynote speech by Michael Jones) Google Earth is by far the most used GIS application in the world. While many people use Google Earth to look for fun things such as a friend's house, crop circles, and other oddities, Google Earth also allows you to add georeferenced images, view parcel data, and find routes.

Georeferencing Photos

One of my favorite things to do is georeference pictures. Georeferencing is the process of giving an image a "place". Using Panoramio this is very easy to do for Google Earth. This is really fun if you took a road trip, or any trip. Going a step beyond that is Photosynth (by Microsoft), where you can not only georeference an image, but also "stitch" images together. There is another free application that provides users a globe, ArcGIS Explorer from ESRI. ESRI, known for its desktop and server GIS applications, has released a free viewer that includes an updated user interface and some great features; I like to think of it as Google Earth on steroids. There are several add-ins you can use to see Bing imagery, Open Street Maps roads, geotweets, and more. Its built-in features include deciding routing, making notes/annotations, and creating presentations.

Even before the average computer user was using GIS on a near daily basis, everyone has benefited from it. The government uses GIS to decide voting districts, analyze demographics, and even time street lights. The real power of GIS is that it is more than a map, it is a map that can show us exactly what we want to see.

How has GIS become such an integral part of society almost seamlessly? Google, Garmin, and others were not creating products with "Hey, the mass public needs GIS" in mind, no, they were meeting needs. Humans think geographically. "Who, What, When, Where, Why, and How" those are the five Ws right? Place is extremely important to people. When studying how human populations have acted over the past millennia it is easy to see how geography dictated culture. Today, place still dictates much of our lives: property values, crime rates, education standards, these can all be classified by place. It is interesting to see when a technology has become so ingrained in a society that people don't consider it when they use it, they just use it; like with cell phones, cars, microwaves, etc. (that list could be very long). Personally, as someone who loves maps and loves computers and works in the GIS field I think it is great that an eight-year-old has the ability to look-up their friends address and show their parents exactly where they are going, or for family members to be able to see pictures of those they love where they were taken, and so many more cool things that GIS allows us to do without thinking.

2. Introduction:

This Roads & Highways Safety Hot Spot Analysis Template is an example of an ArcGIS Desktop geoprocessing workflow that can be used by crash safety analysts in a DOT. This set of models can be used to identify areas with a higher than expected level of crash severity. It includes a map document, sample dataset, and geoprocessing models showing how a hot spot layer can be created.

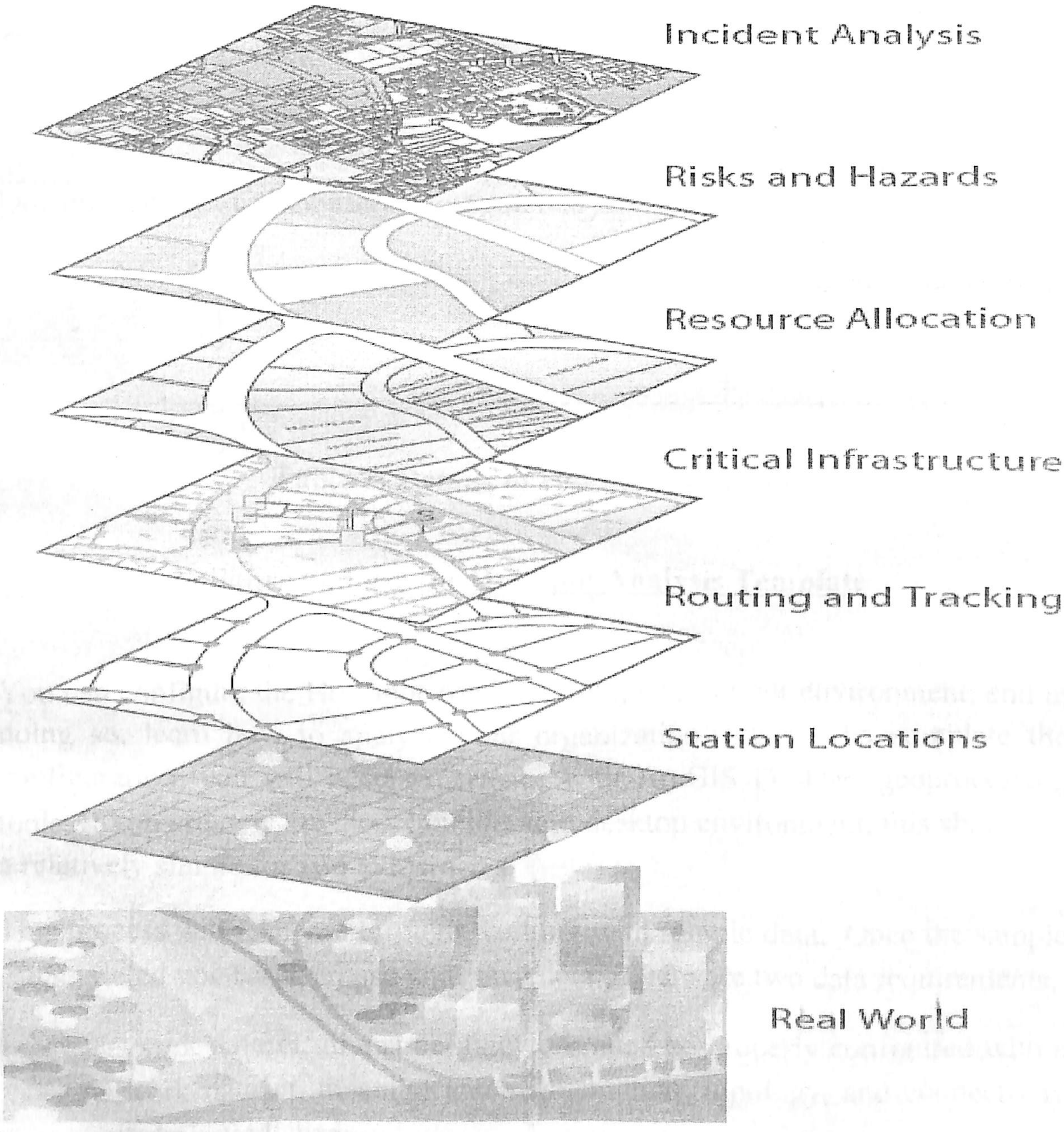


Figure 2: IDENTIFYING GEODATABASES FOR ACCIDENTS

2.1 Template Contents

The following files are provided in the Hotspot Analysis:

Template Directory	**Item**	**Description**
DataMapTools	HotSpotAnalysis.mxd	The map document containing analysis data.
	NetworkMatrixWeight.swm	Network Matrix Weight file for input into model
	HotSpotAnalysis.gdb	File geodatabase containing street network, crash locations, and analysis data.
Documentation	CrashSafety_HotSpotAnalysis	A word document that describes how to implement and configure this template.

Table 1: Details of Hot Spot Analysis

2.2 How to configure and use the Hot Spot Analysis Template

You can configure the Hot Spot Analysis Template in your environment, and in doing so, learn how to analyze your organization's data. To complete the configuration, you will need experience with ArcGIS Desktop geoprocessing tools. If you are new the model builder and desktop environment, this should be a relatively simple for you to learn.

This process will take you through working with sample data. Once the sample is completed you can then use your own data. There are two data requirements:

- Network dataset: the street linework must be properly configured with a network dataset, meaning that segmentation, topology, and connectivity must be established.

- Crash point layer: the crash points need to be coincident with your road linework, and be intersected by roads. Crash severity attributes are optional.

Software Environment

The following software must be installed and configured:

- ArcGIS Desktop 10
- Network Analyst Extension

You can improve critical decision making in a rapidly changing environment and have a direct impact on the safety of your officers and the citizens you serve.

