AUTOMATED CONVERSION OF MILEPOINT DATA TO INTERSECTION/LINK
STRUCTURE: AN APPLICATION OF GIS IN TRANSPORTATION

by

Balakrishna Akundi

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of
Master of Science
in
Civil Engineering

APPROVED:

Dr. Donald R. Drew, Chairman

Dr. Antonio A. Trani

Dr. Wende O'Neill

July 20, 1990
Blacksburg, Virginia
LD
5655
V855
1990
A369
C.2
AUTOMATED CONVERSION OF MILEPOINT DATA TO INTERSECTION/LINK STRUCTURE: AN APPLICATION OF GIS IN TRANSPORTATION

by

Balakrishna Akundi

Dr. Donald R. Drew, Chairman
Civil Engineering

(ABSTRACT)

Network data restructuring is an essential function in Geographic Information Systems (GIS) when adapted to transportation. Implementing effective data restructuring models in GIS allows users to collect and maintain data in the format with which they are most familiar while allowing others to utilize it in a format they require. Aggregation and disaggregation of network data facilitates storage, display and plotting times. There could however be important adverse effects. The effects of generalizing attribute data for aggregation/disaggregation is being researched here.

Milepoint referenced data in road inventory files provides valuable network information for transportation research. Individual records in these files represent variable length sections of roads. A new record is created each time a highway attribute changes. Consequently, a segment of road between two intersections may be represented by several records in a road inventory file. Further, all attributes in these records are associated with both directions of travel along a road.

Many transportation analysis models require networks to be represented by a node-link structure where nodes symbolize an intersection of two or more roads. Further, if a road is two directional, it is represented by two links each of which has its own set of attributes. To utilize road inventory data in these analysis models, network information has to be converted into an intersection/link format. This process
involves aggregating and disaggregating attribute data to represent longer and shorter road segments and also disaggregating data into bi-directional information.

This thesis describes data conversion efforts needed to produce intersection/link network representations. The development of a microcomputer model for data conversion is detailed and application issues and model sensitivities are addressed.
I would like to thank my advisor Dr. Wende O’Neill for her excellent guidance and support. Without her help and encouragement, I would never have been able to finish my Master’s program. I am very thankful to Dr. Donald Drew for having agreed to be the Chairman of my committee and to Dr. Antonio Trani for having agreed to be on my committee. To Dr. Antoine Hobeika for having supported financially for the better part of my masters program. To my wife Shallu for her love and affection. To Shanthi, Chakki and Siva for all their help and good wishes.
List of Tables

Table 1. Conversion Rules Tested in Capacity Calculations .................. 31
Table 2. Road Inventory Variables ............................................. 37
Table 3. Road Inventory Variables in HPMS Sections ....................... 38
Table 4. Number of Non-consecutive Records ................................. 46
Table 5. County Conversion Statistics by Abstraction Levels .............. 52
# List of Illustrations

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Variable classification for road inventory variables</td>
<td>36</td>
</tr>
<tr>
<td>2</td>
<td>Flow-chart for the conversion program</td>
<td>42</td>
</tr>
<tr>
<td>3</td>
<td>Graph of RMSE for 64 conversion rules, Abstraction level &gt; 5000</td>
<td>57</td>
</tr>
<tr>
<td>4</td>
<td>Graph of RMSE for 64 conversion rules, Abstraction level &gt; 1000</td>
<td>58</td>
</tr>
<tr>
<td>5</td>
<td>Graph of RMSE for 64 conversion rules, Abstraction level &gt; 800</td>
<td>59</td>
</tr>
<tr>
<td>6</td>
<td>Graph of RMSE for 64 conversion rules, Abstraction level &gt; 700</td>
<td>60</td>
</tr>
<tr>
<td>7</td>
<td>Distribution of records per link, Abstraction level &gt; 5000</td>
<td>61</td>
</tr>
<tr>
<td>8</td>
<td>Distribution of records per link, Abstraction level &gt; 1000</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>Distribution of records per link, Abstraction level &gt; 800</td>
<td>63</td>
</tr>
<tr>
<td>10</td>
<td>Distribution of records per link, Abstraction level &gt; 700</td>
<td>64</td>
</tr>
<tr>
<td>11</td>
<td>Number of links conversion rules minimizes, Abstraction</td>
<td>65</td>
</tr>
<tr>
<td>12</td>
<td>Number of links conversion rules minimizes, Abstraction</td>
<td>66</td>
</tr>
<tr>
<td>13</td>
<td>Number of links conversion rules minimizes, Abstraction</td>
<td>67</td>
</tr>
<tr>
<td>14</td>
<td>Number of links conversion rules minimizes, Abstraction</td>
<td>68</td>
</tr>
<tr>
<td>15</td>
<td>Number of links conversion rules minimizes, Caroline County</td>
<td>69</td>
</tr>
<tr>
<td>16</td>
<td>Number of links conversion rules minimizes, Hanover County</td>
<td>70</td>
</tr>
<tr>
<td>17</td>
<td>Graph of RMSE for 64 conversion rules, Caroline County</td>
<td>71</td>
</tr>
<tr>
<td>18</td>
<td>Graph of RMSE for 64 conversion rules, Hanover County</td>
<td>72</td>
</tr>
<tr>
<td>19</td>
<td>Graph of RMSE for 64 conversion rules, Charles City County</td>
<td>73</td>
</tr>
<tr>
<td>20</td>
<td>Graph of RMSE for 64 conversion rules, Spotsylvania County</td>
<td>74</td>
</tr>
</tbody>
</table>
# Table of Contents

1.0 Literature Review .................................................. 1

1.1 Introduction ..................................................... 1

1.2 Geographic Information Systems .................................. 2
  1.2.1 Spatial Data .................................................. 3
  1.2.2 Base Map Creation ........................................... 5

1.3 Topologically Integrated Geographic Encoding and Referencing System .................. 6

1.4 GIS Software ..................................................... 7

1.5 GIS in Transportation Agencies ................................... 7

1.6 Highway Data ..................................................... 10

1.7 Duplication of Data ................................................ 10

1.8 Road Inventory Data ............................................... 11

1.9 Road Inventory Files ............................................. 13

1.10 Network Data Aggregation, Disaggregation Issues ............................... 14

1.11 Network Abstraction ............................................ 16

1.12 Highway Location Reference Systems ................................ 18
  1.12.1 Route-Milepoint Method ..................................... 19
    1.12.1.1 Milepost/Milepoint .................................... 20
    1.12.1.2 Control Section ....................................... 20
  1.12.2 VDOT’S Road Inventory Files ................................ 21
  1.12.3 NodeLink Method ........................................... 22
    1.12.3.1 Advantages of Node-Link Method ...................... 23
1.0 Literature Review

1.1 Introduction

Highway data is an essential component of any analysis related to transportation planning or traffic study. Vast amounts of data are collected and stored in different formats and structures by Federal, state, and local transportation agencies. The collection of highway related data involves a wide variety of activities such as traffic counts, sign inventories, skid resistance measurements, photologging, accident investigation, recording of construction and maintenance projects and funding, and roadside obstacles, speed monitoring, right of way surveys, bridge inspection, AADT, geometric characteristics, and other data-collection and maintainence activities.

The following sections describe some of the basic aspects of highway data found in transportation data bases and how such data bases are being used for transportation planning.

Many transportation agencies currently are investigating adaptation of Geographic Information Systems (GIS) to transportation. One issue for consideration focuses on data structures, specifically representation of transportation networks in digital data bases. Solutions requiring distinct groups within planning agencies to conform to one specific structure are unrealistic and unsatisfactory. A model developed here demonstrates that different network structures may be maintained
by individual groups. Further, various data bases may be shared among groups thus increasing information utilization and lessening redundancy.

1.2 Geographic Information Systems

There are many different ways in which Geographic Information Systems are currently being defined. Among them are:

- A Geographic Information System is defined as an integrated data base containing information about georeferenced spatial objects - points, lines and areas, plus the software and hardware used by personnel to manipulate these objects(11).

- A geographic information system (GIS) combines visual information (maps) and descriptive data (attributes). A GIS allows users to relate information from different sources and in different forms on the basis of the geographical location of the features being described (12).

- A Geographic Information System is a computerized data base management system for the capture, storage, retrieval, analysis, and display of spatial or locationally defined data(13).

GIS is a tool whose applications are limited only by the sophistication of the hardware and software, the quality of the data and the imagination of the users.

Many state departments of transportation and planning organizations have realized the value of having a GIS and are in the process of adopting GIS concepts and technology into their organizations. For a long time now, many of these agencies
had been using applications of computer aided drafting and design (CADD) systems and computer assisted cartography.

Geographical information systems differ from computer graphics; the latter are largely concerned with the display and manipulation of visible material. Computer graphics systems do not pay much attention to the non-graphic attributes that the depicted features may have and which might be useful for data analysis. Good computer graphics are essential for a modern geographical information system, but a graphics package by itself is not sufficient for the tasks expected of a GIS(8).

Geographical information systems do have a lot in common with computer-aided design (CAD) systems. Both GIS and CAD systems need to be able to relate objects to a frame of reference, to handle non-graphic attributes and to be able to describe topological relations. The major differences between GIS and CAD systems are the much greater volume and diversity of the data input to GIS systems and specialized nature of the analysis methods used(8).

The GIS functionality needed by transportation organizations as stated by Nyerges and Dueker (11) consists of:

- Geographically structured data, with point, line and area relations, i.e topology;
- linkage of locational and attribute data about transportation facilities.
- analytical map overlay of separate map data

### 1.2.1 Spatial Data

A GIS is capable of topological operations; it understands how elements contained in the data base are related to each other spatially and it can perform spatial manipulations on these elements. For a GIS to be functional, information has
to be available in two different forms: geocoded spatial data and attribute data. Geocoded spatial data is available either by tying the manually digitized data to a spatial location referencing system or by using Topologically Integrated Geographic Encoding and Referencing System (TIGER) files made by the Bureau of the Census. Attribute data is found in the road inventory files. Attributes associated with a street include its width, number of lanes, pavement condition, type of facility, average annual daily traffic etc. Attribute data has to be associated with a topologic object like a point, line or a polygon that has a position somewhere on the surface of the earth.

