Quality and Accuracy of Positional Data in Transportation

DETAILS

0 pages | | PAPERBACK
ISBN 978-0-309-43126-2 | DOI 10.17226/21953

AUTHORS

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

NCHRP REPORT 506

Quality and Accuracy of Positional Data in Transportation

E. S. FEKPE
T. WINDHOLZ
K. BEARD

AND

K. NOVAK
Battelle
Columbus, OH

SUBJECT AREAS
Planning and Administration • Highway and Facility Design

Research Sponsored by the American Association of State Highway and Transportation Officials
in Cooperation with the Federal Highway Administration

TRANSPORTATION RESEARCH BOARD

WASHINGTON, D.C.
2003
www.TRB.org

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

Systematic, well-designed research provides the most effective approach to the solution of many problems facing highway administrators and engineers. Often, highway problems are of local interest and can best be studied by highway departments individually or in cooperation with their state universities and others. However, the accelerating growth of highway transportation develops increasingly complex problems of wide interest to highway authorities. These problems are best studied through a coordinated program of cooperative research.

In recognition of these needs, the highway administrators of the American Association of State Highway and Transportation Officials initiated in 1962 an objective national highway research program employing modern scientific techniques. This program is supported on a continuing basis by funds from participating member states of the Association and it receives the full cooperation and support of the Federal Highway Administration, United States Department of Transportation.

The Transportation Research Board of the National Academies was requested by the Association to administer the research program because of the Board's recognized objectivity and understanding of modern research practices. The Board is uniquely suited for this purpose as it maintains an extensive committee structure from which authorities on any highway transportation subject may be drawn; it possesses avenues of communications and cooperation with federal, state and local governmental agencies, universities, and industry; its relationship to the National Research Council is an insurance of objectivity; it maintains a full-time research correlation staff of specialists in highway transportation matters to bring the findings of research directly to those who are in a position to use them.

The program is developed on the basis of research needs identified by chief administrators of the highway and transportation departments and by committees of AASHTO. Each year, specific areas of research needs to be included in the program are proposed to the National Research Council and the Board by the American Association of State Highway and Transportation Officials. Research projects to fulfill these needs are defined by the Board, and qualified research agencies are selected from those that have submitted proposals. Administration and surveillance of research contracts are the responsibilities of the National Research Council and the Transportation Research Board.

The needs for highway research are many, and the National Cooperative Highway Research Program can make significant contributions to the solution of highway transportation problems of mutual concern to many responsible groups. The program, however, is intended to complement rather than to substitute for or duplicate other highway research programs.

Note: The Transportation Research Board of the National Academies, the National Research Council, the Federal Highway Administration, the American Association of State Highway and Transportation Officials, and the individual states participating in the National Cooperative Highway Research Program do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of this report.

NCHRP REPORT 506

Project SP20-47(01) FY 1998

ISSN 0077-5614

ISBN 0-309-08781-3

Library of Congress Control Number 20033115135

Price $31.00

NOTICE

The project that is the subject of this report was a part of the National Cooperative Highway Research Program conducted by the Transportation Research Board with the approval of the Governing Board of the National Research Council. Such approval reflects the Governing Board's judgment that the program concerned is of national importance and appropriate with respect to both the purposes and resources of the National Research Council.

The members of the technical committee selected to monitor this project and to review this report were chosen for recognized scholarly competence and with due consideration for the balance of disciplines appropriate to the project. The opinions and conclusions expressed or implied are those of the research agency that performed the research, and, while they have been accepted as appropriate by the technical committee, they are not necessarily those of the Transportation Research Board, the National Research Council, the American Association of State Highway and Transportation Officials, or the Federal Highway Administration, U.S. Department of Transportation.

Each report is reviewed and accepted for publication by the technical committee according to procedures established and monitored by the Transportation Research Board Executive Committee and the Governing Board of the National Research Council.

Published reports of the

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

are available from:

Transportation Research Board
Business Office
500 Fifth Street, NW
Washington, DC 20001

and can be ordered through the Internet at:

http://www.national-academies.org/trb/bookstore

Printed in the United States of America

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. On the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. William A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, on its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both the Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. William A. Wulf are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is a division of the National Research Council, which serves the National Academy of Sciences and the National Academy of Engineering. The Board's mission is to promote innovation and progress in transportation through research. In an objective and interdisciplinary setting, the Board facilitates the sharing of information on transportation practice and policy by researchers and practitioners; stimulates research and offers research management services that promote technical excellence; provides expert advice on transportation policy and programs; and disseminates research results broadly and encourages their implementation. The Board's varied activities annually engage more than 4,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation. **www.TRB.org**

www.national-academies.org

COOPERATIVE RESEARCH PROGRAMS STAFF FOR NCHRP REPORT 506

ROBERT J. REILLY, *Director, Cooperative Research Programs*
CRAWFORD F. JENCKS, *Manager, NCHRP*
CHRISTOPHER J. HEDGES, *Senior Program Officer*
EILEEN P. DELANEY, *Managing Editor*
HILARY FREER, *Associate Editor II*

NCHRP PROJECT SP20-47 (1) PANEL
Field of Special Projects—Area of Special Projects

RON CIHON, *Washington State DOT* (Chair)
DAVID W. BARANIAK, *Lambda Tech International, Waukesha, WI*
JONATHAN J. DUCHATEAU, *Wisconsin DOT*
CHRIS A. GLAZIER, *Utah DOT*
MICHAEL F. GOODCHILD, *University of California - Santa Barbara*
ELIZABETH A. HARPER, *Parsons Brinckerhoff, Morrisville, NC*
BRAD JOHNSON, *Fredericksburg, VA*
WENDE A. O'NEILL, *Westat, Inc., Rockville, MD*
DARIS W. ORMESHER, *South Dakota DOT*
MIKE NEATHERY, *FHWA Liaison Representative*
THOMAS PALMERLEE, *TRB Liaison Representative*

AUTHOR ACKNOWLEDGMENTS

The research reported in this interim report was performed under the NCHRP Project 20-47(01) by Battelle, Columbus, Ohio, in association with the Department of Spatial Information Science and Engineering, National Center for Geographic Information and Analysis, University of Maine, Orono, Maine; TRANSMAP Corporation, Columbus, Ohio; and ND-Lea Engineers and Planners, Winnipeg, Canada.

Dr. Edward Fekpe, Senior Research Scientist, Battelle, was the principal investigator. The other authors of this report are Dr. Kate Beard of the Department of Spatial Information Science and Engineering, National Center for Geographic Information and Analysis, University of Maine; Dr. Kurt Novak of TRANSMAP Corporation; and Dr. Thomas Windholz of Idaho State University (formerly of University of Maine). Other research team members who contributed to the development of this report are Mr. Mohammed Alam, Mr. Mohammed Majed, Mr. Deepak Gopalakrishna, and Ms. Melissa Murphy of Battelle; Mr. Tony Zifer and John Yablonski of TRANSMAP Corporation; and Ms. Kerra Mruss of ND-Lea.

FOREWORD

By Christopher J. Hedges
Staff Officer
Transportation Research Board

This report presents guidance for practitioners on the use of positional, or spatial, data in Geographic Information Systems (GIS) for transportation applications. As GIS applications become more common in transportation-system management and decision making, concerns have grown about the accuracy of the data used to populate them. Transportation spatial data usually describes the location of features on the highway system using a one-dimensional linear referencing system. The level of accuracy varies by data source and is affected by the precision of the measurement system used to collect it. Agencies need a way to understand the errors that may result when using these data in GIS applications and how these errors may be compounded when combining data from various sources. This project reviewed the linear referencing systems used by state DOTs, examined the limitations of typical data sources used in these systems, and developed a model to evaluate the effects of varying data accuracy and provide an assessment of the level of confidence in the system outputs. This report will provide valuable information to transportation practitioners who need to understand and account for the level of precision in GIS-based transportation decision tools.

Most transportation data are linearly referenced in a one-dimensional (1D) model. The implications of spatial data quality in the 1D model are not well understood, thereby significantly limiting the value of analyses using these data and the efficacy of subsequent decision making.

Many methods have been used to measure the positions of objects or events relative to the highway network, and technologies such as Global Positioning Systems (GPS) are providing highly efficient means for accurately establishing 2D and 3D positions that can be used to locate point features such as accidents, signs, and intersections rapidly and conveniently. GPS can also be used for locating moving vehicles in real time. Difficulties arise with the use of GPS technology because, traditionally, the location component of a data item is captured in coordinates (2D) that must be transformed to some linear reference such as log-mile point (1D). Moreover, analytical operations on spatial data, in support of transportation applications, are complicated because coordinate geometry cannot be applied to positions referenced in linear space (e.g., the distance from A to B is measured along a path, not along a straight line between coordinates).

Designers and managers of GIS need guidance on the appropriate scales and number of calibration points in formulating DOT base maps. Practitioners using GIS-generated data summaries need to know the bounds on "true" location that can be derived from the integration of diverse data sources (e.g., data collected using distance measuring instruments and GPS). There is also a need for methods that will allow the transformation between location referencing systems in the field and in the office and measures of the confidence limits of these transformations.

Under NCHRP Project 20-47(01), a research team led by Edward Fekpe of Battelle began by describing the characteristics of spatial data, including linear referencing methods and systems. The research team then evaluated the quality and precision of available spatial data sources, given the accuracy capabilities of current spatial data measurement methods. After reviewing transportation applications for spatial data and their sensitivity to data quality, the research team developed a prototype error model that can be used to understand the effects of using and combining typical data sources. The outputs of the model list the confidence values at various probabilities associated with specific data sources for a wide range of transportation applications. The error model is implemented as a software program called GISError. It was developed in Visual Basic with a graphical user interface, and is included with this report on *CRP-CD-41*. A user guide for GISError is included in the report as Appendix A.

CONTENTS

QUALITY AND ACCURACY OF POSITIONAL DATA IN TRANSPORTATION

SUMMARY

Spatial data used by state departments of transportation (DOTs) come from different sources and are used for various applications. Although some states are quite advanced in using new data collecting techniques, others rely on traditional methods. Different users have different perceptions as to the importance of error and accuracy, because the value of spatial data is directly concerned with the fitness of the data for a particular purpose, and the critical measure of that fitness for use is quality. Emerging applications of positional data include emergency evacuation, automated oversize overweight truck permitting and routing, and bus routing. Applications such as transportation planning, commercial vehicle operations, and regulatory and policy analyses are less sensitive to accuracy of positional data than highway inventory, highway design, and construction applications.

There is no uniformity in the linear referencing methods (LRMs) used by state DOTs in collecting and referencing spatial data for transportation applications. Most states use multiple LRMs. When correlating data from various systems or LRMs, state DOTs tend to rely on in-house transformation methods or algorithms built into the geographic information system (GIS) software.

The primary sources of error associated with positional data are acquisition or measurement, processing, transformation, and presentation or visualization. A transformation can be between different reference systems as well as different reference methods. Transformations introduce a degree of uncertainty to the transformed data. A prototype data error model was developed based on an object-oriented approach, where the position of an event can be visualized as an object that depends on (a) an event, (b) a reference system, (c) a network, and (d) a measurement device or methodology. Conceptually, the data error model was designed to handle the uncertainties associated with event data and networks present in transportation applications.

The concept of probability zones to describe the uncertainty of locations was used for visualization of errors. The calculation of a probability zone is based on the measurement error as well as the resolution of the applied measurement system or the embedded reference system. The probabilistic approach assigns n-dimensional probability zones immediately surrounding every n-dimensional measured feature location. The size of each of these n-dimensional zones depends on two components: (1) the

uncertainty arising from imprecise measurements or derived inaccuracy values and (2) a user-selected probability threshold that the *true* feature location is to be found within this probabilistic space. The confidence intervals are based on the χ^2-distribution, where the probability that the measured point location is within the tabulated distance of the true point location can be tested.

The prototype data error model is encapsulated into a software program developed in Visual Basic programming language with graphic user interfaces. The program has a feature that allows results from the error analysis to be exported to other GIS application software. The program was developed to help analyze and visualize the results of data errors and display how errors of different data sources affect each other. The program outputs the results of the error analysis and shows the confidence intervals and buffers around data points and lines. The program also computes the probability of intersection of two features or data sources to determine whether they are compatible and should be used together. The data error model is capable of handling the following types of data elements:

- **Point Feature**: defined by a Cartesian (X, Y) coordinate, such as the Easting and Northing in a State Plane system;
- **Line Feature**: defined by two points (i.e., the starting and end points, also defined in the Cartesian coordinate frame);
- **Points on a Line**: defined by a distance from the starting point. These correspond to the linear reference system used for transportation applications (the distance is the mileage value of the event from the beginning of the route or the nearest intersection); and
- **Error Values**: representing the mean error of the data element (e.g., the error value of a line represents the uncertainty of the location of the whole line as a buffer in the specified coordinate frame).

When two data elements and their corresponding tolerances (error values) are entered, the program generates a graphical display of the situation showing error buffers of various probabilities at different significance levels (i.e., 75%, 80%, 85%, 90%, 95%, 99%) and presented in different colors. The program also generates the Intersection Probability, which is a measure for the likelihood of the two data elements intersecting. This number is important to determine which datasets are actually compatible and which datasets should not be combined.

The model was tested with real-world case study data for a wide range of transportation applications. The case studies demonstrated that the prototype model is sufficiently generic and can be used to evaluate the quality of positional data intended for a wide range of transportation applications. The prototype data error model would allow users of positional data to be aware of the bounds of the “true” location that can be derived from the integration of diverse data sources and the level of certainty that can be associated with spatial data. The model also would allow users to assess the potential quality implications of combining data from different sources and with different qualities. The program offers an efficient way to visualize the quality of the data at different significance levels of confidence.

Recommendations for using the prototype error model and standards for positional data quality are developed. Recommendations for positional data quality standards include metadata documentation for linear datum components to ensure stability and reportability of positional data quality. It is recommended that positional accuracy reports should indicate the positional accuracy of various components. Key components affecting a linear reference are the linear datum components (anchor sections and anchor points), the network, the linear measurement methods involved, and their dependencies. At a minimum, separate positional accuracy reports for anchor sections and network components should be included. The best way to overcome the loss of positional accuracy in transformations from linear referencing to 2D or vice versa is to ensure a consistent matching of accurately measured anchor sections to the corresponding sections of any digital spatial representation.

CHAPTER 1

INTRODUCTION AND RESEARCH APPROACH

1.1 PROBLEM STATEMENT

Transportation agencies use spatial data to locate or describe events on a transportation system. The spatial representation of a network can be expressed in one, two, or three dimensions. Each representation level has sources of uncertainty. Transportation data are usually referenced to highway networks by using a one-dimensional (1D) linear referencing model. With this model, objects along a network are located using a set of known points on the network and distances and directions from the known points to the objects. Linear referencing is currently the most common practice used to locate or describe transportation features.

Many methods have been used to measure the positions of objects or events relative to the highway network. Emerging technologies, such as Global Positioning System (GPS) receivers, are providing highly efficient means for establishing two-dimensional (2D) and three-dimensional (3D) positions that can be used to locate point features, including crashes, signs, and intersections, rapidly and conveniently. GPS also can be used for locating moving vehicles in real time. Recently, data collection vehicles with GPS positioning capability have been acquired by some transportation agencies to support highway inventories and photologging.

Difficulties arise with the use of GPS technology because, traditionally, the location component of a data item is captured in coordinates (2D) that must be transformed to some linear reference such as a log-mile point (1D). Furthermore, analytical operations on spatial data, in support of transportation applications, are complicated because coordinate geometry cannot be applied to positions referenced in linear space (e.g., the distance from A to B is measured along a path, not along a straight line between coordinates).

Spatial data quality is associated with the idea of fitness for use, which refers to the fact that different transportation applications require spatial data at different scales and that no one scale can support all transportation applications. Data quality assesses the degree of uncertainty associated with data. A better understanding and means to assess the quality of positional data (i.e., 1D spatial data) offers various benefits. The implications of spatial data quality in the 1D model are not well understood, placing significant limitations on the value of analyses using these data and the efficacy of subsequent decision making. National spatial data quality standards have been established for 2D and 3D data. These standards allow users to understand the robustness of the data and to make judgments concerning the level of risk in decision making. There is also a need for methods that will allow the transformation between location referencing systems in the field and in the office as well as measures of the confidence limits of these transformations.

1.2 OBJECTIVES AND SCOPE

This research is intended to compile and develop information needed to address issues related to positional data quality. This includes the formulation of methodologies to analyze the effects when considering trade-offs or transforming the location data obtained from different measurement systems. The specific objectives of the project are as follows:

- Identify the positional data quality needs for common transportation applications,
- Document the effectiveness of various techniques for establishing spatial positions,
- Develop methodologies for assessing the effects of positional data accuracy in transformations between measurement techniques and spatial referencing systems, and
- Package the findings into materials that can be readily implemented by DOT personnel.

A primary focus of this project is on linearly referenced data that are predominant in transportation agencies. Also, positional-data quality is intended to include, at least, data accuracy, precision, and resolution.

1.3 REPORT ORGANIZATION

This report is organized into four main chapters, the first of which provides an overview of the problem statement, objectives, and research approach. The second chapter, which describes the research findings, is divided into several sections as follows:

- The first section describes spatial data characteristics, which include linear referencing methods and systems;

- The second section describes spatial data quality, spatial data errors, and data quality standards;
- The third section describes measuring systems and accuracy capabilities; and
- The fourth section identifies the transportation applications of positional data and the applications' sensitivity to positional data quality

These sections also include summaries of data collected from the states in order to reflect practices performed by state departments of transportation (DOTs).

The fifth section of Chapter 2 describes the data error model. This includes a description of the data error modeling concept, sources of error, transformation methodology, presentation and visualization of positional data error, and a prototype error model.

Chapter 3 discusses the results of evaluation of the data error model (i.e., case study analysis), guidelines for incorporating data error indices in GIS applications, and recommendations for positional data quality standards.

Chapter 4 presents the conclusions and recommendations for suggested research.

Appendixes A, B, and C contain a prototype Data Error Model User Guide, results of case studies, and the questionnaire used to interview state DOTs, respectively.

CHAPTER 2

FINDINGS

2.1 SPATIAL DATA CHARACTERISTICS

2.1.1 Introduction

"Spatial data" refer to information that is referenced to a geographic location on the earth and includes the three dimensions of space, time, and theme (where-when-what) *(1, 2)*. Spatial data include information that represents the geographic position of features as well as descriptive information about those features. Nearly all transportation data are, or can be, geographically referenced. Geographic Information Systems (GIS) provide an effective way to manage and integrate the spatial data necessary for the planning, design, construction, analysis, operation, maintenance, and administration of transportation systems and facilities. Transportation agencies use spatial data to locate or describe events on a transportation system. The spatial representation of a network can be expressed in one, two, or three dimensions. All spatial data can be characterized and defined as one of three basic feature types: points, lines, or areas, which are described as follows *(1, 3)*:

- *Points* refer to data associated with a single location in space and, because of the scale of the map, are represented by symbolic points, rather than by an areal dimension. Examples of point data include wells, post boxes, and lampposts.
- *Lines* refer to data represented by a one-dimensional (1D) line and are described by a string of spatial coordinates. Examples of line data include roads, railways, rivers, and pipelines.
- *Areas* refer to data represented by a common string of spatial coordinates, homogeneous zones as defined by natural property or categories, or alternatively artificial units used for thematic representation or management purposes. Areas are also commonly referred to as polygons. Examples of area data include land-use zones, soil classification areas, administrative boundaries, and climate zones.

The demand for spatial data in GIS for transportation applications has grown exponentially since 1990. A lack of spatial data is no longer an issue limiting GIS-T applications. The main issues are quality, sensitivity, and long-term (cumulative error) effect of both transforming a linearly referenced one-dimensional data model to its cartographic representation and transforming two-dimensional (2D) and three-dimensional (3D) data models to linear representations relative to the network.

Every year more spatial data become available. In the past, most spatial data originated from government sources. Recently, however, more and more spatial data are being produced by the private sector, for both specific projects and resale. Some of the existing historical spatial data collected and maintained by government agencies are now considered to be of low quality and inconsistent with data available from newer technologies *(4)*.

Unknown spatial data accuracy is becoming an increasingly common problem. Unknown data quality leads to tentative decisions, increased liability, and loss of productivity. Conversely, decisions based on data of known quality are made with greater confidence and are more easily explained and defended *(5)*. The quality of spatial data is becoming increasingly important as large databases are created for access and exchange by many individuals. The recognition, evaluation, and resolution of errors associated with spatial data are important issues that, until quite recently, have received little attention, as the problems of error and accuracy were largely unknown *(3, 5)*. Users are becoming increasingly concerned about the quality and reliability of spatial data.

There is a need for good quality spatial data, where the term "good quality" is defined by the data's specific application as well as other information concerning the data quality. Different users have different perceptions as to the importance of error and accuracy, as the value of spatial data depends on its fitness for a particular purpose. The critical measure of that fitness for use is quality *(3, 5)*. Despite the importance of having information about data quality, information on the accuracy and reliability of spatial data is generally poor or nonexistent. Unfortunately, determining and ensuring the accuracy and integrity of spatial information is complicated *(6)*.

Quality assurance is a basic requirement for performing an application reliably, and any application performed using spatial data should be accompanied by a detailed evaluation of the quality. This evaluation will help to determine if the data adequately represent the information needed to answer the question raised by the application. The quality of spatial data should be assessed and reported as part of each spatial data file of information. In addition, a comprehensive statement of data quality should accompany the transfer of all spatial data. As well, the quality of spatial data used in any analysis should be passed on to the consumer of that analysis. Different data types will tolerate different margins of error and accuracy depending on their specific application *(3)*.

The concern for spatial data quality has increased in recent years because of factors, such as the following *(2)*:

- Increased data production by the private sector and non-government agencies, which are not governed by uniform quality standards (production of data by national agencies has long been required to conform to national accuracy standards);
- Increased use of GIS for decision support, highlighting the implications of using low-quality data, including the possibility of litigation; and
- Increased reliance on secondary data sources, because of the growth of the Internet, data translators, and data transfer standards, making poor quality data ever easier to get.

2.1.2 Linear Referencing

Transportation data are usually referenced to highway networks by using a 1D linear referencing model. With this model, objects along a network are located using a set of known points on the network and distances and directions from the known points to the objects. All linear referencing methods are based on this concept *(7)*. Many transportation agencies use various spatial measurement techniques to describe events linearly or locate events of the "network profile" and "point profile" spatial components. This technique is commonly known as the Linear Referencing Method (LRM), defined as a "way to identify a specific location with respect to a known point." A Location Referencing System (LRS) is "a set of office and field procedures that include a highway location reference method" *(8)*.

Theoretical models for referencing linear objects typically use combinations of one, two, or three independent concepts that "anchor" the linear objects to reality. These three elements are (1) an identifier, (2) a physical linear extent without a reference point, and (3) a temporal linear extent *(9)*.

The basic structural variations (data elements) of linear referencing methods' theoretical models are largely a function of the event measurement techniques in locating a point or linear object. The event can be identified as an offset from known points or by a series of "control" or reference points *(10)*. Various methods for linear location referencing in transportation have come about because state DOTs need to know where objects and attributes are located on roadways. These roadways can be conveniently modeled as linear features, allowing the application of linear location referencing *(11)*.

2.1.2.1 Linear Referencing Methods

The primary objective of any highway location referencing method is to provide a means for designating and recording the geographic position of specific locations on a highway and for using designations as a key to stored information about locations. A method's planned application determines its most significant characteristics. Three elements common to all location referencing methods are (1) identification of a known point, (2) measurement from the known point, and (3) direction of measurement *(8)*.

Various measuring systems as well as referencing methods are available to state DOTs. The critical difference between various linear referencing methods is their respective measurement techniques. The event can be identified as an offset from a known point or by a series of reference points. Linear location referencing methods commonly used by transportation agencies include route-mile-point, route-reference-post-offset, route-mile-post-offset, and methods based on link-node models *(8, 12)*. Adams et al. *(11)* described the various LRMs as well as their advantages and disadvantages. Data collected using one of the LRMs may not be suited for applications based on another method. The inability to relate and/or cross reference information results in the effective loss of information.

These LRMs use an offset distance along a highway from a known beginning point to define the position of interest. Such items include attributes of the road and features that exist as part of the road or adjacent to it. Typical attributes include speed limit, pavement type, functional class, traffic volume, number of lanes, and jurisdiction. Common features include intersections, bridges, signs, and guardrailing. Both attributes and features may be of the point or linear type. A point data item is located using a single offset distance, while a linear data item is located using a pair of offset distances (beginning and ending).

2.1.2.2 Linear Referencing System (LRS) Data Model

This section provides an overview of existing LRS data models that serve as a guide to state DOTs in developing their LRS.

The NCHRP 20-27(2) LRS Data Model. The NCHRP 20-27(2) linear LRS data model was developed in response to a growing awareness of the need to integrate increasing amounts of linearly referenced data used by the transportation community *(10)*. The NCHRP 20-27(2) data model includes multiple linear location referencing methods, multiple cartographic representations, and multiple network representations. Data integration is supported through transformations among methods, networks, and cartographic representations by association with a central object, referred to as a "linear datum."

The conceptual model for the LRS in NCHRP 20-27(2) *(10)* was designed to meet four basic functional requirements: (1) determination of unknown locations of items of interest in the field, (2) positioning of these items in location-referenced databases, (3) placement of these items of interest in the field at known locations, and (4) transformation of linear location references among various methods. The model was intended to be generic so as to support as many applications as possible. Therefore, a fundamental "lowest common denominator" was sought as the generic basis for multiple LRMs, because it forms the functional "heart" of a true LRS.

The NCHRP 20-27(2) LRS data model was created to facilitate sharing linearly referenced data across modes and agencies. It provides the framework to manage and transform linearly referenced data. The central notion is a linear datum that supports multiple cartographic representations (at any scale) and multiple network models (for various application areas). The datum consists of anchor points and anchor sections connecting these points. It also provides the fundamental referencing space for transformations among various LRMs, network models, and cartographic representations *(10)*.

The Dueker-Butler GIS Data Model. The Dueker-Butler LRS data model covers a broad set of business rules for all modes of transportation and a wide range of applications *(13, 14)*. The LRS components generally fall into four classes: (1) the geographic network, (2) cartography, (3) the transportation network, and (4) transportation topology. The model was designed to accomplish the following goals:

- Accommodate the basic forms supported by GIS: point, line, and area (the model focuses on attributes);
- Express point and area features in terms of their relationship to linear features in transportation databases;
- Support fixed- and variable-length segmentation schema;
- Support the four functional requirements of the 20-27(2) model *(10)* ;
- Express functional requirements as business rules (data and process requirements) *(13, 14, 15, 16, 17)*; and
- Support non-transportation features of the point and area types, and add a mechanism for expressing the location of linear transportation feature attributes using real-world (2D and 3D) coordinate systems *(13)*.

Generalized Model. The generalized model is a simplification of the NCHRP 20-27(2) *(18)* conceptual data model. The lower four levels of the NCHRP 20-27(2) model are all composed of linear elements. LRMs use traversals, networks have links, the linear datum has anchor sections, and the cartographic representation has lines. If the constraint against locating events directly on a link, anchor section, or line is relaxed, the NCHRP 20-27(2) model can be compressed into a two-level model. The generalized model has the following characteristics:

- It de-couples linear element types from measurement methods so that measurement methods may be applied to multiple linear element types.
- It generalizes the concept of a linear element in order to enable links and anchor sections to be treated as traversals.
- It allows event locations to be specified against any linear element.
- It formalizes the concept of a location expression as the combination of an LRM (measurement method and linear element type), linear element instance, and distance expression.
- It formalizes the concept of distance expression.
- It enables the generalization of the translation process between locations, linear elements, or LRMs.

The generalized model has been reduced to two levels by realizing similarities between each of the four lower levels of the NCHRP 20-27(2) model. According to the generalized model, the networks are not required as an intermediary between event locations and the linear datum, as long as tracing is not required or if the linear datum is complete with respect to connectivity. The Network, Linear Datum, and Cartographic Representation levels become LRMs, with linear elements that can directly support event locations. Though not mandatory, the linear datum LRM is still recommended in order to simplify the translation between multiple LRMs *(18)*.

Scarponcini *(19)* suggested the introduction of a more robust location expression (LX) that provides an association of an event to a location by applying LX = (LRM, LE, DX), where LRM is the linear referencing method, LE the linear element, and DX a distance expression native to the referencing method. In this discussion, the author also mentions the possibility of extending the model to support lateral offsets as well as temporal information. This generalized approach is also suitable for the inclusion of an uncertainty component, which is discussed in the subsequent section.

2.1.3 Summary of State Practices

Information presented in this section is based on a survey of state DOTs and other spatial data users. In all, 33 state DOTs and 3 other organizations were surveyed. The response rate to the survey was just under 30 percent. However, only a few of the respondents provided useful information for the purposes of this research. The information from the survey is useful in providing a reflection of the state practices. The following subsections summarize state practices with regard to the characteristics of spatial data, LRMs, and transformation methods. A discussion of current and emerging transportation applications of spatial data by state DOTs is presented in Section 2.3 of this report.

2.1.3.1 General Characteristics of Spatial Data

Spatial data used by state DOTs and other transportation authorities such as Oak Ridge National Laboratory (ORNL) come from different sources and are used for various applications. Although some states are quite advanced in using new data collecting techniques, others rely on traditional methods. The base maps of the highway network maintained by state DOTs were digitized from 1 : 24,000 scale aerial photographs, U.S. Geological Survey (USGS) digital ortho quarter quads (DOQQs), or coordinate geometry (COGO). The following sections summarize information on the characteristics of spatial data maintained by various state DOTs. The databases, as well as line, point, and area data profiles, are described. The data source and the measuring methods are also identified.

Arizona. Arizona Department of Transportation (ADOT) has three main spatial data products, each containing a specific type of information. One database contains the state highway network data at a scale of 1 : 5,000. The data were captured with a GPS device with desired resolution in the range of 1 to 3 m. This database is used for highway inventory, planning, and safety applications. The tolerance level of these applications is 12 m. The second database contains a local streets network for the entire state, at 1 : 6000 to 1 : 24,000 scale. The data were collected using photogrammetry and GPS. The desired resolution is 3 to 15 m with a tolerance of 15 m. This database is used for planning and safety applications. The third database contains highway markers at 1 : 12,000 scale. GPS and photogrammetry are used in collecting these data with desired resolutions in the range of 3 to 7 m and 12 m tolerance. This database is also used for highway inventory, planning, and safety applications.

ADOT's original data from its 1970s system were modified to meet current ESRI software and in-house business needs. ADOT staff are currently using ArcInfo covers and ESRI measured shape files.

Iowa. Line data, such as highway network data, come from digital computer-aided design and drafting (CADD) files/COGO. The original resolution was 1 : 100,000 Digital Line Graph (DLG) data. Information for road plans is obtained from cities and counties, local maps, and local aerial photos. The original resolution of these data was 1 : 24,000. Line data are updated yearly with information that varies in accuracy. Rail network information is derived from USGS digital ortho quarter quads (DOQQs). Iowa DOT is currently exploring the possibility of using GPS to capture rail network data. Currently, only a U.S. DOT Bureau of Transportation Statistics (BTS) layer is available and it is used for reference.

Point data such as crash locations are located using an Iowa DOT-developed location tool. These are located on maps in GIS software with 1 : 100,000 scale. The tools use the map and allow the user to locate the crash on the map. Airport locations are placed on the map cartographically.

Polygon area such as boundary information is obtained from various sources and are usually mapped or digitized along with roads. It is assumed that Iowa state boundary information is 1 : 100,000 scale and the city boundaries are more accurate (approximately 1 : 24,000). The resolution also varies. Boundaries are very difficult to capture from readily available sources (e.g., aerial photos, satellite). Most information comes from boundary descriptions that are then digitized as accurately as possible.

Maine. Examples of events stored in the Maine DOT databases are crash locations, assets (e.g., bridges [>20ft.], struts [10 to 20ft.], culverts [<10ft.]), pavement conditions (roughness, skid resistance), business signs, road signs, spray properties, traffic counts (per min, hour, day, and truck, car), and intersections. Events are usually measured as a distance from a node. The origins (i.e., the nodes) are considered to have no inaccuracies (i.e., they are considered error-free). Thus, any errors attached to events originate from the measurement method. Using accurate distance measuring instruments (DMIs), which yield results that are within the given accuracy requirements, basically eliminates these errors. One exception is crash locations, because they are estimated by the responding police officer (crash locations are also stored as distances from nodes). If a study involves crash locations, the resolution interval of the study is set to 0.3 miles, thus addressing the issue of uncertainty in an indirect way.