Law enforcement often suffers from limited resources. GIS can help you efficiently and effectively match demands for service with service delivery. By knowing where your problems are, it also provides a visual means to proactively combat crime and communicate with citizens to build support.

GIS can help you to

- Understand events and dynamics in a neighbor including persons, events, and crime hazards.
- Identify risk factors including businesses, buildings, or other locations that draw crime.
- Rapidly reconfigure beats and reallocate resources after analyzing crime trends over time.
- Develop plans for special crime abatement teams to address regional or seasonal hot spot locations.
- Capture repeat call-for-service locations to apply additional help and resources.

Officers working the street, or investigations working a case, already know generally where the crime is. In addition to this knowledge, mapping over a period of months can help officers and investigators precisely view and understand underlying crime movements and patterns. For example, can certain types of crime be correlated to this time of day or day of week? Is an open, unmonitored parking structure a draw for criminals? As a crime analyst, you can support your officers in developing more effective tactical approaches and development strategies, ultimately preventing crime by identifying trend.

Crash mapping and analysis have evolved significantly over the past 30 years. In the beginning, many agencies utilized city and precinct maps colored pins to visualize individual crime events and crime plagued areas. Today, with the rapid advancement of technology, computer-based techniques for exploring, visualizing and explaining the occurrences of criminal activity have been essential. One of the more influential tools facilitating exploration of the spatial distribution of crime has been GIS (Ratcliffe and McCullagh, 1999; Harries, 1999). As Murray et al. (2001) note, it is the ability to combine spatial information with other data that makes GIS so valuable. Furthermore, the sheer quantity of information that available to most analysts necessitates an intelligent computational system, able to integrate a wide variety of data and facilitate the identification of patterns with minimal effort.

Crash mapping is used by analysts in law enforcement agencies to map, visualize, and analyze strategy. Mapping crime, using Graphic Information Systems (GIS), allows crime analysts to identify crime hot spots, along with trends and patterns. Using GIS, crime analysts can overlay other datasets causes of crime and help law enforcement administrators to device strategies to deal with the problem. GIS is also useful for law enforcement operations, such as allocating police officers and dispatching to emergencies.

Underlying theories that help explain spatial behavior of crime include environmental criminology, which was devised in the 1980s by Patricia and Paul Brantingham, routine activity rational choice theory, develop by Lawrence Cohen and Marcus Felson and originally published in 1979, and rational choice theory, developed by Ronald V. Clarke and Derek Cornish, and originally published in 1986. In recent years, crime mapping and analysis has incorporated spatial data analysis techniques that add stastical rigor and address inherent limitations of spatial data, including spatial autocorrelation and spatial heterogeneity. Spatial data analysis helps one analyze crime data and better understand why and not just where crime is occurring.

Research into computer-based crime mapping started in 1986, when the National Institute of Justice (NIJ) funded a project in the Chicago Police Department to explore crime mapping as an adjunct to community policing. The success of this project prompted NIJ to initiate the Drug Market Analysis Program (with the appropriate acronym D-MAP) in five cities, and the techniques these efforts developed led to the spread of crime mapping throughout the US and elsewhere, including the New York City Police Department's Compstat.

Whether the problem is tactical, strategic, or administrative, any data containing location information can be displayed and analyzed using geographic information system (GIS) technology. GIS is an essential part of crime analyst's toolkit-your means of creating valuable information for your officers in the field. By incorporating traditional law enforcement data with data such as demographics, infrastructure, and offender tracking, you can use GIS to transform information into actionable intelligence.

Evert crime problem is related to some location, whether it's an address, street, ZIP code, or district. GIS can help you leverage the locational aspect of your data to analyze, understand, and build solutions to the problems you face.
GIS aids crime analysis by

- Identifying and highlighting suspicious incidents and events that may require further investigation.
- Supporting pattern and trend analysis across multiple jurisdictions.
- Enhancing the implementation of various policing methodologies to reduce overall crime and disorder
- Integrating traditional and nontraditional law enforcement data to improve overall analysis.
- Educating the public with visual information to clarify crime concerns and enlist community action.
- Proving tools and techniques to capture crime series and forecast future crime occurrences.

Using GIS, you can model your workflow and capture your best practices. This allows you to determine what works and share it with others. You can build on your existing knowledge and experience collaborating with officers, investigators, and other analysts.

Geography is a natural way to organize information in law enforcement. You can use the geographic advantage in administrative crime mapping to provide an essay-to-understand view into your agency's activities.

GIS mapping and analysis can help you demonstrate strategy to government officials, provide a common operating picture to commanders during strategic Compstat meetings, and communicate with the public through an interactive web site. GIS aids in all aspects of the business of law enforcement.

Every day you face something different-different demands and different crime. Calls for service can quickly translate into calls from the community to take action. Officers and investigators look to you to provide them with information and analysis. GIS supports your needs in tactical crime mapping by providing a platform for

- Analyzing crime data to identify crime patterns and series
- Linking modus operandi and suspect information to specific crime locations
- Forecasting potential crime locations through spatial tools and techniques
- Supporting operations plans and clarifying ideal development locations
- Clearing cases by linking suspects to specific crimes after arrest
- Continuous active monitoring of high-risk and convicted sex offenders

Using GIS, you can create a map that can identify where the crimes are occurring and clarify what crimes are or are not related on your research. This can allow investigators to target their efforts and line officers to patrol and respond to locations while being more fully aware.

You can analyze data to examine an alibi or clarify whether a suspect was in direct proximity to a crime. GIS links data from various datasets on offenders and crime and allows a sophisticated platform for analysis.

Officers working the street, or investigations working a case, already know generally where the crime is. In addition to this knowledge, mapping over a period of months can help officers and investigators precisely view and understand underlying crime movements and patterns. For example, can certain types of crime be correlated to this time of day or day of week? Is an open, unmonitored parking structure a draw for criminals? As a crime analyst, you can support your officers in developing more effective tactical approaches and development strategies, ultimately preventing crime by identifying trend.

Using GIS, you can model your workflow and capture your best practices. This allows you to determine what works and share it with others. You can build on your existing knowledge and experience collaborating with officers, investigators, and other analysts.

Geography is a natural way to organize information in law enforcement. You can use the geographic advantage in administrative crime mapping to provide an essay-to-understand view into your agency's activities. GIS mapping and analysis can help you demonstrate strategy to government officials, provide a common operating picture to commanders during strategic Compstat meetings, and communicate with the public through an interactive web site. GIS aids in all aspects of the business of law enforcement.

Every day you face something different-different demands and different crime. Calls for service can quickly translate into calls from the community to take action. Officers and investigators look to you to provide them with information and analysis. GIS supports your needs in tactical crime mapping by providing a platform for

- Analyzing crime data to identify crime patterns and series
- Linking modus operandi and suspect information to specific crime locations
- Forecasting potential crime locations through spatial tools and techniques
- Supporting operations plans and clarifying ideal development locations
- Clearing cases by linking suspects to specific crimes after arrest
- Continuous active monitoring of high-risk and convicted sex offenders

Using GIS, you can create a map that can identify where the crimes are occurring and clarify what crimes are or are not related on your research. This can allow investigators to target their efforts and line officers to patrol and respond to locations while being more fully aware.

You can analyze data to examine an alibi or clarify whether a suspect was in direct proximity to a crime. GIS links data from various datasets on offenders and crime and allows a sophisticated platform for analysis.

Maps offer crime analysis graphic representations of crime-related issues. An understanding of where and why crimes occur can improve attempts to fight crime. Mapping crime can help police protect citizens more effectively. Simple maps that display the locations where crimes or concentrations of crime have occurred can be used to help direct patrols to places they are most needed. Policy makers in police departments might use more complex maps to observe trends in criminal activity, and maps may prove invaluable in solving criminal cases.