A transportation network/system consists of nodes, links and other entities distributed in two or three dimensional space. A transportation data base like the road inventory files, despite containing a lot of information, does not lend itself easily to plotting and spatial analysis. This is because the data in these files lack topology or in other words they are not spatial. Also they contain a lot of information that is redundant and not required by the GIS. Generally the data is not referenced to a geographical coordinate system (13). With just the traditional milepoint or reference point data base spatial analysis is not possible.

A GIS is supposed to add a degree of intelligence and sophistication to the transportation data base (13). For example with a GIS we can easily determine what routes intersect a particular route, the presence of an actual intersection and its location with respect to a particular place. With a GIS, it is possible to perform geographic queries which earlier was restricted to textual queries.

The distinction between a GIS and other computer aided tools like CAD systems has to be clearly understood. A transportation demand model that has graphical representation is not a GIS. Such models have a rigid file structure suitable only for the task at hand. Such networks, while displaying a graphical abstraction of the nodes
and links, lack in topology or spatial relationship. Without topological intelligence, spatial analysis is difficult, if not impossible (13).

1.2.2 Base Map Creation

A GIS requires digital data with geographic coordinates. A map image of the highway network can be put into the computer by either manual digitizing or automated scanning. While manual digitizing is relatively cheap but very time consuming, automated scanning is very expensive, less time consuming and more accurate.

Manual digitizing is done on a digitizing tablet to which the original map is attached. The original map has to be available on paper or some other physical medium. The digitizing tablet is provided with a cursor or a pointing device which is used to locate points on the source map. The digitizing tablet transfers the position of the cursor to the computer which records the X, Y coordinates and projects the position of the point onto the monitor. Manual geocoding is the slowest method of inputting cartographic base data.

Scanning is the process of electronically reading a map and converting it to a point image. This process is most successful when the map image being scanned is drawn clearly and contains only one map feature.

Existing digital files available from federal agencies are excellent sources of digital data. Digital map products are available from the U. S. Geological survey (USGS). Its National Mapping program produces Digital Line Graph (DLG) files that are topologically structured. DLG’s are geocoded in latitude/longitude. Attribute
data however must be attached to DLG's by the user. This requires the use of sophisticated software.

1.3 Topologically Integrated Geographic Encoding and Referencing System

TIGER, which stands for Topologically Integrated Geographic Encoding and Referencing System, is a topological data base containing every street and block of the United States for use in the 1990 census. TIGER will significantly affect the way transportation planning will be done in the 1990's. A transportation agency will no longer have to go through the expensive and time consuming process of digitizing a network to create a base map. TIGER contains nearly every street digitized to an acceptable level of accuracy.

In order to effectively use TIGER for transportation planning, it is necessary to either develop or acquire software that can use the complex topology implicit in TIGER. These files contain roads, waterways, pipelines, and other facilities that are arranged in random order. The user has to build an ordered, connected transportation network from the data base. Attribute information is not included in these files.
1.4 GIS Software

A GIS requires specialized and sophisticated software to run. A typical highway network will consist of a number of points, intersections and other cartographic and topological features(15). Attribute data such as those found in the road inventory files are associated with the route links. The main advantage of a GIS is its ability to display networks on the screen. The user is able to plot sections of the network and perform spatial analysis on the plotted network. It also allows to zoom in on certain sections to carry out microscopic analysis.

GIS software should be capable of overlaying data to calculate the union or intersection of cartographic objects. It should be able to read data stored in latitude/longitude, state plane coordinates, milepoint or some other reference system and then convert them into the specified GIS system.

1.5 GIS in Transportation Agencies

GIS concepts and technology can help transportation organizations to better deal with information handling and display according to their needs. A GIS will provide access to spatially oriented data for decision making(14).

The U.S. Department of Transportation recently listed six primary benefits that may be derived from the use of a GIS:

- Retrieving transportation operations information on a geographical basis,
- linking data geographically for new applications,
• visualizing relationships through thematic mapping,
• computing spatial relationships using coordinates and transportation feature characteristics,
• building more complete network models for use in analysis, and
• editing data on a geographical basis using both graphical and nongraphical displays for data base management.

Leading the way in the application of Geographical Information Systems to transportation is the Federal Highway Administration (FHWA). FHWA is helping the Census bureau and the U.S. Geological survey in a project to assess the applicability of the Census TIGER file for transportation planning and analysis. A project initiated by the Office of Planning is being handled by Johnson City, Tennessee and the University of Tennessee Department of Civil Engineering in an effort to provide on-line links between computerized traffic demand models and Geographic Information Systems.

The FHWA has also sponsored the development of the Geographic Roadway Information Display System (GRID S). GRIDS is an interactive microcomputer program that accesses and displays data about the U. S. Interstate system. The GRIDS database is a subset of the data contained in the Highway Performance Monitoring System (HPMS) data base. Data can be presented on the map display and band widths used to distinguish between heavily and lightly traveled roads.

The FHWA Office of Policy is also responsible for the development of the Highway Traffic Forecasting System (HTFS). This is basically a decision support system designed primarily to analyze truck size and weight issues. The existing model is being merged with the TransCAD GIS software to provide for greater data entry, editing, display and analysis capabilities. The FHWA Office of Planning is also enhancing the Oak Ridge National Laboratories' (ORNL) National Highway Network
with additional links and attributes. The data base will include socio-economic data from the Census tapes, county business studies collected by the Federal Government and the Bureau of Economic Analysis Forecasts. The system is meant to be used for policy analysis and decision support (13).

The Wisconsin Department of Transportation has been using photolog to assist in the planning, design and operation of the state highway system. The system can be used in the office to verify and edit data in a number of inventory data bases and to build a new data base. WisDot's photolog system creates 35mm camera images every 0.01 mi. The photo images are transferred directly from negatives to a video disk that can hold up to 24,000 frames or 240 miles of highway. A microcomputer can access any frame in less than a second. This method is being used at WisDot to locate any point on the highway network and display associated attribute information.

Apart from the activities of the FHWA and state agencies, there are other tasks that are of significant importance to the transportation planner that can be performed using a GIS. One such task is transportation demand modeling. Transportation demand modeling requires the building of abstract travel networks involving a number of nodes, links and population and employment centroids. All these can be accomplished within a GIS much faster and with greater accuracy. A GIS will also make it easier to perform sub area analysis and to modify the traffic analysis zone structure.
1.6 **Highway Data**

The emphasis of the Nation’s highway program has shifted from the construction of new facilities to the reconstruction and preservation of existing facilities. This needs reliable, relevant, and economical data. In this world of fast changing technology, there has been significant progress in the collection, processing, storage, access, and analysis techniques of highway data. Leading the way in the collection of highway data are the Federal Highway Administration and State Departments of Transportation.

1.7 **Duplication of Data**

Highway data bases are typically developed to meet needs within the various organizational units of each highway agency (1). The lack of a centralised unit to oversee what each data-collection effort accomplished led to duplication of data. Various data bases were created because of requirements of the federal government or of a particular state program. The same data is collected in different formats making it incompatible with other data bases. Multiple data bases also make it difficult to edit data which are in different structures.

Not all duplication of data is wasteful (1). Some duplication or redundancy in data collection and processing often is valuable for verifying data quality. However much of the duplication found among existing data bases serves no useful purpose.
Duplication of data takes place on an interagency basis as well as within agencies. The costs of manual collection and processing of data have soared in recent years, while the costs of computer storage and processing have dropped steadily. It is important that unnecessary redundancy in manual operations and duplication of data be eliminated and better use made of the existing data.

1.8 Road Inventory Data

The lead agencies in the collection of highway related data are the state DOT’s. Every state DOT is responsible for the collection and maintainence of data related to highways falling within that state. Basic to any states data is an inventory containing the physical and geometric attributes of those segments of highway located on the state highway system(1). Before the advent of computers, such data were stored on handwritten inventory sheets or straightline representations of roadway logs. But with computerization, most states turned to digital formats for storing their highway data. In fact, roadway inventory files are among the first files to be converted to computerized formats. Most of these are based on the Highway Performance Monitoring Systems (HPMS) implemented by the Federal Highway Administration (FHWA). The FHWA stipulates that every state DOT report its road inventory in a computerized format every year. The following is a short list of the different kinds of data commonly included in the state and local highway inventory files:

- Functional class
- State System
- Signed route number
- Federal-aid system
- Route type
- Inventory route number
Governmental Control  Domain
Toll considerations  Special route
Length  # of through lanes
Type of surface  Type of pavement section
Access control  Lane width
Approach width  Shoulder width
Shoulder type  Median type
Parking information  AADT

The above listed attributes more or less constitute the entire range of information that is needed to administer a highway system. It is very difficult to find all of the above named attributes associated with a highway in one file. Most of the agencies responsible for handling highway related data maintain more than just one file. This facilitates easier collection and storage of data. For example, the Ohio Department of Transportation maintains a primary inventory file that contains data related to the facilities on the State highway system (1). In addition, it also has a local road inventory file containing data on county and township roads, as well as a municipal street inventory file that carries data on facilities in the various municipalities of the State.

The Iowa DOT maintains a summary of roadway inventory data for state primary routes in Base Record Inventory, which includes traffic and accident data as well as geometric, sufficiency, and administrative data.

The Washington State Department of Transportation (WSDOT) is implementing a roadway file as the first major component of its Transportation Information and Planning Support System (TRIPS), an integrated information system currently being developed to make WSDOT data more accessible to users(1).
Idaho's Milepost and Coded Segment System (MACS) provides a uniform method for cross-referencing various categories of data relating to roads within the state. Data linkage is accomplished by means of a road or ramp segment code, milepost, and time. Roadway features are referenced by the beginning and ending mileposts on a particular ramp or segment(1).

1.9 Road Inventory Files

Road inventory files represent an extensive data base containing highway attribute information. Records in these files represent sections of roadways such that a new record is created each time an attribute changes. The spatial component in this data base limits queries to identification of district, county, city or town to which a highway belongs. No coordinate information, such as latitude and longitude, is provided to assist in locating a link. Milepoint information is provided for each road segment terminus that does not represent a highway intersection. Additional milepoint information in each record gives the distance, in miles, from a set reference point to the beginning of the highway segment represented by the data record.