North Dakota. In addition to its GIS base map data, the North Dakota Department of Transportation (NDDOT) maintains a database inventory, Roadways Information Management System (RIMS), of roadway features. In that system, NDDOT maintains a location inventory (using dynamic segmentation and route-ID/route-measure coding) for all guardrail, fences, lights, signs, roadway geometry, mile markers, roadway construction, pavement conditions, and numerous other themes in the highway system. These data tables are used within ArcView as event themes referenced against the state's centerline roadway coverage, which contains the other half of the dynamic segmentation-coding scheme. The design goal for those collecting RIMS information is that the location accuracy be better than the $^1\!/_{10}$-mile offsets used for the route measure and, using differential GPS, within a few feet for orthogonal displacement from the road centerline for all features.

Ohio. The main type of spatial data used by the Ohio Department of Transportation (ODOT) is the base network. Most of the data are derived from USGS quads and control points for field survey with DMI with scale 1 : 24,000. The current resolution is at 50 ft while the desired resolution is 5 ft. The 1947 highway inventory was collected with mechanical DMI. Since then, stations have been converted to mileposts. The use of electronic DMI is calibrated to base and verified inventories. In addition to the base network, ODOT maintains an inventory of ramps at the same 1 : 24,000 scale that is derived from DOQQs.

Position of points is determined with GPS and traditional methods (spirit and digital leveling instruments). Because coordinates are used, scale is unimportant. Data collected during surveys are archived for each project. Crash location data are derived from descriptions by state and local law enforcement agencies. ODOT also maintains a bridge database that is used for condition assessment and permit vehicle routing.

The majority of data are stored at an accuracy of 52.8 ft. However, much discussion has concerned what accuracy would be appropriate for the DOT considering cost, time, and functionality. The current thoughts are that 52.8 ft should satisfy these requirements. Although more accuracy is usually better, increased accuracy has a cost. The other issue is the recent sharing of data with other state agencies and/or local governments. Most of the data received from other agencies is based on 1 : 24,000 scale USGS quads.

ODOT recently began evaluating the use of GPS for field data collection. Initial results are encouraging. The main issue identified with the use of GPS for the update and maintenance of the LRS has been how to automate the propagation of the changes throughout the highway network as well as the updating of control point information, both attributes as well as location.

Pennsylvania. Pennsylvania maintains two spatial data products, each containing a specific type of information. The first, the Geographic Names Information System (GNIS), containing quad sheets and points, is derived from USGS. The second database, Global Data Technologies (GDT), contains addresses and centerlines. The nominal scale for these data products is 1 : 24,000. The line data profile includes highway network and waterway data. These are derived from USGS DOQQS and quad sheets. Point data, such as the location of airports, are collected with GPS by other agencies. Crash location data are estimated by state police with varying precision. Point data for vertical control are collected with traditional geodetic equipment. Polygon data include boundaries and drainage basins.

Oak Ridge National Laboratory. Oak Ridge National Laboratory (ORNL) relies on spatial data collected by other agencies and maintains four spatial databases. First, the national highway planning network (NHPN) contains the highway network for the entire United States, Canada, and Mexico. This network has a nominal scale of 1 : 100,000. It is used for planning, routing, and infrastructure applications.

Second, the Center for Transportation Analysis railroad network is a rail network for the United States, Canada, and Mexico. The nominal scale for this rail network is 1 : 100,000 and is derived from TIGER, DLG data, and Federal Railroad Administration (FRA) files. This network is also used for planning, routing, and infrastructure applications.

Third, the global seaway database contains the waterway network that covers U.S. coastal and inland waters and the world's oceans. This network is derived from the U.S. Army Corps of Engineers' National Waterway Network (NWN). The nominal scale is 1 : 100,000 (U.S.) and 1 : 3,000,000 (world). This database is used primarily for planning purposes.

Fourth, the terminals database contains intermodal facility locations at 1 : 100,000 scale, derived from surveys. This database covers the United States and is used for planning applications.

2.1.3.2 Linear Referencing Methods and Transformation Methods

This section discusses the LRMs, LRSs, and transformation methods used by state DOTs. There is no uniformity in the LRMs used by state DOTs in collecting and referencing spatial data for transportation applications. Most states use multiple LRMs. State DOTs tend to rely on in-house transformation methods or algorithms that come with the GIS software. Some states use GPS for collecting crash and other point data.

Arizona. The ADOT data model (ATISROADS) has the capabilities of linear referencing, dynamic segmentation, and routing. It does not have address or location geocoding capabilities. The data model is based on an in-house mainframe system and is modified to include ESRI coverage and route systems. Positional data are not currently collected uniquely for LRS. GIS and LRS are used subsequently for display only, or display and analysis. All files relate to road centerline with offsets and lengths varying from 1 ft to 1 mi.

For the purposes of transformation, Arizona builds intersection tables for each route in a system, with all reference markers (MP) and intersecting features (i.e., roads, jurisdiction boundary, drainage, and rail) and the corresponding route measure for each marker or feature. The corresponding route measure, with plus or minus offset, is used to geocode (linear reference) point or length data along the measured route.

Iowa. The Iowa Department of Transportation (Iowa DOT) does not currently have an implemented LRS. Iowa DOT has several different LRMs that are inconsistently used in the field to measure events/features. Iowa DOT is implementing an LRS model based on NCHRP 20-27(2). This model has linear referencing, dynamic segmentation, and routing capabilities. This model does not, however, have address or location geocoding capabilities. Iowa DOT does not currently follow a specific spatial data model or standard. Iowa DOT uses several LRMs for different applications as follows:

- Route-Mile-Point is used for Geographic Information Management System (GIMS) and inventory data for creating Highway Performance Monitoring System (HPMS) data and general road inventory. DMI are used for collecting road inventory data along the centerline with a precision of 0.01 mile. This LRM is also used for videolog inventory for right-of-way data using DMI at a resolution of 0.01 miles. It is also used for referencing pavement management.

- Route-Reference-Post-Offset is used primarily for referencing pavement management data collected with DMI and GPS at a 0.01 mile resolution.
- Stationing is used for sign inventory. Stationing is the measurement technique and the resolution is 10 feet. This information is used also to locate and monitor driveways.
- Coordinate Route using 1 : 100,000 map location is the LRM used for referencing crash locations for crash analysis. GPS is used for data collections with resolution of 0.01 mile.
- Coordinate Route using GPS is another LRM used for video-logging, inventory, and pavement management data collection.

With regard to transformation, most data are referenced back to GIMS using a link to its segments. Segments vary in length and are created based on about 27 different criteria. Any time one of the criteria is met, a new segment is created. Another translation method used is a cross-reference table that links the GIMS milepoint by route and county for the primary system with the Reference Posts. The final method is to conflate data to the GIMS centerline cartography from Coordinate Route (route name and GPS or DGPS coordinates). Given that the GIMS centerline cartography is based on a point of intersection, and, in most cases, does not have multilane facilities mapped separately, some errors occur when this conflation is done. The new LRS project will use a datum concept from the NCHRP 20-27(2) model to do all transformations between the LRMs.

Maine. The Maine DOT is in transforming its old version of the linearly referenced data structure (TINIS) into a future version of a linearly referenced data structure (D-Roads). Both identify a system of links and nodes. Links have a nominal maximum length of 6 miles. The numerical precision of the measured link lengths stored in the two systems was given as ±0.01 mile and ±0.001 mile, respectively. Their approach is to try to measure all new features within the required margin of accuracy and simply add the numerical precision to transformed events (from TINIS to D-Roads). Since re-measurements of all primary links are scheduled to occur every 2 to 4 years, the accuracy requirements of event data should be met within that timeframe. D-Roads is based on control nodes and segments between those nodes. Present and future measurements of the distances between nodes (i.e., segments) are collected via a DMI. The endpoints of the segments are defined by existing features, such as town lines, crossings, bridges, or by the nominal maximum 6-mile length. Updates of segment length are not necessarily constrained by the abovementioned update rate of a 4-year maximum. Segment lengths are seen as constants as long as no significant changes to the network itself are applied (e.g., a new bridge or road realignment).

Currently, no multi-dimensional (i.e., 2D or 3D) reference system is in use. There are plans to incorporate GPS-measured coordinates into very task-specific applications. For example, to enforce “no-spray zones,” a survey crew would first measure the location of a no-spray zone using GPS and then provide the spray crew with the acquired coordinates. In general, the spray planning (e.g., determination of amount of spray) is still based on the linear system. No plans exist to merge these two reference systems. The expectation is that both systems will continue to run in parallel.

Ohio. ODOT uses an in-house LRS standard based on county, route, and log (CRL). The LRM is coordinated with base route, with changes at county boundaries. LRM has dynamic segmentation capabilities, but no routing, address, or location geocoding capabilities. Measured accuracy in the region of $^1/_{1000}$ of a mile is converted to $^1/_{100}$ of a mile. The base files used as the DOT’s LRS were digitized from USGS quads with location and distance collected using a DMI. Much of the DOT’s roadway-based data are currently collected using some type of DMI and/or related back to a set of manuals containing LRS information (straight line diagrams).

Ohio does not have any unique transformation method because transformation is automatically executed in the software.

Pennsylvania. The Pennsylvania Department of Transportation (PennDOT) uses Integragh MGE, which has capabilities for linear referencing and dynamic segmentation, but not for routing or address or location geocoding. PennDOT uses the route offset LRM for all spatial data. Traffic, highway maintenance, and project-related data are estimated from LRS with ±50 feet precision. DMI is used for other business data such as shoulder, pavement roughness, road signs inventory, guardrail, and bridge inspection. The precision of DMI is approximately 1 foot.

ORNL. ORNL uses internal (in-house), ad hoc models. These models have capabilities for linear referencing, dynamic segmentation, routing, and address or location geocoding. All standards are internally driven by applications or else provided. The process of transformation invariably involves identification of “control points” (common locations identifiable in both inventory list and network), route construction between them, and interpolation.

2.2 SPATIAL DATA QUALITY

The issue of data quality is continuing to challenge the spatial data community. Data quality is the relationship of the spatial data to the reality that it is attempting to represent *(20)*. The value of any spatial data depends less on its cost and more on its fitness for a particular purpose. Quality of spatial data, therefore, can simply be defined as its fitness for use. This definition enables users to make a judgement for each specific application, and quality is directly based on the extent to which a data set satisfies the needs of the person judging it *(6)*. Data that are appropriate for use with one application may not be fit for use with another. The quality of spatial data depends fully on the scale, accuracy, and extent of the data set, as well as the quality of other data sets to be used *(1)*.

Different transportation applications require spatial data at different scales, and no one scale can support all transportation applications. The life span and multiple uses of spatial data generally require that quality be assessed repeatedly and from different perspectives depending on the type of transportation analysis. Spatial databases must be properly maintained and upgraded in order to maximize their usefulness (e.g., updated with changes in alignment, topology, and referencing systems) *(7)*. Recent concerns over the accuracy and reliability of spatial information in GIS have raised interest in trying to understand the reliability and uncertainty of GIS information. Because of the variety and amount of linearly referenced data that need to be stored in geographic databases, it is crucial to provide reliable, efficient procedures to link all relevant data sources linearly *(21)*.

When used in GIS analysis, a data set's quality significantly affects confidence in the results. Unknown data quality leads to tentative decisions, increased liability, and loss of productivity *(22)*. The primary objective of data quality standards is to help data recipients and owners evaluate the "fitness for use" of data. Definitions of "fitness for use" vary, based on environment and intended application. Therefore, a definition of "data quality" should include a sufficiently broad set of criteria to address the full range of possible data characteristics that might affect its application. Setting data quality standards and documenting data quality require considerable forethought. The investment pays off, however, when evaluating the data for use, when sharing the data, and when attempting to communicate the benefits and limits of conclusions based on the data *(23)*.

2.2.1 Measures of Quality

Data quality is expressed in terms of precision, accuracy, and resolution. When referencing location, it is important for the field data collector to be aware of the resolution and precision of the offset needed to report locations (e.g., 0.1, 0.01, 0.001 of a mile/kilometer), and the measurement position (e.g., along the centerline, along the shoulder lane, along the median lane). When using referenced locations for analysis, it is important for the analyst to be aware of the location resolution and precision of reference posts, points, markers, and nodes in the field *(11)*.

Previous research *(24)* has identified several parameters (i.e., positional accuracy, thematic accuracy, temporal accuracy, logical consistency, completeness, data status, and lineage) as encompassing the quality aspects of geographic information. Considered together, these characteristics indicate the overall quality of a geographic database.

Accuracy. When referring to geographic data, the term "accuracy" is usually described with two components: (1) positional accuracy and (2) attribute accuracy *(23)*. The positional accuracy of a spatial object, or a digital representation of a feature, can be defined through measures of the difference between the apparent location of the features as recorded in a database and its true location *(25)*. Positional accuracy refers to the amount of offset present within a data set from the true location of the features being represented, that is, how closely the coordinate descriptions of the features compare with their actual location. This type of accuracy is typically measured directly by comparison with data known to be more accurate or by inferring the amount of error introduced from processing the data; for example, a 1 : 24,000 scale road network may be tested against a set of GPS-based control points. If detailed positional accuracy analyses are beyond the reach of the project being performed, the data developer should at least document the processing steps and tolerances used, and the accuracy of any source materials compiled.

Attribute accuracy refers to how well the attribute portion of the database describes the geographic features being represented. That is, how thoroughly and correctly the features in the data set are described. Before assessing attribute accuracy, it is necessary to clearly define the interpretation rules used to represent information in the database. Rigorously determining attribute accuracy requires statistical analysis. At a minimum, data developers should document steps taken to ensure the integrity of attribute data.

Resolution. Resolution, or precision, refers to the amount of detail that can be discerned in space, time, or theme. It is directly linked with accuracy and is also used to determine how useful a given database is for a particular application. Two databases with the same accuracy levels but different levels of resolution do not have the same quality.

Data Status. Data status refers to the "currentness" of the data set. When developing data, it is important to maintain records of source material and observation dates used in the compilation. It is also important to maintain records on update cycles *(23)*.

Completeness. Data completeness refers to the degree to which the data describe the content of the source or phenomena being mapped. Completeness refers to a lack of errors of omission in spatial data. It includes consideration of holes in the data, unclassified areas, and any compilation procedures that may have caused data to be eliminated. Data completeness can be described by listing the features included in the data and whether the data are "completed" or "in progress." One might also consider what might have been omitted. For example, a particular attribute may have been collected for only part of an area, or perhaps paved roads but not gravel roads appear in a layer *(23)*.

Logical Consistency. Consistency refers to the adherence of the data to a given data structure, that is, the decisions that determine what the data set contains. Logical consistency refers to the absence of apparent contradictions in spatial data. Consistency is a measure of the internal validity of the data and is assessed using information that is contained within the data, which typically include spatial data inconsistencies such as incorrect line intersections, duplicate lines or boundaries, or gaps in lines. These are referred to as spatial or topological errors. Consistency measures the extent to which geometric problems and drafting inconsistencies exist within the data set. For example, are attribute tables formatted identically throughout the database? Are minimum feature size criteria consistently applied? Are the data topologically correct? Do features of the same type have the same descriptive data and level of detail? Are naming conventions consistent? *(23)*

Lineage. Lineage refers to a record of all data sources used to construct the spatial data set and all operations that have been taken to process the data. Thorough documentation for all spatial data is essential for determining quality. Information about appropriate ranges of use and scales at which the information is valid should be included with the original spatial data and any derived data sets. Lineage is concerned with historical and compilation aspects of the data, such as source of the data, content of the data, data capture specifications, geographic coverage of the data, compilation method of the data, transformation methods applied to the data, and use of any pertinent algorithms during compilation.

Knowing and documenting the original source of the data and its quality and establishing an audit trail of all transformations and changes that have been applied is essential for evaluating the overall quality of any resulting data set. The same data set that is reasonable for some applications is often not suitable for other applications where high quality is important.

Timeliness. For certain types of spatial data that are constantly changing, such as roads, the quality of the data depends directly on the timeliness of the data. The primary data quality issues are related to authenticating and validating the data and maintaining a detailed historical audit trail of updates for users of the data, so that quality can be verified and publications based on the data can be properly attributed.

2.2.2 Positional Accuracy

Accuracy is often defined generally as a measurement of exactness or correctness. In terms of spatial information, positional accuracy refers to how closely the data represent the real world. Because spatial data usually generalize the real world, it is often difficult to identify a true value. Because the true value of the data is not actually known but can be estimated only, the actual accuracy of the measured quantity is also unknown, and the accuracy of spatial data can only be estimated only. *(26)*

For points, accuracy is defined in terms of the distance between the encoded locations and "actual" location. For lines and areas, the situation is more complex because error is a mixture of positional error (error in the location of points along the line) and generalization error (error in the points selected to represent the line) *(2)*. Positional accuracy has two components: absolute and relative accuracy *(1, 27)*. Absolute accuracy and relative accuracy are considered separately because although spatial data may define a very accurate shape, the shape may not be located correctly *(27)*.

Absolute accuracy involves the accuracy of data elements with respect to a coordinate scheme. Absolute accuracy refers to how close a location on a map or data representation is to its real location on the earth. For example, a claim of absolute accuracy might be that 95 percent of the actual locations of wells in a given area are within 50 meters of their surveyed locations.

Relative accuracy concerns the positioning of spatial features relative to one another. Relative accuracy considers how similar a shape on a map or data representation is to the shape of the object on the earth. For example, cutblock boundaries do not vary by more than 10 meters from their actual shape.

The spatial position of an arbitrary object defined within a GIS data layer has a positional error that can be described by one of the primary parameters of positional accuracy. Table 2-1 shows examples of measures and metrics associated with positional accuracy *(23)*. Accuracy addresses concerns for data quality, error, uncertainty, scale, resolution, and precision in spatial data and affects the ways in which it can be used and interpreted *(28)*. Accuracy is always a relative measure, because it is always measured relative to some specification *(2)*. Two sources of error can reduce positional accuracy *(1):* inherent error, which is the error present in source documents and data, and operational error, which is all introduced error.

2.2.3 Uncertainty

It has been argued that in the context of geographic data, there is a clear distinction between error and uncertainty *(29)*. "Error" implies that some degree of knowledge has been attained about the difference between the results or observations and the truth to which they pertain. "Uncertainty," on the other hand, conveys that it is the lack of knowledge that is responsible for hesitancy in accepting without caution, and often the term "error" is used when it would be more appropriate to use "uncertainty."

The term "uncertainty" has gained recent popularity but suffers from inconsistent and ambiguous usage. Mowrer *(30)*

TABLE 2-1 Measures of positional accuracy *(23)*

Measures	Metrics	Remarks
Absolute accuracy (against reference frame)	• RMSE (root mean square error) • Error ellipse (2D)	• Describe the measurement method and error model • Can be represented by a vector – 2 axles and rotation angle
Relative accuracy (positional relative to adjacent features)	• Error of distance • Relative error ellipse (2D)	• Describe measurement method • Give the random and systematic error components • Describe the method (e.g., adjustment)
Accuracy of pixel position	• RMSE of pixel position	• Describe source of error representation via a real variable or a vector with random and systematic components
Height accuracy	• RMSE of height	• Accuracy of height of a single point

provides a recent compilation of most frequent interpretations. Geographic Information Science *(31, 32, 33)*, a relatively new field, has emerged as a combination of several different scientific fields (e.g., computer science, geography, surveying, and photogrammetry). Each of these scientific fields has a different view of uncertainty. Some claim, for example, that there is a difference between a situation of risk and one of uncertainty. The distinction is that in a risky situation, a random event comes from a known probability distribution, whereas in an uncertain situation the probability distribution is not known.

With any GIS product there is a level of uncertainty about the nature of its quality. It is important to provide the GIS user with the necessary awareness that these problems exist. Although there is a continuing interest in improving data quality standards *(24, 35)*, commercial GIS packages put little or no effort into calculating and communicating the inherent imperfections to the user *(36)*. Several researchers (e.g., *37, 38, 39, 40, 41, 42*) have explored different approaches to handling either a single imperfection (e.g., inaccuracy) or a conglomerate of imperfections (e.g., imprecision and inconsistency).

To improve the management of quality within GIS, it is essential to detect occurrences of imperfections and to clarify some frequently used terms. Steps in this direction have been made over the last several years. The development of a Spatial Data Transfer Standard and other national and international research efforts have been directed at understanding spatial data uncertainty (e.g., *37, 43, 44, 45*). Various approaches to the management of uncertainty have been proposed. For example, the possibility for assessing the fitness for use of spatial information as one form of uncertainty measure has been explored *(46)*. A different approach was offered that emphasized the design of a GIS to avoid misuse of spatial information *(47)*. The development of an intelligent GIS also has been proposed as a possible approach to managing uncertainty *(48)*. Another approach focused on data quality issues with regard to user interface design *(49)*, while another *(50)* discussed the relationship between the advantages of high resolution and the disadvantages of the accompanying high costs in GIS.

2.2.3.1 Measures of Uncertainty

Many methods of measuring geospatial uncertainty, such as positional root mean square error (RMSE), have been adopted more or less directly from traditional statistics. Several problems arise in extending them to complex geospatial objects. For example, how does one measure the positional accuracy of a complex geographic curve like a shoreline in ways that are independent of artifacts like point sampling? Several methods have been proposed recently *(25)*, but these have not been tested or assessed for large, realistic data sets.

To be useful, measures of uncertainty for spatial databases should satisfy certain definable criteria that mirror those underlying such traditional measures as RMSE. Goodchild et al. *(25)* identify the following criteria:

- Insensitivity to implementation details of the digital representation of the feature,
- Insensitivity to outliers,
- Unbiasedness, and
- Minimum variance.

The last two properties can be defined only in the presence of a stochastic model of uncertainty that allows comparison across multiple realizations. For example, if such a model were available for a digitized representation of a line, it would support simulation of a population of realizations of the line, each of which would be equally likely to be observed in reality. The measure of uncertainty would be a parameter of the model and it would be possible to analyze the performance of various procedures for estimating its value. The RMSE satisfies this requirement for a Gaussian model of uncertainty in point position, but there is no comparable theoretical analysis of measures of uncertainty in complex spatial objects.

Although much is known about measuring uncertainty in individual measurements, the problems of uncertainty in geospatial data are exacerbated by the lack of simple lineage between independent measurements and final product. It is currently impossible, for example, to identify the independent measurements responsible for uncertainty in a single elevation value drawn from a Digital Elevation Model (DEM).

Hunter et al. *(51)* analyzed this situation and showed the importance of knowing the spatial dependence among individual uncertainties in many geospatial applications.

Any measures of uncertainty must consider that uncertainty varies through space and time and is context sensitive. This has important implications for modeling. Certain existing measures have limited utility because they do not describe the spatial distribution of uncertainty (e.g., a single RMSE does not capture variation over space). Thus, an important and desirable characteristic for measures of uncertainty is that they do indeed describe the variation in uncertainty over space. There is also the need for measures to be dynamic (e.g., if they are computed and stored in the database, any updates to the database may require updating stored measures). It is clear that there are many elements of uncertainty, each of which is ideally measured individually (e.g., positional uncertainty, topological uncertainty, attribute uncertainty, and temporal uncertainty). Given different user contexts, there may be a need for a combined measure of uncertainty that aggregates the measures of individual elements. For example, users in a rapid decision-making environment may not have the time or interest to request a view of each of the individual measures of uncertainty. They may prefer a combined measure that can inform them of the overall uncertainty of a specific piece of information. Important measurement issues thus relate to aggregating individual measures of uncertainty.

It is presumptuous to give general solutions for the primary parameters of data quality that fulfill each useable occurrence in every GIS application. Therefore, measures and metrics can be defined by the data provider—appropriate for the individual kind of data set and the demands of the user. One might state that the demands of the user cannot be anticipated a priori except in an idealized case. Nevertheless, only the assumption of the ideal situation, or a situation where the provider at least has an idea of the intended GIS application, allows the provider to choose the parameters to give useful information on accuracy. If all possible accuracy values have to be evaluated, the costs of information on accuracy would be too high.

2.2.4 Spatial Data Errors

All spatial data is inherently inaccurate, as it is only a conceptualization of the reality it tries to represent. The degree of uncertainty associated with spatial data is affected by various factors, which range from measurement error, to inherent variability, to instability, to conceptual ambiguity, to over-abstraction, or to simple ignorance of important model parameters *(3)*. Errors also can be introduced by collection methods, data translation, digitizing methods, source material, generalization, symbol interpretation, specifications of aerial photography, aerotriangulation technique, ground control reliability, photogrammetric characteristics, scribing precision, resolution, and processing algorithms *(5)*. The error associated with any one of the potential sources is often small, but together these errors can significantly affect the accuracy of the spatial data, thereby affecting the potential uses of the data *(5)*.

The method of data collection sets limitations on the selection of the measures of uncertainty and their metrics. Both common surveying methods using tacheometer (or GPS) and aerial photography produce positional error. For the tacheometer, sources of error include orientation error, the scale derived from distance measurement error, and errors from adjustment. Attachments to surveyed points could also introduce instrument error, operator error, and other types of error to the point coordinates. In photogrammetry, one has to consider the resolution, distortion, scale (flight height), transformations (picture coordinates), and bundle block adjustment.

Following the initial acquisition of data, a series of cartographic techniques are used to translate this acquired information into mapped information. Errors and inaccuracies introduced at the digitizing stage are largely unpredictable and random in nature. Integrating data from different sources, in different original formats (e.g., points, lines, and areas), at different original scales, and with inherent errors can yield a product of questionable accuracy *(1)*. Common practices in map compilation (e.g., generalization, aggregation, line smoothing, and separation of features) can introduce further inaccuracies *(28)*. Processing of data produces errors such as misuse of logic, generalization, problems of interpretation, mathematical errors, accuracy lost from low precision computations, and rasterization of vector data *(28)*.

The method of processing data also determines the resulting error type and its metric. The use of spatial data in GIS can further reduce the quality of the data. Because most of the spatial data used by GIS is required in predefined formats, the spatial data must be modified to fit the standard. This modification or compression of the data into the acceptable format often reduces the accuracy of the information. Furthermore, once the spatial data are in the acceptable format for use by a GIS, every action that uses the data can generate additional errors and compound existing ones. The errors and inaccuracies associated with spatial data are cumulative and build up through the various processes of data manipulation and analysis *(3)*. Error also can spread to other spatial data that incorporate the data in the GIS *(6)*.

Errors introduced through measurement and processing can be either systematic or random. Examples of more GIS-specific errors are errors of orientation. These include errors from the transformations used while digitizing a paper map or as a result of the orientation of a GIS raster. During the process of conversion of data into a raster map, the level of granularity changes, which is an additional error source *(52, 53)*.

An example of error propagation modeling deals with methods for visualization of the accuracy of geometrical data. Areas are represented by their boundaries. The vertices of these polygons are treated as stochastic information. The mathematical principle is based on the probability of the location of an arbitrary point within a closed polygon. This model can be used to determine the accuracy of an area segment by overlaying two areas with a map overlay operation. The latter quality model combines variances as well as correlation and systematic errors based on proven theoretical methods.

A data quality model (DQM) is one way of integrating and presenting uncertainty information to a GIS user. The DQM is a subschema in the concept of metadata *(35)*. It provides essential additional information to assess the decisions made with the help of a GIS. A model of the real world requires transformations of the data to reduce the information to the essential quantity. During this process, discrete data obtained from continuous reality introduces errors.

2.2.4.1 Scale, Resolution, and Discretization

As noted in Sinton *(54)*, Chrisman *(37)*, and others, when making measurements, resolution is imposed across the three dimensions of space, theme, and time in the form of discretization. Control is a discretization along one or more dimensions so that another dimension can be measured. The imposition of discretization results in a loss of information that contributes to the uncertainty about the variable or phenomena being described. In terms of uncertainty, the effects of discretization are likely to be more substantial than measurement error. Work on uncertainty has tended to focus on measurement errors, and yet the effects of discretization may be more substantial. In other words, the imperfections in the measurements are less cause for concern than that which is not measured.

Several researchers discuss the effects of resolution or scale *(55)* in a broad variety of approaches. Watzek et al. *(56)*, for example, focused on an empirical approach to determine the perceived scale accuracy of computer visual simulations. Bruegger *(57)* proposed spatial theory models for integrating datasets of different levels of resolution in GISs. Cushnie *(58)* discussed the interactive effect of spatial resolution and the degree of internal variability within land-cover types on classification accuracies. Canters et al. *(59)* and Moody et al. *(60)* take a different approach that focuses on the errors introduced in land-cover proportions due to varying scale. Burrough *(61)* and Oliver et al. *(62)* investigated comparable methodologies, concentrating on the influence of variations in a continuous field. An application-specific approach (e.g., road density estimates) of scale-dependent accuracies can be found in Wade et al. *(63)*. On a global scale, Townsend et al. *(64)* elaborate on the effects of resolution in conjunction with a specific application—global monitoring of land transformations. Similar effects such as aggregation and support are discussed in Heuvelink *(65)*. Prisley et al. *(66)* investigated the effects that the underlying variation in the attribute variable had on the GIS-based decisions.

The influence of discretization on the quality of spatial representations has not been addressed in any systematic way. Researchers van Groenigan *(67)* and Burrough et al. *(33)* address a similar problem in a slightly different way. They are interested in optimizing the layout of a sample field. However, in their approach, the underlying variation of the attribute does not play a central role. Their approach is based on an a priori optimization, whereas this study is interested in estimating the loss of information a posteriori. Their approach is directed toward data producers, whereas this study concentrates on providing the user with helpful information on the inherent uncertainty. In general, the overall reliability of a spatial representation is less influenced by the accuracy or precision of a measurement than by the number, density, or spacing interval of the measurements. Accuracy measures are most often associated with well-defined points, which have little to say about unmeasured locations. Discretization is an implicit measure of what is not known or what might be missing as a result of the discretization.

2.2.5 Spatial Data Standards

Quality assurance is a basic requirement for reliably performing an application, and all applications should be accompanied by a detailed evaluation of the fitness-of-use of the data used (to examine whether the data represent the information needed to answer the question raised by the application). A statement of accuracy generally includes a statistical determination of uncertainty and variation, as well as how and when the information was collected *(27)*. Often a statement of accuracy is accompanied by the confidence level of the spatial data, which is defined as the probability that the true value of the data falls within a range of given values *(26)*.

Standards provide for consistency among data, users, and systems. Most accuracy standards for spatial data require a standard for the horizontal component of accuracy and another standard for the vertical component of accuracy, as well as a description of the method used to evaluate the accuracy *(26)*. The reporting standard in the *horizontal* component is the radius of a circle of uncertainty, such that the true or theoretical location of the point falls within that circle 95 percent of the time. The reporting standard in the *vertical* component is a linear uncertainty value, such that the true or theoretical location of the point falls within plus or minus of that linear uncertainty value 95 percent of the time.

The method used to evaluate accuracy (e.g., statistical testing, least squares adjustment results, comparison with values of higher accuracy, repeat measurements, or estimation) should be described.

Comprehensive statements of spatial data quality should accompany the use or transfer of all spatial data, because it is not feasible to remove error entirely from spatial data sets, although a reduction of error is possible. The introduction and adoption of spatial data standards addresses the issue of spatial data quality, but heavy reliance on the fitness for use of the data means that most of the responsibility remains in the hands of spatial data users. An awareness of the accuracy of spatial data allows users to make a subjective statement on the quality and reliability of the information *(1)*. Spatial data error cannot be predicted, neither can it be entirely prevented; at best, it can only be coped with *(3)*.