Display spatial patterns of events: Digital maps are the quickest means of visualizing the entire crime scenario. The locations of crime events, arrests, etc., can be routinely displayed on maps. This provides an essay method of viewing activities in an area rather than searching through a listing of events. Maps can also be used to convey more than one type of information at a time. Crime locations can be symbolized according to the day of week, type of crime, modus operandi (a particular suspect's method of operation when committing a crime) or frequency.

Integrate community characteristics: Community characteristics (for e.g. ., slums, markets, colleges, parks, alcohol permit locations, red light area, etc.) can be routinely displayed on maps while analyzing crime patterns to interpret relationship between these characteristics and the crime. For example the locations of the aggravated assaults, robberies and alcohol permits can be displayed to see if crime is clustering around locations that sell alcohol. Other mapping data such as bus routed and public housing can also be displayed at the same time to analyze relationships between neighbourhood characteristics and crime.

Produce thematic maps: Maps can be produced at any geographic level (e.g. police stations, divisions or zones) to aid in the analysis of crime patterns. Each response area can be shaded to represent the number of crime that occurred in that area during a specific time frame. The darker the shade the more events that occurred within the response area. These thematic maps can also be used to show the change in the area's crime rate. The percent change in number of crime incidents can be displayed by shading each area according to whether there was an increase, decrease, or no change.

Crime analysis is defined as a system of systematic, analytical processes directed at proving timely and pertinent information relative to crime patterns and trend correlations to assist the operational and administrative personnel in planning the deployment of resources for the prevention and suppression of criminal activities, aiding the investigation process, and increasing apprehensions and the clearance of cases. It supports a number of department functions including patrol development, special operations, and tactical units, investigations, planning and research, crime prevention, and administrative services.

Crash analysis can be divided into three categories:

Tactical: An analytical process that provides information used to assist operations personnel (patrol and investigative officers) in identifying specific and immediate crime trends, patterns, series, sprees and hotspots, providing investigate leads and clearing cases. Analytical includes associating criminal activity by method of the crime, time, date, location, suspect vehicle, and other types of information.

Strategic: Concerned with long-range problems and projections of long term increases or decreases in crime stastical summaries, resource acquisition and allocation studies.

Administrative: Focuses on provision of economic, geographic, or social information to administration.

Historically, the causes and origins of crime have been the subject of investigation. Some factors known to affect the volume and type of crime occurring from place to place are:

- Population density and degree of urbanization
- Racial heterogeneity
- Variations in composition of the population, particularly youth concentration
- Stability of population with respect to residents mobility, commuting patterns, and transient factors
- Modes of transportation and highway system
- Economic conditions, including median income, poverty level, and job availability
- Cultural factors and educational, recreational, and religious characteristics
- Family conditions with respect to divorce and family cohesiveness
- Climate
- Effective strength of law enforcement agencies
- Administrative and investigate emphases of law enforcement
- Policies of other components of the criminal justice system (e.g. prosecutorial, judicial, correctional, and probability)
- Citizens attitudes toward crime
- Crime reporting practices of the citizenry

GIS helps identify potential suspects to investigate suspect base when no leads are evident

Investigate serial offences: The home addresses, work addresses, and places frequented by victims and suspects of serial offences can be displayed. When details of both victims and suspects are viewed in this manner, intersections between the two can be seen. The spatial patterns of less serious offences that often lead up to more serious crimes can be examined. This is especially important since serial rapists often begin their criminal carrier's committing other offences such as peeping tom incidents, indecent exposures, or residential burglaries.

Predict Behavior: Based on the crime history, models can be developed to predict behavior of criminals. Using these models, the probable location of an offender's home can be identified. The geographic area where the next crime most likely to occur can also be predicted.

Identify suspects: crime data consisting of spatial information and methods of operation characteristics can be combined with arrest locations to provide investigators a tool to identify potential suspects. Investigators can quickly see hotspot locations and the people who have been previously arrested for the particular crime around these locations.

3. HOTSPOT ANALYSIS:

This tool mainly useful to identifies statistically significant spatial clusters of high values (hotspots) and low values (cold spots). It creates a new output feature class with a z-score and p-score value for each feature in the Input Feature Class. It also returns the z-score and p-value field names as derived output values for potential use in custom models and scripts.

The z-score and p-values are measure of statistical significance which tells you whether or not to reject the null hypothesis, feature by feature. In effect, they observe whether the observed spatial clustering of high and low values is more pronounced than one would expect in a random distribution of those same values.

A high z-score and small p-value for a feature indicates a spatial clustering of high values. A low z-score and small p-values indicates the spatial clustering of low values. The higher (or lower) the z-score, the more intense the clustering. A z-score near zero indicates no apparent spatial clustering. The z-score is based on the randomization of computation.

3.1 Z-SCORE AND P-VALUE:

Most of the statistical tests begin by identifying a null hypothesis. The null hypothesis for the pattern analysis tools is complete spatial randomness (CSR), either of the features themselves or the values associated with those features.
The z-scores and p-values returned by the pattern analysis tools tell you whether you can reject that null hypothesis or not.

Often, you will run one of the pattern analysis tools, hoping that the z-score and p-value that you can reject that null hypothesis, because it would indicate that rather than a random pattern, your features exhibit statistically significant clustering or dispersion. Whenever you see spatial structure, like clustering in the landscape, you are seeing evidence of some underlying spatial process at work, and as a geographer or GIS analyst, this is often what you are most interested in.

The p-value is a probability. For the pattern analysis tool, it is the probability that the observed spatial pattern was created by some random process. When the p-value is very small, it means it is very unlikely that the observed spatial pattern is the result of random processes, so you can reject the null hypothesis.

Z-scores are simply standard deviation. If, for example, a tool returns a z-score of +2.5, you would say that the result is 2.5 standard deviations. Both z-scores and p-values are associated with the standard normal distribution.

3.2 PROCEDURE TO GENERATE HOTSPOTS:

Calculation based on either Euclidean or Manhattan distance require projected data to accurately measure distances. For line and polygon features, feature centroids are used in distance computations. For multi points, polylines, or polygons with multiple parts, the centroid is computed using the weighted mean center of all feature parts. The weighting for point features in 1, for line features is length, and for polygon features is area.

The input field should contain a variety of values. The math for this statistic requires some variation in the variable being analyzed; it cannot solve if all input values are 1, for example. If you want to use this tool to analyze the spatial pattern of incident data, consider aggregating your incident data.

Your choice for the Conceptualization of Spatial Relationships parameter should reflect inherent relationships among the features you are analyzing. The more realistically you can model how features are interact with each other in space, the more accurate your results are.

Additional options for the Conceptualization of Spatial Relationships parameter are available using the Generate Spatial Weights Matrix or Generate Network Spatial Weights tools. To take advantage of these additional options, use one of these tools to construct the Spatial Weights Matrix File prior to analysis.

Maps layers can be used to define the Input Feature Class. When using a layer with a selection, only the selected features are included in the analysis. If you provide a Weights Matrix File with a .swm extension, this tool is expecting a spatial weights matrix file created using the Generate Spatial Weights Matrix tool. For analysis more than 5000 features, .swm formatted file is used. When this tool runs in ArcMAP, the Output Feature Class is automatically added to the Table of Contents (TOC) with default rendering applied to the z-score field.

4. ABOUT THE MODEL

Step 1: Collect events

Converts event data such as accidents or disease incidents, to weighted point data.