Transportation planning models typically require a network data structure in which links represent one directional sections of roads. Nodes distinguish the beginning and end of each link and represent an intersection where change in travel direction may occur. Attributes, such as those found in road inventory files, are associated with each link. Depending on the nature of the application, an intersection/link network may encompass every road in a region or simply a subset of roads.
To make use of road inventory attribute information in studies requiring an intersection/link data structure, aggregation and disaggregation of network data must take place. The manual process for performing this task is labor intensive and error prone. An automated procedure has been developed to simplify this data structure conversion process. As a result, road inventory files may be utilized in a broad range of applications.

This work describes theoretical and practical issues related to conversion from one network data structure to another. Simple rules are defined in the model to achieve an intersection/link data base. Aggregation and disaggregation impacts the accuracy of information. The sensitivity of transportation models to such a conversion is also addressed.

1.10 Network Data Aggregation, Disaggregation Issues

Aggregation and disaggregation of spatial data has been addressed primarily by geographers for polygonal data structures. Spatial data is being used in a wide variety of fields with development of data entry and handling capabilities.

In transportation sketch planning and network abstraction issues are relevant to aggregation and disaggregation of linear data. Literature on GIS applications in transportation discusses data base design, methodologies for attaching attribute data to lines, and the need for flexibility in transportation data bases to meet a variety of objectives. Unfortunately few have researched the types of errors produced from aggregating and disaggregating network data and sensitivity of transportation models to these errors.
Most urban transportation analyses rely heavily on network models; these network models are structured in the form of nodes and links with nodes representing street intersections; the path connecting two nodes is a link. The user can build a network of his choice depending on the kinds of analysis for which the network will be used (9). A transportation analysis related information base must have the capability to update the data base, edit the network, extract subsets of the network and condense the extracted network down to the level of detail required for a given analysis. Network editing consists of selective addition or deletion of network links/segments without compromising data accuracy. A detailed corridor analysis might need a subset of the network data base for which data has to be disaggregated. These different functions facilitate the user to design a network and its associated data base tailored to his needs.

Although conversion is taking place in transportation agencies, the underlying issue of how to attach attributes to a new structure without compromising accuracy of the information has not been sufficiently researched. An typical example of the situation faced in data structure conversion is shown in Figure 1. These roads are in James City County, Virginia. Note that a single line has been indicated for the link structure when in fact each line represents two links, one for each travel direction. Only three geographical locations of intersections and milestone terminus correspond. Consequently, network attribute information found in segments 1, 2, and 3 must be mapped to links 1 and 2. Data relevant to segment 5 must be reflected in attributes of links 3, 4 and 5.
1.11 Network Abstraction

The level of detail for a highway network depends on the situation for which the network is built and the desired accuracy of the solutions (10).

Transportation models need varying levels of detail for their networks depending on the task at hand. This requires building the network selectively from among all the highways available, choosing only those routes and links which could be of use to the model.

Haghani and Daskin (10) define network aggregation as “the process of condensing a given network into another one which is 1) small enough to be managed efficiently and effectively; and 2) which preserves some desired characteristics of the original network or satisfies certain objectives.”

Most transportation models require aggregate networks. The data in the road inventory files as such is not compatible with an aggregated or abstracted network. The attributes attached to each link of the network have also got to be aggregated. This could lead to aggregation errors. The study by Haghani and Daskin suggests that “as long as long as the level of user equilibrium flows are preserved on the links of the network, very good solutions could be obtained using highly aggregate networks.” For the conversion program, capacity has been used as a measure to evaluate aggregation errors.

Network abstraction in the algorithm proposed by Haghani and Daskin is based on the extraction of links and nodes. The user equilibrium flow level of the links is used as a criterion by which links are chosen for extraction from a given network. The network abstraction done during conversion is based on route number designation.
For a given abstraction value, all routes designated above that value are eliminated from the conversion process.

Some network attribute data are used in transportation analysis to represent supply available for travel. These values determine level of service of a network which influences the amount of travel predicted as well as the distribution of travel demand through the system. Certain transportation applications, such as evaluating the impact on travel demand of increasing capacity on network links, require models to be sensitive to changes in network attributes, like increases in the number of lanes.

For example, when investigating a sub-region of a network, an abstracted representation of the network outside the study area is desired. To achieve this abstraction, several links will be replaced by one link which, ideally, represents the same amount of supply available in original data. This approach of minimizing the effects of altering a network is required for converting data structures too. A new network representation should accurately reflect transportation supply found in the existing structure.

An initial survey has been done of State DOT's to determine methods employed in practice for data structure conversion. The Virginia DOT is working with consultants to convert milepoint data to an intersection/link structure. All data is being stored as offset information and pointers are used to indicate direction of travel to which some attributes apply. It is left to the user of the system to aggregate and disaggregate information to fit the particular application. (16). The Maryland DOT has developed a data base structure that accommodates both milepoint and intersection/link queries. The Florida DOT recently completed conversion of milepoint data to an intersection/link structure.
1.12 Highway Location Reference Systems

Highway location reference systems are field and office procedures that include a method for identifying and recording a specific location on a highway (Highway Location Reference Methods). There are a number of methods for identifying and recording a specific location on a highway. These methods consists of:

- Identification of a point with a known location;
- Measurement of the distance from that known point;
- Observation of the direction of measurement.

The Michigan Department of Transportation compiled a list of 38 reference systems that were being adopted within that state, each with its own reference method(2). This was done as part of their effort to choose a reference system for the state transportation department. Among the systems considered are:

- Coordinate Systems
- Area Identifier Systems
- Segmental Location Systems
- Point Location Systems

From among the above mentioned systems, Coordinate systems and Segmental location systems include methods that would be classed as highway location reference methods.

Michigan Department of Transportation finally chose the control section/milestone method as the standard for inventory and other data collection efforts. To allow for greater accuracy and more detail, a modification of the original method was suggested by way of assigning control section numbers to each roadway of a divided highway where the roadways in opposite directions are widely separated or of

Literature Review
different lengths. It would be appropriate to mention at this time that most road
inventory files have a single record for both directions of travel. Most transportation
problems require highway data to be bi-directional. The conversion program being
developed here is not equipped to disaggregate data for the two directions of travel.

As time goes by, there might be better and more efficient location reference
methods, but as of now the Michigan department of transportation feels that the
Control section/Milepoint method is the most useful. It is also supposed to provide a
good base for conversion to other methods in the future(4).

In general one of two types of methods are used by state departments or other
data collection agencies. These are the route-milepoint and nodelink methods.

1.12.1 Route-Milepoint Method

Linearity is a significant characteristics of highways Any two locations referenced
to the same point are related linearly to each other; it is known how far apart they are.
Highway authorities commonly use the technique of distance and direction from a
known point to identify highway locations or sections in their data files. The "known
point" may be a State line, a county line, a control section terminus or some other
fixed location. This form of location identification is called the "route
number-milepoint" identification. It is the most commonly used and easiest way of
relating any point on a highway with all other points. Most of the location reference
systems now in use are based on route number-milepoint identification in the data
records, either directly or through some variation of this principle.
Most of the states use mileposts other than any other reference method. So it follows that the route-milepoint method would be the ideal choice for highway inventory data in many states.

The route number/milepoint identification method facilitates the collection of highway data in an unambiguous and easy way. It provides the data collector a fixed way for creating a new record each time an attribute value changes. This method is also responsible for the size of the database with significant data redundancy.

1.12.1.1 Milepost/Milepoint

A physical entity, ordinarily a sign placed beside a highway and containing a number that indicates the mileage to that point from some zero point on the highway (3).

1.12.1.2 Control Section

A Control Section reference method identifies a length of highway such that the types of data to be associated with the control section are homogeneous for that length(5). The homogeneity depends on the application context. A control section is based on the assumption that length of a highway, as short as one tenth of a mile or as long as a couple of miles, can be described as having homogeneous characteristics between milepoints.

The Highway Performance Monitoring Systems, which are databases used to describe performance characteristics of a highway, use a control section reference system.
A control section would be of use for a certain application for which the section is devised. It will however be inadequate for general referencing of highway information because control sections tend to encourage neglect of some information due to the objectives for homogeneity of attributes on the section.

This method would work best if each variable is collected separately for a section and then overlapped to completely describe the segment. This process of overlapping attribute data is called Dynamic Segmentation which is fast gaining popularity in Geographical Information Systems.

1.12.2 VDOT’S Road Inventory Files

VDOT’s road inventory files adopt the route-milepoint method. In this method a unique route number is assigned to a continuous section of the highway. This number could be the same as the one on the highway map or some other number used exclusively for inventory purposes. A zero mileage point is chosen on each route. There are certain standard conventions used in locating the zero mileage point for a route. Typically it is the western most terminus for each east-west route and the southern most terminus for north-south routes(6). The mileage measured along the route from that point forms the unique location along the particular route.

Virginia Department of Transportation’s Road Inventory Files form a data base of immense size and content. They contain information about all the highways falling within the State of Virginia. All data is stored in two files: a primary data file and a secondary data file. The primary data file contains attribute data on all highways classified as primary highways such as interstates and state primary highways. For this analysis, primary data files have not been used as these require a slightly
detailed study of ramps and exists. Primary highways also involve overlapping routes, that is routes that have more than one route number for a particular stretch of the facility.

Secondary files form the bulk of the road inventory files. These contain data for the rest of the highways.

The data found in the road inventory files is non-graphic. They do not contain any spatial data in the form of coordinates that could be used to plot networks. The spatial content in these files is restricted to the route number and the county within which the route lies.

1.12.3 Nodelink Method

Another commonly used method employs a nodelink notation which is what most transportation analysis models need. Under this notation, each intersection, change in highway direction or other critical location along a highway is identified as a node and is assigned a unique node number. Each node is connected to at least one other node by a section of road called a link. A node signifies a physical intersection where a change in the direction of travel can be made possible. The nodelink method enjoys an advantage in that it lends itself more readily to the automatic plotting of data by location(1). But despite all the advantages of this method, only a few states, Maine and New York among them have chosen to organize their data files using this methodology. The Highway Departments in Virginia, Maryland and Florida are in the process of converting their road inventory files to a link-node structure.
The main disadvantage with the node/link structure lies in the collection of data in this format. The data collector is forced to use his judgement in specifying a value for an attribute that varies between two nodes.