2.3 APPLICATIONS OF POSITIONAL DATA

2.3.1 Introduction

Spatial data has little or no value to transportation applications without any attribute data attached to it. Each spatial data element (a line, a point, or a polygon) has a cartographic representation as well as a unique identifier to associate attribute information with that data element. In contrast, data collected by transportation agencies for their facilities may not have any cartographic representation (i.e., geo-referenced). Data are collected in a network model, a theoretical framework that is applied to and depends on the functionality of different LRMs. Given that the network does not require any cartographic representation (i.e., spatial data element), and attribute data are collected independently from the cartographic representation of the transportation element (i.e., highway segment), it is important to address the issue of sensitivity of applications in transforming various LRM data to the linear datum (cartographic) representation. Different applications require spatial data at different scales. Vonderohe et al. *(68)* suggested the use of four spatial database scales for DOT activities. As noted in Table 2-2, the transportation applications of GIS can be divided into three primary functional groups: planning, management, and engineering. Planning applications are usually at statewide and regional levels and do not require highly precise locational data. Spatial databases for these applications are at 1 : 500,000 to 1 : 000,000 scales. Management applications often require more detailed locational data that are available at regional or district levels. The spatial databases are usually in the 1 : 100,000 to 1 : 24,000 range. Engineering applications require a high level of spatial accuracy and these applications are restricted to project or corridor level. The preferred scales for engineering applications are 1 : 12,000 to 1 : 24,000. This grouping suggests that engineering applications are more sensitive to positional data quality than management applications.

A different way of grouping the current and emerging applications of GIS is by transportation subject area. This concept of grouping recognizes that applications within a subject area may include planning, management, and engineering functions. Moreover, grouping by the three functional classes may conflict with the sensitivity of the individual applications to spatial data quality. For example, while crash reporting may not be classified as an engineering activity, identifying crash-prone locations is sensitive to the data quality. Similarly, highway infrastructure management may be classified erroneously as a management function, when it actually involves engineering applications. Table 2-3 shows the current and emerging uses of spatial data in transportation as well as the levels of sensitivity of transportation applications to spatial data quality. These levels are based on state DOT perceptions of the sensitivity of the various applications to positional data quality.

2.3.2 Examples of Applications

Pittman et al. *(69)* provided an overview of the various transportation applications of spatial data and observed that GIS-T are being effectively used to do the following:

- Provide support for making quality decisions on maintaining the transportation infrastructure,
- Design efficient routes for maintenance operations and serving the riders of transit systems,
- Manage traffic and incidents, and
- Develop multi-year improvement plans that take into account existing roadway characteristics and conditions and crash record information.

O’Neill et al. *(70)* identified emerging applications of positional data to include field crew scheduling, customer complaint and response, decision support system, facility management, and policy analysis.

The following are examples of specific projects that demonstrate the applications of GIS in the various subject areas identified in Table 2-3. These examples are provided to illustrate the range of current and emerging applications by state DOTs. These examples do not exhaust the full range of possible current and future applications. Some of the applications overlap

TABLE 2-2 Scales and typical applications *(68)*

Scale of Spatial Database	Precision of Spatial Database (ft)	Typical Activities or Applications
1:500,000	830	Statewide planning
1:100,000	170	District-level planning and facilities management
1:12,000 – 1:24,000	30 – 40	Engineering
1:120 – 1:1,200	0.33 – 3	Project-level activities

TABLE 2-3 Applications of positional data in transportation

Subject Area	Applications	Sensitivity		
		L	M	H
Safety	- Crash reporting		•	
	- Black spot/ crash prone location identification		•	
	- Traffic safety investigation		•	
	- Rail crossing safety analysis		•	
	- Pedestrian and bicycle safety analysis		•	
	- Incident management			•
	- 911 emergency planning and response			•
Transportation Planning, Impact Analysis, Policy Analysis	- Travel demand modeling	•		
	- Multi-modal freight modeling	•		
	- Hazardous materials routing		•	
	- Traffic impact analysis		•	
Transit and Public Transport Planning and Operations	- Transit planning	•		
	- Transit routing		•	
	- Handi-transit		•	
	- Real-time tracking and scheduling of buses		•	
Transportation Infrastructure Management and Operations	- Location of facilities (road, highway, airport, port) inventory	•		
	- Pavement management system		•	
	- Asset management		•	
	- Operation (congestion, service)		•	
	- Corridor analysis (rail, road, highway)	•		
	- Rail/highway information system management	•		
Transportation Design and Construction Planning	- Sources of construction materials	•		
	- Right of way			•
	- Road closure and detour		•	
	- Construction information	•		
	- Field crew scheduling	•		
	- Maintenance and operation		•	
	- snow plowing	•		
	- garbage collection	•		
	- street sweeping	•		
ITS Applications	- Traveler Information System		•	
	- Integrated Highway Information System (IHIS)		•	
	- Integrated Traffic Monitoring System (ITMS)		•	
	- Web-based road condition reporting system		•	
	- Vehicle Navigation System			•
	- Applications to commercial vehicle operations regulatory enforcement activities	•		
Freight Analysis and Commercial Vehicle Operations	- Fleet management	•		
	- Vehicle tracking, guidance, dispatching, and other		•	
	routing applications		•	
	- Permitting	•		
	- Freight movement	•		

two or more application or subject areas or can be classified in more than one category.

2.3.2.1 *Safety*

- North Carolina DOT used a GIS-based referencing system to identify locations with a high probability of truck crashes on truck corridors. The framework allowed visualization of geographic patterns of land use activities associated with frequent crash locations *(71)*.
- Iowa DOT developed a GIS-based crash location and analysis system designed to manage crash data retrieval and analysis. The system also allows analysis of implications of crash location characteristics for emergency response services *(72)*.

2.3.2.2 *Transportation Planning, Impact Analysis, and Policy Analysis*

- A GIS software application with transportation demand modeling capability is used for freight modeling to help identify highway capacity problems of the national freight transportation system. This study was conducted for the FHWA's Office of Freight Management and Operations. The primary objective of the highway freight capacity analysis is to develop a policy tool for analyzing potential freight-related policies and examining the sufficiency of capacity of the transportation system in meeting forecast freight demand *(73)*.
- Florida DOT's office of system planning uses GIS-T in an ad hoc production of maps used to manage and develop the Florida interstate highway systems *(69)*.

2.3.2.3 Transit and Public Transport Planning and Operations

- A prototype decision support system was developed in GIS for the Cape Cod Regional Transit Authority. The tool was designed to support operational decisions, which integrate paratransit ridership with regional and community-based fixed-route transit, and planning decisions regarding intermodal transit connections throughout the Cape Cod regions *(74)*.
- The Delaware DOT examined the use of GIS to better understand travel demand and to identify opportunities for transit in New Castle County. GIS was a valuable tool in demonstrating the relationship between transit markets and existing transit service, providing a method to describe travel demand at a very detailed level, and suggesting the best location for park-and-ride facilities and transit centers *(75)*.
- Research was carried out by the Orange County Transportation Authority that proved GIS to be a useful tool to project transit passengers' mobility patterns with greater accuracy, consequently strengthening the validation database for travel demand forecasting analysis with respect to transit planning *(76)*.

2.3.2.4 Transportation Infrastructure Management and Operations

- For a highway infrastructure management application, a road centerline base map and inventory of transportation infrastructure in Seneca County, Ohio, allowed the county to improve the maintenance of traffic signs, bridges, guardrails, and culverts *(77)*.
- In a pavement management application, dynamic segmentation was used to project transit passengers' mobility patterns with greater accuracy, consequently strengthening the validation database for travel demand forecasting analysis with respect to transit planning. The necessary data were collected on I-85 in South Carolina, but the study was sponsored by NCDOT *(78)*.
- The New Jersey Turnpike Authority implemented a system integrating Automatic Traffic Surveillance and Control System (ATSCS) technology with GIS-T to improve its transportation operation activities. This example also can be classified under ITS applications.

2.3.2.5 Transportation Design and Construction Planning

- The Maryland State Highway Administration sponsored the use of a GIS model to optimize the selection of geometric designs for highways. GIS was integrated with a Highway Design Optimization Model (HDOM) to compute geographically sensitive costs to be used with an iterative optimization scheme. It was shown that the GIS model provides accurate geographical features, computes location-dependent costs, and transmits these costs to an external program. An example study was carried out for Talbot County, Maryland *(79)*.

2.3.2.6 ITS Applications

- The NJDOT and NJ Transit sponsored a study that investigated the use of an Automatic Vehicle Location (AVL) system to monitor the locations of buses. Information from the AVL is displayed in a GIS that contains data on bus routes, bus stops, intersections, and landmarks. The system required that the positions of the features be accurately determined. The system was tested in a densely built urban area with high-rise buildings, tunnels, and overpasses. Accuracy of the results was within the 30-ft tolerance limit *(80)*.
- Through a public-private partnership between Mobility Technologies (formerly Traffic.com), Pennsylvania DOT, and U.S. DOT, an Integrated Surveillance and Data Management Infrastructure (ISDMI) program was implemented in Pittsburgh and Philadelphia. Real-time traffic data are integrated with a GIS-based freeway management system that stakeholders can readily access. This system is expected to enhance traffic and incident management. The system also provides traveler information to road users *(81)*.

2.3.2.7 Freight Analysis and Commercial Vehicle Operations

- Freight flow characteristics were integrated with GIS for the identification and analysis of the location of transportation facilities and freight generators, freight movement patterns, variation in truck traffic mix by configuration and body type, and truck travel time. The purpose is to examine the specific details of policy options and how these options may affect the operation, modal competition, equipment selection, and response of primary decision-making groups. The study develops a set of metrics that will allow examination of implications of possible federal truck size and weight policy *(82)*.

2.3.3 Emerging and Future GIS-T Applications by State DOTs

Table 2-4 summarizes information from state DOTs as well as the Oak Ridge National Laboratory (ORNL) on current, emerging, and future GIS-T applications. Both current and anticipated future applications vary from state to state. However, several current applications, such as a highway

TABLE 2-4 Summary of state DOT applications

State/ Organization	Current Applications	Future/Potential Applications
Arizona	• Planning • Safety analysis • Incident detours • Highway closures and restrictions	• Asset management • Feature inventory • Detour routing
Iowa	• Road inventory • Pavement management • Crash analysis and reporting • Highway inventory • Travel demand modeling	• Automated overweight/oversized truck routing • Safety inventory • Sign inventory • Automated traveler information system (ATIS) • ITS (emergency and construction routing) • Automatic vehicle location (AVL)
Ohio	• Planning applications: - environmental impact studies (wetland studies) - historical and archeological studies - highway safety (location of crashes) - congestion management - level of service - statewide travel demand modeling • Pavement management • Traffic studies (impact, design) • ITS applications	• Bus routing • Intermodal • Freight analysis • Emergency evacuation • Traffic demand modeling
Pennsylvania	• Crash analysis • Right of way • Pavement management • Traffic studies	• Linear reference control • Environmental reviews including cultural resource • Environmental permitting • Wetland mitigation • Address matching
ORNL	• Planning • Routing • Infrastructure management	• Development of hierarchical networks to integrate functions e.g., inventory, navigation, strategic routing

inventory, pavement management, traffic studies, and crash analysis, are common to all states. Vehicle routing (e.g., bus, truck, or permit vehicles) and detour routing are common future applications identified by state DOTs.

2.3.4 Sensitivity of Applications to Positional Data Quality

Knowledge of the uncertainty associated with geographic information is critical to the effective use and credibility of GIS and GIS outputs. The key components of a research agenda for uncertainty have been identified as modeling, propagation, communication, fitness-for-use assessment, and uncertainty absorption *(46)*. The "truth in labeling" concept is aimed at providing users with information to help assess fitness for use of data. However, the lack of actual procedures for this assessment means that, in many cases, valuable data quality statements remain under-utilized. Agumya et al. *(83)* discussed risk management techniques in assessing fitness for use of geographic information by translating uncertainty in the information into risk in the decision.

The sensitivity of transportation applications to positional data accuracy can be assessed either by standards-based methods or by a risk-based approach. The traditional method to assess the acceptability or fitness of use—the standards-based method—compares data uncertainty with a set of standards that defines acceptable levels of uncertainty in the data *(36)*. This approach measures the sensitivity of the positional data for a particular application by directly comparing the quality elements of information against a set of standards or error benchmarks that represent the acceptability of the data components. Although uncertainty in spatial data is composed of several well-known elements *(84)*, the obvious measurable ones are map scale (resolution), currency, attribute accuracy, and percentage of completeness. However, measures of these elements are difficult to combine into a single, meaningful, composite unit *(85)* and require testing the sensitivity of the application to error associated with each element. A typical example would be U.S. census TIGER street centerline spatial data, which are used for urban transportation modeling applications. There is no means of separating the individual error effects of poor map scale (e.g., positional accuracy of the street segments), logical consistency (e.g., street network topology), attribute accuracy (e.g., travel time), or completeness (e.g., missing street segments) *(86)*.

A risk-based approach, in which the sensitivity of an application is measured against the adverse effect of the ultimate decision, is based on the results of the analysis. Agumya et al. *(86)* stated that the "risk-based approach is a technique based on risk management practices, in which a study is made of the effect that uncertainty in the data has upon the ultimate decision to be made with it. In turn the adverse consequences of making a poor decision are quantified, and it is this information which enables a user to determine whether a data set is fit for use or not." Risk analysis has already been suggested as a plausible basis for characterizing and estimating the consequences of uncertainty in spatial data *(38)*. In the earlier example, the risk-based approach would have determined the consequence and liability associated with this particular application, by using the TIGER street line spatial data and formulating a strategy for reducing this liability or consequence in the most cost-effective manner.

The sensitivity assessment of positional data under this approach would require addressing two fundamental questions *(83)*:

- What are the consequences associated with the decision, in terms of risk, in using a particular set of spatial data with error in different transportation applications?
- What are the acceptable consequences of uncertainty in terms of risk?

The first question entails the partition of spatial data error for a particular dataset into its various elements, the determination of the risk a transportation analyst may incur by making the decision based on the dataset, and the extent to which this dataset influences the decisions. If the positional accuracy of the dataset has the lesser effect on the decision, such as traffic or freight assignment using a TIGER street file, then it is reasonable to accept the risk and uncertainty associated with this particular application. However, for vehicle navigation purposes, the risk may still be too high to be acceptable.

The second question entails establishing a threshold for the risk that is considered acceptable. The acceptability of risk may vary widely among the data users and depend on the nature of the applications. The acceptability of project-level analysis or a decision is more conservative than the planning-level transportation application. For a given spatial dataset (e.g., TIGER street file), acceptability of the positional accuracy is much higher.

2.3.4.1 State Practices—Sensitivity of Applications to Positional Data Quality

This section discusses perceptions of the sensitivity of positional data quality on various transportation applications. Most respondents were unable to provide any meaningful responses to the question of how the quality of data affects or is taken into account in various transportation applications. Applications such as commercial vehicle operations or regulatory and policy analyses are less sensitive to the accuracy of positional data than highway design, construction planning, and infrastructure management applications.

Iowa. The Iowa DOT is creating an LRS, based on a datum as part of the LRS Development Project Pilot (scheduled to be completed in June 2001). A needs assessment was completed as part of that project. Part of that assessment identified user accuracy requirements. These accuracy requirements were quite diverse, even for events/features in the same database. The consensus (including cost considerations) was that the achievable accuracy was 10 meters along the roadway. Given that location is the basis for integrating the data, the accuracy along the centerline becomes one of the most important aspects. As technology improves and becomes more economical, Iowa will no doubt increase the accuracy of the datum locations. This will be necessary so that the business data mapped against the datum will not be degraded if the business data are more accurate than the datum.

ORNL. The accuracy has to be better than the size of the objects. Roadway segments are rarely less than 40 meters, so there is little benefit for accuracy better than 20 meters. Nevertheless, 100-meter accuracy is still useable if that is all that is available. ORNL's experience has been that other sources of ambiguity dominate locational error such as unequal spacing of mileposts.

2.3.4.2 State Practices—Effects of Data Quality on Decisions

The quality of positional data influences decisions relating to different applications. For planning and management applications that do not require high accuracy of positional data, a general idea of the quality of data may be sufficient to make decisions. However, for engineering applications where specificity is critical, the quality of data receives more emphasis in making decisions. In the absence of knowledge of the quality of positional data, states tend to rely on the standards to guide the assessment of the data quality. Further applications are designed around available accuracy or quality of data.

Arizona. Accuracy of positional data is adequate for planning, statistics, and inventory. ADOT noted that one adverse effect of using spatial data that do not meet the minimum quality standards, or data with uncertain accuracy, is difficulty in coordinating with other data.

Most decisions are not currently made on readily available spatial data. Initial analysis may be performed so that more exact field surveys can be obtained. At that point, engineering-accurate surveys provide the spatial information needed to make decisions. Even in the areas of pavement management, crash analysis, and ITS, the current spatial accuracy is used mostly for a general description of the location, not as an engineering decision-making tool. As the LRS is developed and uses the location to integrate the data, more dependence will be placed on the location/linear accuracy.

In general, the USGS ortho photos (1-meter pixels) will probably meet most accuracy requirements. ADOT is getting hard measurements to confirm that the expected accuracy in a “flat” state like Iowa will be substantially better than the nominal accuracy stated by the USGS. ADOT is also acquiring higher accuracy orthos from local governments, as they become available, with 6 inches, 1 foot, and 2 feet pixels. The 1½ feet pixel sizes will definitely meet all but the most stringent requirements. These sources vary in spatial resolution from 1 : 1,000 to 1 : 12,000.

Obviously, spatial data that fail to meet minimum accuracy standards can cause incorrect decisions to be made or require that analyses be verified using costly fieldwork. In some cases, limited accuracies will mean that the data are not useable (e.g., 15-meter panchromatic spatial images are too coarse for most transportation needs). In some instances, ADOT receives data from other state agencies with a resolution of 1 : 1,000,000 or less. Such data are only useful for very macro-level analysis.

Ohio. When data fail to meet minimum quality standards, it is evident during processing when the coordinates do not fit. A decision has to be made whether to use existing data or new data. That decision depends on the project. For example, in culvert replacement, vertical alignment accuracy is critical, while horizontal alignment is not so critical. For bridges, the position of piers and elevation require higher levels of accuracy, while in boundary work accuracy is not very important, so they use the state minimum as a guide.

The standard used depends on the type of survey. National Geodetic Survey (NGS) specifications are used for certain types of surveys and second-order NGS specifications are used for the control of engineering designs, for example, center line points (1 : 50,000). The state minimum is 1 : 5000. However, the NGS specification is not always followed.

ORNL. Applications are designed around available accuracy; the need for more accuracy seldom arises. Applications are more dependent on attribute accuracy and currency, where the scale of the objects is substantially under geographic accuracy. All applications have error rates, and more accuracy will reduce these. The biggest problems have been in facility locations on networks such as bridges and railroad grade crossings. An improvement from 100 meters to 20 meters of maximum error would reduce location-caused error rates from 10 percent to near zero.

2.4 MEASUREMENT SYSTEMS

2.4.1 Introduction

Several techniques are available to measure the positions of objects or events to be mapped to the highway network. Examples include milepost-referencing, distance measuring instruments (DMIs), surveying, aerial and satellite imagery, and GPS. Techniques such as milepost-referencing and DMI techniques measure the positions of objects along linear paths directly. In many applications, however, use of these techniques is either not possible or practical. Examples include real-time emergency vehicle routing, automatic vehicle location (AVL), and monitoring of construction equipment. For these applications, techniques such as aerial/satellite imagery and GPS techniques are more feasible. Difficulties arise with the use of these techniques, however, because the 2D (or 3D) positional data must be mapped to a 1D linear reference. In most cases, this data mapping is done with the help of a GIS *(21)*.

This section summarizes the different methods used in transportation for measuring positions and for locating vehicles. The objective is to describe each system and indicate the levels of accuracy it can achieve. Section 2.4.2 covers measuring methods that deal with locating roadway features for creating maps or geographic databases. Section 2.4.3 discusses positioning methods commonly used to determine the current location of a vehicle in real time. In this case, the actual measurement is immediately used for navigation or vehicle tracking. In each case, the measuring device and measuring method are described. In addition to the descriptions in the following sections, details of the measuring and positioning methods are summarized in Table 2-5.

2.4.2 Measuring Methods

2.4.2.1 Aerial

Photogrammetry. The fundamental principle used in aerial photogrammetry is triangulation. Aerial photographs are taken with an airplane or a helicopter. By taking photographs from at least two different locations, so-called “lines of sight” can be developed from each camera to point on the object. These lines of sight (sometimes called rays because of their optical nature) are mathematically intersected to produce the 3D coordinates of the points of interest. At a minimum, one needs two different photographs to reconstruct the 3D world. To triangulate a set of points, one must also know the camera position and aiming angles (together called the orientation) for all the pictures in the set. The orientation can be computed using ground control points or by installing survey-grade GPS in the aircraft. Aerial triangulation ties blocks of aerial photos together and simultaneously computes the orientation parameters of all photographs.

TABLE 2-5 Characteristics of measuring methods

Methods	Description	Measurements	Applications	Accuracy
Measuring				
Aerial				
Photogrammetry	Stereo-Plotting – Digital and Analog	x,y,z	Engineering, design, GIS basemapping	3-5 inches
Orthophotos	On-screen digitizing	x,y	Design, GIS – direct basemapping	1.5 ft w/ 0.5 ft resolution 6 ft w/3 ft resolution
LIDAR	Automatic height measurement using laser	z(x,y)	Engineering, design, digital elevation models	4 inches
Ground: Vehicle Based				
Mobile Mapping	Global Positioning System (GPS) / Inertial Navigation System (INS)/ Digital Stereo Measuring	x,y,z	GIS- asset inventory, mapping, engineering	< 1 meter
Video-Logging	DMI w/ GPS/ Single Video Camera	x,y of vehicle distance (D), offset (ΔO)	Inventorying, pavement condition analysis	3-10 meters
Distance Measuring Instruments (DMIs)	DMI w/ data logger	D	Asset inventorying	> 1 meter (% of distance
Ground: Surveying				
Wheel	Operator walks w/ wheel (like DMI) and measures distances relative to stations defined in a map	Relative Distance (ΔD)	Crash investigation, local surveys for maintenance and planning	2 feet +2% of (Δd)
Kinematic GPS	Dual frequency carrier phase with base stations	x,y,z	Engineering design, property surveys	1-5 inches
Differential GPS	Pseudo ranges w/ real time differential conditions	x,y (z)	Asset inventorying	5-10 feet
Laser Ranging	Laser gun with compass and inclinometer to determine location of objects	Δd, angle (α), (x,y,z)	Asset inventorying	1 inch + % of (Δd)
Total Stations (theodolite)	Land surveying weith theodolite/electronic distance measuring system (EDMs)	x,y,z	Engineering design, property surveys	1-5 inches
Map Digitizing	Paper maps are placed on digitizing tablet	x,y	GIS-basemapping, legacy data conversion	5-50 feet
Positioning				
Qualitative/Approximate Locating				
Distance from landmark	Estimate distance from landmark	Δd	Crash reports and investigation, emergency response (EMS-911), roadway maintenance crews	100-300 feet
Distance from intersection	Estimate distance from intersection	Δd		
Distance from milepost marker	Estimate distance from milepost marker	Δd		
Address	Address number	Address number (#n)		
Automatic Vehicle Location				
GPS for car navigation	GPS, compass and odometer data are merged with street maps to keep track of vehicle and show its current location	x,y, α	Vehicle tracking, routing, car navigation, emergency dispatch	10-50 feet
Compass		α		>10 degrees
Odometer		Δd		> 20 ft (+ 5% of distance)

(*continued on next page*)

Once the orientation is available, analytical stereo plotters or digital photogrammetric workstations are used to extract spatial data. The stereo plotter operator views a 3D model in his or her workstation and, using a 3D cursor, traces the lines to be added to the map (e.g., road centerlines, intersections, or contours).

Aerial photogrammetry is used to take measurements in x,y,z coordinates. High-resolution aerial photos are used for highway engineering and roadway design, while lower-resolution photos can be used for GIS base mapping (to extract road centerlines). The accuracy of measurements with aerial photogrammetry depends on the image scale at which aerial photos were collected and on the pixel resolution in the case of digital images. Accuracy of 3 to 5 inches and an image collection scale of 1 inch = 100 feet can be achieved with regular aerial photographs and high-quality stereo plotters. The productivity of this measuring method is limited by the capabilities of the stereo plotter operator. The data extraction process is mostly manual, although there is promising research for automatically extracting road centerlines and road edges. Data with high-resolution photos are used for project-level applications, while network-level applications require data with lower-resolution photos.

Ground Control Points (GCPs) are typically established using GPS to serve as location reference points; however, aerial GPS is becoming more popular and reduces the number of GCPs required.

Orthophotography. A digital orthophoto is a rasterized (scanned) aerial photograph, which is fully rectified to remove all of the distortions that occur in the original image: the pitch and roll of the aircraft, the radial distortion from the camera lens, and the image displacement from the topography. The removal of these distortions results in the imagery becoming a true scale representation of the ground. The orthophotos can be used for 2D digitizing on a computer screen. With digital orthophotos, all of the information on the original photograph is on the rectified image and is located in its true position.

The standard DOQs (digital ortho quad) produced by the U.S. Geological Survey (USGS) are either gray-scale or color-infrared (CIR) images with a 1-meter ground resolution; they cover an area measuring 3.75 minutes longitude by 3.75 minutes latitude, or approximately 5 miles on each side. Each DOQ has between 50 and 300 meters of overedge image beyond the latitude and longitude corner crosses embedded in

TABLE 2-5 ***(Continued)***

Methods	Reference Point Locations	Data Collection Vehicle	Productivity	Level of Data
Measuring				
Aerial				
Photogrammetry	Ground Control Points (GCP), aerial GPS	Airplane/helicopter	Manual post processing	Project
Orthophotos	GCP and aerial GPS	Airplane	Direct use as basemap	Network
LIDAR	Aerial GPS and inertial system	Helicopter/airplane	Real-Time Heights, post processing for DEM	Project
Ground: Vehicle Based				
Mobile Mapping	GPS base stations - HARN	Van	Manual and semi-automatic post processing	Network, Project
Video-Logging	COARSE – Coast Guard GPS reference stations	Van	Visual image inspection	Network
Distance Measuring Instruments (DMIs)	Intersections (anchor points, nodes)	Van/Car	Real-time data logging in vehicle	Network
Ground: Surveying				
Wheel	Stationing along roads	Person walking	Measurements recorded on printed map or notepad	Project
Kinematic GPS	HARN, first-order GPS reference stations	Person walking (tripod)	Data collector connected to receiver	Project
Differential GPS	COARSE- Coast Guard GPS reference stations	Person walking	Data collector connected to receiver	Project/Network
Laser Ranging	Local Reference Points, GCPs or GPS	Person walking (bipod)	Data collector connected to laser	Project/Network
Total Stations (theodolite)	HARN, first-order GPS reference stations	Instrument static on tripod	Data collector connected to total station	Project
Map Digitizing	GCPs	Person working in the office	Recorded on computer, cleanup in CAD system	Network
Positioning				
Qualitative/Approximate Locating				
Distance from landmark	Landmark	Persons walking or driving in a car records data on paper	Recorded in the field in real-time	Network
Distance from intersection	Intersection			
Distance from milepost marker	Milepost marker			
Address	Street segment, block			
Automatic Vehicle Location				
GPS for car navigation	GPS satellites	Sensor(s) installed in car or truck	Data are recorded and merged with map in real-time to continuously show the location of the vehicle	Network
Compass	Magnetic north			
Odometer	Start of travel			

the image. All DOQs are referenced to the North American Datum of 1983 (NAD 83) and cast on the Universal Transverse Mercator (UTM) projection. The file size of a gray-scale DOQ is 40 to 45 megabytes, and a CIR DOQ can be three times this size.

Digital orthophotos are a standard product commonly used as a base map for GIS. Typical resolutions for county-wide mapping projects are 2 to 3 feet. In cities, resolutions of 6 inches and 1 foot are used. From these orthos a limited number of roadway features can be extracted.

Digital orthophotos are incorporated in GIS. They function as a cartographic base for displaying, generating, and modifying associated digital planimetric data. Other applications include vegetation and timber management, routing and habitat analysis, environmental impact assessments, emergency evacuation planning, flood analysis, soil erosion assessment, facility management, and groundwater and watershed analysis. Orthos created from satellite images are sometimes used to create statewide road centerline maps.

The accuracy of orthophotos ranges from 1.5 feet with 0.5-foot pixel ground resolution to 6 feet with 3-foot ground resolution. The horizontal accuracy of a DOQ is typically around 3 meters (i.e., orthophoto error is typically three times the pixel resolution). Similar to aerial photogrammetry, GCPs are typically established using GPS and aerial GPS. Digitizing is manual, however, and no special equipment is needed, as the orthophoto can be directly used as the base map. Orthophotos are mostly used for network-level applications.

LIDAR. This system automatically measures elevations using laser technology–Light Detection and Ranging (LIDAR). The laser system is mounted on an aircraft, along with a GPS and an Inertial Measuring Unit (IMU). The GPS derives the laser's latitude, longitude, and height. The IMU provides information on the aircraft's roll, pitch, and yaw. Using these measurements, a computer can calculate the position of the laser as a function of time.

As the aircraft proceeds along the flight path, the laser oscillates back and forth perpendicular to the aircraft's direction, while rapidly sending and receiving laser pulses that reflect off the earth's surface. Utilizing the information on the position and attitude of the sensor, the elapsed time between laser pulse and sensor retrieval, and the speed of light constant, a large series of x, y, and z ground surface points are collected. These points are then transformed into a regular digital elevation model (DEM).

LIDAR creates three-dimensional surface points. However, because these points do not correspond to a specific feature, the horizontal component is of limited value. LIDAR is used for engineering design projects as well as DEMs along roadways.

The accuracy level of this measuring method is about 4 inches in the vertical direction. Typically, the laser is flown at altitudes of 5,000 to 8,000 feet above the ground surface. Theoretically, this can produce a horizontal accuracy of ±0.4 meters and a vertical accuracy of ±0.15 meters. Part of the accuracy equation is the accuracy of the GPS used to locate each LIDAR pulse return point. GPS is usually accurate to 5 or 7 centimeters.

Aerial GPS and inertial systems are used as the reference points of control. The data collection vehicle is either a helicopter or an airplane. LIDAR can be flown by either type of aircraft; the selection usually depends on the altitude of the flight. The system is fully automatic; an operator is not needed when processing DEMs. The large quantity of data created and the narrow swath of the LIDAR system limits its application to the project level.

2.4.2.2 Ground Vehicle-Based

Mobile Mapping. A mobile mapping van is equipped with a survey-grade, kinematic GPS receiver; an INS (Inertial Navigation System) unit; and up to five digital cameras. GPS data are used to determine the position of the van at any time, while the digital cameras capture high-resolution color images pointing forward and to the road right of way, showing a "windshield view" of roadside assets and condition. Each pair of digital cameras can be used to measure the spatial locations of roadway features.

This method of inventorying highway infrastructure and integrating the images into Infrastructure Management System (IMS) databases is considerably more efficient than traditional approaches. This system allows users to create GIS base maps and infrastructure management systems at an affordable cost and with a short turn-around time. Measurements can be made in x,y,z coordinates—this system creates real 3D coordinates. Data collected with this method are used in GIS for various applications, including asset inventory, mapping, and engineering. The accuracy of measurements is less than 1 meter. GPS base stations are usually set up at High Accuracy Reference Network (HARN) points or other first-order reference points. In addition, data are manually and semi-automatically post-processed. The system can be used in both network- and project-level data collection; however, it is most efficient if the roadway mileage of a project is more than 20 miles.

Video-Logging. For video-logging applications, digital or analog right-of-way images are captured in a single pass driving along a roadway. Some vendors offer multiple cameras configured to provide a 130-degree panoramic view, similar to a driver's view. Other agencies configure the right-side camera to provide a roadside view for environmental applications. Images are typically captured at predetermined intervals, usually 100 frames per mile, which equals a spacing of 53 feet between images. Images are usually stored on videotapes; newer systems deliver digital video. A video banner describing the roadway ID, date, time, and milepost can be optionally burned to the images. The location of the vehicle is determined by distance measurement instruments (DMIs) and/or real-time, differential GPS.

This method is used to collect data on a vehicle in x,y coordinates: distance (D) and offset (ΔO). Accuracy of data collection is 3 to 10 meters. Images are collected in real time; visual image inspection is used to extract asset information. The data collected with this method are used in inventorying and pavement condition analysis. Continuously Operating Reference Stations (CORS) are the reference point locations.