- Collect event combines coincident points: it creates a new output feature class containing all of the unique locations found in the input feature class. It then adds a field named ICOUNT to hold the sum of all incidents at each unique location.
- This tool will only combine features that have the exact same X and Y centroid coordinates. You may want to use the integrate tool to snap nearby features together pripr to running the collect events tool.
- The integrate tool permanently alters feature geometry; always make a backup copy of your feature class prior to using integrate.
- The hot spot analysis (Getis-ord Gi*), cluster and outlier analysis (Local Moran's I), and spatial autocorrelation tools, for example, require weighted points rather than individual incidents. Collect events can be used to create weights when the input feature class contains coincident features.
- Although this tool will work with polygon or line data, it is really only appropriate for event, incident, or other point feature data. For line and polygon features, feature coincidence is based on feature true geometric centroids. For multipoint, polyline, or polygons with multiple parts, the centroid is computed using the weighted mean center of all feature parts. The weighting for point features is 1, for line features is length, and for polygon features is area.
- If you want each individual point/part of multipoint/multipart data treated as single part features, run the multipart to single part tool, then run collect events on the single part feature class. For mare information, see processing multipoint data.
- In addition to the Output Feature Class, this function passes, as derived output values, the name of the count field and the maximum count value encountered for any one location. These derived output values are helpful when you use this tool in models or scripts.

Step 2: Generate network spatial weight

Constructs a spatial weights matrix file (.swm) using a Network dataset, defining feature spatial relationships in terms of the underlying network structure.

Usage

- Output from this tool is a spatial weight matrix file (.swm). tools that require you to specify a conceptualization of spatial relationships option will accept a spatial weights matrix file; select GET SPATIAL WEIGHT FROM FILE for the conceptualization of spatial relationships parameter and, for the weight matrix file parameter, specify the full pathname to the spatial weights file created using this tool.
- This tool was designed to work with point input feature class data only.
- A spatial weight matrix quantifies the spatial relationships that exist among the features in your dataset. Many tools in the spatial statistics toolbox evaluate each feature within the context of its neighboring features. The spatial weights matrix file defines those neighbor relationships. For this tool, neighbor relationships are based on the time or distance between features, in the case where travel is restricted to a network. For more information about spatial weights and spatial weight matrix files, see spatial weights.

Step 3: Hot spot analysis

Calculates the Getis-Ord Gi* statistic for hot spot analysis and then applies a cold-to-hot type of rendering to the output z-scores.

Usage

- FIXED DISTANCE BAND

  The default value for the Distance Band or Threshold Distance parameter ensures that each feature has at least one neighbor, and this is important. But often, this default will **not** be the most appropriate distance to use for your analysis.

- <u>INVERSE DISTANCE or INVERSE DISTANCE SQUARED</u>

  When 0 is entered for the Distance Band or Threshold Distance parameter, all features are considered neighbours of all other features; when this parameter is left blank, the default threshold distance is applied.

  Weights for distances less than 1 become unstable. The weighting for features separated by less than one unit of distance (common with geographic coordinate system projections) is 1.

  Analysis on features with a geographic coordinate system projection is not recommended when you select any of the inverse distance-based spatial conceptualization methods (INVERSE DISTANCE, INVERSE DISTANCE SQUARED, or ZONE OF INDIFFERENCE).

For these Inverse Distance options, any two points that are coincident are given a weight of 1 to avoid zero division. This ensures features are not excluded from analysis.

- Additional options for the Conceptualization of Spatial Relationships parameter are available using the Generate Spatial Weights Matrix or Generate Network Spatial Weights tools. To take advantage of these additional options, use one of these tools to construct the spatial weights matrix file prior to analysis; select GET SPATIAL WEIGHTS FROM FILE for the Conceptualization of Spatial Relationships parameter; and, for the Weights Matrix File parameter, specify the path to the spatial weights file you created.
- Map layers can be used to define the Input Feature Class. When using a layer with a selection, only the selected features are included in the analysis.
- If you provide a Weights Matrix File with a .SWM or .swm extention, this tool is expecting a spatial weights matrix file created using either the Generate Spatial Weights Matrix or Generate Network Spatial Weights tools. Otherwise this tool is expecting an ASCII formatted spatial weights matrix file. In some cases, behavior is different depending on which type of spatial weights matrix file you use:

  -

- **ASCII formatted spatial weights matrix files:**
  - Weights are used AS-IS. Missing feature-to-feature relationships are treated as zeros.
  - The default weight for self-potential is zero, unless you specify a Self-Potential field value, or include self-potential weights explicitly.
  - If the weights are row standardized, results will likely be incorrect for analyses on selection sets. If you need to run your analysis on a selection set, convert the ASCII spatial weights file to a .swm file by reading the ASCII data into a table, then using the CONVERT_TABLE option with the Generate Spatial Weights Matrix tool.
- **.SWM formatted spatial weights matrix file**
  - If the weights are row standardized, they will be re-standardized for selection sets. Otherwise weights are used AS-IS.
  - The default weight for self-potential is one, unless you specify a Self-Potential field value.

- Running your analysis with an ASCII formatted spatial weights matrix file is memory intensive. For analyses on more than about 5000 features, consider converting your ASCII formatted spatial weights matrix file into a .swm formatted file. First put your ASCII weights into a formatted table (using Excel, for example). Next run the Generate Spatial Weights Matrix tool using CONVERT_TABLE for the Conceptualization of Spatial Relationships parameter. The output will be a .swm formatted spatial weights matrix file.
- The Modeling Spatial Relationships help topic provides additional information about this tool's parameters.

5. PROCEDURE:

Step 1:

- The satellite imagery data is taken that is IKONOS with spatial resolution 1m and QUICKBIRD of 0.7m resolution are taken for creating the data required for the project.
- The satellite imagery is digitized in order to get the major roads and the other roads required for the project.
- After digitization the attribute data is added from the field data that is collected from the SURVEY OF INDIA.
- This area falls under the area of the HYDERABAD covering the toposheet of no: 56/K/11/NW in 1:25,000 scales. The coordinate system assigned to this data base is UTM ZONE (44) of WGS 1984 datum.