Some state departments use both methods of location reference for their highway related data. For example, Iowa Department of Transportation uses a route-milepoint location reference for its base record inventory. For accident data it uses a node reference method. They have also developed software to convert node locations to route-milepoint locations so that accident data can be used with highway inventory and traffic data from the base record inventory file(7).

1.12.3.1 Advantages of Node-Link Method

- A nodelink reference scheme attempts to model the topology(5).
- This method provides access to both the coordinate geometry and topology of the physical transportation network.
- Lends itself more readily to the automatic plotting of data by location(1).
- Road data in this format is acceptable to many transportation analysis problems and software.
2.0 Capacity

Capacity reflects the transportation supply that is available on a highway. For this study, capacity has been used as a measure of effectiveness to test the sensitivity of the conversion model. The following paragraphs deal with the general concept of capacity followed by a section on the methodology used for calculating capacity for the conversion program.

2.1 Two-lane Highways

The data base used to test the conversion program is drawn from the secondary files of the Virginia Department of Transportation. These files are primarily made up of road inventory data related to two-lane highways. The Highway Capacity Manual has a fixed set of procedures for calculating capacity on two-lane facilities. The following sections briefly discuss the method used to calculate capacity on two-lane facilities.

A two-lane highway may be defined as a two-lane roadway having one lane for use by traffic in each direction (18). Passing of slower vehicles requires the use of the opposing lane where sight distance and gaps in the opposing traffic stream permit. Two-lane highways compose the predominant mileage of most national
highway systems. They are used for a variety of functions, are located in all geographic areas, and serve a wide range of traffic requirements.

2.1.1 Ideal Conditions

Ideal conditions for two-lane highways are defined as no restrictive geometric, traffic, or environmental conditions. Some of the ideal conditions have been used as default values for those parameters that were not available from the road inventory tiles. Specifically, they include:

- Design speed greater than or equal to 60 mph;
- Lane widths greater than or equal to 12 ft.;
- Clear shoulders wider than or equal to 6 ft.;
- No "no passing zones" on the highway;
- All passenger cars in the traffic stream;
- A 50/50 directional split of traffic;
- No impediments to through traffic due to traffic control or turning vehicles;
- Level terrain.

The capacity of two-lane rural highways under these ideal conditions is 2,800 pcph, total, in both directions. This capacity reflects the impact of opposing vehicles on the ability to efficiently fill gaps in the traffic stream. This phenomenon restricts capacity to a lower value than the 2,000 pcphpl which may be accommodated on multilane uninterrupted flow facilities.
2.1.2 General relationship

The general relationship describing traffic operations on general terrain segments is as follows:

\[ SF_i = 2800 \left( \frac{V}{C} \right)_i f_o f_w f_{nv} \]

where:

\[ SF_i = \text{total service flow rate in both directions for prevailing roadway and traffic conditions, for level of service } i, \text{ in vph} \]

\[ \left( \frac{V}{C} \right)_i = \text{ratio of flow rate to ideal capacity for level of service } i \]

\[ f_o = \text{adjustment factor for directional distribution of traffic} \]

\[ f_w = \text{adjustment factor for narrow lanes and restricted shoulder width} \]

\[ f_{nv} = \text{adjustment factor for the presence of heavy vehicles in the traffic stream, computed as:} \]

\[ f_{nv} = \frac{1}{1 + P_T (E_T - 1) + P_R (E_R - 1) - P_o (E_o - 1)} \]

where:

\[ P_T = \text{proportion of trucks in the traffic stream, expressed as a decimal} \]
$P_r = \text{proportion of recreational vehicles in the traffic stream}$

$P_b = \text{proportion of buses in the traffic stream}$

$E_r = \text{passenger-car equivalent for trucks}$

$E_s = \text{passenger-car equivalent for recreational vehicles}$

$E_b = \text{passenger-car equivalent for buses}$

The adjustment factors for directional distribution of traffic and adjustment factor for narrow lanes and restricted shoulder width are found in tables in the Highway Capacity Manual. So are the passenger-car equivalents for trucks, recreational vehicles, and buses.

The highest service flow attainable under level-of-service E defines the capacity of the highway. Under ideal conditions, capacity is 2800 pcph, total in both directions.

### 2.2 Capacity Calculation For the Conversion Program

Calculation of capacity using the Highway Capacity Manual's procedures are very lengthy, data intensive and complex. Attributes such as % trucks, terrain type, and speeds are not available in the road inventory files of the Virginia Department of Transportation. All these attributes are essential for the calculation of capacity using the highway capacity software that replicates the procedures of the manual. So
certain assumptions had to be made in order for the conversion program to calculate capacity.

2.2.1 Assumptions

The following general assumptions were made while calculating capacity on the links using certain rules. As described earlier, the equation for service flow rate (SF) contains adjustment factors for

- Directional distribution of traffic;
- narrow lanes and restricted shoulder width;
- presence of heavy vehicles.

The first assumption has to do with the directional distribution of traffic. The directional split was assumed to be 50/50 and from the Highway Capacity Manual, the adjustment factor fixed at 1.00.

The second assumption is with regard to the presence of heavy vehicles in the traffic stream. The Highway Capacity Manual has set down the following default values when estimates of the traffic mix are not available. These values are

- Percent Trucks = 0.14;
- Percent Recreational Vehicles = 0.04;
- Percent Buses = 0.00.

The heavy vehicle adjustment factor is dependent on the passenger car equivalents for each of the above three categories which are in turn dependent on the type of terrain.
The third assumption follows from the above requirement. For lack of information on terrain type in the road inventory files, it has been assumed that the terrain over all the highway segments is rolling.

The adjustment factor for narrow lanes and restricted shoulder width was based on Table 8-5 in the Highway Capacity Manual which gives the adjustment factor for a range of shoulder widths and lane widths.

The analysis of extended specific grades on two-lane highways is more complex than for general terrain segments and the road inventory files do not contain any information about specific grades. So for the present, analysis of grades is not included in this work.

Using a set of derived and default values, the conversion program uses the basic formula for the Service flow rate as given in the Highway Capacity Manual.

2.3 Rules for Capacity Calculation

From the basic equation for Service flow, it is quite clear that capacity is a function of lane width, shoulder width and the number of lanes in each direction. Lane width is obtained from surface width and number of lanes. Since the main focus of this research has been on two-lane highways, the number of lanes does not affect the results of this analysis. But for the sake of future research, the number of lanes has been included in the conversion program. It also makes sure that the conversion process does not fail when a third lane is encountered in a segment which is in transition between two other segments that have two lanes.
Capacity is a function of the above mentioned three variables for a link as well as the milepoint segments that make up the link. The values for the variables may not remain constant throughout the link. Capacity on a link is calculated using the following four values for each of the three factors that affect capacity. The four values that each of the variables take on are:

- Maximum variable value;
- Minimum variable value;
- Standard average value;
- Weighted average value.

So with the three variables and four values that each could take on, there are 64 possible combinations which form the rule base for the capacity calculations in the conversion program. Table 1 lists all possible combinations of rules used to calculate link capacity.

2.3.1 Travel Impedance

Capacity calculated from the aggregated variables should reflect capacity of the individual records that make up the link. But capacity, unlike the measured variables cannot be summed. If two records form a link, the sum of their capacities is not equal to the capacity of the link. So travel impedance has been used to measure capacity differences resulting from conversion.