Distance Measuring Instruments (DMIs). DMIs are used for measuring distances and are installed in a car or van, combined with a data logger. The DMI needs to be initialized at a known reference point: an intersection or other log point. When the vehicle moves along the road, the accurate distance is recorded. For example, the NITESTAR® distance measuring instrument can measure distances to ±1 foot over the course of 1 mile. NITESTAR® has been designed to make distance measuring easy, and it is linked to a special keyboard for data logging. NITESTAR® has internal memory to store numerous events along with the distance at which they occur.

DMIs are used to collect data for asset inventorying (e.g., paint line length, guide rail length, pole or sign spacing, cable or pipeline length, truck, bus, or postal routes, E-911 address locating, crash reconstruction, and roadway and railway lengths.) The accuracy level is greater than 1 meter, based on a percentage of distance, ±1 foot per mile. Intersections (anchor points, nodes) serve as reference points.

2.4.2.3 Ground Surveying

Wheel. In ground surveying, the operator walks with a measuring wheel (like DMI) and measures distances relative to stations (visible reference points) defined on a map. Wheels range in size from 4 to 25 inches in radius. This method is used for measuring relative distance (ΔD); the counter measures up to 100,000 units (feet or meters). It is used for crash investigation, local surveys for maintenance, planning at the city and county level, and telecommunications inventorying. The accuracy is around 2 feet plus 2 percent change in distance. Stations along roads are used as reference points. Distance measurements have to be recorded manually on a printed map or notepad. The data are used for project-level applications.

GPS. Recent significant advances in roadway mapping reflect the use of combined technologies (e.g., GPS, dead reckoning technique). GPS is increasingly being used to obtain coordinate data associated with events and to generate GIS-based vector drawings to map those events to the network.

Positional accuracy varies depending on the data collection equipment used. GPS positional accuracy is much finer than those obtained with traditional maps (e.g., with TIGER files) and maintains tighter control for the location of linear features and events. GPS data positional accuracy is typically expressed in 2D or 3D, for example, in terms of circular error probability (CEP) or spherical error probability (SEP). Thus, it is necessary to transform these accuracy measures into 1D measures to make them comparable with linear feature distance accuracy measures.

Many states are collecting inventory and pavement condition data using vans equipped with videos, digital cameras, computers, and GPS receivers. Several states have experimented with the use of GPS for collection of incident data. GPS technology has rapidly matured to the point where, using differential GPS, sub-meter accuracy is technologically possible. Once 3D data are collected, DOTs are left to deal with determining the relationship with the associated cartographic centerline in their GIS spatial database. NCHRP Project 15-15 is evaluating various technologies for cost-effective ways to collect data on physical attributes of highway facilities and display them in straight-line diagrams.

Many states recognize considerable practical applications for using GPS in the field, and many state experts are exploring the options daily. GPS technology will likely dramatically affect how future GIS systems are built. The availability of highly accurate, 3D measurements makes it possible to calculate locations and distances more easily than with some of the linear location referencing methods currently in use. However, in the absence of 100-percent accuracy in both the spatial database and the GPS-collected data, there is still error in relating a GPS point or linear event to its accurate location on the associated centerline representation. That is, GPS, in itself, does not solve the conflation problem. Each time data are collected along the same roadway with a GPS van, a different string of coordinates will be obtained. These coordinate strings must be related before the data can be integrated.

Kinematic GPS. Kinematic GPS deals with dual frequency, carrier-phase data processing. The basis of GPS is the measurement of distances to GPS satellites using the travel time of radio signals. Only carrier-phase processing can provide millimeter-level accuracy; code-phase processing using single-frequency signals can yield only meter-level accuracy. The combination of two frequencies removes the effects of the ionosphere. Under heavy foliage or when satellite signals pass through light trees, the signal strength is greatly diminished. The receivers that have better sensitivity can track signals more reliably under such adverse conditions. The accuracy, reliability, and speed of obtaining results increase with the number of satellites. Five satellites are the minimum for obtaining a reliable position.

Measurement can be made in x,y,z coordinates with an accuracy of 1 to 5 inches. Positional data collected with GPS are used for various applications, including engineering design and property surveys. The location points of reference for this device are the HARN and first-order GPS reference stations. Data are collected by a person walking with a tripod; the data collection equipment is connected to a receiver. The kinds of positional data collected with this method are used for project-level applications.

Differential GPS. Differential GPS (DGPS) is a technique used to improve positioning or navigation accuracy. It is performed by determining the positioning error at a known location and subsequently incorporating a differential correction factor (by real-time transmission of corrections or by post-processing) into the position calculations of another receiver operating in the same area and simultaneously tracking the same satellites. Differential GPS is based on processing of pseudo-range (distances) between receiver and satellite using a ground reference station to provide corrections of atmospheric effects on the signals. One (fixed) receiver measures the timing errors and then provides correction information to the other (roving) receivers.

Measurement can be made in x,y,(z) (elevations are not very accurate) coordinates with accuracy of 5 to 10 feet. The location point of reference for this device is Continuously Operating Reference Stations (CORS). Data are collected by a person walking, with equipment connected to a receiver. Data collected with DGPS are used for project- and network-level applications including asset inventorying.

Laser Ranging. A common tool for inventorying assets is a laser range finder with integrated compass and inclinometer to determine locations of objects. Typically infrared, GaAs laser diodes are used for distance measurement. The generated light energy has a wavelength of approximately 900 nanometers, with a beam divergence of 3 milliradians, equal to a beam width of about 3 meters at 1000 meters. The target acquisition times range from 0.3 to 0.7 seconds. These lasers are completely eye safe, meeting FDA Class 1 specifications, which means that a person could stare directly into the laser for 3 hours without any harm to the eyesight. The radiated light power is in the order of 50 microwatts; it outputs only 5 percent of the light power of a typical TV remote control, and far less than a flashlight. Laser range finders calculate distance by measuring the time of flight of very short pulses of infrared light. This method differs from the traditional surveying instrument method of measuring phase shifts by comparing the incoming wavelength with the phase of the reflected light. Any solid object will reflect back a certain percentage of the emitted light energy. The instrument measures the time it takes a laser pulse to travel to the target and back with a precision, crystal-controlled time base. Knowing the speed of light, the distance is calculated. To increase accuracy, the laser measures as many as 60 pulses, utilizing the average to determine the range.

Using this method, measurements of Δd, angles (α, ζ) (azimuth, inclination), and x,y,z can be computed from angle and range, if the location of the laser gun is known. The level of accuracy is 1 inch plus a percentage of Δd. The location points of reference for this device are local reference points, GCPs or GPS. Data are collected by a person walking with a bi-pod, and with equipment connected to a laser for real-time data collection in the field. Data collected with this device are used for project- and network-level applications, including asset inventorying, surveying, and construction.

Total Stations (Theodolites). Land surveying with theodolites combined with an electronic distance measuring (EDM) system is the preferred method for project-level, high-accuracy mapping of small project areas. These instruments are also called total stations. They need to be set up on tripods and leveled by the surveyor. Measurements consist of a distance to a reflector, as well as a horizontal and vertical angle. The 3D location of the object point is computed immediately and stored on a data collector. There are total stations that work without reflectors and some that automatically trace the reflector (basically reducing the total station crew to the person holding the reflector).

Measurement can be made in x, y, z coordinates with accuracy of 0.5 to 3 inches. The location points of reference for this device are HARN or first-order GPS reference stations. Data collected with theodolites are used for project-level applications, including engineering design and property surveys.

Map Digitizing. Paper maps are attached to a digitizing tablet and lines are traced with a mouse or cursor directly on top of the map. An advanced approach is based on scanning the map and digitizing the lines on the computer screen using the mouse and computer cursor. There are automated programs for digitizing specific map elements, such as contours and road centerlines. In order to convert the digitized lines into a real-world coordinate system, control points are needed.

Measurements are typically in x,y coordinates (except when contour lines are digitized). Digitized maps are used in GIS for base mapping, legacy data conversion (parcel maps, utility drawings), engineering design, and property surveys. Accuracy of digitized maps depends on the quality and scale of the maps. It can be no better than the nominal accuracy of the original map. Typically the positional accuracy of any measurement represented on the map can be anywhere between 5 to 50 feet from its true position. Data are directly recorded on the computer. Clean up of data in a CAD system is necessary; some automated digitizing programs are available.

2.4.3 Positioning

2.4.3.1 Qualitative/Approximate Locating

The methods described in this section are commonly used to determine the current location of a vehicle, person, or feature in real time with measurement tools that are available to the average consumer or vehicle operator.

Distance from Landmark. The current position is determined as the estimated distance from a landmark (e.g., church, easily identifiable building, or roadside object). There is no offset, however, and the side of the road (e.g., north, south) is typically known. Typical applications are crash reports and investigation, police, emergency response systems (EMS-911), and roadway maintenance. Accuracy of measurements is in the range of 100 to 300 feet. Landmarks serve as the reference points. Data are collected by a person walking or driving in a car, recording data on paper.

Distance from Intersection. The current position is determined as the estimated distance from the nearest intersection. The side of the road (e.g., north, south) is typically known. These are distance, Δd, measurements and used in crash reports and investigation, police, emergency response systems (EMS-911), and roadway maintenance. Accuracy is in the range of 100 to 300 feet. Reference points are usually intersections. A person walking or driving in a car records the data on paper. Data are recorded in the field in real time by reading a vehicle odometer.

Distance from Milepost Marker. The current position is determined as the estimated distance from a milepost marker along the roadway. The side of the road (e.g., north, south) is typically known. These are distance, (d, measurements and used in crash reports and investigation, police, emergency response systems (EMS-911), and roadway maintenance. Accuracy is in the range of 100 to 300 feet. The reference point is the milepost marker. A person walking or driving in a car records the data on paper. Data are recorded in the field in real time by reading a vehicle odometer.

Address. Address numbers are defined in a grid system over a city or county, or as a function of the distance along a roadway, relative to a starting point. Often address ranges provided in TIGER files are used to estimate the location. Addresses are difficult to use, as they may not appear on a building, and they may be different in postal, county, and utility databases. Address numbers (#n) are recorded by a person walking or driving in a car. Data are recorded in the field in real time by reading the odometer. Accuracy is in the range of 100 to 300 feet. The reference point is a street segment or block.

2.4.3.2 Automatic Vehicle Location

GPS for Car Navigation. A GPS, compass, and odometer are often used in an integrated system. The measurements are automatically merged with street maps to keep track of a vehicle and show its current location. A navigation system needs to know where the vehicle is on a map. Correlating the raw data from the sensors to a navigable map database enables meaningful map display of the car's location, calculation of distances between possible destinations and turns, and route calculation. These functions are only as good as the map database on which they rely—accuracy, detail, and coverage are crucial to satisfactory performance.

The type of measurements are x,y, and azimuth angle (α) coordinates of the location and driving direction of the vehicle. This method is used in vehicle tracking, truck routing, car navigation, and emergency dispatch. Accuracy is around 10 to 50 feet, and GPS satellites are the reference points. Data are recorded and merged with a map in real time to continuously show the location of the vehicle.

Compass. A compass is an instrument that indicates direction. Two fundamental types of compass are used: the magnetic compass, which probably originated in ancient China, and the gyrocompass, a device developed at the beginning of the 20th century. In the magnetic compass, directions are obtained by means of one or more magnetic needles pointing in the general direction of the magnetic North Pole under the influence of the magnetic field of the earth. The gyrocompass, which is unaffected by the magnetism of the earth, consists of a gyroscope, with the spinning wheel on an axis confined to the horizontal plane so that its axle aligns itself with the north-south line parallel to the axis of the rotation of the earth, thereby indicating true north.

The compass is used to measure the azimuth angle (α). Accuracy of this device is better than 2 degrees. This device is used for vehicle tracking, truck routing, car navigation, and emergency dispatch. The reference point is the magnetic or true north. Data are recorded and merged with a map in real time to continuously show the location of the vehicle.

Odometer. An odometer is an instrument in automotive vehicles to indicate the total number of miles that have been traveled. The odometer generally shares housing with the vehicle's speedometer and is driven by a cable that the two share. When the vehicle is in motion, this cable moves a series of gears in the odometer, turning a set of numbered drums that count the miles traveled. Some odometers, called trip meters, can be manually reset to zero to measure the lengths of individual trips.

This device is used to measure distance, Δd, with accuracy greater than 20 feet plus 5 percent of distance. The data are used for vehicle tracking, truck routing, car navigation, and emergency dispatch. The reference is the start of travel. Data are recorded and merged with a map in real time to continuously show the location of the vehicle.

2.5 MODELING DATA ERROR

2.5.1 Introduction

This section describes a conceptual error model for assessing the effects of data uncertainty in measurement techniques applied to transportation phenomena and transformations between spatial referencing systems. The fundamental question of interest is the positional accuracy of a recorded position for any transportation feature or event. The error model would allow users of GIS data to be aware of the bounds on the "true" locations of transportation features and events, whether these are independently arrived at or derived from the integration of diverse data sources. To assess positional errors, there must be an understanding of how a recorded position for a transportation feature or event is determined. A starting point for the development of the conceptual error model is a model of the transportation system. The 20-27(3) data model is a comprehensive and well-developed model of transportation phenomena and it contains many of the relations necessary to make the above determination. However, it falls short in supporting a comprehensive view and hence management strategy for positional accuracy, as some important contributing error sources are not modeled. The most critical components for developing the error model lie in the relationships of transportation features to spatial objects and spatial objects to spatial referencing systems. Before describing the error model, some relevant terms and issues are defined and discussed in the following section.

2.5.2 Review of 20-27(3) Data Model and Clarification of Terms

This section defines and clarifies terms pertinent to the development of the conceptual error model and indicates their overlap or deviation from the 20-27(3) data model. The terms of interest include transportation feature, event, physical roadway, roadway section, link, node, network, spatial reference systems, reference objects, locational reference, and anchor section.

In the 20-27(3) data model, a key object is the **transportation feature**. It is defined as a non-decomposable phenomenon in the transportation domain. Examples of transportation features include roads, routes, ramps, bridge abutments, culverts, maintenance management zones (e.g., spray zones and no sand/salt sections), and pavement management zones. Another object in the 20-27(3) data model for which positional accuracy issues are of concern is the **event** object. Events refer either to occurrences or changes of state to features on or along a roadway. Events can be traffic crashes, construction, or repair activities applied to transportation features.

In the 20-27(3) data model, both transportation features and events are associated with spatio-temporal objects. For this project, the interest is only in the spatial dimension. Transportation features and events are modeled in the 20-27(3) study as being represented by spatial objects and associated with spatial reference systems. According to the 20-27(3) data model, each transportation feature or event can have zero-to-many associated spatial objects, and each spatial object can be associated with zero-to-many topological or zero-to-many geometric objects. The topological objects serve to model the connectivity among spatial objects. Each geometric object serves to represent the position and possibly size and shape of a transportation feature at some point in time. Each geometric object has one or more associated spatial reference systems that allow a transportation feature to be spatially positioned.

For the purpose of having some distinct transportation features to refer to in developing the conceptual error model, a few of them (e.g., physical roadway and roadway section) are distinguished and their associated spatial objects (e.g., link, node, and network) discussed.

Physical roadways are the connected set of transportation features, such as highways, streets, roads, and exit ramps, that have a real-world presence. Because the physical roadway is a complex connected system, it is frequently of interest to be able to identify and refer to its sub-sections. A road section is a sub-unit defined as the portion of physical roadway between intersections. A roadway section is a transportation feature and a section in the 20-27(3) data model.

A physical roadway and hence roadway section can have multiple associated digital representations that will vary in spatial detail and hence positional accuracy. For most transportation applications, a centerline representation serves as the geometric representation of the physical roadway. There are two commonly available public centerline digital representations for most major roads in the United States: the USGS 1:24,000 scale DLG and the Census TIGER file roads (nominally of 1:100,000 scale heritage). Another possible geometric representation is the edge of pavement as often captured from aerial photography. This results in multiple digital spatial representations for a single physical roadway section, as indicated in Figure 2-1.

The term **link** refers to the digital spatial representation of a **roadway section** centerline. It corresponds to a spatial object in 20-27(3) and is defined as a spatial object that represents the section of roadway between intersections. A **node** is a spatial object that represents the road intersection. A **link** has one topological representation but may have multiple geometries. The geometry of a link is typically a set of (x, y, and sometimes z) coordinates for a road centerline. The geometry of a node is a (x, y, and sometimes z) coordinate for a road intersection.

A **network**, a complex spatial object in 20-27(3), is defined as a set of connected links and nodes. A network may have topology and/or geometry. The topological representation captures roadway connectivity and typically indicates the bounding nodes for each link and the incident links for each node. A network is a key component and concept in linear referencing systems that are one form of **spatial referencing systems**.

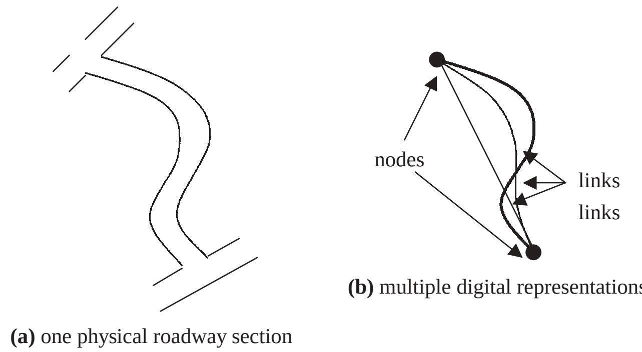


Figure 2-1. Representations of physical roadway.

The 20-27(3) data model identifies spatial referencing systems and it is agreed that there are multiple spatial referencing systems that differ primarily with respect to their dimensions. A referencing system for any dimension (i.e., space, theme, or time) is defined as a framework for a set of measurements where a measurement is the assignment of class or score to a phenomenon based on a set of rules. A spatial reference system defines the parameters and rules to situate a measurement in space. The essential parameters for any spatial reference system are an origin and units (the required parameters for a linear spatial referencing system). A 2D system further requires specification and orientation of two axes and possibly location and relation of the origin and axes to a geometric body. A 3D system requires specification of a geometric body and orientation of the origin and three axes with respect to this body. Figure 2-2 illustrates components of these systems.

The parameters and rules required for each dimension correspond to the datum object specified in the 20-27(3) data model. To generate a measurement in one of these systems involves any number of different measurement methods for distances, angles, or times. **Reference objects** specified in the 20-27(3) data model are an important concept within reference systems. These objects are measured typically to well-defined standards such that additional measurements can refer to these measured positions rather than to the original system parameters. For example, mile markers can be reference objects in the linear system and new measured positions can be based on the measured mile-marker locations rather than with respect to the system origin. As shown in Figure 2-3, the

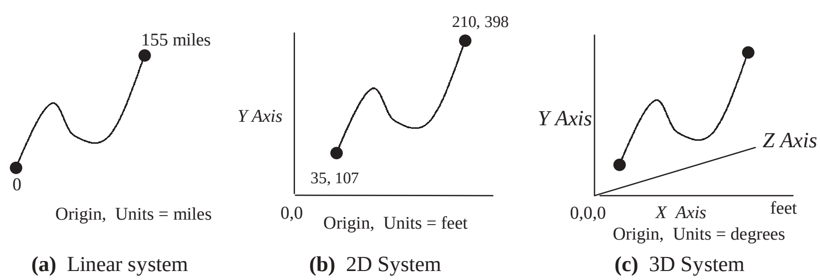
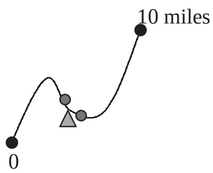


Figure 2-2. Components of spatial reference systems.



NOTE: In this case, mile markers along a road can be used to generate measures for unmeasured transportation features or events.

Figure 2-3. Reference objects.

position of a transportation feature, represented by the triangle, can be determined by the two bounding mile markers rather than by the origin.

Any transportation feature or event may be associated with one or more linear or higher dimensional spatial reference datums and one or more different measuring methods (e.g., photogrammetry or GPS). A further important distinction is that they also can be associated with different orders of measurement (i.e., measured directly according to the system parameters or measured with respect to one or more reference objects). This distinction captures the situation in which a roadway inventory project uses DMI to measure both road centerlines and all the assets along the roadway at the same time. All of the transportation features in this situation would have directly measured positions, rather than positions measured through reference objects (mile markers). The 20-27(3) data model indicates that geometric objects are only linked to a spatial reference system through reference objects.

Based on these associations, every transportation feature or event has one or more **locational references**. A locational reference is a term not used in 20-27(3). It refers to the information stored in the database that provides the spatial location description for any transportation feature or event. Figure 2-4 illustrates locational references for 2D and linear reference systems.

As mentioned above, **links** may have multiple geometries and, for each geometry, there may exist one or more spatial reference systems. Multiple linear reference systems may exist for a set of links due to the passage of time. A linear reference system, for example, might have been put in place in 1980 and re-measured in 1999 using a new measurement technology, so that for some transition period, two linear measurement systems will coexist. Linear measurements will also typically coexist (on links) with one or more 2D or 3D measurement systems.

The multiple spatial reference systems attached to links may be dependent on or independent of each other. An example of an independent case is a situation where the link geometry is 1 : 24,000 scale DLG data in a 2D reference system but with a linear measured distance for the link captured by odometer or DMI. In this case the linear measured distance is not dependent on the 2D geometry. A dependent scenario occurs when the linear measurement is computed directly from a 2D or 3D measurement system, say by computational geometry. As an example, GPS might be used to measure coordinates (longitude, latitude, and height) for road centerlines and these measures might subsequently be used to compute a 3D distance measure for the road centerline. As an extension to the 20-27(3) data model, the dependencies among measurement systems should be accounted for, as they are pertinent to the error model. In the dependent case, the linear measured distance will be affected by the error characteristics of the 2D or 3D reference system.

Under a linear reference system, the distance measure is applied to an **anchor section**, which is a set of connected roadway sections or links. The measured distance of the anchor section is used to reference other transportation features on or along the roadway. In some cases, the linear distances to transportation features along the roadway are measured at the same time as the measures are applied to the anchor section. These measurements all have the same measurement characteristics and hence the same error characteristics. The anchor section is described and other transportation features measured simultaneously as having direct linear measured positions. Any transportation features subsequently referenced to the anchor section will have an indirect linear measured position and hence different error characteristics.

Similar distinctions may apply in the 2D and 3D cases. So as a refinement to the 20-27(3) data model *(87)*, it is suggested that locational references for transportation features be categorized as follows:

- Direct linear measured position,
- Indirect linear measured position using linear reference objects,
- Indirect measured position using 2D reference objects, and
- Indirect measured position using 3D reference objects.

(a) Independent locational reference **(b)** Dependent locational reference

Figure 2-4. Distinction between locational references.

Positional accuracy depends on the measurement methods applied in each case and also on the accuracy of reference object measurements where these apply.

The above qualifications and identified dependencies are not explicit in 20-27(3), yet they have implications for understanding the positional error characteristics. These dependencies and their effects on positional error lay the foundation for development of the error model.

2.5.3 Sources of Error in Spatial Data

The locations of transportation features are typically collected, analyzed, operated on, transformed, and compared relative to other transportation feature locations without regard for positional accuracy or the quality aspects of the data. Positional errors arising from imperfect measurements are inherent in data. Also, certain operations on data, such as transformation among spatial referencing systems, introduce additional persistent spatial distortions. Such errors propagate through spatial analytical processes and are embedded in applications that manipulate data in various ways to produce results used in decision making (*87*).

The first step in developing the conceptual data error model is to identify the various sources of error. The primary sources of error associated with positional data are acquisition or measurement, processing, and presentation or visualization. Regardless of the measurement technique and referencing system, data will be observed with error. As discussed in Section 2.4 of this report, the method of data collection sets limitations on the selection of the measures and their metrics.

As described in the preceding section, every transportation feature is or can be associated with one or more spatial referencing systems. Depending on the measurement techniques used by a referencing system, each recorded location reference will have different error characteristics. Figure 2-5 outlines the error sources associated with different spatial referencing systems. The important difference in the linear referencing system is its dependency on a path definition. The path can be the physical roadway and the measurement method may be applied to the physical roadway (e.g., using DMI). Alternatively, a path can be a digital representation of the physical roadway, in which case, the linear measurement may be computed from the digital representation. In this latter case, the level of the network's spatial detail (i.e., topological and geometric) and the measurement technique will affect the measured distance and any subsequent locational references that employ this representation and measurement.

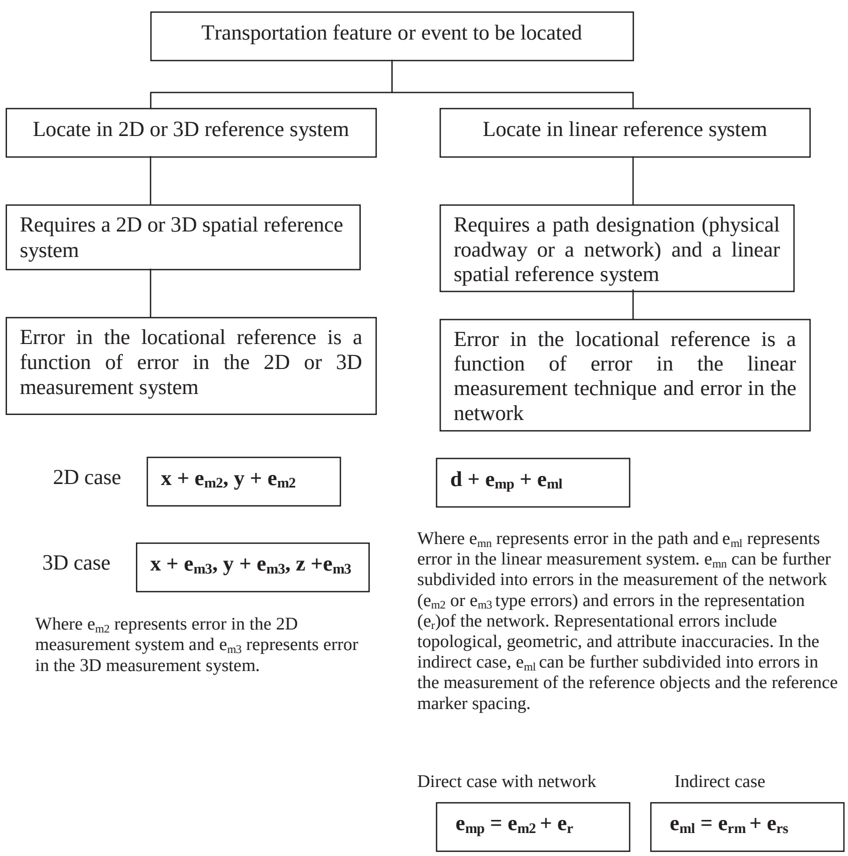


Figure 2-5. Outline of error sources associated with the process of assigning locational references to transportation features.

A transportation feature or event, whose location is measured by a 2D or 3D measurement system, such as photogrammetry or GPS, is independent of the road geometry. For example, a hazardous waste spill from an overturned truck can be captured and recorded by GPS without reference to the adjacent roadway. As another example, a traffic crash may be reported as cell phone coordinates (x, y) using any of a number of cell phone locational methods.

Linearly referenced transportation features or events, because of their potential dependency on a network (i.e., digital representation of the physical roadway), are subject to inaccuracies in the network as well as characteristics of the measurement methodology. As indicated in the previous section, it is also important to distinguish a direct linear measured location from an indirect measured location. The direct linear measures typically apply to the path and physical transportation assets along the path. The accuracy of indirect measurements depends on reference objects and will be influenced by the measurement errors in the reference objects and the spacing of the reference objects as indicated in Figure 2-5.

The linear spacing of the reference objects associated with the linear reference system can substantially affect positional accuracy. The spacing between linear reference markers is a form of resolution and the coarser the spacing or resolution, the less accurate the locational reference. If, for example, a crash is reported as between Exits 49 and 50, the accuracy of the event is a function of the distance between exits (approximately ±5 miles for a 10-mile spacing between the exits). A crash referenced as just south of Mile Marker 315 can have a higher accuracy because of the finer spacing (i.e., resolution) of mile markers. The error model must consider the positional accuracy of a linear locational reference as a function of the type and resolution of the linear reference method.

Where a digital representation substitutes for the physical path, the results are multiple possible topological and geometric representations of the physical roadway, some of which may be substantially less accurate than others. Characteristics of a network that affect the quality of referencing a transportation feature or event include topological completeness, geometric accuracy and detail, and attribute accuracy and consistency.

To understand the role of geometric accuracy and detail, consider the case in which GPS is used to position a road centerline. Each recorded coordinate might have centimeter-level accuracy. However, the number of coordinates collected and their ability to capture the geometry of the physical roadway will have a sizeable effect on the accuracy of the linear measured distance generated from these coordinates. This is another instance in which resolution has a significant effect on positional accuracy.

Attribute accuracy and consistency play a role given that, as noted in the previous section, a linear location reference includes a route or similar reference that must ultimately provide a link to an anchor section. Relationships need to be established among several objects across the database to make connections from route identifiers to links to an anchor section. If a name or identifier is incorrect or inconsistent somewhere in the database, misconnections will occur, resulting often in gross inaccuracies in a referenced position. The problem is most likely to occur with indirect linear measurements.

From a practical point, the location of objects relative to a network is of great importance. Practically, it does not matter if the road network is a meter off, as long as one can find the location of a certain feature. Networks with low positional accuracy can be used, as long as they are complete relative to the log points and intersections.

Although several sources of error are involved in generating a locational reference for transportation features or events, the transformations between spatial reference systems are another source of positional error. In the transformation process, either two independent reference systems have to be combined into one new system, or one system must be transformed to the other. Both approaches raise issues of uncertainty.

Figure 2-6 is a schematic representation of three sources of errors involved in this context. Figure 2-6a illustrates an example of 2D measurement error. Because the measurement is independent of the road network, the measurement may be off the roadway even though, in reality, it is on the roadway. Transforming the 2D reference to a linear reference will place the location on the roadway but with some error that is a function of the 2D measurement error plus a linear measurement error. Conceptually, the 2D-measured location moves to the closest point on the roadway. However, given the error in the measurement, there are multiple closest points represented by the normal vectors from the circular error bound to the road centerline.

Figure 2-6b illustrates error in the network representation. Given that the centerline position has error, the set of closest points extends to positions represented by the network error buffer. Figure 2-6c represents the cumulative error from these sources. Finally, Figure 2-6d illustrates the errors that might be present in the linear referencing system. Figure 2-6d illustrates potential bounds on the transformed linear position. The specific error value depends on the errors in each of the respective referencing systems. The effective result is that the 2D error transforms to a linear error in the linear referencing system. Figure 2-7 shows an example for a 2D error ellipse.

2.5.4 Transformation Methodology

Transformation of data provides the necessary key for the interoperability of data sets. Many transportation agencies recognize a need to be able to translate location references between spatial referencing systems. Some agencies establish one referencing system as the primary system and derive the locations in other systems from the primary system. For example, the primary location referencing method at the Virginia Department of Transportation (VDOT) is link-node. VDOT

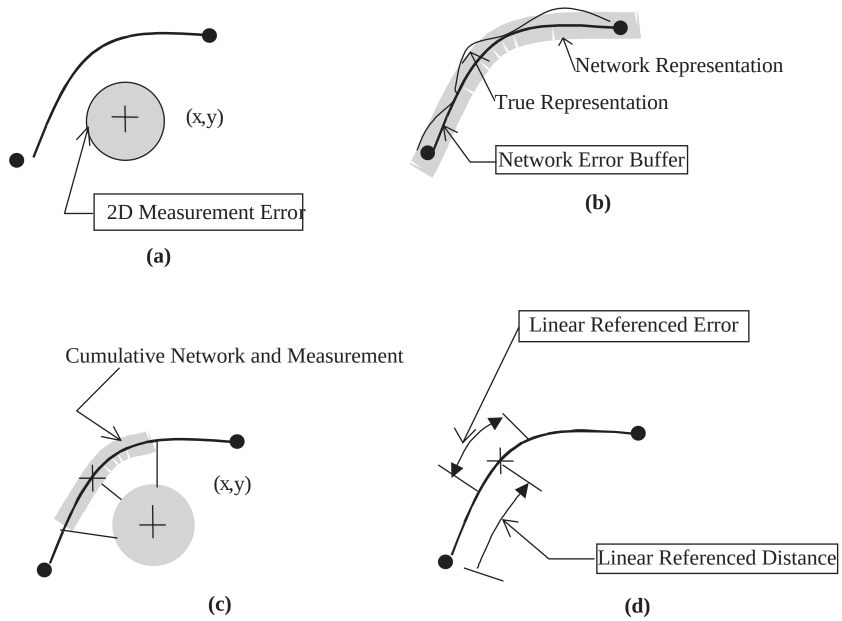


Figure 2-6. Schematic representation of positional data error.

also uses mile points derived from link lengths between nodes with known mile-point locations. Other states establish a location control mechanism that is independent of any LRM. For example, Wisconsin DOT has a Location Control Management System, which is used for conversion between different linear referencing systems (e.g., link/site and reference point) *(11)*. The main types of transformation are defined and illustrated in the following sub-sections.