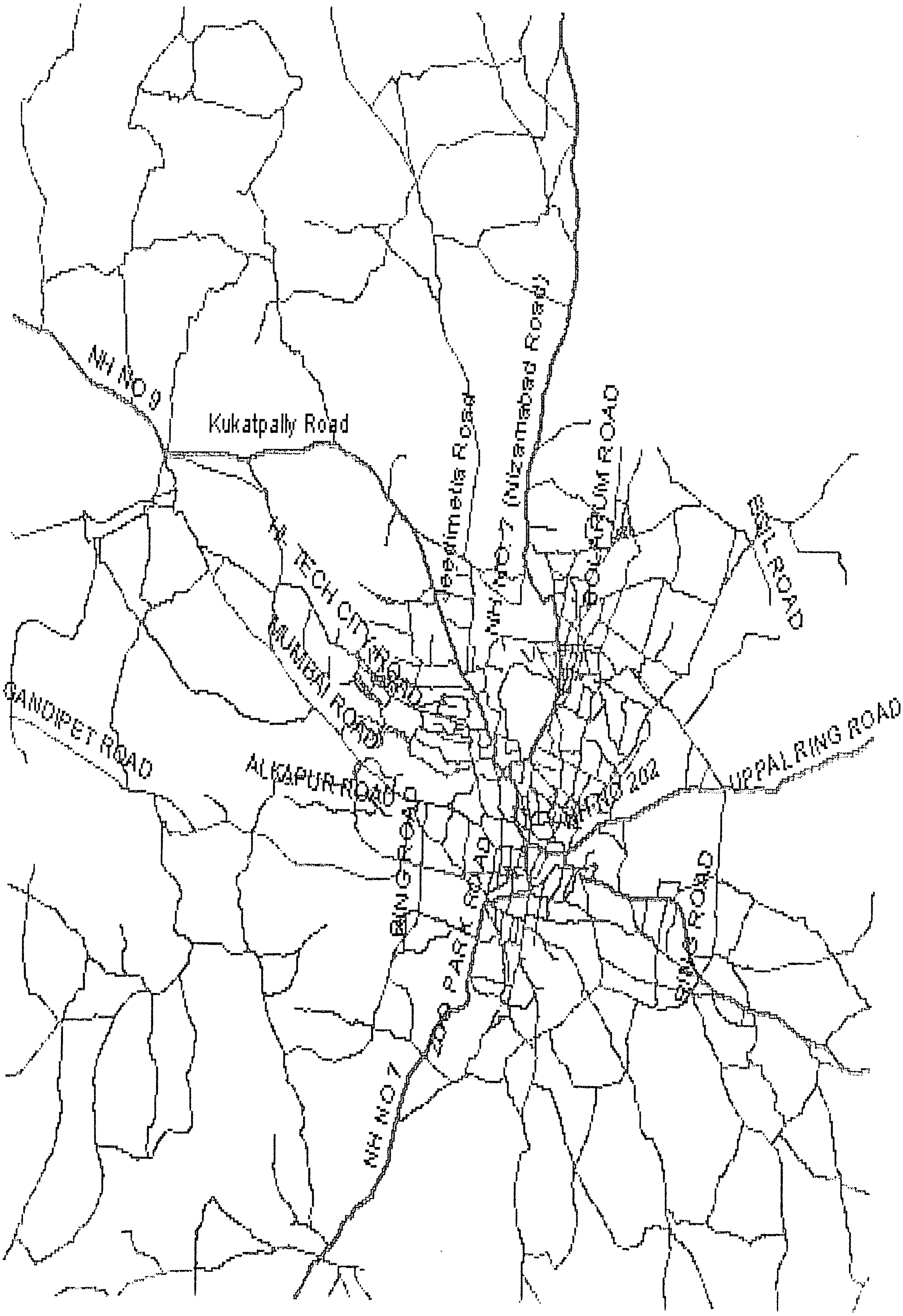


Figure 3: Study Area Taken of Hyderabad City

Step 2:

- The second part of the data is we have collected the date regarding the accidents that have taken place in Hyderabad during 2010 and 2011.

- These points are collected and they have been projected on the major roads data that has already created in the previous step.

- The same projection is also assigned to the points so they fit exactly on the location.

- The data has been created in the excel sheet and has been imported to the ArcGIS software and the layer has been created.

- This is the major data that is required for the project. From this data the crime hot spots are generated to know the area with high accident rate.

Figure 4: Complete Road Networks of Hyderabad City

Step 3 – Set up file and folder structure

- Extract the hot spot zip file you downloaded from the Resource Center and place the contents of the zip file in the following directory on your computer.
    - <your_directory>\ArcGIS Templates\Highways Roads\HotSpotAnalysis

  Note: If you do not have this folder structure on your computer, you should create it.

Step 4 – Explore the ArcMap document contents

- Open the HotSpotAnalysis.mxd.
- In the the ArcMap Catalog window, notice the contents that are included. Your Catalog window should appear as below.
    -
    - Note: in order to see the NetworkMatrixWeight.swm file, you will have to add the .swm file type to the Catalog options.
    - To do this open ArcCatalog and Go to Customize → ArcCatalog Options. On the File Types tab, you will need to specify a New Type for .swm.

Step 5: creating network dataset

- A separate geodatabase is created and the major roads, national high ways are imported as feature dataset (roads) and the crime report collected have been imported as a feature class.
- For the generation of the junction points of major roads and national high ways a network dataset is created. In order to generate hot spots these network data set is necessary so it is created.

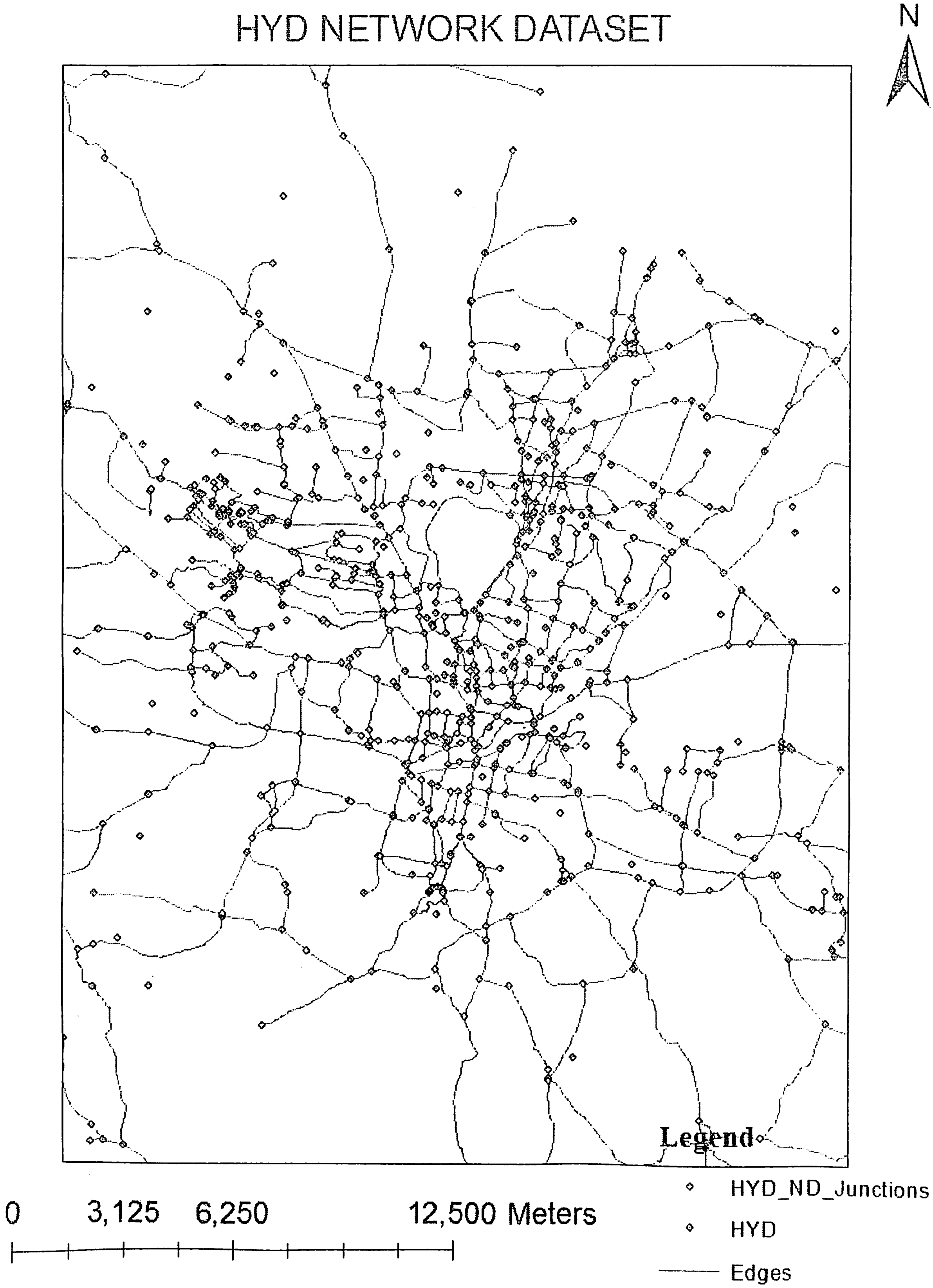


Figure 5: Hyderabad Road Network Dataset

Step 6:

- A series of connections between the different types of the tools is necessary to generate the hot spots.
- So a model is generated so that all the tools are joined in an order to complete the work.
- The following figure gives an overview of the model that has been generated and the inputs that are required to it.
- In this model for each tool different input fields and a specified output features are assigned.
- Thus the model is made ready to run.