The equation used to calculate travel time or impedance is a simplified function developed by the U.S. Bureau of Public Roads (BPR). The equation is given by:
<table>
<thead>
<tr>
<th>#</th>
<th>Surface Width</th>
<th>No. of Lanes</th>
<th>Shoulder Width</th>
<th>#</th>
<th>Surface Width</th>
<th>No. of Lanes</th>
<th>Shoulder Width</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>max</td>
<td>max</td>
<td>max</td>
<td>33</td>
<td>s. avg.</td>
<td>max</td>
<td>max</td>
</tr>
<tr>
<td>2</td>
<td>max</td>
<td>max</td>
<td>min</td>
<td>34</td>
<td>s. avg.</td>
<td>max</td>
<td>min</td>
</tr>
<tr>
<td>3</td>
<td>max</td>
<td>max</td>
<td>s. avg.</td>
<td>35</td>
<td>s. avg.</td>
<td>max</td>
<td>s. avg.</td>
</tr>
<tr>
<td>4</td>
<td>max</td>
<td>max</td>
<td>w. avg.</td>
<td>36</td>
<td>s. avg.</td>
<td>max</td>
<td>w. avg.</td>
</tr>
<tr>
<td>5</td>
<td>max</td>
<td>min</td>
<td>max</td>
<td>37</td>
<td>s. avg.</td>
<td>min</td>
<td>max</td>
</tr>
<tr>
<td>6</td>
<td>max</td>
<td>min</td>
<td>min</td>
<td>38</td>
<td>s. avg.</td>
<td>min</td>
<td>min</td>
</tr>
<tr>
<td>7</td>
<td>max</td>
<td>min</td>
<td>s. avg.</td>
<td>39</td>
<td>s. avg.</td>
<td>min</td>
<td>s. avg.</td>
</tr>
<tr>
<td>8</td>
<td>max</td>
<td>min</td>
<td>w. avg.</td>
<td>40</td>
<td>s. avg.</td>
<td>min</td>
<td>w. avg.</td>
</tr>
<tr>
<td>9</td>
<td>s. avg.</td>
<td>max</td>
<td>max</td>
<td>41</td>
<td>s. avg.</td>
<td>s. avg.</td>
<td>max</td>
</tr>
<tr>
<td>10</td>
<td>s. avg.</td>
<td>min</td>
<td>s. avg.</td>
<td>42</td>
<td>s. avg.</td>
<td>s. avg.</td>
<td>min</td>
</tr>
<tr>
<td>11</td>
<td>s. avg.</td>
<td>s. avg.</td>
<td>s. avg.</td>
<td>43</td>
<td>s. avg.</td>
<td>s. avg.</td>
<td>s. avg.</td>
</tr>
<tr>
<td>12</td>
<td>s. avg.</td>
<td>w. avg.</td>
<td>s. avg.</td>
<td>44</td>
<td>s. avg.</td>
<td>s. avg.</td>
<td>w. avg.</td>
</tr>
<tr>
<td>13</td>
<td>w. avg.</td>
<td>max</td>
<td>w. avg.</td>
<td>45</td>
<td>s. avg.</td>
<td>w. avg.</td>
<td>max</td>
</tr>
<tr>
<td>14</td>
<td>w. avg.</td>
<td>min</td>
<td>w. avg.</td>
<td>46</td>
<td>s. avg.</td>
<td>w. avg.</td>
<td>min</td>
</tr>
<tr>
<td>15</td>
<td>w. avg.</td>
<td>s. avg.</td>
<td>w. avg.</td>
<td>47</td>
<td>s. avg.</td>
<td>w. avg.</td>
<td>s. avg.</td>
</tr>
<tr>
<td>16</td>
<td>w. avg.</td>
<td>w. avg.</td>
<td>w. avg.</td>
<td>48</td>
<td>s. avg.</td>
<td>w. avg.</td>
<td>w. avg.</td>
</tr>
<tr>
<td>17</td>
<td>min</td>
<td>max</td>
<td>max</td>
<td>49</td>
<td>w. avg.</td>
<td>max</td>
<td>max</td>
</tr>
<tr>
<td>18</td>
<td>min</td>
<td>max</td>
<td>min</td>
<td>50</td>
<td>w. avg.</td>
<td>max</td>
<td>min</td>
</tr>
<tr>
<td>19</td>
<td>min</td>
<td>max</td>
<td>s. avg.</td>
<td>51</td>
<td>w. avg.</td>
<td>max</td>
<td>s. avg.</td>
</tr>
<tr>
<td>20</td>
<td>min</td>
<td>max</td>
<td>w. avg.</td>
<td>52</td>
<td>w. avg.</td>
<td>max</td>
<td>w. avg.</td>
</tr>
<tr>
<td>21</td>
<td>min</td>
<td>min</td>
<td>max</td>
<td>53</td>
<td>w. avg.</td>
<td>min</td>
<td>max</td>
</tr>
<tr>
<td>22</td>
<td>min</td>
<td>min</td>
<td>min</td>
<td>54</td>
<td>w. avg.</td>
<td>min</td>
<td>min</td>
</tr>
<tr>
<td>23</td>
<td>min</td>
<td>min</td>
<td>s. avg.</td>
<td>55</td>
<td>w. avg.</td>
<td>min</td>
<td>s. avg.</td>
</tr>
<tr>
<td>24</td>
<td>min</td>
<td>min</td>
<td>w. avg.</td>
<td>56</td>
<td>w. avg.</td>
<td>min</td>
<td>w. avg.</td>
</tr>
<tr>
<td>25</td>
<td>s. avg.</td>
<td>max</td>
<td>max</td>
<td>57</td>
<td>w. avg.</td>
<td>s. avg.</td>
<td>max</td>
</tr>
<tr>
<td>26</td>
<td>s. avg.</td>
<td>min</td>
<td>s. avg.</td>
<td>58</td>
<td>w. avg.</td>
<td>s. avg.</td>
<td>min</td>
</tr>
<tr>
<td>27</td>
<td>s. avg.</td>
<td>s. avg.</td>
<td>s. avg.</td>
<td>59</td>
<td>w. avg.</td>
<td>s. avg.</td>
<td>s. avg.</td>
</tr>
<tr>
<td>28</td>
<td>s. avg.</td>
<td>w. avg.</td>
<td>s. avg.</td>
<td>60</td>
<td>w. avg.</td>
<td>s. avg.</td>
<td>w. avg.</td>
</tr>
<tr>
<td>29</td>
<td>w. avg.</td>
<td>max</td>
<td>w. avg.</td>
<td>61</td>
<td>w. avg.</td>
<td>w. avg.</td>
<td>max</td>
</tr>
<tr>
<td>30</td>
<td>w. avg.</td>
<td>min</td>
<td>w. avg.</td>
<td>62</td>
<td>w. avg.</td>
<td>w. avg.</td>
<td>min</td>
</tr>
<tr>
<td>31</td>
<td>w. avg.</td>
<td>s. avg.</td>
<td>w. avg.</td>
<td>63</td>
<td>w. avg.</td>
<td>w. avg.</td>
<td>s. avg.</td>
</tr>
<tr>
<td>32</td>
<td>w. avg.</td>
<td>w. avg.</td>
<td>w. avg.</td>
<td>64</td>
<td>w. avg.</td>
<td>w. avg.</td>
<td>w. avg.</td>
</tr>
</tbody>
</table>

s. avg. = Standard Average  
w. avg. = Weighted Average
\[ t_a = t_a^0 \left[ 1 + \alpha \left( \frac{x_a}{c_a} \right)^\beta \right] \]

In this formula, \( t_a \) and \( x_a \) are the travel time and flow, respectively, on link \( a \). \( t_a^0 \) is the free-flow travel time, and \( c_a \) is the "practical capacity" of link \( a \). The quantities \( \alpha \) and \( \beta \) are model parameters whose typical values are 0.15 min and 4.0 respectively.

Free-flow impedance is calculated as a function of distance and speed. Speed is assumed to be equal for all records as the data base being used does not contain speed information; so the sum of free flow impedences of all segments in a link is equal to the free flow impedance of the link.

Equating the travel time over the link to the travel time calculated from the milepoint segments, we derive the following expression:

\[
\sum_{r=1}^{\alpha} \left[ U^0(r) + (\alpha \cdot U^0(r) \cdot v^0(r)) / c^0(r) \right]
\]

\[ = U^0(k) + (\alpha \cdot U^0(k) \cdot v^0(k)) / c^0(k) + \epsilon \]

where:

\( U^0(r) \) = free-flow impedance on link or record \( r \)

\( v(r) \) = traffic volume or flow on link or record \( r \)

\( c(r) \) = capacity of link or record \( r \)

\( \epsilon \) = error introduced by conversion rule
The Root Mean Square Error (RMSE) is used to measure the variance between network impedance calculated from the milepoint structure and impedance determined for the converted link structure. RMSE is calculated as follows:

\[
RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^{m} [(l(i, R) - l(i, L)]^2}
\]

where:

\(l(i, R)\) = sum of impedances calculated for each record in a link, \(i\)

\(l(i, L)\) = impedance calculated for link \(i\) using rule \(j\)

\(m\) = total number of links in the network
3.0 Conversion Methodology

3.1 Data Structure Conversion

A two step process is undertaken for converting milepoint information to an intersection/link structure. The first step involves variable or attribute definition. The second step converts data base on rules defined in the previous step. Details on each of these procedures follows.

3.1.1 Variable Definition

Two types of variables are found in road inventory files and may be classified as measured and descriptive. A measured variable is one which may be aggregated or disaggregated using mathematical equations without significant loss in accuracy. This assumption is not valid for descriptive variables, aggregation of information will result in unacceptable loss in accuracy.

An example will clarify the variable classification methodology required to distinguish data types. Figure 1 shows a link that is composed of three milepoint road segments. Each segment has a variable or attribute associated with it that takes on values A, B, and C. Suppose the variable is system domain and a value of A
represents private land, B represents State agencies and C represents National Park Service. If this information is aggregated based on a weighting formula, the value of the variable stored in the intersection/link structure is C since it is the value associated with the longest segment. When the new data base is queried to identify all links with value A, this link would not be selected since this information is lost from the system. However, if for each of these segments the values of A, B, and C represent the number of structures within a certain distance from the shoulder, these values may be summed into a single link value without loss of accuracy.

Examples of measured data in road inventory files include length, number of at-grade railroad crossings, and number of structures. Descriptive data include surface and base type, route signing, and functional class. A group of variables, like surface and shoulder width, number of lanes, and AADT can not be classified strictly as measured or descriptive. Mathematically aggregating values of these variables over a link is appropriate for some applications and inappropriate for others. A complete list of road inventory variables are listed in Table 2.

Each variable has been labeled as descriptive or measured. Some records in road inventory files represent sample sections for the Federal Highway Administration’s Highway Performance Monitoring System (HPMS). For these sample sections, much more comprehensive data are maintained. A list of additional variables found in HPMS sections is provided in Table 3 as well as default classifications as measured or descriptive variables.

In the computer model, users are allowed to select which variables in the road inventory files to convert to an intersection/link structure. The specific application for which data are required will dictate this selection process. For instance, applications using data to calculate link capacity from the Highway Capacity Manual formulation
may utilize lane and shoulder width, number of lanes, and percent trucks as opposed to those deriving link capacity from the UTPS look-up table which utilize facility and area type data.