2.5.4.1 Types of Transformations

The main types of transformations for transportation applications involving linear reference systems are as follows:

- **Transformation Type 1**—transformation of a 2D (or 3D) location expression to a linear location expression. This might occur when new data are collected using GPS and these GPS coordinates need to be converted to a linear referenced position to integrate with legacy data already linearly referenced (Figure 2-8a).
- **Transformation Type 2**—transformation of a linear location expression to a 2D (or 3D) expression. This might occur when linearly referenced data need to be converted to 2D coordinates for analytical purposes such as finding all crashes within 2 miles of an intersection for all intersections in a jurisdiction (Figure 2-8b).
- **Transformation Type 3**—transformation of one linear location reference to another linear location reference. This might occur if transportation features referenced in a legacy linear system need to be updated to a new linear system or if more than one linear system exists within an organization and data need to be integrated across these systems (Figure 2-8c).

Currently, locational references, regardless of the type of spatial reference system, are not reported with error. In terms of an error model, for the 2D case, assume a coordinate (x,y) with error such that the expression is as follows:

$$(x + \gamma_X, y + \gamma_Y)$$

where γ_X, γ_Y are the errors in the x, y values respectively**.**

In the linear case, it is assumed that there is some error associated with the distance measure *d*. If *A* is the anchor

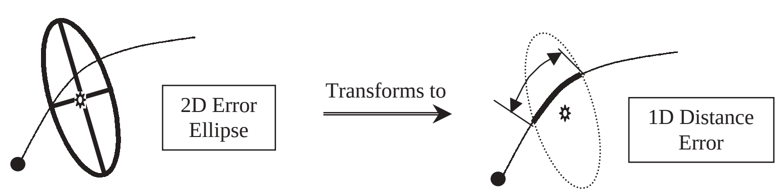
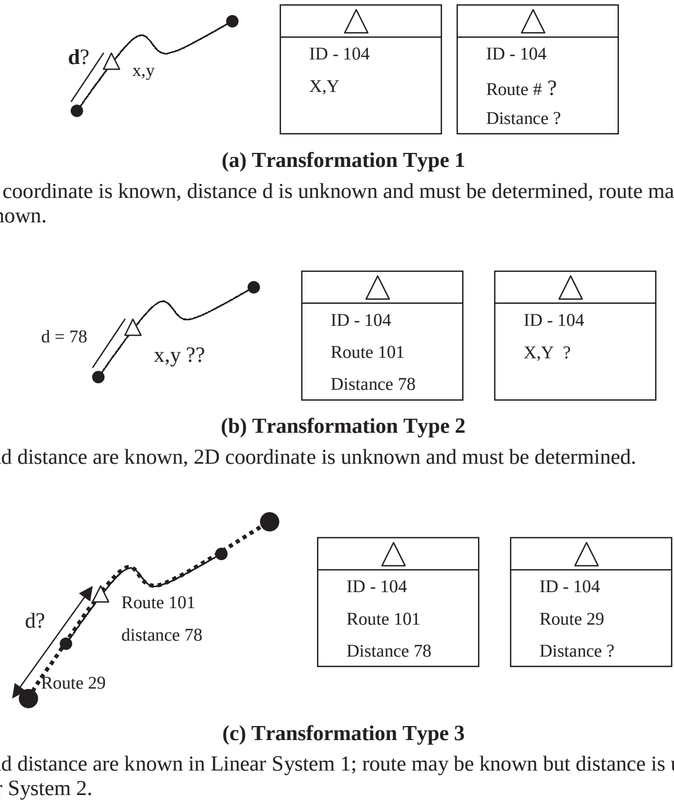


Figure 2-7. Example of 1D error generation.



(a) Transformation Type 1

2D (x,y) coordinate is known, distance d is unknown and must be determined, route may or may not be known.

(b) Transformation Type 2

Route and distance are known, 2D coordinate is unknown and must be determined.

(c) Transformation Type 3

Route and distance are known in Linear System 1; route may be known but distance is unknown in Linear System 2.

Figure 2-8. Illustration of transformation types.

point or origin for a distance measure, then the expression for the distance measure with error is as follows:

$d_A + e_A$

where d_A is the measured distance relative to the anchor point *A* and e_A is the error associated the measurement. Using these error expressions, errors in the three main transformation types are illustrated in Figures 2-9a through 2-9c.

Several variations of these three main transformation types are possible. Illustrations of some specific transformation cases for Types 1 and 3 are considered in the next section.

2.5.4.2 Transformation between GPS and LRS (Type 1)

Transformation between GPS and LRS is an example of a roadway inventory project conducted by a contractor for a state DOT. In this example, coordinates of transportation features are captured using GPS and transformed to UTM or State Plane. Both road centerline and assets are measured with the same system (e.g., a stereo imaging system, which bases its locations on GPS). This is an example of the direct linear measurement case and, therefore, both road centerline and roadway features are of the same accuracy (<3 ft). The

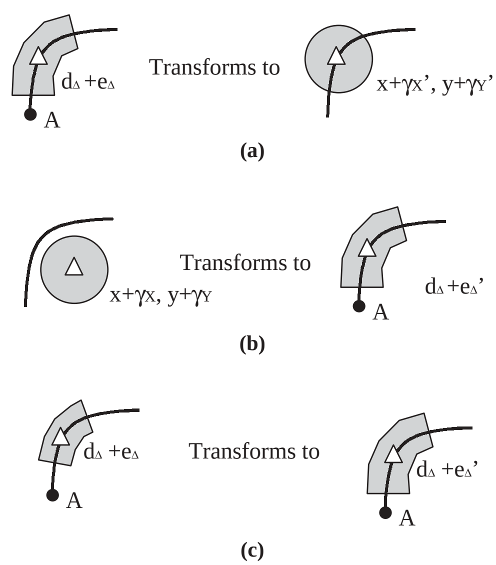


Figure 2-9. Examples of errors associated with the transformation types.

state DOT wants the asset information converted to a linear reference system, so it is necessary to transform the GPS data. The contractor uses the following steps to accomplish this transformation (The steps are illustrated in Figure 2-10):

1. Compute the 3D (i.e., slope) distance of the road centerline, starting at the beginning of the road (there are certain rules defining the beginning and end mileposts, e.g., north–south).
2. Compute the mileposts of all intersections (i.e., log points and anchor points). This road centerline, together with the distance references of the log points, serves as the network.
3. Find the closest point on the road centerline of the roadway assets inventoried, then compute the milepost (i.e., distance from the start of the route) and the offset of the feature from the centerline.

The result is two measures (distance [*D*] and offset [*O*]) for each transportation asset. This method also allows the positioning of linear features if the beginning and end points of the feature were measured.

2.5.4.3 Transformation Between Two Linear Reference Systems (Type 3)

Often state DOTs have legacy data that are positioned using a form of linear referencing system. For instance, a DOT may have a road centerline network available that was digitized from geocoded aerial photos and of questionable accuracy; however, the data are consistent with all other GIS data layers in the DOT system. In many instances, the DOT may not want to add inventory points created with GPS, because they would not overlay with the legacy data, even if they are more accurate. One approach to this problem is to transform the transportation feature or event data captured with GPS to a linear reference system (LRS-1) and then convert this system to the customer's legacy linear reference system (LRS-2). This can be accomplished with the following steps as illustrated in Figure 2-11:

1. Compute the distance to all intersections and the end point of LRS-1 relative to the origin or starting point of the legacy road centerline LRS-2.
2. Take the transportation feature data referenced in LRS-1 and, using the anchor points (i.e., intersections) as reference objects, squeeze or stretch the distances to match the measured distances of the legacy system, LRS-2. The desired transformations can be accomplished with most standard GIS programs using dynamic segmentation routines.

The result is the new feature inventory referenced to the old centerline. This allows the user to combine new roadway features with legacy data without having to change the existing system completely.

The opposite transformation also may occur where legacy data are transformed to a newer linear referencing system. A question of interest for DOTs may be whether there is a significant accuracy difference between converting new inventory features to a legacy linear referencing system and converting features referenced in the legacy system to a new linear system.

Assuming that the same reference system was used, one approach would be to accept the data as is, without considering the consistency of topology and the differences in the precision of the measurements and the resolution of the reference methods. Another approach would be to use redundant information by comparing locations of identical events and to stretch or shrink the historical data set. In the latter case, the uncertainty information attached to (or assumed for) historical events would have to be transformed as well.

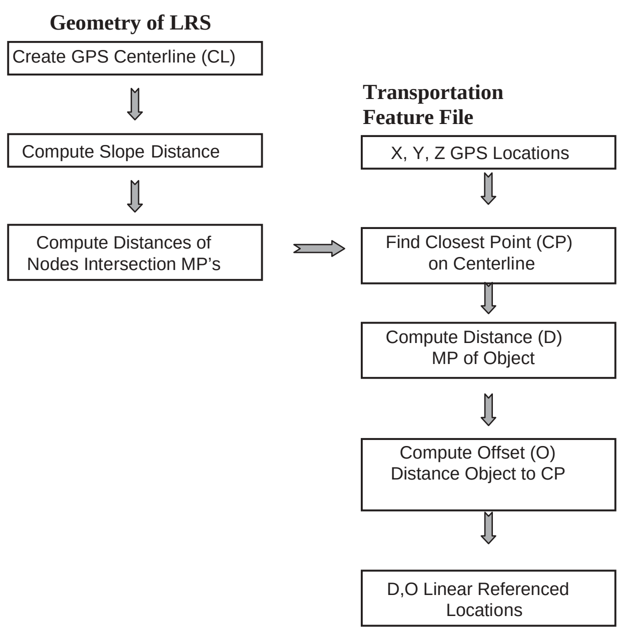


Figure 2-10. Steps in transformation of data for GPS to LRS.

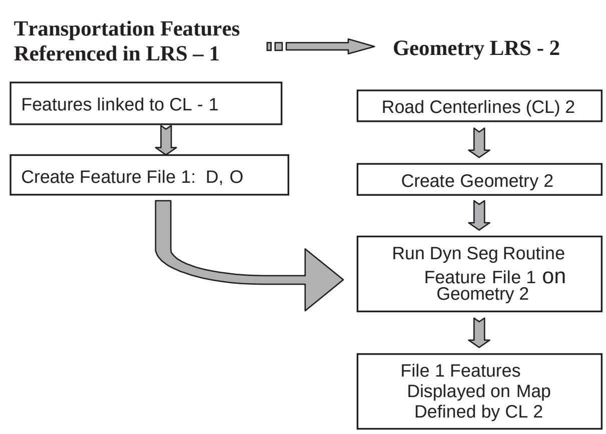


Figure 2-11. Steps in transformation of data from LRS-1 to LRS-2.

2.5.5 Model Concept

A formal approach in developing the error model considers all the components of uncertainty mentioned above, as well as transformation-specific properties. This includes recorded measurement precision, accuracy of the network, and issues of scale and resolution. Key issues in the model concern errors initiated in the measurement system followed by errors in the transformations between the different reference systems. A location and its associated uncertainties are the central objects of interest. The conceptual data-error model is designed to handle the uncertainties associated with the locations of transportation features and events present in transportation-related applications.

A location can first be influenced by the definition of a feature or event (e.g., a crash location can be seen as the location where the crash started or where involved cars stopped). Once a definition has been established, a transportation feature or event is located in a specific reference system by a particular measurement method and measurement device and the level of uncertainty depends on the reference system characteristics. In a 2D (or 3D) reference system, the positional characteristics of the error component will be dictated primarily by the precision of the measurement device. A linear reference system is particularly prone to accumulating systematic errors. Additionally, in a linear system, the resolution and accuracy used to record the network, as well as the method and device used to acquire its location, affect the quality of the data.

The goal is to isolate the location and formalize a more abstract model for the related parameters (i.e., feature or event, network, measurement system characteristics and dependencies). This approach allows a generalization of transformation procedures and, thus, builds the basis for an error model formulation that will allow development of an uncertainty expression for a location.

The three main components of the conceptual model define the input data, the desired output, and the requirements needed to achieve the desired outputs:

- Inputs (or information that exists)
  - Reference systems (1D, 2D, and 3D)
  - Event data (1D, 2D, and 3D) including networks and event data
  - Associated accuracy or uncertainty information (e.g., measurement error)
- Output (i.e., estimation of errors)
  - Estimates of errors associated with the transformation between reference systems
  - Estimates of errors in the combination of data from different dimensions (e.g., 1D network with 2D event, such as highway: pavement status with crash location)
- Requirements (or processes to use the available information to obtain the desired outputs)
  - Knowledge of all involved reference systems
  - Transformation methods between reference systems for events and their associated uncertainties
  - Means to combine uncertainties associated with different data.

These components define the structure of the data error model. This is an object-oriented concept, where the position of an event can be visualized as an object that depends on (a) an event, (b) a reference system, (c) a network, and (d) a measurement device or methodology. The focus of the error model is methods of transforming between reference systems and the associated uncertainties, as well as a means to combine uncertainties associated with different data. Figures 2-12 and 2-13 illustrate the dependencies of the error model. Figure 2-12 shows the relationships among event, measurement method and device, and reference systems. Figure 2-13 shows the transformation between two referencing systems. A measurement device (e.g., DMI versus photogrammetry) or a measurement method (e.g., linear distancing versus 2D measurements) introduces uncertainties. The level of uncertainty also depends on the reference system itself. For example, in a 2D reference system, the positional characteristics of the error component of the uncertainty of an event will be primarily dictated by the precision of the measurement device. A 1D reference system is particularly prone to accumulating systematic errors. Thus, the positional error component of an event depends increasingly on the systematic errors inherent in an existing linear network. The transformation methodology dictates the transformation of associated uncertainties.

The dependencies shown in Figure 2-12 can be compared with the model outlined in 20-27(3) data model *(87)*. The given terminology, however, varies slightly. The essential parallels are that events are directly linked to a location and that the location is directly linked to a reference method and a reference system. The addition of the direct link to the information on the measurement device as well as the network (or, to be more precise, the uncertainty of the network) is an additional requirement of the conceptual data error model. This outline, however, fulfills the purpose of enhancing the visualization of the uncertainty portion of the concept. It should be emphasized that access to information regarding the reference system, the network, the measurement method and device, as well as the event, is essential for an error model. The means of getting this information is secondary. For the purpose of retrieving this information, the data model described in Adams et al. *(87)* can be used as a basis. Additional objects (e.g., an uncertainty object), however, have to be introduced.

2.5.3.1 Model Formulation

A mathematical formulation of a conceptual model that incorporates error into positional data can be written as follows:

$$L = TL + EL_1 + EL_2 + \ldots + EL_k,$$

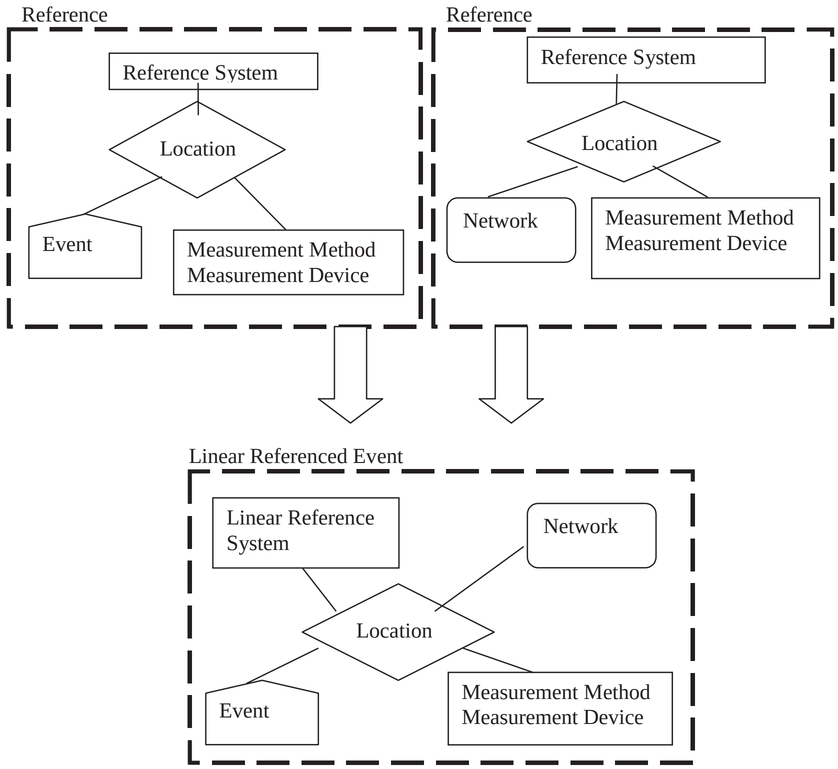


Figure 2-12. Data error model—combination of event and network.

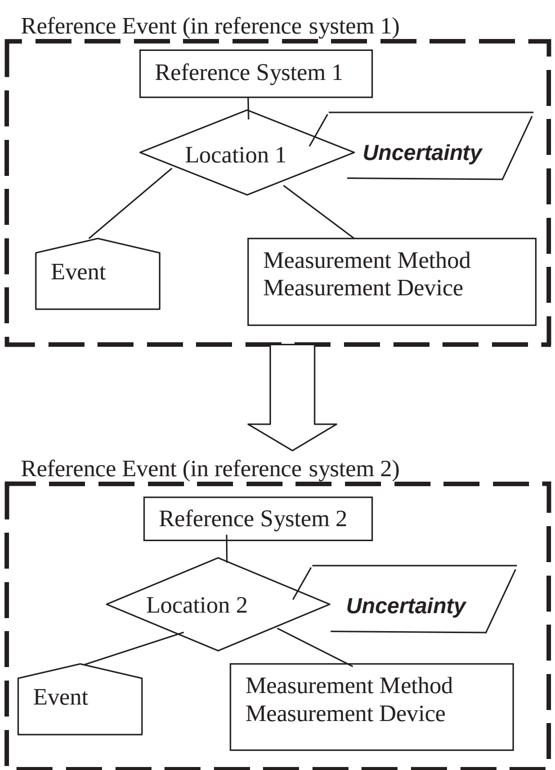


Figure 2-13. Data error model—transformation between referencing systems.

where L is a recorded location, TL is the true location, and the EL_i are errors associated with a location from *k* different sources. Thus, each measured location is the sum of the true location and a number of error terms. Each of the error terms has an associated probability distribution that describes the likelihood of errors over its range of plausible values. A number of the error terms were outlined in Figure 2-5.

The data-error model, however, can be simplified by combining the errors into a single term. The resulting statistical model is written as follows:

$$L = TL + EL,$$

where EL is the overall error term at location L. The probability distribution of EL is determined by combining the probability distributions of the individual error sources, which may be correlated.

An extension to the generalized model *(19)* can be represented by the modified location expression:

$$LX = (LRM, LE, DX, EL).$$

where LX is a linear location expression composed of linear referencing method (LRM), linear element (LE), and distance expression (DX). The additional term EL for the overall error term of the linear location expression LX can be specified by the probability distribution around the true location. This term, however, is not measured or given as the other terms of the location expression. Calculations are required to acquire an estimate of the associated error term EL.

These models are general and can be applied to data collected using any location measurement technique and any referencing system. An appropriate model for each location measurement technique will be chosen to determine plausible probability distributions for the error components and total data error.

Transforming data from one referencing system to another will result in transformed locations that also have errors (i.e., errors propagated from the original measurements and errors introduced by the transformation process). These errors will be present, regardless of whether the user is converting from 2D or 3D data to 1D data or from a particular dimension to the same dimension. A conceptual model for the errors associated with transformation can be written as follows:

$$g(L) = g(TL) + FL,$$

where g is the transformation function and FL is the error associated with the transformed location. As in the model for the errors in observed data, F will have a probability distribution that must be determined. The transformation function itself may have a systematic bias as a result of the transformation or the referencing system.

For simpler transformations (such as 1D to 1D), the probability distribution of FL may be determined by mathematical derivation, such as the "delta" method. However, if the transformation is more complex (e.g., across dimensions or using a map-matching algorithm), the probability distribution of FL probably will need to be obtained by numerical methods.

In formalizing the conceptual error model, an "uncertainty" object with knowledge or stored attributes of, for example, the resolution of the measurement system, active scale, and measurement error (or legacy if transformed) is added to the 20-27(3) data model *(87)*. Possible uncertainty attributes are listed in Table 2-6. Modifications of the uncertainty object for different objects (e.g., geometric location, temporal time stamp, and network) are advisable. This would require implementing an uncertainty object, for example, for a stored spatial location, time stamp, measurement method, and network. It is optional to store one uncertainty object with all possible attributes and use it to store spatial, temporal, and network-specific uncertainties. The option chosen will be determined by the status of the currently implemented system (DOT-specific). For an existing system, it might be easier to implement a single additional object, rather than to add attributes and functions to a multitude of existing objects. Furthermore, Table 2-6 is not necessarily a complete representation of all attributes and functions. Additional attributes and functions can be added as needed.

TABLE 2-6 Uncertainty object

	Object: Uncertainty
Attributes	Resolution
	Active Scale
	Measurement Error
	Network Accuracy
	Topological Completeness
	Lineage (of Previous Transformations)
	Probability Zone
	Temporal Uncertainty
Functions	Visualize
	Transform
	Merge

The functions of the uncertainty object use the information stored by the attributes of the object. The functions of the object help to communicate the inherent uncertainties. For example, one can visualize the uncertainty of a spatial event by presenting the probability zone around the specified event. The transform function has several variations, one for each possible transformation with a corresponding metric that has knowledge of how the probability zone of an event (i.e., location error) has to be adjusted to reflect the performed transformation. The merge function accounts for the combination of two or more objects with properties of different reference systems. These can be events, events and a network, or networks.

2.5.6 Presentation of Positional Data Error

Error can be represented by either a description or an error map. Hansen *(88)* noted that data quality standards on positional accuracy emphasize the accuracy of the coordinate values in the x, y, z plane. Error estimates with confidence intervals for these coordinate values are not explicitly described as elements, nor is the precision of the coordinate values delineated.

Agumya et al. *(83)* noted that the primary concern that end-users have regarding uncertainty in data is its potential effect on their decisions. The intention of assessing fitness for use is to avoid the application of data whose uncertainty may cause unacceptable results. The traditional method to assess the acceptability or fitness of use—the standards-based method—compares data uncertainty with a set of standard methods that reflect the acceptability levels of uncertainty in the data *(36)*. With this technique, fitness for use is assessed by directly comparing the quality elements of information against a set of standards that represent the corresponding acceptable quality components. To facilitate the comparison, the standards are defined using the same elements as those used for describing data quality. These may include scale (of the source document), root mean square error (RMSE), resolution, percentage of correctly classified pixel (PCCP), currency, and percentage completeness *(83)*.

The three main ways of presenting uncertainty associated with positional data for two-dimensional GIS are as follows: (1) a confidence region model based on a rigorous statistical model, (2) error band models derived from the error propagation law in statistics and stochastic approaches, and (3) reliability of linear measures based on simulation and statistical techniques. Analytical and simulation techniques were used to investigate positional error. It was concluded that both techniques provided approximations of the error with identical results. The simulation technique was found to be time-consuming compared with an analytical method *(89).*

It is commonly assumed that a node is distributed within an ellipsoid, centered at its corresponding true location *(90).* Modeling positional error assumes that the error of each node is normally distributed within an error ellipsoid centered at its true location. GPS data positional accuracy is typically expressed in 2D or 3D, for example, in terms of circular error probability (CEP) or spherical error probability (SEP). Thus, it is necessary to transform these accuracy measures into 1D measures to make them comparable with linear feature distance accuracy measures. As DOTs deploy GPS in positional data collection, the need for integrating GPS data into existing LRS increases. As the GPS technology improves, accuracy increases, although the data captured are not entirely error-free. Accuracy measures of GPS readings can be shown by the probability distributions of error *(9)*, which involves identifying the location with a band of probable variations based on the error.

The uncertainty model that has evolved can be defined as a stochastic process capable of generating a population of distorted versions of the same reality (such as a map), with each version being a sample from the same population. The traditional Gaussian model, where the mean of the population estimates the true value and the standard deviation is a measure of variation in the observations, is one approach to describing error. Nevertheless, the Gaussian model is global in nature and says nothing about the processes by which error may be accumulated *(29).*

McGranaghan *(92)* discussed various techniques for displaying the uncertainty of the location of a spatial feature, the distinctness of boundaries, and the relative size of the features. Beard et al. *(93)* presented methods of using exploratory data analysis in a spatial context where quantitative methods are not available. These methods illustrate the reliability in the classification of features based on the size of a feature. Hansen *(20)* noted that defining these spatial characteristics forms a basis from which one can begin to model error. This approach includes identifying the type of error distribution and methods of estimation for a spatial characteristic of a feature. Measurement-based systems develop error estimates derived from a normal distribution of error for repeated measurements and redundant measurements, which permits correction for distortions introduced by map projections, the differences in actual elevation, and the spheroid surface to length and area measurements of survey data *(94, 95).* Other spatial characteristics may require the use of another error distribution.

2.5.6.1 Conceptual Approach to Presentation of Error

One of the primary objectives of this project is to develop an approach for presenting positional data error. The concept is to introduce probability zones around features (e.g., an event) to describe the uncertainty of their locations. The calculation of a probability zone is based on the measurement error as well as the resolution of the applied measurement system or the embedded reference system. The goal of the probabilistic approach is to assign n-dimensional probability zones immediately surrounding every n-dimensional *measured* feature location. The size of each of these n-dimensional zones depends directly on two components: (1) the uncertainty arising from imprecise measurements expressed in imprecision measures or derived inaccuracy values (e.g., ±5m) and (2) a user-selected probability threshold (e.g., 95 percent) that the *true* feature location is to be found within this probabilistic space.

The basic idea is to transform, for example, an accuracy value of (x meters) into a statistical probability that a point can be found in its neighborhood based on the normalized normal distribution and the present resolution. Subsequently, each feature or event is assigned a probability space. The probability zones are confidence intervals indicating the confidence or the probability that a specific measured event is actually located within a given area. For example, a surveyed point location is known with a spatial accuracy measure of ±1 meter. Thus, one can assume, with a probability of about 68 percent that the actual location of this point is within a circle of radius 1 meter. Assuming, however, that one would like to be 95 percent confident that the point is located within a specified area one would have to use a circular area with radius of 1.96 meters according to the normal distribution. Applying this principle, one can now translate error measures of ±x units into probability zones. Additionally, this allows one to overlay two such generated probability zones to gain information on the possibility (given in percentages of probability) that two locations are congruent.

Probability zones can either be binary and continuous. These are both indicators for the probability that a specific GIS feature is located within an estimated probabilistic area. These are described below.

Binary Zone. In a binary probability zone, the GIS feature of interest is a subset of a single unit of the measurement system. In this case, the resolution of the measurement system dictates the resulting uncertainty values. The binary zone is a rather simple approach. The basic idea is to determine the probability that a sub area is selected. The term binary is assigned because no distinction is made as to what degree the sub area is selected. The possible result set is: {selected, not selected}. Furthermore, this approach results in a single value for the entire measurement unit. Hence, it is termed the binary zone. Consider the following scenario:

> Assume that a feature (e.g., a parking lot measuring 10 m by 10 m) is smaller than the atomic unit of the measurement system (e.g., a 30 m by 30 m pixel) and is positioned somewhere within a specific unit (e.g., pixel x, x). If one chooses to walk to the real location of this unit (e.g., the 30 m by 30 m area) 100 times, how many times would one actually stand on the parking lot? The result can be obtained by simply calculating the percentage of the sub-area in comparison to the unit. Thus, it can be derived that one would stand approximately 11 out of the 100 times on the parking lot. In other words, the percentage indicates the probability that the subset of interest is selected (e.g., ~11 percent). It is a measure for the degree of uncertainty that one actually selects the desired sub area. Or one could state that one can select the desired sub area with a probability or with a certainty of about 11 percent.

The above example illustrates the case of 2D raster-based imagery. The measurement system also could be linear. For example, the police record a crash location based on the nearest milepost. In this case the feature extension would be about 75 yards; however, the resolution of the measurement system is based on 1-mile segments (or 1,028 yards). Thus, if one would visit the location based on the nearest milepost, the probability of standing somewhere within the 75 yards of the actual crash would be only 7.3 percent.

This approach further assumes that the actual value of the sub area is known (e.g., one knows the size of the parking lot). This approach requires some sort of external information source or the implementation of one of the above-mentioned approaches such as discussed in Ehlschlaeger *(96).* An extension to the binary zone can be applied for unions of multiple atomic values. In this case, one would assign a new atomic value equal to the sum of all previous atomic values in the union.

Continuous Zones. Continuous probability zones, on the other hand, are mostly independent of the resolution of the measurement system. In the continuous case, the decisive factor is the measurement error, or to be more precise, the resulting variance associated with a measured location. Continuous zones can be calculated for any geometric object embedded in n-dimensional space. Shi *(97)* provided generic derivations for the geometric objects of a point, line segment, and line. The model assumes that measured locations (X_n) of an n-dimensional feature are based on a normal distribution with variance (σ^2) around the true location (μ_n). Furthermore, Shi *(97)* describes the calculation of confidence intervals based on the confidence level itself, the geometric feature (e.g., point and line), and the n-dimensionality of space. The confidence intervals are based on the χ^2-distribution, where the probability that the measured point location is within the tabulated distance of the true point location can be tested.

The approach used in this project makes two adjustments to the general approach discussed by Shi *(97).* First, it is assumed that equality exists among the n variances associated with each of the cardinal directions, resulting in Equations 1 and 2:

$$\sigma^2 = \sigma_x^2 = \sigma_y^2 = \ldots \sigma_n^2 \qquad (1)$$

$$\textit{Measured location} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \sim N_n\left[\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ & & \\ 0 & 0 & \sigma^2 \end{bmatrix}\right] \qquad (2)$$

Usually, a single accuracy value (i.e., σ^2) is provided, if at all. The requirement of explicit specifications for the variances in all cardinal directions is the ideal scenario; however, it is unlikely to be found in practical applications. Equation 3 is a general descriptor for the spatial extent of the probability zone around an n-dimensional point location:

$$\begin{aligned} &P(\textit{True location} \in \textit{Confidence space}) > \gamma \\ &X_i - b_{ni} \le x \le X_i + b_{ni} \\ &\text{with } b_{ni} = \sigma_i \sqrt{\chi^2_{1;(n-1+\gamma)/n}} \end{aligned} \qquad (3)$$

Similarly, for a link node system we can derive Equation 4

$$\begin{aligned} \textit{Measured location} &= \begin{bmatrix} X_{1r} \\ X_{2r} \\ \vdots \\ X_{nr} \end{bmatrix} \\ &\sim N_n\left[\begin{bmatrix} (1-r)\mu_{11} + r\mu_{12} \\ (1-r)\mu_{21} + r\mu_{22} \\ \vdots \\ (1-r)\mu_{n1} + r\mu_{n2} \end{bmatrix},\right. \\ &\left.((1-r)^2 + r^2)\begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ & & \\ 0 & 0 & \sigma^2 \end{bmatrix}\right] \end{aligned} \qquad (4)$$

with an interval of $X_i - b_{ni} \le x \le X_i + b_{ni}$

with $b_{ni} = \sigma_i \sqrt{\chi^2_{2;(n-1+\gamma)/n} * ((1-r)^2 + r^2)}$

In the second modification, several layers of probability zones are generated, rather than a single zone, allowing for a more detailed representation of the validity for the subsequent discussion on the combination of two or more features. For any GIS feature in an n-dimensional space, the probability zones can be calculated based on preset confidence levels, for example, $PZ_{.75}$, for the 75 percent probability zone, to $PZ_{.99}$, for the 99 percent probability zone. Each of the probability zones covers the continuous space immediately adjacent to its neighboring zones. This principle is best explained by using an example. In the case of an n-dimensional point location, one can calculate the probability zones in the following manner: for example, $PZ_{.75}$: $dPZ_{.75} = \pm\sigma \cdot \sqrt{\chi^2_{1;(n-0.25)/n}}$, which indicates that the true point location lies, with a probability of 75 percent, within the zone outlined by the shape formed at a distance $dPZ_{.75}$. For the 2D scenario, this would result in a square with dimensions $dPZ_{.75}$ by $dPZ_{.75}$ with the point feature located at the intersection of the diagonals of the square.