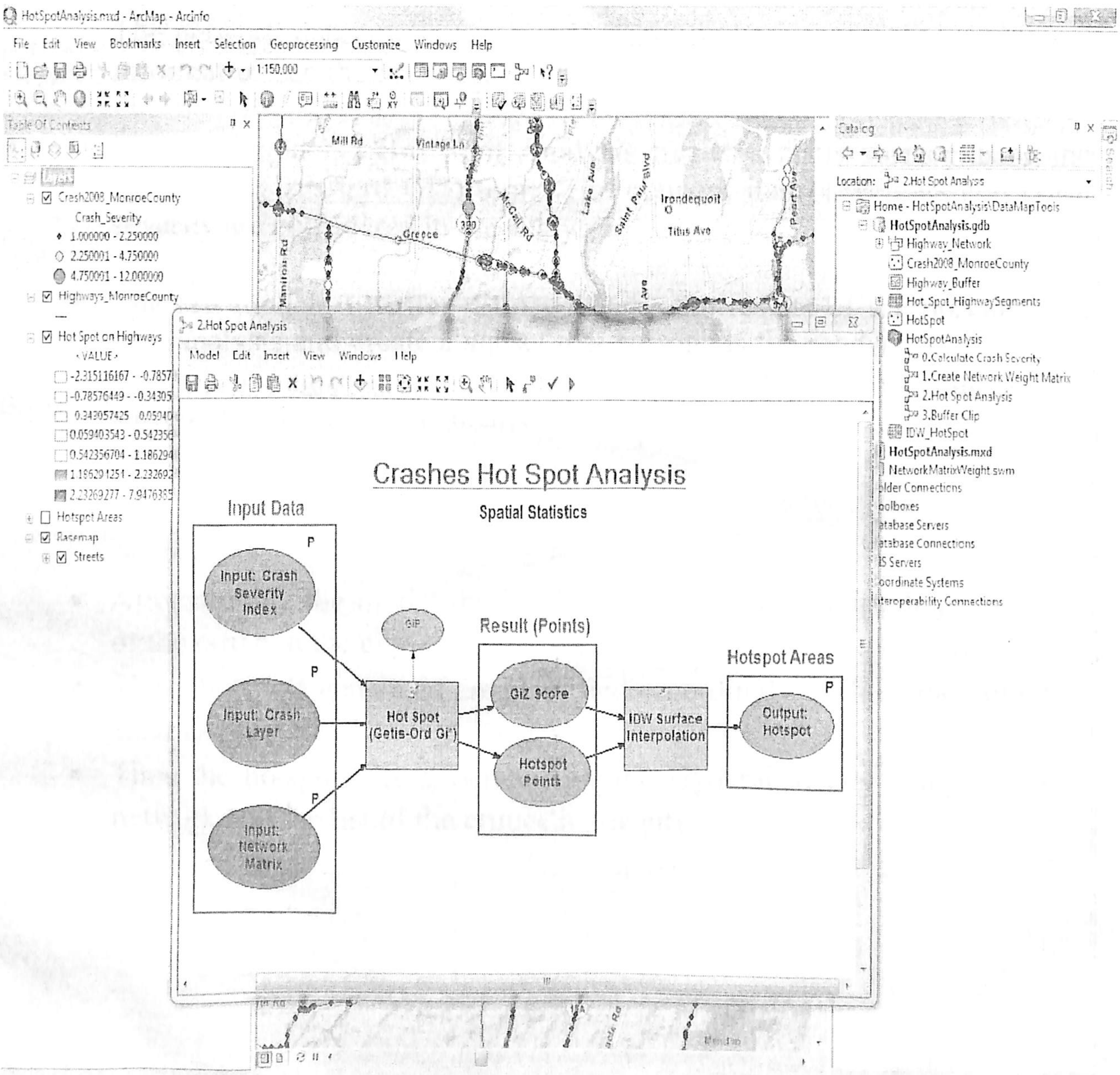


Figure 6: Building Crash Hot-Spot Analysis Model

Step 7 – Run through the Models

- The map document already shows the results of the analysis. As you go through the process of running each of the 4 models, the existing file geodatabase results will be overwritten.

- First, open the **0.Calculate Crash Severity** model. This model takes several existing attributes of the crash feature class (fatalities, injuries, and vehicles involved) and calculates a severity value. This equation may be modified to best suit your needs.

- Second, open the **1.Create Network Weight Matrix** model. This model creates the .swm file that is needed for the next Hot Spot Analysis model. The network dataset and crash points are used as inputs. For additional information see the help file.

- The third model, **2.Hot Spot Analysis** creates a raster surface using the Hot Spot (Getis-Ord GI*) tool. The network matrix weights and crash severity are considered in the analysis.

- The last model, **3.Buffer Clip** will take the results of the previous hot spot analysis and create a subset that surrounds the roadways. This is to simplify visualization of the results for only those areas immediately surrounding the street segments.

Step 8:

- After running the model the HOT SPOTS that are having high intensity of the crime in the city are generated.
- These hotspots generated are classified accordingly to show the output as clear as possible.
- Thus the hotspots are generated for the Hyderabad city using the road network and the list of the crimes in the city.

Crash hyd_data

OBJECTID *	SHAPE *	Case Number	Case Year	Reference Marker	Accident Date	Road System	Number of Fatalities	Number of Injuries	Report
1	Point	324619245623	2008	390I43031001	1/24/2008	01	0	0	Y
2	Point	327819240785	2008	15 43031001	10/31/2008	01	0	0	Y
3	Point	326619246252	2008	15 43031001	7/20/2008	01	0	0	Y
4	Point	325319246546	2008	64 43021000	3/12/2008	01	0	0	Y
5	Point	324719241817	2008	390I43031002	1/28/2008	01	0	0	Y
6	Point	327019247809	2008	390I43031002	7/21/2008	01	0	0	Y
7	Point	327119241539	2008	15 43031002	8/30/2008	01	0	2	Y
8	Point	327119248081	2008	64 43021001	9/8/2008	01	0	0	Y
9	Point	328619240388	2008	64 43021002	11/30/2008	01	0	0	Y
10	Point	326819249397	2008	15 43031004	8/3/2008	01	0	0	Y
11	Point	328119243419	2008	64 43021003	11/29/2008	01	0	0	Y
12	Point	328419243933	2008	<Null>	12/17/2008	01	0	0	Y
13	Point	327119246024	2008	15A43041002	9/7/2008	01	0	0	Y
14	Point	328319249654	2008	15A43041002	12/2/2008	01	0	2	Y
15	Point	328219242741	2008	15A43041002	12/4/2008	01	0	2	Y
16	Point	328419240225	2008	15A43041002	12/13/2008	02	0	0	Y
17	Point	324519246162	2008	940D43011002	1/8/2008	09	0	0	Y
18	Point	325119245106	2008	940D43011002	2/26/2008	09	0	0	Y
19	Point	325819245758	2008	15 43031006	4/26/2008	01	0	0	Y
20	Point	325619240319	2008	15 43031006	3/25/2008	01	0	0	Y
21	Point	324519245603	2008	15 43031006	1/15/2008	01	0	1	Y
22	Point	325319246260	2008	15 43031007	3/22/2008	01	0	0	Y
23	Point	326419243266	2008	15 43031007	7/1/2008	01	0	1	Y
24	Point	326519245119	2008	390I43031007	7/5/2008	01	0	0	Y
25	Point	325719243733	2008	390I43031007	2/22/2008	01	0	0	Y
26	Point	326419248559	2008	390I43031007	7/2/2008	01	0	0	Y
27	Point	326419247708	2008	<Null>	7/1/2008	01	0	0	Y
28	Point	326319241252	2008	65 43031004	6/10/2008	01	0	0	Y
29	Point	327919249661	2008	15 43031007	11/17/2008	01	0	1	Y
30	Point	327119248830	2008	15A43041005	9/10/2008	01	0	0	Y
31	Point	324619245981	2008	390I43031008	1/25/2008	01	0	1	Y
32	Point	325019247927	2008	65 43031005	2/6/2008	01	0	0	Y
33	Point	325919249938	2008	64 43021007	5/8/2008	01	0	0	Y
34	Point	324419248228	2008	65 43031006	1/7/2008	01	0	1	Y
35	Point	328419247768	2008	64 43021007	12/17/2008	01	0	0	Y
36	Point	327019243139	2008	15 43031009	8/18/2008	01	0	0	Y
37	Point	326619242014	2008	15 43031009	7/10/2008	01	0	0	Y