Users may also change default classifications of selected variables. Classification of a variable dictates how and what type of information is kept in the new data base. If a variable is classified as descriptive, each link contains offset information along with the value of the variable associated with each segment. For example, referring to the link in Figure 1, attached to this link in the new data base is information indicating that this variable has value A for 0.4 miles, value B for 0.5 miles, and value C for 0.6 miles. If, however, the value of a descriptive variable does not change between segments, redundant information is not stored. For example, if the variable in Figure 1 had value A in segment 2, the new data base saves this information as follows: value A for 0.7 miles, value C for 0.6 miles. Inclusion of a large number of descriptive variables significantly increases data base size. To simplify manipulation of information, separate data files are created for each descriptive variable containing pointers to network links.
<table>
<thead>
<tr>
<th>COLUMN(S)</th>
<th>VARIABLE</th>
<th>CLASSIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>46-50</td>
<td>Length</td>
<td>Measured</td>
</tr>
<tr>
<td>51-52</td>
<td>Surface Width</td>
<td>Measured/Descriptive</td>
</tr>
<tr>
<td>53-54</td>
<td>Shoulder Width</td>
<td>Measured/Descriptive</td>
</tr>
<tr>
<td>55</td>
<td>Surface Type</td>
<td>Descriptive</td>
</tr>
<tr>
<td>56</td>
<td>Base Type</td>
<td>Descriptive</td>
</tr>
<tr>
<td>57-58</td>
<td>Kind of Highway</td>
<td>Measured/Descriptive</td>
</tr>
<tr>
<td></td>
<td># Lanes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Type of Facility &amp; Access Control</td>
<td>Descriptive</td>
</tr>
<tr>
<td>59-60</td>
<td>Rural &amp; Urban Public at Grade Railroad Crossings</td>
<td>Measured</td>
</tr>
<tr>
<td>61-66</td>
<td>AADT</td>
<td>Measured/Descriptive</td>
</tr>
<tr>
<td>67</td>
<td>Type of Section Identification</td>
<td>Descriptive</td>
</tr>
<tr>
<td>68-72</td>
<td>Milepoint</td>
<td>Measured</td>
</tr>
<tr>
<td>73</td>
<td>Route Signing</td>
<td>Descriptive</td>
</tr>
<tr>
<td>74-79</td>
<td>System</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Governmental Level of Control Location</td>
<td>Descriptive</td>
</tr>
<tr>
<td></td>
<td>Federal Aid System</td>
<td>Descriptive</td>
</tr>
<tr>
<td></td>
<td>Federal, State &amp; Local Domain</td>
<td>Descriptive</td>
</tr>
<tr>
<td></td>
<td>Special System</td>
<td>Descriptive</td>
</tr>
<tr>
<td></td>
<td>Public Road</td>
<td>Descriptive</td>
</tr>
<tr>
<td>80-83</td>
<td>Operation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reversible Lanes</td>
<td>Descriptive</td>
</tr>
<tr>
<td></td>
<td>Trucks/Commercial Vehicles</td>
<td>Descriptive</td>
</tr>
<tr>
<td></td>
<td>HOV Lanes</td>
<td>Descriptive</td>
</tr>
<tr>
<td></td>
<td>Toll</td>
<td>Descriptive</td>
</tr>
<tr>
<td>84</td>
<td>Functional Class</td>
<td>Descriptive</td>
</tr>
<tr>
<td>85</td>
<td>Special Code</td>
<td>Descriptive</td>
</tr>
<tr>
<td>COLUMN(S)</td>
<td>VARIABLE</td>
<td>CLASSIFICATION</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>97-98</td>
<td>Surface Width</td>
<td>Measured/Descriptive</td>
</tr>
<tr>
<td>99-101</td>
<td>Approach Width</td>
<td>Measured/Descriptive</td>
</tr>
<tr>
<td>102</td>
<td>Shoulder Type</td>
<td>Descriptive</td>
</tr>
<tr>
<td>103-106</td>
<td>Shoulder Width</td>
<td>Measured/Descriptive</td>
</tr>
<tr>
<td>107</td>
<td>Median Type</td>
<td>Descriptive</td>
</tr>
<tr>
<td>108-109</td>
<td>Median Width</td>
<td>Measured/Descriptive</td>
</tr>
<tr>
<td>110</td>
<td>Approach Width</td>
<td>Measured/Descriptive</td>
</tr>
<tr>
<td>111</td>
<td>Horizontal Alignment Adequacy</td>
<td>Descriptive</td>
</tr>
<tr>
<td>112</td>
<td>Vertical Alignment Adequacy</td>
<td>Descriptive</td>
</tr>
<tr>
<td>113-115</td>
<td>% Passing Sight Distance</td>
<td>Measured/Descriptive</td>
</tr>
<tr>
<td>116-117</td>
<td>Speed Limit</td>
<td>Measured/Descriptive</td>
</tr>
<tr>
<td>118-119</td>
<td>Average Highway Speed</td>
<td>Measured/Descriptive</td>
</tr>
<tr>
<td>120</td>
<td>Signal Type</td>
<td>Descriptive</td>
</tr>
<tr>
<td>121-122</td>
<td>% Green Time</td>
<td>Measured</td>
</tr>
<tr>
<td>123-124</td>
<td>Parking</td>
<td>Descriptive</td>
</tr>
<tr>
<td>125</td>
<td>Terrain Type</td>
<td>Descriptive</td>
</tr>
<tr>
<td>126</td>
<td>Type of Development</td>
<td>Descriptive</td>
</tr>
<tr>
<td>127</td>
<td>Urban Location</td>
<td>Descriptive</td>
</tr>
<tr>
<td>128-129</td>
<td># Grade Separated Interchanges</td>
<td>Measured</td>
</tr>
<tr>
<td></td>
<td>At Grade Intersection</td>
<td>Measured</td>
</tr>
<tr>
<td>130-131</td>
<td>Signals</td>
<td>Measured</td>
</tr>
<tr>
<td>132-133</td>
<td>Stop Signs</td>
<td>Measured</td>
</tr>
<tr>
<td>134-135</td>
<td>Other or no Control</td>
<td>Measured</td>
</tr>
<tr>
<td>136-137</td>
<td># Commercial Access Points</td>
<td>Measured</td>
</tr>
<tr>
<td>138-139</td>
<td># Structures</td>
<td>Measured</td>
</tr>
<tr>
<td>140-141</td>
<td># At-Grade Railroad Crossings</td>
<td>Measured</td>
</tr>
<tr>
<td>142-143</td>
<td>Type of Improvement</td>
<td>Descriptive</td>
</tr>
<tr>
<td>144-155</td>
<td>Sample Number Identification</td>
<td>Measured</td>
</tr>
<tr>
<td>156</td>
<td>Sample Section Subdivision</td>
<td>Measured</td>
</tr>
<tr>
<td>157-161</td>
<td>State HPMS Sample Number</td>
<td>Measured</td>
</tr>
<tr>
<td>162-163</td>
<td>AADT Volume Group Identifier</td>
<td>Descriptive</td>
</tr>
<tr>
<td>164-168</td>
<td>Expansion Factor</td>
<td>Measured</td>
</tr>
</tbody>
</table>
Variables classified as measured have default aggregation/disaggregation rules associated with them. For instance, length, number of structures, and number of RR crossings for each segment are summed. If AADT is aggregated, a weighted averaging method is provided. As this system is enhanced, users will be allowed to enter conversion equations as well as define new variables as mathematical and logical combinations of road inventory variables.

3.1.2 Data Conversion

Ideally, if all road segments fit perfectly into a link, conversion would be rather simple. Complications arise from the two input sources namely, the road inventory files and the intersection/link data base. In road inventory files, segments do not necessarily terminate at all intersections. Further, users may not have selected all roads in the network as part of their digital data base. Consequently, users must be queried on how to handle special cases that occur during link building.

Before conversion can take place, road inventory files must be downloaded from the mainframe to a PC. Due to the size of these files, it is recommended that data are sorted first by county and stored in separate files. Users may then download any county of interest when needed.

Primary information in road inventory files utilized in the search algorithm are beginning and end termini labels. If a sequential data conversion technique is utilized, all links in a route identify a path from reference point zero, through the county to a second zero reference point. Links are processed along this path by matching record end termini labels to cross-route labels at intersections. If an intersection occurs between the endpoints of a road segment, distance information
is utilized to disaggregate data in this segment and assign converted data to the two links. When multiple road segments comprise a link, selected variables are aggregated using rules defined by users. As records in road inventory files are being assigned to network links, they are displayed on the terminal to facilitate user intervention.

The data base for the conversion program is a subset of Virginia Department of Transportation's Road Inventory Files (RIF's). Information on all the highways within the State of Virginia are contained in these files. These files are structured in the route-milepoint format, which uses a unique route number and measured distance along the route from a known point.

Data within the Road Inventory Files are sorted and arranged by county. For the sake of analysis, eleven counties were selected and downloaded to the PC from the mainframe. All the files selected for this study are secondary highway files. For purposes of this analysis, only certain attributes have been extracted from the road inventory files. These are:

- Route Number;
- From termini;
- To termini;
- Distance;
- Surface width;
- Shoulder width;
- Number of lanes;
- AADT;
- HPMS.

The From and To termini labels form the backbone for the conversion process.
The From and To termini labels are used as pointers for the formation of links. The program uses these labels while selecting segments for conversion based on a few simple rules. These rules are described under the section of True Link. Distance is the only direct additive variable being used in the conversion process. The distances of all the segments that make up a link are added to give the link length. This particular variable is useful in computing the total mileage along a particular highway. Surface width, shoulder width and the number of lanes are the key variables in the capacity calculations. These are the only variables that could be used in the equations for transportation supply from among the attributes in the road inventory files. Variables like percent trucks and terrain type are not available for all segments of the road inventory files. Only those segments on the Highway Performance Monitoring System contain these variables. Unfortunately there are not too many of these segments in VDOT’S files. As more HPMS sections become available, the conversion program could be updated to incorporate the variables for which default values have been assigned.

3.2 Steps in the Conversion Process

Figure 2 depicts a flow chart of the conversion process.
Figure 2. Flow-chart for the conversion program
3.3 File Name

In this study, the program was tested by converting road data by county. The database for this analysis consists of twenty two different files of which eleven are primary data files and the other eleven are secondary. These are for the eleven counties in the state of Virginia that have been chosen out of VDOT'S road inventory files.

3.4 Abstraction Value

This value is provided to the program by the user to selectively eliminate route numbers designated above a certain value from the conversion process. For this analysis four levels of abstraction, namely, 5000, 1000, 800 and 700 are provided. When provided with an abstraction value of 1000, the program eliminates all records with route numbers designated greater than 1000 from link construction.

Once the abstraction value is given, the conversion process begins. Route number and the from and to labels are input. If the record already exists in link structure, its capacity is calculated and output to a file. Also attached to the output are the various attributes that are associated with the record for verification purposes.
3.5 True Link

A true link is one that has either of the following beginning and ending termini labels:

- begins and ends with an Rt in the from and to labels; eg. Rt 600 from Rt 714 to Rt 636
- Begins with an Rt and ends with a directional Rt; the direction being north, south, east or west. eg Rt 600 from Rt 714 to N Rt 636
- Begins and ends with a directional Rt; eg Rt 600 from S Rt 636 to N Rt 636
- Begins with an directional Rt and ends with an Rt; eg Rt 600 from S Rt 636 to Rt 714.

3.6 Checking for Closure

If the end termini label is not one of the above mentioned, the program proceeds to the next record, for the same route number. Once it finds the proper ending termini label, it calculates the capacity of the link using the 64 rules and updates the link attributes. The updated results are then output to a file.
3.7 Consecutive Record

If the current record being processed is not the first record in a link, a check is made to determine if the record is consecutive i.e., whether the milepoint of the previous end termini label matches the milepoint of the current from termini label of the next record. If the From termini label is consecutive, and the End termini label of this record closes a link, capacity is calculated for this link and the link attributes updated. If a perfect match between labels cannot be made, an error is printed indicating non-consecutive records. Further the program prompts users to either join or split a link. Terminal labels for the previous and current records are printed on the screen. Users can decide if a typographical error exists in the file or consult a highway map to make the decision to join or split these records.