The distance dPZ from the point in each of the cardinal directions x . . . n and consequently the intervals for the probability zones PZ are as follows:

$$-dPZ_{.75} \quad \text{to} \quad +\left(dPZ_{.75} = \sigma \cdot \sqrt{\chi^2_{1;(n-0.25)/n}}\right)$$

$$-dPZ_{.80} \quad \text{to} \quad -dPZ_{.75} \quad \text{and}$$
$$+dPZ_{.75} \quad \text{to} \quad +\left(dPZ_{.80} = \sigma \cdot \sqrt{\chi^2_{1;(n-0.2)/n}}\right)$$

$$-dPZ_{.85} \quad \text{to} \quad -dPZ_{.80} \quad \text{and}$$
$$+dPZ_{.80} \quad \text{to} \quad +dPZ_{.85}$$

$$-dPZ_{.90} \quad \text{to} \quad -dPZ_{.85} \quad \text{and}$$
$$+dPZ_{.85} \quad \text{to} \quad +dPZ_{.90}$$

$$-dPZ_{.95} \quad \text{to} \quad -dPZ_{.90} \quad \text{and}$$
$$+dPZ_{.90} \quad \text{to} \quad +dPZ_{.95}$$

$$-dPZ_{.999} \quad \text{to} \quad -dPZ_{.95} \quad \text{and}$$
$$+dPZ_{.95} \quad \text{to} \quad +\left(dPZ_{.99} = \sigma \cdot \sqrt{\chi^2_{1;(n-0.01)/n}}\right)$$

Figure 2-14 depicts an example of a 1D point feature. The left side shows a single probability zone at the 75 percent confidence interval; the right side shows multiple probability zones according to the previous example ($dPZ_{.75}$ to $dPZ_{.99}$). As noted in Figure 2-14, the width or radius of a probability zone increases as the distance from the measured location increases, keeping in mind that the gained probability increase is constant (with the exception of the last interval where it is decreased). This makes the gain of additional confidence at higher confidence levels rather costly because of exponentially increasing the borders of the area of uncertainty.

Subdividing the probability zones in such a way helps to describe the different stages of confidence levels. Another advantage of this procedure is the more detailed gain of confidence per unit (e.g., linear distance or square units for the 2D case) information, which is desired as outlined in the subsequent discussion on the combination of two or more features. Table 2-7 illustrates a comparison of gained confidence

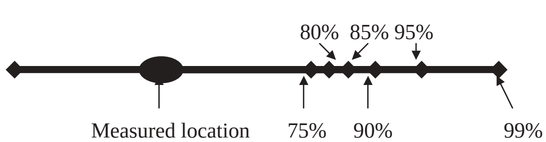


Figure 2-14. Probability zones of a one-dimensional point feature.

versus units added to the uncertainty interval for a point feature. In Table 2-7, for the 2D point feature, a 75 percent confidence interval is equal to an area of uncertainty of 1.32 σ square units. The increase from a 95 percent to a 99 percent confidence interval is more costly, i.e., a gain of 4 percent confidence costs an additional 2.97 σ square units of uncertainty area.

2.5.6.2 *Combination of Two or More Features*

This section discusses the combination of two or more features resulting in an estimate for the probability of congruency. In other words, this approach calculates the probability that, for example, two points are identical or that two lines intersect. This section introduces the principle using the example of a test of congruency for two 2D point features.

Figure 2-15 illustrates the two measured point locations (P1 and P2) along with their confidence intervals. For illustrative purposes, the number of probability zones is reduced to two for each of the point locations. The two point locations have associated standard deviations of σ_1 and σ_2 (with $\sigma_2 < \sigma_1$), respectively. The inner probability zone is a $PZ_{.75}$ and the outer one a $PZ_{.95}$. For simplicity, both probability zones of Point 2 are located completely within one zone of Point 1.

To calculate the probability that the two points are congruent, one needs to calculate the probability that the true point locations of Point 1 and Point 2, respectively, are in Area A and Area B (see shaded areas in Figure 2-15). First, calculate the probability that the true location of P2 is within

TABLE 2-7 Confidence versus units added to the uncertainty interval for a point feature

Probability Zone	One-Dimensional Point	Two-Dimensional Point
$PZ_{.75}$	2.30 σ	1.32 σ
$PZ_{.80}$	0.26 σ	0.32 σ
$PZ_{.85}$	0.32 σ	0.43 σ
$PZ_{.90}$	0.41 σ	0.64 σ
$PZ_{.95}$	0.63 σ	1.13 σ
$PZ_{.99}$	1.23 σ	2.79 σ

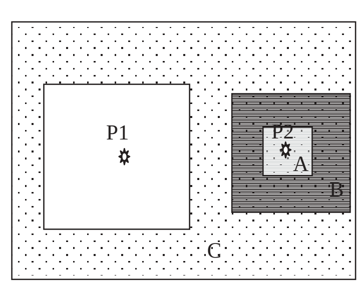


Figure 2-15. Two point locations along with their confidence intervals.

Area A or B. The probability that P2 is in Area A is $P2_A$ = 75% and that it is in Area B is $P2_B$ = 95% – 75% = 20%. For Point P1, calculate the portion of Area C (there is a 20 percent probability that P1 is located in C) covered by Area A and Area B. The probability that the true location of Point P1 is within area A can be derived as $P1_A = C/A \cdot 20\%$. Similarly, the probability that Point P1 is within Area B can be surmised as $P1_B = C/B \cdot 20\%$. Having derived the probabilities that each of the two point locations is in Area A or in Area B, the probability that both events occur can be calculated. The probability that both true point locations (i.e., P1 and P2) are positioned within Area A is $P_A(P1 \cap P2) = P1_A \cdot P2_A$. Similarly, for Area B, $P_B(P1 \cap P2) = P1_B \cdot P2_B$. The probability that Point P1 and Point P2 are congruent can be calculated as $P(P_A \cup P_B) = P_A + P_B$.

2.5.7 Prototype Data Error Model

This section describes the prototype error model and illustrates its application. The prototype model essentially represents the uncertainty object. Fundamentally, this is a metadata reporting approach, which contains uncertainty information about the event and its history. The prototype model is designed as a stand-alone product where all the required input information has to be provided by the user. For a full implementation into existing transportation databases this information can be retrieved from existing data. The challenge in developing a data error model stems from two facts: there are a multitude of data sources (e.g., photogrammetry and distance measuring instruments) and multiple reference systems in multiple dimensions.

These facts were considered in developing the prototype error model to meet the two main goals of estimating the uncertainty (1) associated with data collection, network, and referenced features; and (2) in combining different data sources (e.g., 2D network with 1D event digitized road location with crash site). The first task is to find a common descriptor for positional uncertainty inherent in the spatial data specific to transportation features.

2.5.7.1 Input Requirements

This subsection describes the input requirements for the prototype model. The dependencies of a linear or point feature, shown previously in Figure 2-12, are implemented in the prototype model and require user input for estimates of the standard deviations of each component. For example, the measurement method of recording a crash site on a highway could be by a handheld GPS or by measuring the distance to the nearest milepost via the odometer in a police car. Each method, however, has a known standard deviation, which is used to estimate the associated probability zones. In the latter case, network errors of the milepost system are also a required input for the prototype.

Figure 2-16 shows the required inputs for the uncertainty object and the relationship between the measured object and the sources of error. As noted above, the prototype is essentially an implementation of the uncertainty object itself. Instead of retrieving the necessary input (right side of Figure 2-16) from the system, the user is asked to provide these data.

To avoid crowding the presentation with too much information, each point feature and each network have an individually associated raster. For example, if the input consists of point features and one network of 15 link nodes, the uncertainties of this system are stored in two individual layers (i.e., one for the point feature and one for the network). Each intersection chosen adds raster maps.

Functions within the uncertainty object are used to calculate uncertainties and their propagation and then store individual uncertainties. Specifically, the prototype requires the following inputs:

- Type of feature (i.e., point or line).
- Spatial locations of events and link nodes (i.e., coordinates of the events, such as line event and start- and end-coordinates). In a situation where two lines intersect, for example, the coordinates of the start and end points of each line will be required.
- Estimated imprecision of relative or absolute network errors (i.e., estimates of level of precision of how the event was measured).
- Resolution of measurement system (i.e., resolution of the measuring devices and referencing system used).
- Precision of event measurements (e.g., estimated precision in locating a crash site).
- Extent or description of the event (e.g., crash site versus business sign).

Depending on the feature or event of interest, the error value would reflect the estimated precision of the network, resolution of the measurement system, and/or systematic network error. For example, for a 2D line event, the error value depicts the estimated imprecision and resolution associated with the development of the line or network (or base map) from which

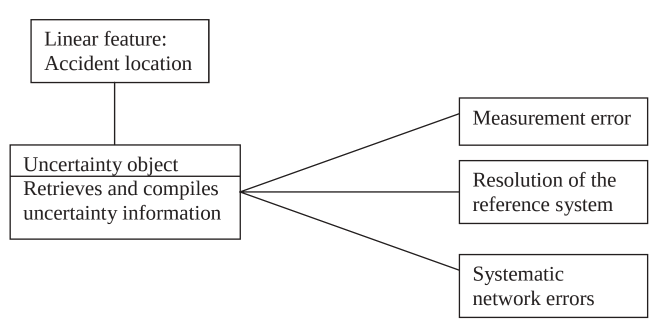


Figure 2-16. Detailed input requirements.

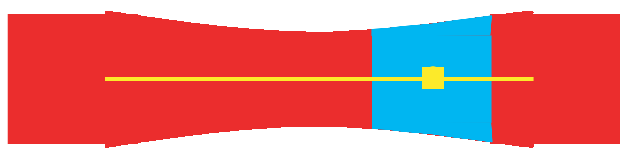
the line was derived. In such a case, the errors associated with digitizing the map or converting aerial photographs into maps will be the input into the error program. On the other hand, for a point event, the error value reflects the precision of the measuring instrument or method. The extent of the event is just a descriptor for the event under consideration.

As discussed in Section 2.5.4, in order to visualize a 1D error in 2D, information on the linear distance from a known point is required. The delta method (where the error in the direction of the line as well as perpendicular to the line) is used. The error value in the direction of the line can be calculated using the chi-squared table, while the width of the line will be used to represent the width or the error.

To illustrate the application of the prototype error model, consider a simple example of crash location data. In this example, the prototype model was used to estimate and display the combined errors associated with recording a crash site located on a highway segment. The crash site was recorded by referencing the nearest milepost (i.e., the measurement method could be handheld GPS or measuring the distance to the nearest milepost via the odometer in a police car) and the road network was digitized from aerial photographs, which requires a transformation from one system to the other. The uncertainty of the linear feature is transformed into 2D space. In applying the model, it is assumed that each measuring method has a known standard deviation of measurement error, which is used to estimate the associated probability zones. The model was used to estimate the combined errors from line and point events.

2.5.7.2 Calculations

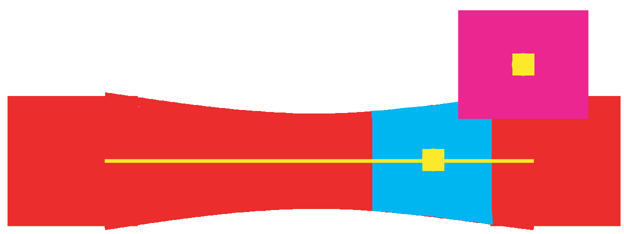
First, the probability zones associated with each feature are calculated and stored in separate geo-referenced Arc ASCII rasters. The zones are based on assumed uniformly distributed intervals of the χ^2-distribution. The zones represented in the prototype are 0–75, 75–85, 85–90, 90–95, and 95-90 percent significance intervals of the associated χ^2-distributions. Thus, each spatial location (added by the user) has five discrete buffers around it, where the closest represents 75 percent probability that the true point location is within the buffer, and the second through the fifth each represents an additional 5 percent. Each pixel in those three zones receives a proportional probability that the location is exactly in that pixel according to the assumption of uniformity within each zone. For example, the uncertainty of a crash site that was linearly referenced with an independently produced 2D representation of the same road can now be combined. Assume that the crash site was recorded by referencing the nearest milepost and that the road



Key:

- Yellow line = road (the true location of the line event)
- Yellow circle = linear referenced point event (crash site)
- Red rectangles (at the ends) = 2D uncertainty of the road centerline
- Red middle portion = extent to which one can visualize error in 2D (this is the result of transformation from 1D to 2D)
- Blue trapezoid = linear referenced error visualized in 2D

Figure 2-17. Combination of linear feature and 2D network.



Key:

- Yellow circle in purple square = location of business event (e.g., sign post)
- Purple square = business event in 2D (event that is not referenced to the road)
- Intersection between the blue and pink areas = the chance that the 2D event is actually on the road (i.e., the data quality of a 2D event with a linear referenced event)

Figure 2-18. Intersection of linear event with independent 2D event.

network was digitized from aerial photographs. In the prototype model, the uncertainty of the linear feature is transformed into 2D space.

Second, the program combines the different uncertainty zones. Following from the example under consideration, one can now combine the uncertainty of a crash site that was linearly referenced with an independently produced 2D representation of the same road. Figure 2-17 shows a sketch of the outcome after the two features are combined. In this specific example, the 2D uncertainty zone is based on the following: (1) the width is based on the linear feature's network error, resolution, and measurement error; and (2) the height is based on the digitization error of the 2D network.

The next step is to calculate the probability that the linear event feature (which now has an associated 2D uncertainty) intersects with a 2D point location. As can be seen in Figure 2-18, it is not necessary to assume that this point location has to be related to the 2D road network.

Based on the individually stored uncertainty layers, one can now calculate the intersection of these layers by intersecting the probability of each pixel in each associated raster layer. The result is stored in a new Arc ASCII raster.

The prototype error model is encapsulated into a software program called GISError developed in Visual Basic programming language with graphic user interfaces (GUI). The GUI facilitates data input and visualization of outputs. A user guide for the application of the prototype error model is included in Appendix A of this report. The program has a feature that allows results from the error analysis to be exported to other GIS application software. The prototype model was applied to case study data obtained from the Ohio DOT. The results of the case study are presented in the next chapter.

CHAPTER 3

INTERPRETATION, APPRAISAL, AND APPLICATIONS

3.1 DISCUSSION OF CASE STUDIES

This section describes the results of applying the prototype model to case study data. A comprehensive set of tests was conducted on the prototype data error model software (i.e., GISError). The GISError program was developed to help analyze and visualize the results of data errors and display how errors of different data sources affect each other. The program outputs the results of the error analysis and shows the confidence intervals and buffers around data points and lines. The program computes the probability of the intersection of two features or data sources to determine whether they are compatible and should be used together. The following data elements could be entered into the program by the user:

- Points: defined by a Cartesian (x, y) coordinate, such as the Easting and Northing in a state plane system.
- Lines: defined by two points (i.e. the starting and end points, which also are defined in the Cartesian coordinate frame).
- Points on a line: defined by a distance from the starting point. These correspond to the linear reference system used for transportation applications, where the line represents the road centerline and the point is an event on the centerline, such as a crash. The distance is the mileage value of the event from the beginning of the route or the nearest intersection.
- Error values: The user also can enter an error value for both points and lines. This value represents the mean error (1 sigma) of the data element. For example, a point whose position was determined with a GPS receiver may have a sigma of 3 feet. The error value of a line shows the uncertainty of the location of the whole line as a buffer in the specified coordinate frame.

When two data elements and their corresponding tolerances (error values) are entered, the program generates a graphical display of the situation showing error buffers of various probabilities at different significance levels and presented in different colors. The program also displays the analytical results of this computation by listing the minimum and maximum values of the confidence radius for specific probabilities (i.e., 75%, 80%, 85%, 90%, 95%, and 99%). Finally, the program displays the Intersection Probability, which is a measure of the likelihood of the two data elements intersecting. This number is important to determine which datasets are actually compatible and which datasets should not be combined.

3.1.1 Case Study Datasets

Three real-life datasets were used to test the prototype data error model. Each dataset consisted of a few data elements, which the user wishes to merge with any of the other data elements. All possible and reasonable combinations of the data elements from different datasets were evaluated. The results, presented in this section, form the basis of the guidelines recommended in a later section of this report.

The first dataset was provided by the Ohio Department of Transportation (ODOT). This dataset contained three data elements, which are used by ODOT:

- Road Centerlines were digitized by ODOT for USGS 1 : 24,000 quad sheets. These centerlines have been edited and annotated by ODOT. They form a statewide network. A linear reference system is defined along these centerlines. The mean error of this data element is approximately 50 feet, which is mostly due to digitizing errors and map generalization.
- Crash Data are represented by mileposts along the road centerlines. The mean error of the crash locations is specified to 52 feet by ODOT, which corresponds to $^1/_{100}$ of a mile. It is unclear how this dataset was collected. Probably it comes from the crash reports submitted by police officers, which are either distances measured to the nearest milepost or, in urban areas, crash locations linked to a physical street address. The latter case would actually lead to a much lower accuracy of this data element.
- Video Log Data are captured with a Mandli video-logging van. This system uses real-time, differential GPS with an estimated accuracy of 12 feet. This data element precisely traces the roadways and is usually less than 2 years old. However, it consists of a collection of points where video log images were captured. Therefore, an individual point does not correspond to a physical location on the roadway. The GPS trace clearly outlines the lane in which the video-logging van was driven.

The second dataset comes from TRANSMAP Corporation. It was collected with its ON-SIGHT mobile mapping van, which features digital stereo cameras, kinematic GPS, and an inertial navigation system. Roadway features and road centerlines are extracted from the stereo images. This dataset was captured in the same area as the ODOT dataset. Therefore, a direct comparison is possible. This dataset contains the following three data elements:

- A Road Centerline measured at the actual location of the center of the road. Typically, the road centerline is visible in the images and defined as the pavement striping in the middle of the roadway or as the center of a middle lane. This centerline network also contains a linear reference system, and, as intersections are directly related to the intersections in the ODOT centerline dataset, they can be directly compared. The mean error of TRANSMAP's road centerline network is 3 feet.
- Roadway Features were extracted from the same set of stereo images. They include signs, pavement markings, light poles, signals, guardrails, and many more. Typically, 35 to 50 different feature types are inventoried from the digital stereo images to create an accurate and complete infrastructure inventory along the roadway. These features are measured in the same stereo images as the centerlines, and, therefore, have the same mean error of 3 feet.
- GPS Image Locations are the third data element of the TRANSMAP dataset. These points represent the locations where images were captured along the roadway. Although these points are called GPS Image Locations, they are determined by integrating kinematic GPS with inertial navigation data.

The final dataset comes from GDT. GDT is a commercial company providing street maps all over the country. GDT's data are based on TIGER files, a set of digitized road centerlines maintained by the Census Bureau and updated and corrected by GDT, especially in urban areas. The dataset, which is about 40-feet accurate, also represents a network of centerlines, similar to ODOT's or TRANSMAP's road centerlines. However, it does not have an attached linear reference system. On the other hand, GDT's centerlines consist of street segments that contain highly accurate street address ranges. This information is critically important for capturing certain types of data, such as traffic crashes, if no other means of measurement are available to the police officer recording the crash.

3.1.2 Methodology

The case study attempted to test as many different combinations of data elements as practical. Table 3-1 shows the matrix of test combinations with the ODOT data elements on the horizontal axis and the TRANSMAP and GDT data elements on the vertical. All combinations of data elements shown in this matrix, with the exception of a comparison of

TABLE 3-1 Case study combinations matrix

Datasets	ODOT		
	Centerline - 50ft	Crash Data - 52ft	GPS Video Log - 12ft
TRANSMAP Centerline – 3ft	B-1, B-2, C-1,C-2	D-1	E-1
TRANSMAP Feature Points – 3ft	H-1	K-1, K-2	F-1, F-2
TRANSMAP GPS Image Loc – 3ft	G-1	I-1, I-2	J-1, J-2
GDT Centerline – 40ft		L-1	A-1
ODOT Centerline – 50ft		M-1, N-1	

NOTES:
Each of the case study examples (A to N) represents a combination of two different data elements provided by TRANSMAP or ODOT as described below:

A: GDT Road Centerline – ODOT GPS Point from video-logging van
B: TRANSMAP Centerline – ODOT Centerline
C: TRANSMAP Centerline – ODOT Centerline
D: TRANSMAP Centerline – ODOT Crash Data
E: TRANSMAP Centerline – ODOT GPS Point from video-logging van
F: TRANSMAP Feature Points – ODOT GPS Point from video-logging van
G: TRANSMAP GPS Image Location – ODOT Centerlines
H: TRANSMAP Feature Points – ODOT Centerlines
I: TRANSMAP GPS Image Location – ODOT Crash Data
J: TRANSMAP GPS Image Location – ODOT GPS Point from video-logging van
K: TRANSMAP Feature Points – ODOT Crash Data
L: GDT Road Centerline – ODOT Crash Data
M: ODOT Road Centerline – ODOT Crash Data
N: ODOT Road Centerline – ODOT Crash Data

the GDT centerlines with the ODOT centerlines, were tested. As these two data elements are based on the same source (i.e., USGS 1 : 24,000 quads) and are of similar accuracy, it was concluded that typically an agency would use either of these datasets, but not both of them together.

For each combination of two data elements, four different scenarios were created and the graphical results generated using the software. These results show the following important information for each of the combinations tested:

- The Data Input screen of the GISError program shows the information entered into the test program (e.g., the exact coordinate values for both data elements and the error values).
- The Error of GIS Feature screen shows analytical results and lists the errors of both the first and the second feature for different probabilities. The results represent the minimum and maximum error radius for a certain probability. This window also shows the intersection probability of the two data elements.
- The graphic of the Error of the GIS Feature displays as concentric circles around a point or a buffer that is rounded at the ends of a line segment. Both data elements are shown in the same window, so the user can make an empirical decision as to whether or not the two features would actually intersect.
- The corresponding datasets in a GIS format (the last window) shows the two datasets in the ArcView GIS and a screen shot of the area used for the analysis. This gives the user a good understanding of how these datasets and data elements look in the real world.

The case study results are discussed in the next section.

3.1.3 Discussion of Case Study Results

This section describes the results of the different case studies. Each case study represents a combination of two different data elements provided by ODOT, TRANSMAP or GDT. This section provides descriptive interpretation of the results and includes samples of the accompanying graphics. The graphics of the remaining case study runs are presented in Appendix B of this report.

The GISError program shows the buffers of the two datasets on top of each other. The second feature (dataset) is always shown on top of the first feature. The intersection probability shows the probability with which the second dataset will fall within the confidence buffer of the first dataset. For example, if the first dataset is inaccurate and has a large error, and the second dataset is very accurate and has a small error then, if the two data elements are separated by less than two times the error of the first dataset, it is very likely that the second one will be within the first one's buffer. However, if the order of the datasets were reversed, which means that the accurate one comes first, then it is unlikely that the second dataset will be within the buffer of the first.

Example A: GDT Road Centerline—ODOT GPS Point from Video-Logging Van

This example illustrates the intersection of a line feature and a point feature (i.e., GDT centerline [error, 40 feet] with ODOT centerline from video-logging van [error, 12 feet]). The graphic clearly shows the better detail and higher resolution of the GPS points. They are captured at distances of around 50 feet, while the GDT centerline consists of straight road segments that are at least 10 times as long. This by itself leads to a significant dilution of accuracy. The intersection probability of the two datasets is 89 percent, which means that a GPS location of an image captured with the video-logging van can be correctly associated to a road segment. Therefore, the video-logging data, which are captured every other year, could be used to update GDT's or ODOT's centerlines to achieve better accuracy overall.

Example B: TRANSMAP Intersection—ODOT Intersection

These examples illustrate the intersection of two point features (i.e., intersections from ODOT dataset [error, 50 feet] and TRANSMAP dataset [error, 3 feet]). The graphic shows a constant (systematic) offset between the two datasets. There is a significant difference in accuracy between the two datasets (the error of ODOT's centerline is almost 20 times larger than TRANSMAP's). The errors can be seen at intersections, which were used for comparison in this example. B-1 shows a 99 percent intersection probability, which means that TRANSMAP's intersection falls within the range of the ODOT intersection with a high probability. On the other hand, it is rather unlikely that ODOT's point is within TRANSMAP's buffer (B-2 intersection probability of 0.33 percent).

Example C: TRANSMAP Centerline—ODOT Centerline

While example B compared intersections, this example illustrates the intersection of two line features (i.e., ODOT centerline [error, 50 feet] with TRANSMAP centerline from video-logging van [error, 3 feet]). The map graphic in Figure 3-1 shows the significant offset between the two datasets and the lack of resolution of the ODOT dataset. TRANSMAP's centerline is created by a very dense sequence of points (every 25 to 50 feet), while ODOT's centerline consists of 100- to 500-foot-long line segments. The probability of intersection between the two datasets is generally very low (Figure 3-1).

Feature 1
ODOT Road Section

Feature 2
TRANSMAP Road Section

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,270,000
y-shift: -1,170,000

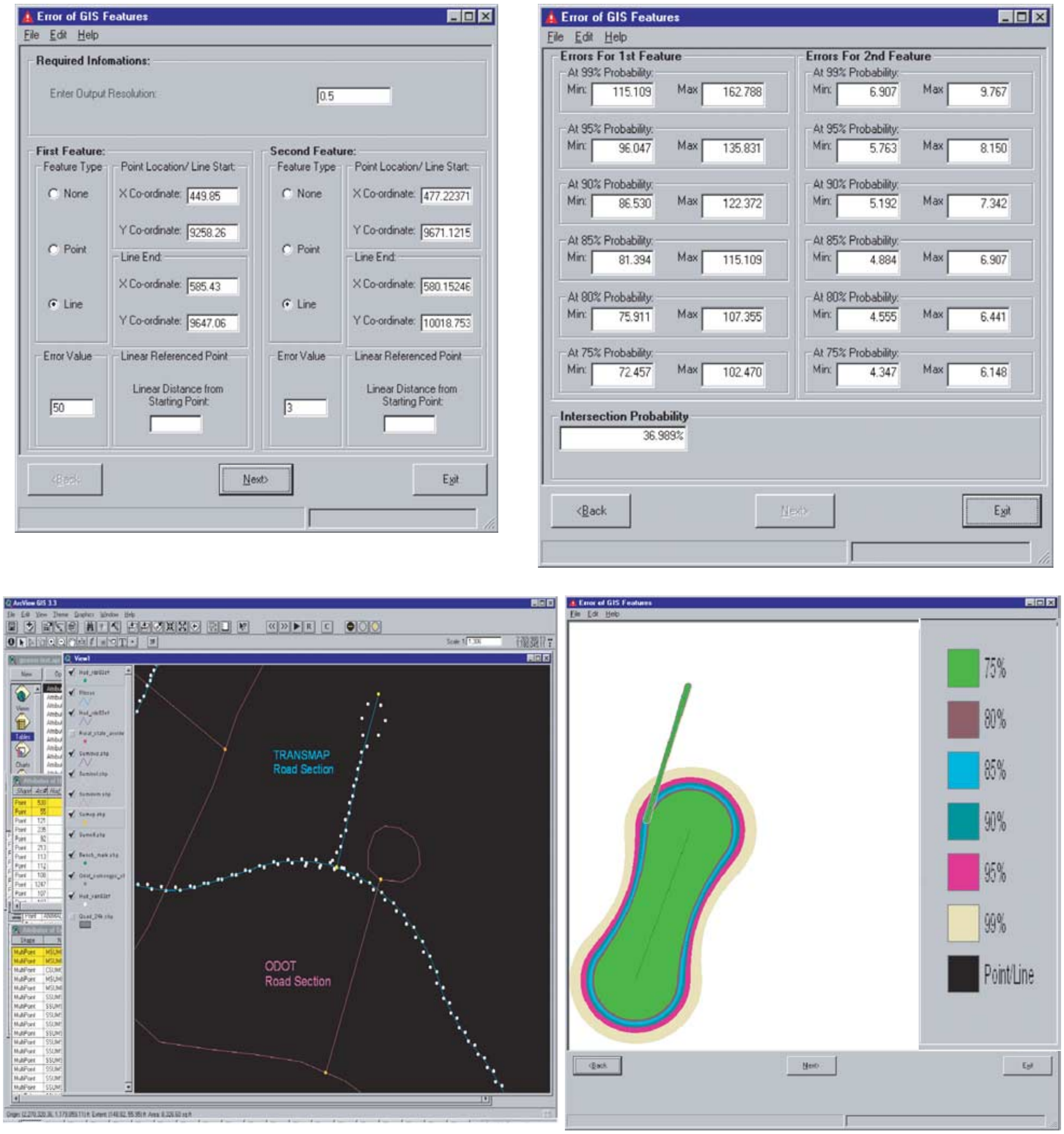


Figure 3-1. Example of intersection of two line features.

Example D: TRANSMAP Centerline—ODOT Crash Data

This example illustrates the intersection of a point feature with a line feature (i.e., crash location from ODOT dataset [error, 52 feet] with TRANSMAP centerline from video-logging van [error, 2 feet]). The ODOT crash data were plotted on top of the ODOT centerline using their known milepost value (Figure 3-2). Because of the inherent error of the ODOT centerline, they are significantly offset from TRANSMAP's centerline; therefore, the intersection probability is low. It would be possible to display the crash data on top of the TRANSMAP centerlines, if corresponding linear reference systems can be defined. This would reduce the uncertainty to only one dimension (along the roadway) and also would make the crash data usable in a more accurate environment (Figure 3-2).

Example E: TRANSMAP Centerline—ODOT GPS Point from Video-Logging Van

This example also illustrates the intersection of a line feature and a point feature (i.e., TRANSMAP centerline [error, 3 feet] with ODOT GPS points from video-logging van [error, 12 feet]). These two datasets are of much higher accuracy than ODOT's centerlines and crash data. They fit together well, if interpreted visually. However, the intersection probability is only 25 percent, because they are separated by more than their mean error (one sigma). In general, these datasets are of similar resolution and show similar detail.

Example F: TRANSMAP Feature Points—ODOT GPS Point from Video-Logging Van

These examples also illustrate the intersection of two point features (i.e., a GPS point from ODOT's video-logging van with a feature point created from TRANSMAP's stereo-imaging system). The examples, F-1 and F-2, illustrate the effects of turning the sequence of datasets around. While F-1 lists the ODOT GPS point first, which results in an intersection probability of 23 percent, F-2 lists TRANSMAP's point first (which is more accurate than ODOT's), which leads to an intersection probability of only 1.3 percent.

Example G: TRANSMAP GPS Image Location—ODOT Centerlines

This example demonstrates that TRANSMAP's GPS Image Locations will fall within the ODOT centerline buffer with a very high degree of probability (99 percent). This is similar to Example A. The graphic shows that the TRANSMAP GPS Location is fully within the ODOT centerline in the dataset chosen.

Example H: TRANSMAP Feature Points—ODOT Centerlines

This example yielded results similar to those of Example G. In this example, a GPS point located by TRANSMAP's mapping van was compared with a feature located on ODOT's centerlines using the linear reference system. The fit is very good because of the large buffer around ODOT's centerline.

Example I: TRANSMAP GPS Image Location—ODOT Crash Data

This example compares the locations of two coordinates. ODOT's crash data are shown as a location and compared with TRANSMAP's GPS Image location. Examples I-1 and I-2 again show the opposite effects of changing the sequence of datasets.

Example J: TRANSMAP GPS Image Location—ODOT GPS Point from Video-Logging Van

In this example, the differences in buffer sizes are not as large, and therefore, the datasets appear closer together. However, there is still a significant effect of sequencing the data and interpreting the results correctly.