Figure 7: List of Crashes Taken place in the city

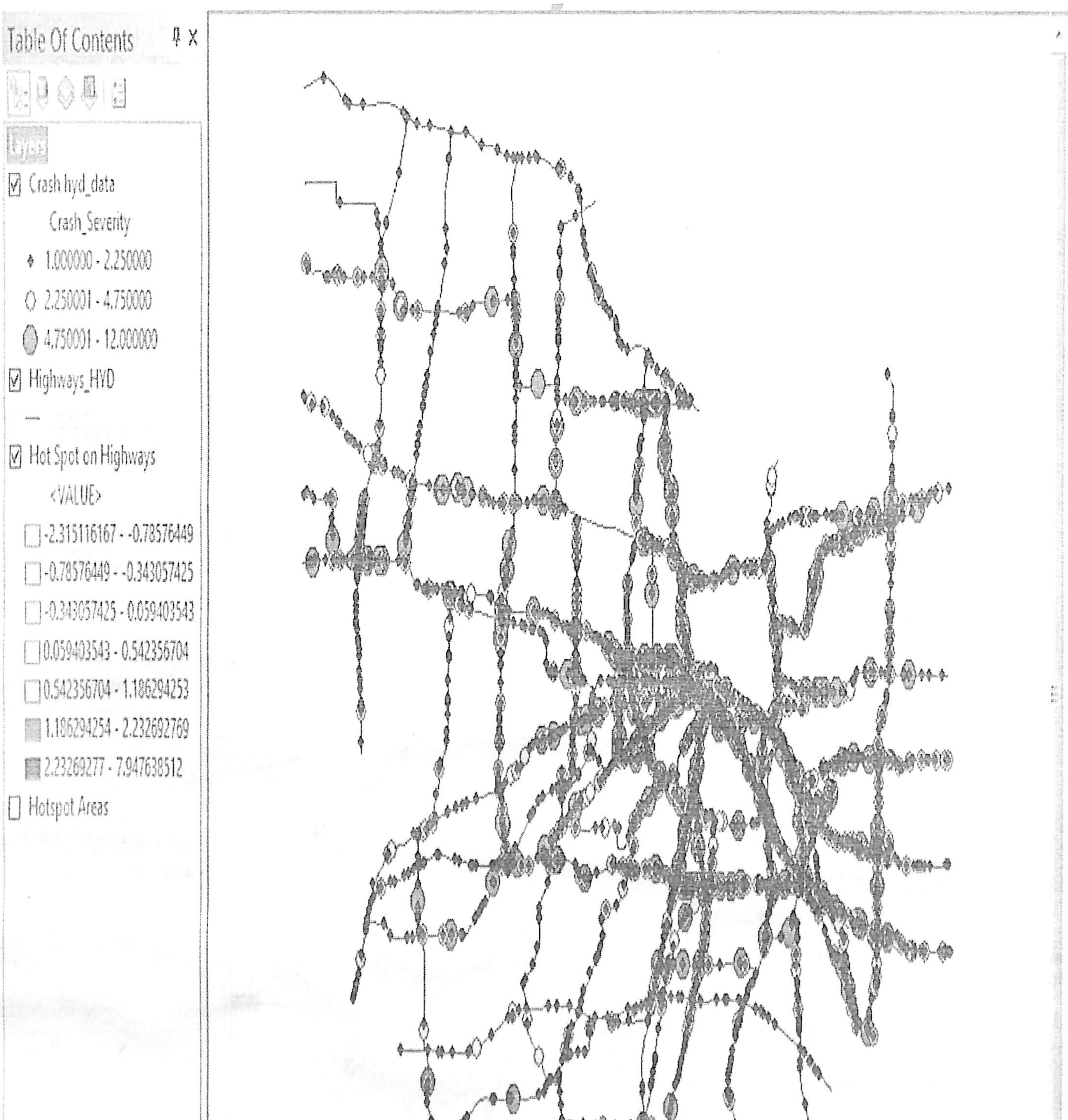


Figure 8: Finding Hot-Spots

Step 9 –Try on your own data

- Finally, try performing this analysis on your own data. Be sure that your street network is topologically connected and network connectivity is complete.

Also, your crash points must sit directly on top of your road network.