3.8 Non-Consecutive Record

When the program encounters a non-consecutive record, it outputs that record to an error file. This file can be checked with the data file to determine the source of the error. Usually these are typographical errors leading to miscoding of data. Table 4 lists the total number of non-consecutive errors and typos in each of the data files.
<table>
<thead>
<tr>
<th>ABSTRACTION LEVEL</th>
<th>5000</th>
<th>1000</th>
<th>800</th>
<th>700</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2/1</td>
<td>2/1</td>
<td>2/1</td>
<td>3/1</td>
</tr>
<tr>
<td>2</td>
<td>1/1</td>
<td>1/1</td>
<td>1/1</td>
<td>1/1</td>
</tr>
<tr>
<td>C</td>
<td>8/8</td>
<td>11/11</td>
<td>12/12</td>
<td>12/9</td>
</tr>
<tr>
<td>O</td>
<td>2/2</td>
<td>2/2</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>U</td>
<td>4/2</td>
<td>5/3</td>
<td>5/3</td>
<td>3/2</td>
</tr>
<tr>
<td>N</td>
<td>5/4</td>
<td>5/4</td>
<td>5/4</td>
<td>5/4</td>
</tr>
<tr>
<td>T</td>
<td>3/3</td>
<td>3/3</td>
<td>3/3</td>
<td>2/2</td>
</tr>
<tr>
<td>Y</td>
<td>10/5</td>
<td>12/5</td>
<td>12/5</td>
<td>10/3</td>
</tr>
<tr>
<td>9</td>
<td>5/4</td>
<td>5/4</td>
<td>5/4</td>
<td>9/4</td>
</tr>
<tr>
<td>10</td>
<td>9/5</td>
<td>10/6</td>
<td>10/6</td>
<td>10/6</td>
</tr>
<tr>
<td>11</td>
<td>1/1</td>
<td>3/3</td>
<td>3/3</td>
<td>10/5</td>
</tr>
</tbody>
</table>

The fraction in each cell represents the **Number of non-consecutive record errors** / the **Number of typos or non-graphical fixes**.
4.0 Analysis of Results

A total of eleven data files were run through the conversion program. This makes up a sizeable data base on which to evaluate the performance of this model and address sensitivity issues. In this work, greater emphasis has been laid on secondary highways which are primarily two-lane highways (one lane in each direction). Two-lane highways constitute a major portion of any state's highway network. They also provide a greater percentage of records that need to be converted to the node-link structure due to the frequent changes in attributes along these routes and the subsequent creation of milepoint segments.

4.1 The data base

Highways from eleven counties in the State of Virginia have been chosen to form the data base for the conversion process. The eleven counties are:

- Caroline co.
- Charles City co.
- Hanover co.
- Isle of Wight Co.
- James City Co.
• Louisa Co.
• Montgomery Co.
• New Kent Co.
• Spotsylvania Co.
• Surry Co.
• York Co.

4.2 Data base Size

The total number of highway records for the eleven counties totals 9098. This includes only the secondary highways. The breakdown of the number of records in each county is shown in Table 5. As described earlier, the road inventory files provided by the Virginia Department of Transportation (VDOT) contain a lot of information placed in a number of columns. The records in these files are sorted by county. These files were then downloaded to the PC. Only those attributes required for route identification and capacity calculation have been extracted from this large data base using a small program in Statistical Applications Software (SAS).

4.3 Results of Conversion

Table 5 gives a complete summary of the results of conversion. It shows:
• The county for which the conversion has been done;
• Number of records in each file;
• Number of links;
• Number of true single links;
• Multiple record links;
• Number of records eliminated during abstraction;
• Multiple record links with capacity change;
• Multiple record links w/o capacity change;

4.4 Number of Records

These are the total number of records in milepost form for a given county as found in the road inventory files.

4.5 Number of links

These are the number of highway links in each county after conversion. A link is a segment of the highway between two nodes which represent physical intersections where a change in the direction of travel can occur.
4.6 True Single Links

These are records within the database that already exist in the link structure i.e. links containing only one record that begins and ends with RT or directional RT characters. These do not need any further conversion. True single links form a major portion of the database as can be seen from their share of the total number of links in Table 5.

4.7 Multiple Record Links

These are links formed as a result of conversion. These links contain more than one segment from the road inventory files.

4.8 Number of Records Eliminated

These are the records eliminated from the conversion process as a result of abstraction.
4.9 *Multiple Record Links with Capacity Change*

These are links over which the milepoint segment capacities differ. This reflects the difference in the maximum, minimum, standard average, and weighted average of variables. These links are used in plotting the RMSE graphs to show capacity variation.

4.10 *Multiple Record Links w/o Capacity Change*

These are links on which the milepoint segment capacities are the same for all the segments in the link.

4.10.1 *Summary Statistics*

Summary statistics by county are shown in Table 5. Each county file was analyzed at four levels of abstraction. These are:

- 5000
- 1000
- 800
- 700
<table>
<thead>
<tr>
<th>County</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Records</td>
<td>721</td>
<td>169</td>
<td>1962</td>
<td>674</td>
<td>754</td>
<td>666</td>
<td>712</td>
<td>341</td>
<td>1484</td>
</tr>
<tr>
<td>Number of links</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 5000</td>
<td>491</td>
<td>128</td>
<td>714</td>
<td>413</td>
<td>367</td>
<td>435</td>
<td>484</td>
<td>191</td>
<td>484</td>
</tr>
<tr>
<td>&lt; 1000</td>
<td>486</td>
<td>128</td>
<td>566</td>
<td>376</td>
<td>333</td>
<td>422</td>
<td>449</td>
<td>173</td>
<td>377</td>
</tr>
<tr>
<td>&lt; 800</td>
<td>486</td>
<td>128</td>
<td>520</td>
<td>376</td>
<td>333</td>
<td>417</td>
<td>392</td>
<td>173</td>
<td>377</td>
</tr>
<tr>
<td>&lt; 700</td>
<td>291</td>
<td>128</td>
<td>294</td>
<td>340</td>
<td>195</td>
<td>241</td>
<td>212</td>
<td>172</td>
<td>257</td>
</tr>
<tr>
<td>True Single Links</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 5000</td>
<td>328</td>
<td>59</td>
<td>490</td>
<td>295</td>
<td>239</td>
<td>249</td>
<td>247</td>
<td>112</td>
<td>304</td>
</tr>
<tr>
<td>&lt; 1000</td>
<td>297</td>
<td>59</td>
<td>315</td>
<td>259</td>
<td>198</td>
<td>237</td>
<td>206</td>
<td>92</td>
<td>176</td>
</tr>
<tr>
<td>&lt; 800</td>
<td>297</td>
<td>59</td>
<td>279</td>
<td>259</td>
<td>198</td>
<td>235</td>
<td>175</td>
<td>92</td>
<td>176</td>
</tr>
<tr>
<td>&lt; 700</td>
<td>196</td>
<td>59</td>
<td>132</td>
<td>229</td>
<td>104</td>
<td>130</td>
<td>79</td>
<td>92</td>
<td>106</td>
</tr>
<tr>
<td>Multiple Record Links</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 5000</td>
<td>62</td>
<td>30</td>
<td>96</td>
<td>68</td>
<td>38</td>
<td>61</td>
<td>91</td>
<td>33</td>
<td>86</td>
</tr>
<tr>
<td>&lt; 1000</td>
<td>69</td>
<td>30</td>
<td>135</td>
<td>68</td>
<td>49</td>
<td>61</td>
<td>109</td>
<td>36</td>
<td>115</td>
</tr>
<tr>
<td>&lt; 800</td>
<td>69</td>
<td>30</td>
<td>142</td>
<td>68</td>
<td>49</td>
<td>61</td>
<td>109</td>
<td>36</td>
<td>115</td>
</tr>
<tr>
<td>&lt; 700</td>
<td>95</td>
<td>30</td>
<td>129</td>
<td>71</td>
<td>50</td>
<td>76</td>
<td>93</td>
<td>36</td>
<td>107</td>
</tr>
<tr>
<td>No. of Records Eliminated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 1000</td>
<td>160</td>
<td>6</td>
<td>1109</td>
<td>176</td>
<td>347</td>
<td>154</td>
<td>115</td>
<td>107</td>
<td>890</td>
</tr>
<tr>
<td>&lt; 800</td>
<td>160</td>
<td>6</td>
<td>1143</td>
<td>176</td>
<td>347</td>
<td>159</td>
<td>157</td>
<td>107</td>
<td>890</td>
</tr>
<tr>
<td>&lt; 700</td>
<td>292</td>
<td>6</td>
<td>1352</td>
<td>208</td>
<td>467</td>
<td>282</td>
<td>320</td>
<td>108</td>
<td>1018</td>
</tr>
<tr>
<td>MR Links with Capacity Change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 5000</td>
<td>56</td>
<td>24</td>
<td>77</td>
<td>48</td>
<td>26</td>
<td>47</td>
<td>77</td>
<td>28</td>
<td>58</td>
</tr>
<tr>
<td>&lt; 1000</td>
<td>58</td>
<td>24</td>
<td>82</td>
<td>46</td>
<td>28</td>
<td>45</td>
<td>79</td>
<td>29</td>
<td>57</td>
</tr>
<tr>
<td>&lt; 800</td>
<td>58</td>
<td>24</td>
<td>83</td>
<td>46</td>
<td>28</td>
<td>44</td>
<td>75</td>
<td>29</td>
<td>57</td>
</tr>
<tr>
<td>&lt; 700</td>
<td>56</td>
<td>24</td>
<td>80</td>
<td>46</td>
<td>32</td>
<td>44</td>
<td>61</td>
<td>29</td>
<td>52</td>
</tr>
<tr>
<td>MR Links w/o Capacity Change</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 5000</td>
<td>6</td>
<td>6</td>
<td>19</td>
<td>20</td>
<td>12</td>
<td>14</td>
<td>14</td>
<td>5</td>
<td>27</td>
</tr>
<tr>
<td>&lt; 1000</td>
<td>11</td>
<td>6</td>
<td>53</td>
<td>22</td>
<td>21</td>
<td>16</td>
<td>30</td>
<td>7</td>
<td>59</td>
</tr>
<tr>
<td>&lt; 800</td>
<td>11</td>
<td>6</td>
<td>59</td>
<td>22</td>
<td>21</td>
<td>17</td>
<td>34</td>
<td>7</td>
<td>59</td>
</tr>
<tr>
<td>&lt; 700</td>
<td>39</td>
<td>6</td>
<td>49</td>
<td>25</td>
<td>18</td>
<td>32</td>
<td>32</td>
<td>7</td>
<td>55</td>
</tr>
<tr>
<td>County</td>
<td>10</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>----</td>
<td>----</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Number of Records</strong></td>
<td>311</td>
<td>1304</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Number of links</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 5000</td>
<td>208</td>
<td>605</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 1000</td>
<td>194</td>
<td>472</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 800</td>
<td>194</td>
<td>470</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 700</td>
<td>190</td>
<td>212</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>True Single Links</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 5000</td>
<td>122</td>
<td>419</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 1000</td>
<td>106</td>
<td>252</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 800</td>
<td>106</td>
<td>249</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 700</td>
<td>104</td>
<td>76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Multiple Record Links</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 5000</td>
<td>40</td>
<td>58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 1000</td>
<td>42</td>
<td>108</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 800</td>
<td>42</td>
<td>109</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 700</td>
<td>41</td>
<td>92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>No. of Records Eliminated</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 1000</td>
<td>54</td>
<td>627</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 800</td>
<td>54</td>
<td>628</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 700</td>
<td>56</td>
<td>897</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MR Links with Capacity Change</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 5000</td>
<td>25</td>
<td>43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 1000</td>
<td>26</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 800</td>
<td>26</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 700</td>
<td>25</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MR Links w/o Capacity Change</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 5000</td>
<td>15</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 1000</td>
<td>16</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 800</td>
<td>16</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 700</td>
<td>16</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**COUNTY KEY**
1 = Caroline County
2 = Charles City County
3 = Hanover County
4 = Isle of Wight County
5 = James City County
6 = Louisa County
7 = Montgomery County
8 = New Kent County
9 = Spotsylvania County
10 = Surry County
11 = York County
At any given abstraction level, any route designated greater than the abstraction value is not converted from milepoint to node-link structure. Also any intersection with a route designated higher than the abstraction value is not considered a valid node. Abstraction of a highway network in this way eliminates routes not required for analysis.