Example K: TRANSMAP Feature Points—ODOT Crash Data

In this example, an ODOT crash location was compared with a TRANSMAP feature point. The error of the ODOT crash point is 20 times larger than the feature point. As ODOT's crashes are tied to their centerlines, the results are similar to those of Example H.

Example L: GDT Road Centerline—ODOT Crash Data

This example compares two datasets of roughly similar accuracy. The intersection probability turns out to be 52 percent, which is caused by a consistent shift between ODOT and GDT data (Figure 3-3).

Example M: ODOT Road Centerline—ODOT Crash Data

This example uses two ODOT datasets with similar uncertainties: the crashes and the road centerlines. They are very likely to intersect (99 percent), as both are defined in the same reference frame (Figure 3-4).

Linear Referenced Point Feature 1
TRANSMAP Road Section
& ODOT Accident Linear Distance

Point Feature 2
ODOT Accident Point

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,250,000
y-shift: -1,190,000

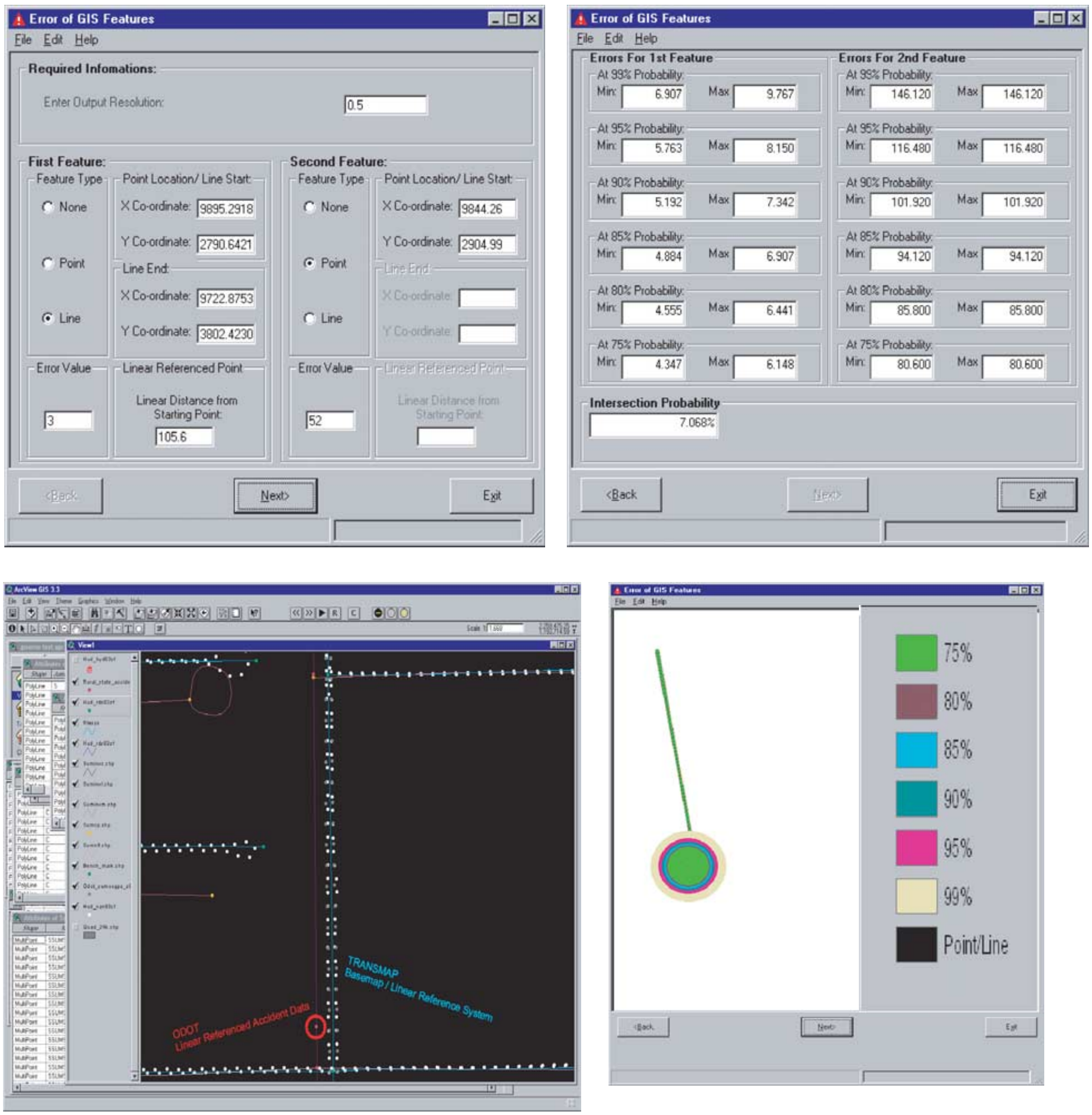


Figure 3-2. Example of intersection of a point and line features.

Point Feature 1
ODOT Accident

Line Feature 2
GDT Road Section and ODOT Accident Linear Distance

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,250,000
y-shift: -1,310,000

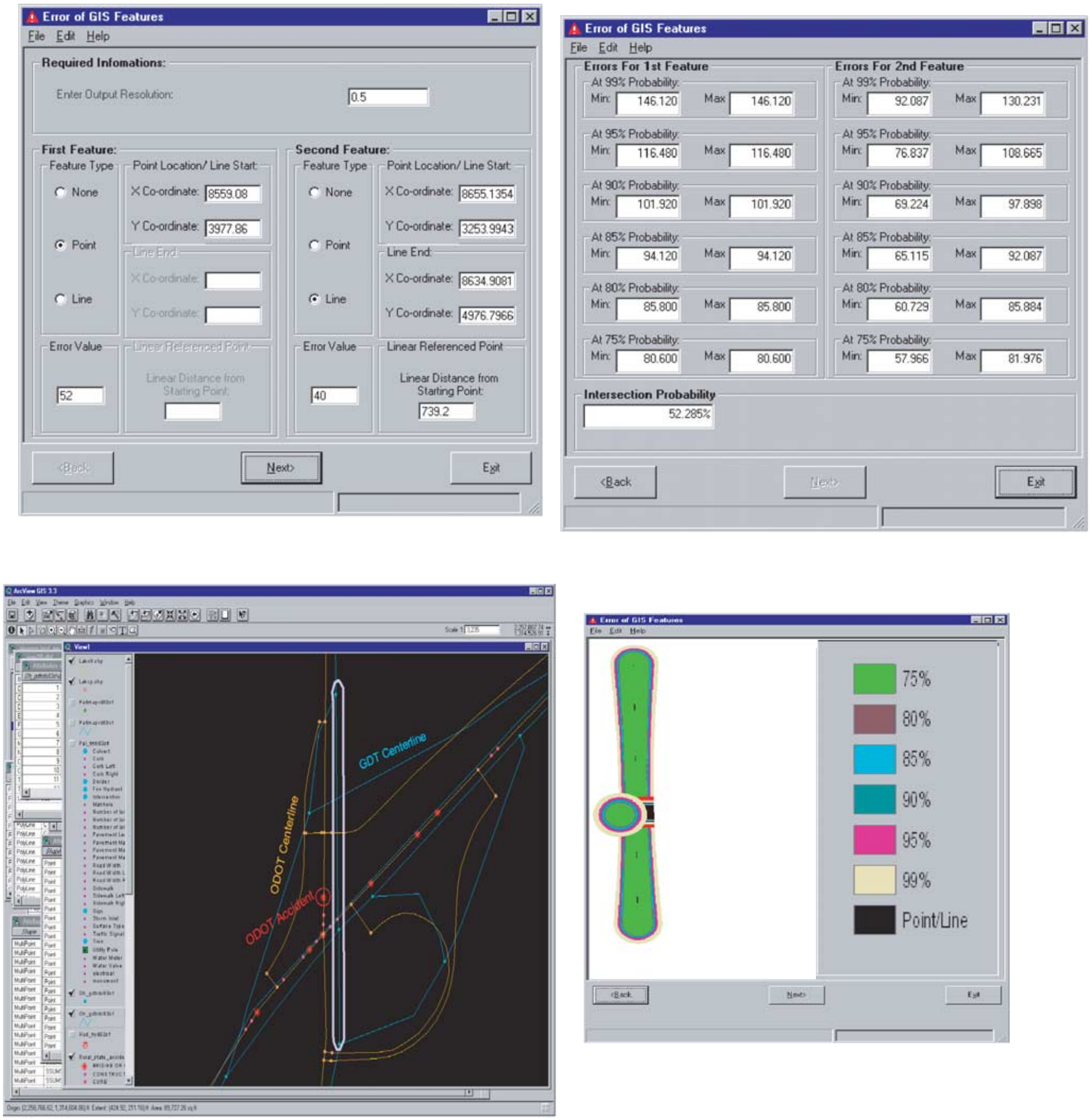


Figure 3-3. Comparison of two data sets with similar accuracy.

Line Feature1
ODOT Road Section

Point Feature2
ODOT Accident

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,300,000
y-shift: -1,340,000

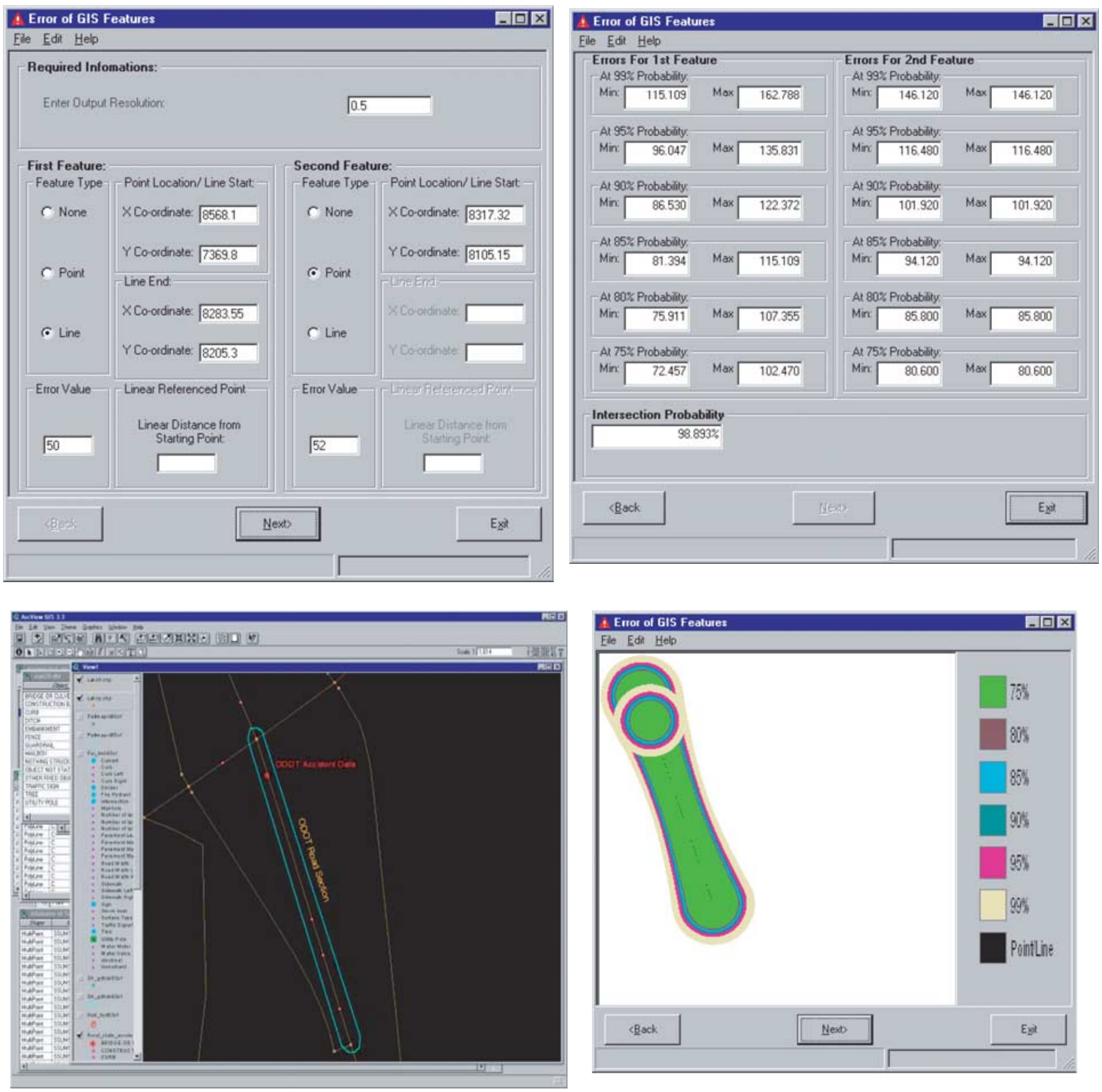


Figure 3-4. Example of crash location defined by milepost.

Example N: ODOT Road Centerline—ODOT Crash Data (Milepost)

This example is similar to Example M except that the crash location (point) was defined by a milepost versus the same crash defined by a coordinate. As shown in the graphics, the locations fit together well and the intersection probability is high (94 percent).

3.1.4 Effect of the Error Model on Applications

The major functional areas of transportation for GIS applications and the appropriate scales were shown in Table 2-2. To demonstrate the applicability of the prototype model to the various groups of transportation applications identified in Table 2-2, the case study examples described previously were mapped to the typical applications. Table 3-2 shows the scales of spatial databases and the types of activities or applications for which they can be used.

Table 2-3 presented a list of transportation applications and their sensitivities to positional errors. Table 3-3 show the case studies associated with practical applications. Table 3-3 relates the effects of data errors to real-life transportation applications. The error model can be used to assess the quality of positional data for all possible transportation-related applications.

3.2 INDICES OF POSITIONAL DATA QUALITY

The prototype error model is designed to help evaluate the margin of error associated with positional datasets and in transforming data from spatial referencing systems. The model is expected to serve as a tool for assessing the quality of positional data. Indices of data quality are probability zones at difference significance levels that provide indications of the level of confidence associated with positional data of different types, collected by different measuring systems and with different levels of accuracy. In assessing the quality of any positional data, the user can quickly assess its quality relative to the intended application. In this way, the user has an idea of the confidence that can be placed in decisions based on the data. This section presents recommendations for incorporating the positional data quality indices in GIS applications.

3.2.1 Indices

The following indices can be used to describe positional data quality for transportation features.

3.2.1.1 Mean Error of a Point

The mean error (one sigma) of a point that is defined by a pair of cartesian coordinates (x, y) represents a 67 percent confidence buffer around this location, which means that the point is inside this buffer with a probability of 67 percent. Although this is a measure commonly used by surveyors and mappers, many practitioners do not understand the exact meaning of mean error. Most users try to specify accuracy as 95 percent (or a similar value in the 90s) of points being inside a buffer. As can be seen by using the GISError program, this value actually corresponds to twice the mean error.

3.2.1.2 Mean Errors of Lines

In some way the mean error of a point also can be used to describe the accuracy of a street centerline network. One can simply assign the mean error to all intersections and vertices along a centerline. However, this may not reflect the complete error of the line, which may additionally be affected by inaccuracies resulting from data capture (i.e., how the line was created). For example, a street centerline can be digitized from an existing map, adding the generalization errors of this map, as well as operator errors that occur during digitizing when following the line on the map. These errors are on top of the coordinate errors affecting points and vertices.

The GISError program offers an accurate and efficient way to visualize the different buffer sizes that are associated with certain probabilities. It is a promising tool for users that specify accuracy as the mean error to see the effect of this error on other datasets. However, this requires access to a computer. Therefore, the use of the program is restricted to the office until a version for handheld devices is implemented. The

TABLE 3-2 Typical application areas and case study examples

Scale of Spatial Database	Precision of Spatial Database (ft)	Typical Activities or Applications	Case Study Examples
1:500,000	830	Statewide Planning	B, C, L, M, N
1:100,000	170	District-level Planning and Facilities Management	G, H, L, M, N
1:12,000 - 1:24,000	30 - 40	Engineering	A, D, I, L, M, N
1:120 - 1:1,200	0.33 - 3	Project-level Activities	E, F, J, K

TABLE 3-3 Relationship between transportation applications of positional data and case study examples

Subject Area	Applications	Sensitivity	Case Study Examples
Safety	Crash Reporting	Medium	D,I,M,N
	Black Spot / Crash-Prone Location Identification	Medium	E,G,H,M,N
	Traffic Safety Investigation	Medium	D,E,M,N
	Rail-Crossing Safety Analysis	Medium	D,E
	Incident Management	High	I,J,K
	911 Emergency Planning and Response	High	E,M,N
Transportation Planning, Impact Analysis, Policy Analysis	Travel Demand Modeling	Low	A,E,L
	Multi-modal Freight Modeling	Low	A,E,L
	Hazardous Materials Routing	Medium	A,E,L
	Traffic Impact Analysis	Medium	A,E,L
Transit and Public Transport Planning and Operations	Transit Planning	Low	A,E,L
	Transit Routing	Medium	E,G
	Handi-transit	Medium	E,G
	Real-time Tracking and Scheduling of Buses	Medium	E,G
Transportation Infrastructure Management and Operations	Location of Facilities (road, highway, airport, port) Inventory	Low	H
	Pavement Management System	Medium	E,G,H
	Asset Management	Medium	H,K
	Operation (congestion, service)	Medium	L
	Corridor Analysis (rail, road, highway)	Low	C,L
	Rail / Highway Information System Management	Low	B,C,L
Transportation Design and Construction Planning	Sources of Construction Materials	Low	A,H,L
	Right of Way	High	B,C,E
	Road Closure and Detour	Medium	D,L
	Construction Information	Low	L
	Field Crew Scheduling	Low	I,K
	Maintenance and Operation	Medium	A,G,H
	- Snow Plowing	Low	A,G,H
	- Garbage Collection	Low	A,G,H
	- Street Sweeping	Low	A,G,H
Intelligent Transportation Systems Applications	Traveler Information System	Medium	D,H,L
	Integrated Highway Information System (IHIS)	Medium	D,H,L
	Integrated Traffic Monitoring System (ITMS)	Medium	E,F,I
	Web-based Road Condition Reporting System	Medium	D,L
	Vehicle Navigation System	High	E,F
	Applications to CVO Regulatory Enforcement Activities	Low	A,G,L,M,N
Freight Analysis and Commercial Vehicle Operations	Fleet Management	Low	A
	Vehicle Tracking, Guidance, Dispatching, and Other Routing Applications	Medium	A,E
	Permitting	Low	L
	Freight Movement	Low	A

alternative approach to use these indices is by following the rules of thumb outlined below.

3.2.2 Rules of Thumb

The following guidelines can be applied in the field without a calculator or computer program to estimate whether two datasets are compatible:

- **Datasets of Similar Accuracy.** If the errors of two data elements (points or lines) are approximately the same (e.g., Examples L, M, and N), the two datasets are typically compatible. The research team's tests verified that the intersection probability is between 50 percent and 99 percent in these cases. This is true only if the two datasets are not affected by a systematic error (e.g., a shift) and are based on the same coordinate frame.
- **One Dataset Is More Accurate than the Other One.** If the error of one dataset is three to five times larger than that of the other dataset (e.g. Examples J-1 and J-2), then Dataset 1 (small error) will usually be within the range of Dataset 2 (large error). However, to ensure that Dataset 2 is within the range of Dataset 1, the two datasets cannot be offset by more than the mean error of Dataset 1 (i.e. the error of the more accurate dataset). This is a practical assumption and should be the case for many datasets.

- **One Dataset Is Significantly More Accurate than the Other One.** If the error of one dataset is 15 to 20 times larger than the other one (e.g., datasets comparing ODOT data with TRANSMAP data), the accurate dataset is typically within the range of the less accurate dataset. However, it is highly unlikely (<1 percent) that the inaccurate dataset falls within the range of the accurate dataset.
- **Data on a Linear Reference System.** Any feature defined by a milepost value is automatically linked to a road centerline. The only error affecting the feature is the distance error along the road centerline. The overall accuracy of these features is not important, because they are always on the roadway. For all practical purposes, the practitioner is interested in the distances from and to the nearest intersection. The overall location accuracy of the road centerline is of little value to linear datasets.

3.2.3 Positional Data Accuracy Guidelines

The following recommendations are intended for the practical use of positional accuracy guidelines and for the combination of transportation datasets in general:

1. Avoid combining datasets with error differences larger than a factor of five. These datasets simply do not fit together and the results are unpredictable.
2. Select the appropriate dataset for the application. Certain accuracies are not usable for some of the applications listed in Section 2.3 of this report.
3. For most transportation applications, an accuracy of 3 feet is sufficient. Unfortunately, in the real world, most datasets are of much lower accuracy. Although it seems like a significant step for many agencies, upgrading from 50-foot accurate data to 3-foot accurate data is feasible and affordable with today's technology. This upgrade will become even more important as agencies use GPS in their day-to-day operations. Hand-held and real-time differential GPS receivers yield accuracies of 3 to 10 feet. The research shows this is compatible with 3-foot road centerlines; however, it is incompatible with 50-foot centerlines currently used by most agencies.
4. It is highly recommended that agencies maintain linear reference systems along road centerlines. The linear reference system can be easily transferred from the inaccurate road centerline to a more accurate road centerline without expensive re-mapping of features. This means that the integration of legacy data related to mileposts on a linear reference system is much easier than matching new coordinates to old feature points and centerlines.
5. Roadway information is always related to a road centerline. For most applications, the location of a feature relative to the centerline is much more important than its absolute location on a map. Therefore, it is recommended that users compute mileposts and offsets (the parameters that relate a point to a road centerline) for any feature inventoried and used by a transportation agency.

3.3 RECOMMENDATIONS FOR POSITIONAL DATA QUALITY STANDARDS

The primary objectives of spatial data quality standards are to help data recipients and owners evaluate the fitness for use of data. Definitions of fitness for use vary, based on environment and intended application. Therefore, a definition of "data quality" should include a sufficiently broad set of criteria to address the full range of possible data characteristics that might affect its application. Recently, paper and digital map products from federal agencies and agencies using federal money have been subject to the National Standard for Spatial Data Accuracy (NSSDA). Positional accuracy using the NSSDA recommends a testing and reporting procedure for determining the horizontal and vertical accuracy of maps and digital spatial data. The accuracy statistic allows users to determine if a set of data is appropriate for a given application. Agencies are encouraged to specify their own thresholds for given applications. The NSSDA is a quality indicator of a map's accuracy *(98)*.

This section first provides recommendations on additions to suggested data models. It further includes recommendations on metadata content standard modifications to address linear referencing issues and offers recommendations on procedural approaches to minimize positional accuracy degradation in transformation procedures.

3.3.1 Recommendations on Data Models

Metadata reports are required of all federal datasets and many state and local governments are developing their data in compliance with these standards as well. Metadata are essential for sharing information across agencies. Metadata reports are typically generated and maintained as files separate from the data, and in the data-sharing environment of a clearinghouse, this separation is appropriate. For routine operations within an organization, metadata are more useful when they are integrated with the data. For example when data quality information is an integral component of the data model, the information can be incorporated within processing routines to track and monitor quality aspects of the data. The following sections indicate modifications in data model components that address documentation of quality information. Modifications are suggested for anchor sections, anchor points, and linear-referenced features.

Metadata documentation of linear datum components is as important for positional quality assessment of linear-referenced features as the documentation of a geodetic datum is for quality assessment of 2D or 3D spatially referenced features. Linear datum components include anchor sections and anchor points (Figure 3-5). Metadata descriptions for

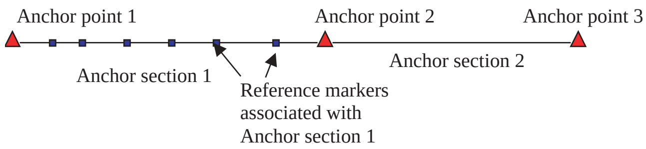


Figure 3-5. Components of linear datum.

these components will help to ensure stability and reportability of positional data quality. To support transformations between referencing systems, ideally the anchor points should maintain precise 2D (or 3D) as well as 1D measured positions. As shown below, the attributes listed for anchor points and anchor sections are the same as those recommended by Vonderohe et al. *(10)* with the addition of a metadata attribute in the form of the measurement method. Because it is likely that over time different measurement methods will be associated with different features (e.g., anchor sections), it becomes important to associate this metadata (measurement method) with the relevant feature (e.g., anchor section).

Anchor Section
Anchor_section_ID
From_Anchor_point_ID
To_Anchor_point_ID
Anchor Section Length
Measurement Method_ID

Anchor Point
Anchor_point_ID
Physical location description (intersection of Grant and Main)
2D position
2D Measurement Method_ID

Reference Mark (Traversal Reference Point)
Reference Mark_ID
Anchor Section_ID
Anchor Point_ID
Linear referenced position of reference mark

A reference mark is assumed to be any physical mark such as a milepost marker or station that is likely to have a precisely measured distance associated with it so that it can be used to reference other points. Anchor points are assumed to be origins or termini for one or more anchor sections. As such, they have no associated 1D referenced positions.

A linear-referenced feature should be associated with the datum components and information that indicate how it was referenced. At a minimum, the information should indicate the identifier of a reference marker (i.e., an anchor point or a reference mark), an anchor section or traversal identifier, the measured distance of the feature from the reference mark, and the measurement method (e.g., odometer or estimate).

Attributes for a linear-referenced feature-point or linear event
Feature_ID
Reference_ID
Traversal or Anchor Section_ID
Measured Distance from Reference Mark
Measurement Method_ID

The preceding feature attribute descriptions included the metadata element: measurement method_ID. Agencies should construct a list and standardized set of codes or unique identifiers for the measurement methods they employ and associate these with estimated or calibrated measures of positional accuracy (Table 3-4). Maintenance of such a table provides important metadata to document how measurements were carried out as well as the imprecision measures or derived inaccuracy values (e.g., ±3 meters) of the various measurement methods in readily accessible form for uncertainty assessment using the proposed error model. For example, a pilot study Washington State DOT indicated 3- to 5-foot accuracy for GPS-measured anchor section points.

3.3.2 Metadata Content Standard Revision Recommendations

Spatial data quality reports provide information that enables users to evaluate how the data fit their application requirements. This information includes descriptions of the source material from which the data were compiled, accuracy of measurement and compilation methods, and processing procedures used in production.

The Federal Geographic Data Committee (FGDC) *(99)* supports a common methodology for defining how to report the positional accuracy for geospatial data collected, produced, or disseminated by federal agencies. The National Standard for Spatial Data Accuracy (NSSDA) implements a statistical and testing methodology for estimating the positional accuracy of points on maps and in digital geospatial databases with respect to georeferenced ground positions of

TABLE 3-4 Standardized codes for measurement methods

Code	Name	Description	Nominal Error
	DMI		±1 foot per mile
	GPS		
	Inertial Navigation		
	Video Photologging		
	Orthophoto Digitization		

higher accuracy. National data quality standards have been specifically developed for 2D and 3D data, but specifications for quality reporting with respect to 1D referencing systems have not been fully developed. To fulfill national standards and the expectations of FGDC, there should be recommendations for reporting metadata for linear-referenced data, including a reporting strategy for positional accuracy. This section outlines metadata components and recommendations for amending spatial data quality metadata elements to cover linear-referenced data.

The NSSDA standard does not define threshold accuracy values. Agencies are encouraged to establish thresholds for their product specifications and applications. Data producers are expected to determine what accuracy exists or is achievable for their data and report it according to NSSDA.

The NSSDA uses RMSE to estimate positional accuracy. Accuracy is typically reported in ground distances at the 95 percent confidence level. The reported accuracy value reflects all uncertainties, including those introduced, for example, by geodetic control coordinates, compilation, and final computation of ground coordinate values in the product.

The Spatial Data Transfer Standard (SDTS) identifies four methods for determining positional accuracy. The preferred method is accuracy testing using an independent source of high accuracy. The other methods include deductive estimates, internal evidence, and comparison with source.

Whether data are tested by an independent source of higher accuracy or evaluated for accuracy by alternative means, metadata should describe how the test results were determined. For linear reference system elements, different accuracy testing procedures will apply. More rigorous tests are advisable for the datum components that serve as the foundation of the system. For anchor sections and anchor points, it may be advisable to test their positions with independent sources of higher accuracy. For example, anchor sections measured by photogrammetric or digital image processing methods could be tested against DMI. The positional accuracy of linear-referenced events is not likely to warrant tests against independent sources of higher accuracy. Their accuracy assessment is most logically based on deductive estimates using the metadata documentation for the referencing components described above in conjunction with the proposed error model.

3.3.3 Recommendations for Metadata Reporting

Any transportation data developed for distribution through the National Spatial Data Infrastructure (NSDI) is expected to include metadata. Whether linearly referenced data are to be distributed through NSDI or simply maintained internally by transportation agencies, there should be acceptable standard methods for reporting metadata for data positioned by linear referencing. Currently, metadata elements for documenting linear-referenced components have not been specified.

The Content Standard for Digital Geospatial Metadata (CSDGM) has three sections that could be modified or amended to incorporate metadata elements for linear-referenced data, specifically to address positional accuracy. These include the Spatial Data Organization and Spatial Referencing sections and the positional accuracy section under data quality. A key feature of the CSDGM Version 2 is the ability of geospatial data communities to develop profiles of the base standard. Many of these profiles have extended the base standard by adding metadata elements to meet specific community metadata requirements. Some considerations for possible content standard adjustments for documenting linear-referenced data are addressed in the next sections. These are preliminary suggestions that require broader community discussion.

Under the Spatial Data Organization section of CSDGM, there is an element for specifying a direct or indirect spatial reference. For this section, any linearly referenced dataset should indicate an indirect spatial reference. An indirect spatial reference element should indicate the type of geographic feature and the means by which locations are referenced in the data. Indirect spatial reference methods use various geographic features, such as a county, state, township, or section of the Public Land Survey System (PLSS); a road; or street address, to identify a place uniquely. The reference may use the name of the feature (e.g., “Westmoreland County”) or a code that identifies the feature (e.g., a county Federal Information Processing System [FIPS] code). If a dataset uses several forms of linear referencing (e.g., a dataset on traffic crashes where some are reported by mile marker, others by exit ramp, and others by station), this section should simply indicate the indirect method in linear referencing. If all features in a dataset have been referenced by a common method, this method could be specified (e.g., State Route Mile Post [SRMP]) as shown below:

Spatial_Data_Organization_Information:
Indirect_Spatial_Reference_Method: linear reference–State Route Mile Post

The fourth section of CSDGM is the Spatial Reference Section, which describes the reference frame and the means to encode coordinate information. Currently, it allows specification of Horizontal Coordinate System Definitions or Vertical Coordinate System Definitions, each of which has elements for describing respective datum information. Currently, no elements allow specification of a linear reference system and linear datum elements.

In a linear referencing system, as in 2D or 3D systems, a datum serves as the basis for locating the linear referencing system in the real world. This is critical information and thus there should be metadata elements to document a linear datum in a manner analogous to 2D and 3D geodetic datum. Currently, the metadata elements for Horizontal Coordinate System Definition are as shown in Figure 3-6. A recommendation

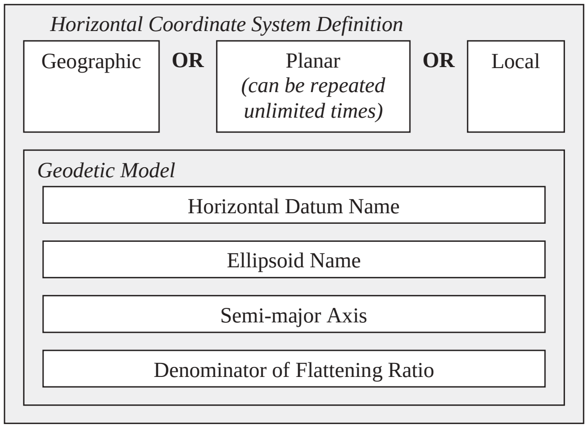


Figure 3-6. Current CSDGM horizontal coordinate system definition elements.

is to change the element Horizontal Coordinate System Definition to Horizontal Reference System Definition and add linear system as a choice with associated linear datum elements. These modifications are shown in Figure 3-7.