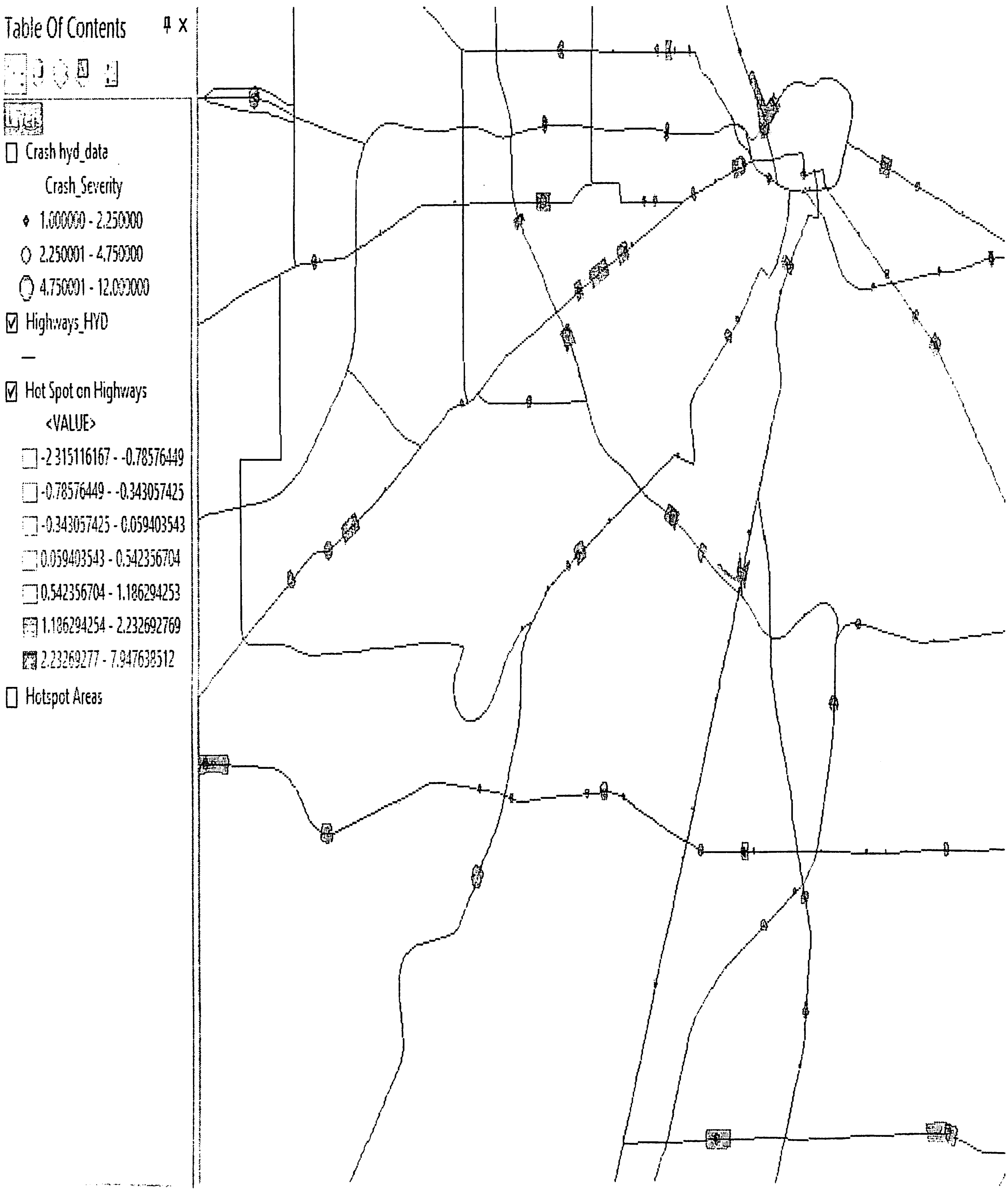


Figure 9: Areas of Maximum Accidents Rate

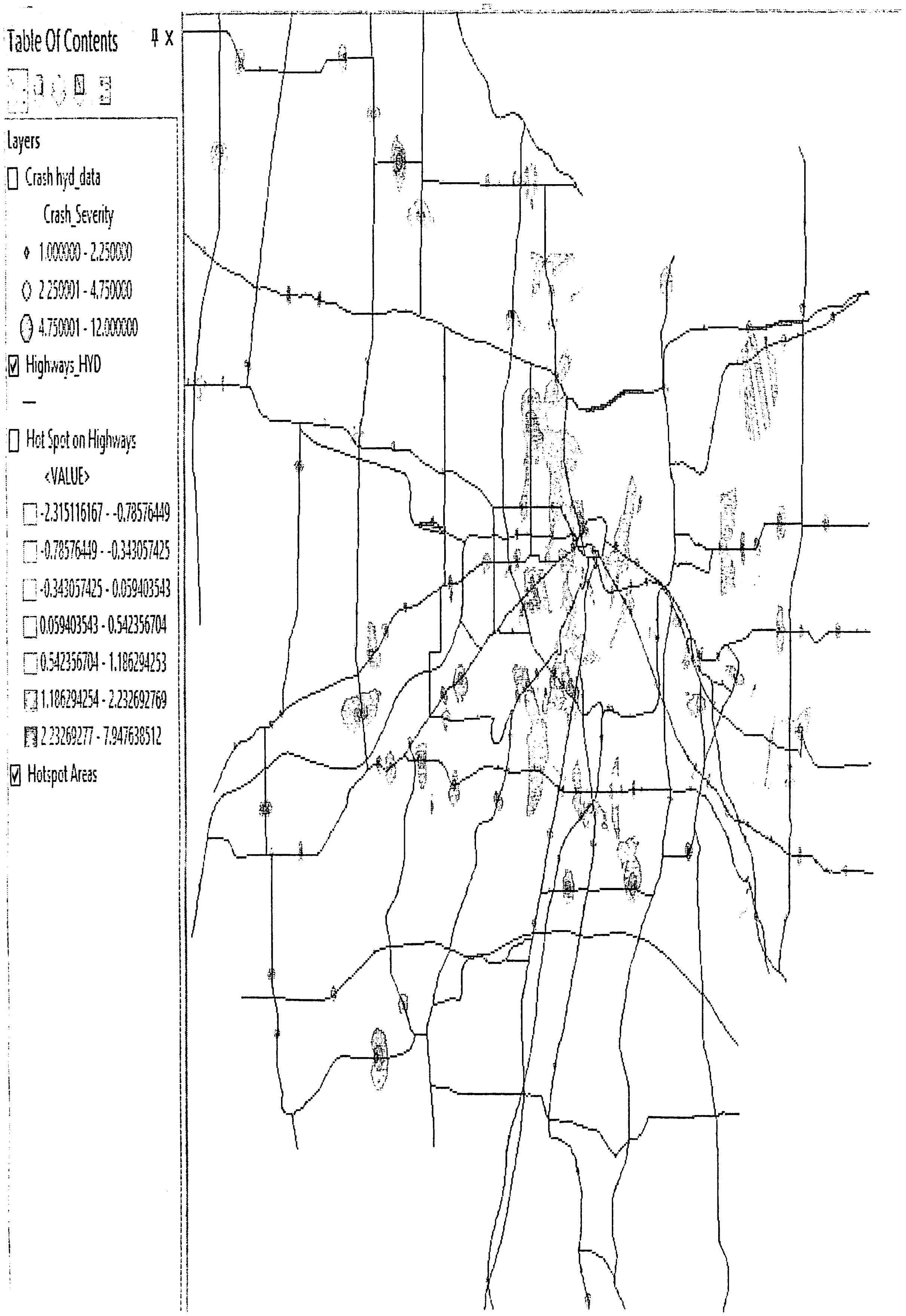


Figure 10: Generating IDW

6. Conclusion:

The overall goal of this project was to illustrate the advantages of using a GIS-based crash referencing system for high-crash identification and analysis. There are several that were applied during the conduct of this study:

Locational Referencing The spatial referencing capabilities of a GIS to locate links and nodes of a highway system in two-dimensional coordinates and connect roadways together is what made the development of the corridor analysis program feasible. As previously described, the program designed for this effort traversed a given route, located each intersection, and traveled up to 3 miles from that intersection in all directions on all possible routes. This type of analysis would have been virtually impossible to do manually.

Incorporation of Nontraditional Data Bases Data bases that are not traditionally used in crash analyses (e.g., census data) can be incorporated and examined for possible relationships to crashes. Data sets that can be translated to spatial coordinates, such as the trucking terminal data used above, may also be incorporated for this type of analysis.
Visual Analysis One of the most powerful features of a GIS based system is the capability to produce figures or maps that can be used to visually assess the problem. This fact was demonstrated in this effort in several ways:

1) The plot of high-crash corridor segments overlaid on the route system quickly showed that 50 percent of the segments were not on designated truck routes.

2) Plotting the truck crashes on the high-crash zone map quickly showed the predominate routes and minor intersections within each zone where crashes were occurring most often.

3) The trucking terminal map overlaid on the route system indicates which terminals are inside and outside the 3-mile drivable buffer. For both present and proposed terminals outside the buffer, this system would allow one to determine the most appropriate routes on which to allow those trucks. For example, the lane widths, shoulder types, and crash histories could be shown for all routes being considered and factored into the decision. Overall, this project produced results which were both useful in terms of the topic analyzed, i.e., high-truck-crash locations, and with respect to illustrating how a GIS crash-based referencing and analysis system can be used

7. REFERENCES

1. *Referencing and Analysis System,* HSIS Summary Report, Publication No. FHWA-RD-99-081. Federal Highway Administration, Washington, DC, February, 1999.

2. *Crash Referencing and Analysis System—User Manual*, North Carolina Center for Geographic Information and Analysis, Raleigh, NC, 1996.

3. D.L. Harkey, "Evaluation of Truck Crashes Using a GIS Based Crash Referencing and Analysis System," presented at the TRB Annual Meeting, January 1999, to be published by the Transportation Research Board, Washington, DC, 1999.