The database has a total of 9098 records. These are reduced to 4520 links at the 5000 level, 3956 links at the 1000 level, 3846 links at the 800 level and 2532 links at the 700 level. But for each of these levels, 3745, 3745, 3827, 5006 records are eliminated from conversion, respectively. These figures indicate a reduction in database size by 16 percent at the 5000 and 1000 abstraction levels, 27 percent at the 800 level, and a maximum reduction of 38 percent at the 700 level of abstraction.

Figures 3 through 6 show the Root Mean Square Error (RMSE) calculated for the 64 rules used for capacity calculation for each level of abstraction. In all these graphs, there is a repeating pattern of RMSE values. This repeating pattern is observed for rules 1, 5, 9, 13; rules 2, 6, 10, 14 and so on. This can be explained by the fact that the number of lanes remains constant throughout the database. So the values for the maximum, minimum, standard average and weighted average are all the same. It is only the values for surface width and shoulder width that vary from a maximum to a weighted average.

From the above four graphs, it is seen that as the network becomes abstract, the RMSE of each rule increases in value. This is because as the network becomes more abstract, there are more links made of multiple records. Also as the abstraction increases, the pattern of rules get more scattered. This is when the effects of individual rules become pronounced.

Rules using a weighted average perform the best at all levels of abstraction. These are rules 52, 56, 60, and 64. Distances of the milepoint segments are used as
weights to calculate averages. These rules have the least RMSE among all the 64 rules. These four rules are followed by those that use weighted averages for surface width and standard average values for shoulder width. These are rules 51, 55, 59, and 63. While rules which utilize minimum surface width and minimum shoulder width also perform well, rules with maximum values have the maximum RMSE.

Figures 7 to 10 show the frequency of links for which each conversion rule was found to minimize the difference between impedance calculated from milepoint records and impedance calculated from rule-based link capacity. Similar patterns are observed for all levels of abstraction. From these plots it can be observed that rules employing minimum surface width and weighted average for shoulder width minimize the error of impedance estimates for more links than rules based on weighted average values which seemed to perform the best from the RMSE plots. Also rules 4, 8, 12, and 16 which employ maximum surface width and weighted average shoulder width minimize error for a large number of links. These results are contradictory to results found from the RMSE plots and gives support to the accepted contention that RMSE is not a good measure for comparison. However for the lack of a better statistical tool, the results for the conversion program have to be based on this statistic alone.

All the above results were for the complete data base, i.e. the eleven data files combined together. Results of RMSE tests on four separate counties are shown in figures 17 to 20. These are Caroline County, Hanover County, Charles City County, Spotsylvania County.

Figure 15 for Caroline County shows that rules that employ minimum surface width and weighted average shoulder width performed best.

Figure 18 for Hanover county deviates from the pattern demonstrated earlier by the complete data base. This also happens for Spotsylvania county and is caused by
erroneous values for the number of lanes. Rules employing minimum values for number of lanes with minimum surface width perform the best. Rule 64, using weighted average values for all the three variables does the best.

The plot for Charles City County is very different from the rest of the RMSE patterns. Rules 2-4 which employ maximum surface width values do better than rules 17-20 and 33-36 which use minimum and standard average values for surface width. However rules using weighted average values of surface and shoulder width minimize impedance error.
Figure 3. Graph of RMSE for 64 conversion rules. Abstraction level > 5000
Figure 4. Graph of RMSE for 64 conversion rules. Abstraction level > 1000
Figure 5. Graph of RMSE for 64 conversion rules. Abstraction level > 800
Figure 6. Graph of RMSE for 14 conversion rules, Abstraction level > 700
Figure 7. Distribution of records per link, Abstraction level > 5000
Figure 8. Distribution of records per link, Abstraction level > 1000
Figure 9. Distribution of records per link. Abstraction level > 800
Figure 10. Distribution of records per link, Abstraction level > 700
Figure 11. Number of links conversion rules minimizes, Abstraction level > 5000
Figure 12. Number of links conversion rules minimizes, Abstraction level > 1000.
Figure 13. Number of links conversion rules minimizes, Abstraction level > 800
Figure 15. Number of links conversion rules minimizes, Caroline County
Figure 16. Number of links conversion rules minimizes, Hanover County
Figure 17. Graph of RMSE for 64 conversion rules, Caroline County
Figure 18. Graph of RMSE for 64 conversion rules, Hanover County
Figure 19. Graph of RMSE for 54 conversion rules, Charles City County
Figure 20. Graph of RMSE for 64 conversion rules, Spotsylvania County
5.0 Conclusions and Recommendations

5.1 Conclusions

For all the databases that were run through the conversion program, rules employing weighted averages for the variables that determine capacity produced the best results in terms of RMSE. But the frequency figures show that it might be more appropriate to use a variety of rules to optimize each link. RMSE is not a good measure of variance because of the tendency for large positive errors and large negative errors to cancel each other's impact.

5.2 Recommendations for Future Research

Graphic display of a network is becoming the norm for many transportation analysis models. Inherent in the definition of Geographic Information Systems (GIS) is the ability of a system to graphically display and manipulate data. To enhance userfriendliness and simplify computational complexity of the conversion procedure described here, graphic images of highway networks are employed by the microcomputer model. However, road inventory files contain highway attribute or
descriptive data from which it is very difficult to construct a map. Locational information found in road inventory files is summarized as follows. Each record contains 1) district, county, city or town identifiers, 2) the route number of the road segment, and 3) twelve character labels identifying termini-from and termini-to locations of the segment. No digital coordinates are provided in these files. Consequently, visual representation of the underlying network is not easily obtainable from this information.

A digital data base representing a map of the study area network is utilized to simplify data conversion. This digital representation conforms to an intersection/link network structure. Information on route number or name of each link accompanies the x, y coordinates of each node. Route numbers of links intersecting at the beginning and end node of each link are necessary. Depending on the nature of the application, other locational information may include county and analysis zone identifiers.

Several options are available for creating or obtaining a digital map to be used in the conversion process. Two procedures are described below. If the study area is small and a digitizing table is available, software is provided in the data conversion system to create the required data base. Several base maps, such as county highway maps and USGS 7.5 minute quadrangle maps, may be used to identify an intersection/link structure which is matched with the location referencing method applied in road inventory files. Users may capture x, y coordinates for just network nodes or digitize several points to represent curves in each link. A label, entered during digitizing, is associated with each link in the file. This label is similar to identifiers found in road inventory files. It indicates route name and the names of intersecting routes. Providing names of intersecting routes during digitizing is not necessary but may be done easily when study area is small. Otherwise, this
information can be constructed from the data base assuming each link has been labeled. Finally, a county code may be entered which matches county code numbers found in road inventory files. County codes are used in the search process to eliminate sections of a road outside the area of interest.

An alternate method for obtaining a digital representation of the study area is to use an existing data base such as HNET files containing link and node information, or TIGER files created by the Census Bureau, or possibly Digital Line Graphs (DLGs) produced by the USGS. Any existing data base will require editing to incorporate link label information and county codes if this information is not currently present. In urbanized study areas, HNET files represent an ideal source as long as network information has been maintained. However, in many instances HNET files are incompatible with road inventory files which contain statewide data on primary and secondary roads, interstates and any road with a federal project number. In this event, TIGER or DIME files or DLGs may be employed. These files require some restructuring to achieve an intersection/link highway network data structure. Further, users should be aware that a common criticism of purchased or externally acquired data bases is that information often is outdated so, for example, newly constructed roads may not be in the file.
Bibliography


9. UTPS Highway Network Development Guide


12. GIS Design Guidelines, Florida Department of Transportation, Dec 1988


Balakrishna Akundi was born on October 14, 1965 in Vijayawada, Andhra Pradesh, India. He did his schooling in Vijayawada and New Delhi. He obtained his Bachelor’s degree in Civil Engineering in July 1988 from Osmania University, Hyderabad, India. He joined Virginia Polytechnic Institute and State University in Fall 1988, and completed his Master of Science degree in Civil Engineering, specializing in Transportation Engineering, in July 1990. His career interests are in Transportation Planning/Traffic Engineering.