The linear datum consists of a connected set of anchor sections with anchor points at their junctions and termini. Initial suggestions for linear datum descriptions include information on the number of anchor sections; maximum, minimum and average section lengths; anchor point characterization (e.g., road centerline intersection points), and measurement method description.

The CSDGM currently has the following items for reporting positional accuracy in the Data Quality section (the illustration below includes example values):

Positional_Accuracy:
Horizontal_Positional_Accuracy:
Horizontal_Positional_Accuracy_Report: Digital data are tested by visual comparison with source mapping
Quantitative_Horizontal_Positional_Accuracy_Assessment:
Horizontal_Positional_Accuracy_Value: 12.19
Horizontal_Positional_Accuracy_Explanation: No quantitative tests

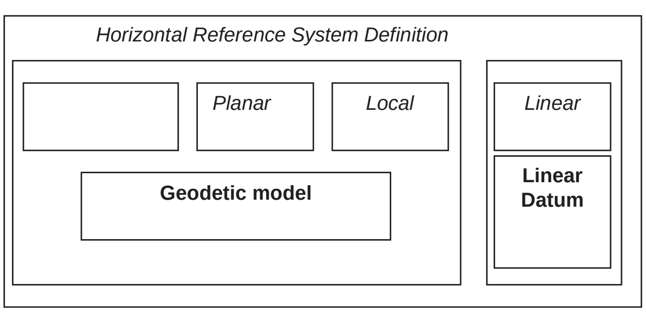


Figure 3-7. Recommended horizontal reference system.

Comparable elements apply for reporting linear-referenced positional accuracy. However, metadata for linearly referenced information should make reference to the components involved in the linear referencing process. Linear referenced positions have several dependencies, as indicated in Section 2.5.2 of this report. Therefore, positional accuracy reports should indicate the positional accuracy of various components. Key components affecting a linear reference are the linear datum components (i.e., anchor sections and anchor points), the network, the linear measurement methods involved, and their dependencies. At a minimum, separate positional accuracy reports for anchor sections and network components should be included, as illustrated below:

Positional_Accuracy:
Linear Datum_Positional_Accuracy:
Linear_Datum_Positional_Accuracy_Report: 20 percent of anchor section lengths measured photogrammetrically were tested by comparison with independent DMI measured lengths
Quantitative_Linear_Datum _Positional_Accuracy_Assessment:
Horizontal_Positional_Accuracy_Value: 1.29 feet
Horizontal_Positional_Accuracy_Explanation:
Positional_Accuracy:
Network_Positional_Accuracy:
Network_Positional_Accuracy_Report:
Network_Positional_Accuracy_Assessment:
Network_Positional_Accuracy_Value: 1.29 feet
Network_Positional_Accuracy_Explanation:

There are then several methods by which features (e.g., business data) are referenced to anchor points or sections. The lineage section of a metadata report should describe the linear referencing methods and processes applied.

One suggestion is to specify accuracy classes for the various components involved in a linear reference. These accuracy classes provide input at a general level for deductive estimates on the accuracy of a linear-referenced position. Example accuracy classes for network, measurement methods, and reference markers are shown below. These serve only as examples and would require further discussion with the broader community.

Accuracy classes for the network (digital spatial representation of centerline)
Class 1: Network centerline coordinates measured by GPS and inertial navigation
Class 2: Centerline measured photogrammetrically
Class 3: USGS 1:24,000 scale cartographic-based centerline

Accuracy classes for distance measurement methods
Class 1: DMI measured distances
Class 2: Photogrammetrically derived distances
Class 3: Over-the-surface distance computation from network
Class 4: Planametric distance computation from network

Accuracy classes of reference markers

Class 1: Absolute measured referents
GPS measured referent
Photogrammetrically measured referent
Class 2: Direct distance measured referents
Class 3: Indirect distance measured referents

As an example of the distinctions among the last set of accuracy classes, consider the following. Assume Crash 105 is reported with a linear-referenced position of 15.7 miles from Mile Marker 54, Route 209. Milepost 54 could be positioned using GPS, in which case it would have an accurate 2D or 3D position. It could have an accurately measured distance from an anchor point using DMI (a direct distance measure). Lastly, the marker might have been measured by odometer from the preceding mile marker. In each case, the accuracy of the distance that is then measured from the marker to Crash 105 is affected.

3.3.4 Transformation-Related Issues and Recommendations

The weakest link in the overall system is the spatial representation of the physical roadway (the network). The problem lies in the use of digital representations that rely on some level of discrete sampling to represent a continuous feature. Distance-measured locations, independent of the spatial representation, can be very accurate, as in those measured by DMI. Similarly, 2D positions determined independently of the spatial representation of the roadway can be very accurate (e.g., positions measured by GPS). If these accurate LR- or 2D-measured locations are maintained independently of a spatial representation of the roadway, derivative independently measured positions will remain stable with respect to positional accuracy. In the case of a linear-referenced system, using an accurate linear measurement method, such as DMI, where measured reference markers are accurate to 0.01 mile, even odometer-measured positions based on these measured reference points will retain a high level of accuracy.

Whenever a low-accuracy spatial representation of the roadway is involved, there can be a substantial loss of accuracy. The degradation in accuracy is a function of the accuracy and resolution of the spatial representation. Up until recently, most spatial representations used by transportation agencies have been derivatives of either the 1 : 100,000 or 1 : 24,000 scale DLG or TIGER data, which have low positional accuracy. Newer spatial representations based on GPS-measured centerlines will reduce the problem. When 2D and LRM positions need to be integrated, a spatial representation of the roadway is required, and if this is a low-accuracy representation, positional accuracy will be compromised.

In the example shown in Figure 3-8, a 2D-measured position is shown as Object A. The circle around A represents its 2D positional accuracy. The solid line represents a digital

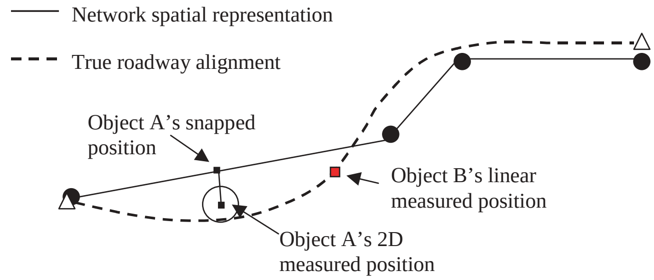


Figure 3-8. Transformation of 2D-measured position to LR position.

spatial representation of the roadway. The dashed line represents the actual roadway (for which there is no spatial representation) with an accurately measured distance based on DMI. To transform the Object A to an LR-measured position for integration with Object B measured using DMI, A's position must first be snapped to the spatial representation of the roadway. It is clear that the new snapped position for A on the spatial representation has a loss of accuracy exceeding the nominal or estimated 2D measurement accuracy represented by the circle. In a pilot study conducted by Iowa DOT, the nominal accuracy of 2D-positioned features using GPS was determined to be 3.29 feet. The distances required to snap these 2D-measured locations to the network representation ranged from 6.9 to 41.9 feet *(100)* with a mean of 23.3 and a standard deviation of 7.67, representing a sizeable loss in positional accuracy. For events on or immediately adjacent to the roadway and accurately measured using GPS, one might assume that they represent reasonable measurements of the roadway location and that the snap distances represent the loss of accuracy. Reported snap distances can be used to estimate positional inaccuracies in the digital spatial representation.

A transformation report should include a report of the snapping distances required to snap 2D-measured positions to a spatial representation. In the representation shown in Figure 3-9, the nominal accuracy of Object A is transformed to a distance error along the digital spatial representation. It was assumed that no more detailed spatial representation than that shown exists. If the distance is now measured along the spatial link to the snapped location, a linear distance measure

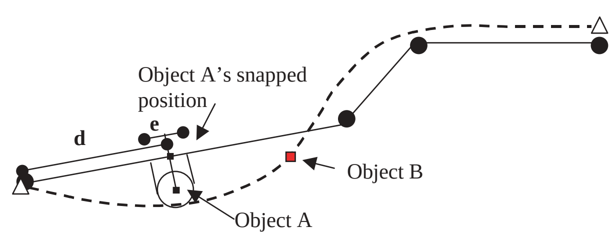


Figure 3-9. Transformation of nominal accuracy to distance error.

for Object A can be generated plus or minus the transformed 2D error ("e").

If this derived linear distance were now projected along the accurately DMI-measured road centerline as shown in Figure 3-10, it would (in this example) reflect a shorter distance than if one were able to position Object A accurately on the roadway.

If the digital spatial representation segment M and the true roadway segment N are known to correspond (i.e., represent the roadway between two intersections), then the linear distance measure along the digital representation can be calibrated according to the approach suggested by Reis et al. *(101)* and the formula

$$d' = \frac{d * N_{\text{dist}}}{M_{\text{dist}}}$$

where

d′ is the calibrated distance
d is the measured distance along segment M
N_{dist} is the DMI measured distance along the roadway
M_{dist} is the computed distance for segment M.

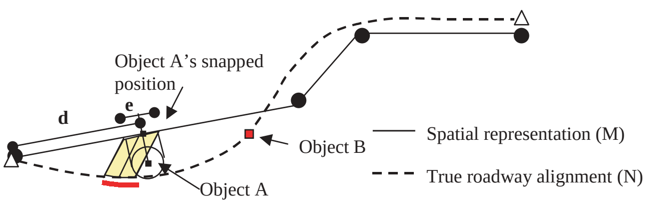


Figure 3-10. Projection of derived linear distance.

The best way to overcome the loss of positional accuracy in transformations from LR to 2D or vice versa is to ensure a consistent matching of accurately measured anchor sections to the corresponding sections of any digital spatial representation. By establishing such a correspondence, a linear-measured distance can be calibrated using the relationship and an accurately measured roadway length.

CHAPTER 4

CONCLUSIONS AND SUGGESTED RESEARCH

4.1 CONCLUSIONS

This project compiled and developed information on positional data quality, including a prototype data error model for analyzing the effects when considering trade-offs or transforming the location data obtained from different sources and measurement systems. Information compiled includes spatial data characteristics, linear referencing systems, spatial data quality, accuracy capabilities of measuring techniques, and applications of spatial data.

Spatial data used by state DOTs come from different sources and are used for various applications. Although some states are quite advanced in using new data collecting techniques, others rely on traditional methods. Different users have different perceptions as to the importance of error and accuracy, because the value of spatial data is directly concerned with the fitness of the data for a particular purpose, and the critical measure of that fitness for use is quality.

Linear referencing methods (LRMs) used by state DOTs in collecting and referencing spatial data for transportation applications are not uniform. Most states use multiple LRMs. When correlating data from various systems or LRMs, state DOTs tend to rely on in-house transformation methods or algorithms built into the GIS software.

Positional data are used for a wide range of transportation applications. These include safety (crash) analysis, transportation demand modeling, infrastructure management, transportation policy analysis, commercial vehicle operations, transit operations, and intelligent transportation systems. Emerging applications of positional data include emergency evacuation, automated oversize/overweight truck permitting and routing, and bus routing. Applications such as transportation planning, commercial vehicle operations, and regulatory and policy analyses are less sensitive to accuracy of positional data than highway inventory, highway design, and construction applications.

The primary sources of error associated with positional data are acquisition or measurement, processing, transformation, and presentation or visualization. Regardless of the measurement technique and referencing system, data will be observed with error. The method of data collection sets limitations on the selection of the measures and their metrics. A transformation can be made between different reference systems as well as different reference methods. These transformations introduce a degree of uncertainty to the transformed data.

Conceptually, the data error model is designed to handle the uncertainties associated with the locations of transportation features and events present in transportation-related applications. The uncertainties relate to recorded measurement precision, accuracy of the network, and issues of scale and resolution. A prototype model was developed that essentially represents the uncertainty object. The model was encapsulated into a software program with a graphic user interface that facilitates its use. The program computes the probability of the intersection of two features or data sources to determine whether they are compatible and if they should be used together. The program offers an efficient way to visualize the quality of the data at different significance levels of confidence. The model was tested with real-world case-study data for a wide range of transportation applications. The case studies demonstrated that the prototype model is sufficiently generic and can be used to evaluate the quality of positional data intended for a wide range of transportation applications. The prototype data error model would allow users of positional data to be aware of the bounds of the “true” location that can be derived from the integration of diverse data sources and the level of certainty that can be associated with spatial data. The model also allows the user to assess the potential quality implications of combining data from different sources and with different qualities.

Recommendations for using the prototype error model and standards for positional data quality are developed. Recommendations for positional data quality standards include metadata documentation for linear datum components to ensure stability and reportability of positional data quality. It is recommended that positional accuracy reports indicate the positional accuracy of various components. Key components affecting a linear reference are the linear datum components (i.e., anchor sections and anchor points), the network, the linear measurement methods involved, and their dependencies. At a minimum, separate positional accuracy reports for anchor sections and network components should be included. The best way to overcome the loss of positional accuracy in transformations from linear referencing to 2D or vice versa is to ensure a consistent matching of accurately measured anchor sections to the corresponding sections of any digital spatial representation.

4.2 SUGGESTED RESEARCH

The prototype error model, however, has certain limitations that can be addressed through further research. The following are the suggested main extensions that can enhance the usefulness of the prototype data error model:

- The prototype data error model requires the user to input the information required to assess the quality of the positional data directly. Further research is needed to enable the model to access stored information (e.g., metadata on the errors associated with data).
- The prototype model is designed as a stand-alone product where all the required input information has to be provided by the user. It is recommended that further research be conducted to implement the prototype model as an integral part of GIS applications. Such a model should be generic enough to be easily integrated with any kind of GIS software application.
- The prototype model in its present form is useful in evaluating the quality of data from data sources as well as the effect of combining data from different sources and with different qualities. In order to evaluate the quality of the application itself, further research is needed. For example, when positional data with given error is used for a certain transportation application, the prototype error model will not be able to estimate the overall effect on the application. Further research is required to determine the extent of utility of the model.

REFERENCES

1. Buckley, D.J. "The GIS Primer, The Nature of Geographic Information." Innovative GIS Solutions, Inc., An Introduction to Geographic Information Systems Website. February 1997. http://www.innovativegis.com/education/primer/nature.html
2. Veregin, H. "Data Quality Measurement and Assessment." Unit 100, NCGIA Core Curriculum in Geographic Information Science, Department of Geography, University of Minnesota, 1998. http://www.ncgia.ucsb.edu/education/curricula/giscc/units/u100/u100.html
3. Rybaczuk, K. "Error and Accuracy in Spatial Data Bases." Department of Geography, University of Dublin Trinity College, Association for Geographic Information Website, June 1993. http://www.agi.org.uk/pages/freepubs.html
4. Natural Resources Canada, "The Geospatial Standards Thrust of the Canadian Geospatial Data Infrastructure," GeoConnections. Canada's Geographic Information on the Internet, Canadian Geospatial Data Infrastructure, Natural Resources Canada Website, 2001. http://www.geoconnections.org/english/publications/reports/idon/idon_report.html
5. Minnesota DOT. "Positional Accuracy Handbook: Using The National Standard for Spatial Data Accuracy to Measure and Report Geographic Data Quality." Minnesota Planning, St. Paul, MN, 1999.
6. McGlamery, P. "Issues of Authenticity of Spatial Data," Proceedings of the 66th IFLA Council and General Conference, Jerusalem, Israel, International Federation of Library Associations and Institutions. August 2000. http://ifla.inist.fr/IV/ifla66/papers/142-124e.htm
7. Vonderohe, A.P., et al., "Adaptation of Geographic Information Systems for Transportation." *Transportation Research Record 359*, Transportation Research Board, National Research Council, Washington, D.C., 1993.
8. Baker, W.T. and W.E. Blessing. "Highway Location Reference Methods." *NCHRP Synthesis of Highway Practice 21*. Transportation Research Board, National Research Council, Washington, D.C., 1974.
9. Rowell, R. "Theory and Practice: Linear Referencing at the Idaho Transportation Department." *GIS-T '96 Proceedings*. American Association of State Highway and Transportation Officials (AASHTO), Washington D.C.,1996, pp. 1–9.
10. Vonderohe, A.P., C.L. Chou, F. Sun, and T.M. Adams. "A Generic Data Model for Linear Referencing Systems." *NCHRP Research Results Digest 218*. Transportation Research Board, National Research Council, Washington, D.C., 1997.
11. Adams, T.M., A.P. Vonderohe, J.A. Butler, and N. Koncz. "Model Considerations for a Multimodal, Multidimensional Location Referencing System." Proceedings GIS-T Symposium, San Diego, CA, 1999.
12. O'Neill, W. and E. Harper. "Resource Guide on the Implementation of Linear Referencing Systems in Geographic Information Systems." Report BTS-CD-22. Bureau of Transportation Statistics, U.S. Department of Transportation, 1998.
13. Dueker, K. J. and J. A. Butler. "GIS-T Enterprise Data Model with Suggested Implementation Choices." *Journal of the Urban and Regional Information Systems Association*, Vol. 10, No. 1, 1998. Also http://www.upa.pdx.edu/CUS/
14. Butler, J. A. "Addendum 2: ARC/INFO Base Map—Final Design." Hamilton County General Government. Chattanooga, TN, 1998a.
15. Vonderohe, A.P. and T. Hepworth. "A Methodology for Design of a Linear Referencing System for Surface Transportation." Report prepared for Sandia National Laboratories, Project AT-4567. University of Wisconsin, Madison, WI, 1996.
16. Butler, J.A. and K. J. Dueker. "A Proposed Method of Transportation Feature Identification." Proceedings, TRB Annual Meeting. Washington, D.C., 1998. Also http://www.upa.pdx.edu/CUS/
17. Butler, J.A. "A Comprehensive Plan to Establish and Operate a Regional Geographic Data Network (GeoNet)." Hamilton County General Government. Chattanooga, TN. 1998b.
18. Scarponcini, P. "Generalized LRS Approach for Interoperability of As-is Legacy Data." Proceedings GIS-T Symposium. http://www.bts.gov/gis/BISWEB/gist99index.html. Accessed April 11, 1999.
19. Scarponcini, P. "Generalized Model for Linear Referencing." Proceedings of the 7th International Symposium on GIS. Kansas City, 1999, pp. 53–59.
20. Hansen, D.T. "Assessing and Reporting Positional Accuracy for Geospatial Data of Mixed Lineage." http://www.mp.usbr.gov/geospat/p131.html. Accessed July 9, 1997.
21. Quiroga, C.A. "Accuracy of Linearly Referenced Data Using GIS." 78th Transportation Research Board Annual Meeting, Washington, DC, 1999. Paper #990050.
22. Minnesota Planning, Land Management Information Center. "Positional Accuracy Handbook, Using the National Standard for Spatial Data Accuracy to Measure and Report Geographic Data Quality." October 1999. http://www.mnplan.state.mn.us/press/accurate.html
23. Minnesota DOT. "Standards for GIS Users—Data Quality Standards." http://www.gis.state.mn.us/committee/stand/New_dataqual.html. Accessed May 6, 2001.
24. FGDC. "Content Standards for Geospatial Metadata." Reston, VA: U.S. Geological Survey, 1994.
25. Goodchild, M.F. and G.J. Hunter. "A Simple Positional Accuracy Measure for Linear Features." *International Journal of Geographical Information Science*. Vol. 11, March 1997, pp. 299–306.
26. Federal Geographic Data Committee, "Geospatial Positioning Accuracy Standards, Part 1: Reporting Methodology." FGDC-STD-007.1-1998, Federal Geodetic Control Subcommittee, Federal Geographic Data Committee, National Spatial Data Infrastructure, 1998.
27. British Columbia Ministry of Environment, Lands & Parks "Scale, Accuracy, and Resolution in a GIS." Information Systems Branch Website, May 1998. http://www.env.gov.bc.ca/gis/gisscale.html
28. Goodchild, M.F. and K.K. Kemps. "Unit 45—Accuracy of Spatial Databases," Department of Geography, University of British Columbia, eds., NCGIA Core Curriculum in GIS, National Center for Geographic Information and Analysis, University of California, Santa Barbara, CA, 1990. http://www.geog.ubc.ca/courses/klink/gis.notes/ncgia/u45.html

29. Hunter, G. J. and M. F. Goodchild. "Modeling the Uncertainty of Vector Data in Geographic Information Systems." Proceedings of ESRI International User Conference, Palm Springs, CA, 1995. (CD-ROM and WWW).
30. Mowrer, H.T. "Accuracy (Re)assurance: Selling Uncertainty Assessment to the Uncertain." *Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources*, Ann Arbor Press, 1999, pp. 3–10.
31. Chrisman, N. *Exploring Geographic Information Systems*, New York: John Wiley & Sons, 1997.
32. Clarke, K. *Getting Started with Geographic Information Systems*, Prentice Hall, 1997.
33. Burrough, P. and R. McDonnell. *Principles of Geographical Information Systems*, Oxford University Press, 1998.
34. Joslyn, C.A. "Web Dictionary of Cybernetics and Systems." 1992. http://pespmc1.vub.ac.be/ASC/uncertainty.html
35. FGDC. "Draft Geospatial Positioning Accuracy Standards (Standards for Geodetic Networks)," ftp://fgdc.er.usgs.gov, 1996.
36. Frank, U. A. "Metamodels for Data Quality Description." *Data Quality in Geographic Information—From Error to Uncertainty*. R. Jeansoulin and M.F. Goodchild. Editions Hermès, Paris, 1998, pp. 15–29.
37. Chrisman, N. "The Role of Quality in Information in the Long Term Functioning of a Geographic Information System." Proceedings of Auto Carto 6, 1983, pp. 303–312.
38. Goodchild, M.F., G. Sun and S. Yang. "Development and Test of an Error Model for Categorical Data." *International Journal of Geographical Information Science*. Vol. 6, 1992, pp. 87–104.
39. Hunter, G.J. and M. F. Goodchild. "Managing Uncertainty in Spatial Databases: Putting Theory into Practice." *Journal of the Urban and Regional Information Systems Association*. Vol. 5, No. 2, 1993, pp. 55–62.
40. Beard, M.K. "A Structure for Organizing Metadata Collection." Proceedings of 3rd International Conference on Integrating GIS and Environmental Modeling, Santa Fe, NM, 1996, pp. 21–26.
41. Parsons, S. "Current Approaches to Handling Imperfect Information in Data and Knowledge Bases." *IEEE Transactions on Knowledge and Data Engineering*. Vol. 8, No. 3, 1996, pp. 353–372.
42. Heuvelink, G.B.M. *Aggregation and Error Propagation in GIS. Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources*. Ann Arbor Press, 1999, pp. 219–225.
43. Goodchild, M.F. and S. Gopal. *Accuracy of Spatial Databases*. Taylor & Francis, London, 1989.
44. Nijkamp, P. and H.J. Scholten. "Information Systems: Caveats in Design and Use." Proceedings of the 2nd European Conference on GIS, Brussels, Belgium, 1991, pp. 737–746.
45. Hunter, G. J. and M. F. Goodchild. "Modeling the Uncertainty of Slope and Aspect Estimates Derived from Spatial Databases." *Geographical Analysis*, Vol. 29, 1997, pp. 35–49.
46. Agumya, A. and G.J. Hunter. "Assessing Fitness for Use of Spatial Information: Information Utilization and Decision Uncertainty." Proceedings of the GIS/LIS '96 Conference, Denver, CO, 1996, pp. 359–370.
47. Beard, M. K. "Designing GIS to Control Misuse of Spatial Information." Proceedings of the URISA '89 Conference, Boston, MA, 1989, pp. 245–255.
48. Burrough, P.A. "The Development on Intelligent Geographical Information Systems." Proceedings of the 2nd European Conference on GIS, Brussels, Belgium, 1991, pp. 165–174.
49. Elmes, G.A. and C. Cai. "Data Quality Issues in User Interface Design for a Knowledge-based Decision Support System." Proceedings of the 5th International Symposium on Spatial Data Handling, Charleston, SC, 1992, pp. 303–312.
50. Blakemore, M. "High or Low Resolution? Conflicts of Accuracy, Cost, Quality, and Application in Computer Mapping." *Computers and Geosciences*, Vol. 11, No. 3, 1985, pp. 345–348.
51. Hunter, G.J. and M.F. Goodchild. "A New Model for Handling Vector Data Uncertainty in GIS." *Journal of the Urban and Regional Information Systems Association*, Vol. 8, No. 1, 1996, pp. 51–57.
52. Goodchild, M. F. "Effects of Generalization in Geographical Data Encoding." *Map Data Processing*. H. Freeman and G.G. Pieroni. Academic Press, New York, 1979, pp. 191–206.
53. Veregin, H. "Error Modeling for the Map Overlay Operation." *Accuracy of Spatial Databases*. M. Goodchild and S. Gopal. Taylor & Francis, London, 1989, pp. 3–18.
54. Sinton, D. F. "The Inherent Structure of Information as a Constraint to Analysis: Mapped Thematic Data as a Case Study." Harvard Papers on Geographic Information Systems. Vol. 6, 1978, pp. 43–59.
55. Goodchild, M.F. and J. Proctor. "Scale in a Digital Geographic World." *Geographical & Environmental Modeling 1*, Vol. 1, 1997, pp. 5–23.
56. Watzek, K. A. and J. C. Ellsworth. "Perceived Scale Accuracy of Computer Visual Simulations." *Landscape Journal*, 1992, pp. 21–36.
57. Bruegger, B.P. "Spatial Theory for the Integration of Resolution-Limited Data." Orono, ME, Dissertation. 1994.
58. Cushnie, J. L. "The Interactive Effect of Spatial Resolution and Degree of Internal Variability within Land-Cover Types on Classification Accuracies." *International Journal of Remote Sensing*. Vol. 8, No. 1, 1987, pp. 15–29.
59. Canters et al. "Estimation of Land-Cover Proportions from Aggregated Medium-Resolution Satellite Data." *Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources*, Ann Arbor Press, MI 1999, pp. 263–270.
60. Moody, A. and C. E. Woodcock. "Scale-Dependent Errors in the Estimation of Land-Cover Proportions: Implications for Global Land-Cover Datasets." *Photogrammetric Engineering and Remote Sensing*, Vol. 8, No. 5, 1994, pp. 585–594.
61. Burrough, P.A. "Multiscale Sources of Spatial Variation in Soil. II. A Non-Brownian Fractal Model and Its Application in Soil Survey." *Journal of Soil Science*. Vol. 34, 1983, pp. 599–620.
62. Oliver, M. A. and R. Webster. "Combining Nested and Linear Sampling for Determining the Scale and Form of Spatial Variation of Regionalized Variables." *Geographical Analysis*. Vol. 18, No. 3, 1986, pp. 227–242.
63. Wade, T. G., J. D. Wickham, and D. F. Bradford. "Accuracy of Road Density Estimates Derived from USGS DLG Data for Use in Environmental Applications." *Photogrammetric Engineering & Remote Sensing*. Vol 65, No. 12, 1999, pp. 1419–1425.
64. Townsend, J. R. G. and C. O. Justice. "Selecting the Spatial Resolution of Satellite Sensors Required for Global Monitoring

of Land Transformations." *International Journal of Remote Sensing*. Vol. 9, No. 2, 1988, pp. 187–236.

65. Heuvelink, G. B. M. "Aggregation and Error Propagation in GIS." *Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources,* Ann Arbor Press, MI 1999, pp. 219–225.
66. Prisley, S. P. and J. L. Smith. "The Effects of Spatial Data Variability on Decisions Reached in a GIS Environment." *GIS Applications in Natural Resources*. M. Heit and A. Shortread, 1991, pp. 167–170.
67. van Groenigan, J. W. and A. Stein. "Constrained Optimization of Spatial Sampling in a Model-based Setting using SANOS Software." *Accuracy 2000*, Amsterdam, Delft University Press, 2000, pp. 679–685.
68. Vonderohe, A.P., L. Travis, R.L. Smith, and V. Tsai. "Adaptation of Geographic Information Systems for Transportation." *NCHRP Report 359,* Transportation Research Board, National Research Council, Washington, D.C., 1993.
69. Pittman, R. and Tucker, J.R. "GIS Brings Value to the Transportation Market." Proceedings GIS-T Symposium, Salt Lake City, UT, 1998, pp. 254– 269.
70. O'Neill, W. and E. Harper. "Linear Location Translation with GIS." Proceedings, ESRI User Conference. 1997.
71. Souleyrette, R., T. Strauss, and M. Pawlovich, "GIS-ALAS: The Integrated Analysis of Highway Crash Data in a GIS Environment." Proceedings of GIS-T Symposium, Salt Lake City, Utah, 1998, pp. 411–428.
72. Miller, S., D. L. Harkey, and T. Johnson. "The Truck Corridor Analysis Using a GIS-Based Crash Referencing System." Proceedings of GIS-T Symposium, Greensboro, NC, 1997, pp. 61–76.
73. Battelle, "Draft Methodology Report," Freight Analysis Framework Capacity and Needs Analysis for the National Strategic Multimodal Freight Transportation System. Project Report submitted to FHWA Office of Freight Management and Operations, Battelle, Columbus, OH, 2002.
74. Dorsey, K., D. Siegel, and L.J. Harman. "Developing a Transit Information Decision Support Environment for the Cape Cod Regional Transit Authority, Dennis, MA," Proceedings GIS-T Symposium, Greensboro, NC, 1997, pp. 308–312.
75. Racca, D.P. "Using GIS to Identify Markets for Transit in Delaware." Proceedings GIS-T, 1998, Salt Lake City, UT, pp. 270–291.
76. Hsiao, S., J. Boberg, B. Barua, and X. Zhang. "Assisting Transit Survey. Integrating GIS with Transit Survey Methodology." 80th Annual TRB Meeting, Washington, D.C., 2001. Paper # 01-0360.
77. Novak, K. and J. Nimz. "Transportation Infrastructure Management—Seneca County Creates a Complete Inventory of Its Transportation Infrastructure." Proceedings GIS-T Symposium, Greensboro, NC. 1997. pp. 301–307.
78. Sarasua, W., e Y. Wang, D.B. Clark, J. Tropea, and C. Collins, "Multi-criteria Dynamic Segmentation: A GIS Application for Managing Pavement Marking Retro Reflectivity." 80th Annual TRB Meeting, Washington, D.C., 2001. Paper # 01-3195.
79. Jha, M.K and P. Schonfeld. "Integrating Genetic Algorithms and GIS to Optimize Highway Alignments." 79th Annual TRB Meeting, Washington, D.C., 2000. Paper # 00-1060.
80. Greenfeld, J. "Digital Map Requirements for Automatic Vehicle Location." *2000 Research at National Center for Transportation and Industrial Productivity (NCTIP): New Jersey Department of Transportation Sponsored Research Projects,* New Jersey Institute of Technology, November 2000.
81. Kolton, G. "Traffic.com to Provide Data for Traffic Reports in Tampa." *The Urban Transportation Monitor*, March 2001.
82. Alam, M. and E. Fekpe. "Freight Data Analysis Application of GIS-T in Freight Data Analysis: A Case Study of I-90/I-94 Corridor Analyses." *Transportation Research Record 1625*, Transportation Research Board, National Research Council, Washington, D.C., pp. 173–183, 1998.
83. Agumya A. and G.J. Hunter, "Estimating Risk in GIS-Supported Decisions," Proceedings presented in CD format. Urban and Regional Information Systems Association (URISA), Sallis M. (ed), URISA 1997, Toronto, Canada: URISA.
84. Guptill, S. C. and J. L. Morrison. Eds., 1995, *Elements of Spatial Data Quality*. Oxford: Elsevier Science.
85. Veregin, H. and P. Hargitai. "An Evaluation Matrix for Geographical Data Quality." *Elements of Data Quality*, Eds. Guptill, S. C. and J. L. Morrison, Oxford, Elsevier Science, 1995, pp. 167–188.
86. Agumya, A. and G. J. Hunter. "The Standard Based Approach to Assessing the Fitness for Use." *URISA Journal*, Vol 11, No 1., Spring 1999.
87. Adams, T. M., N.A. Koncz, and A.P. Vonderohe, "Guidelines for the Implementation of Multimodal Transportation Location Referencing Systems." NCHRP 20-27(3) Final Report, Jan. 2001.
88. Hansen, D.T. "Spatial Data Coordinate Quality and ArcInfo Data Model." Proceedings of ESRI International User Conference, Palm Springs, CA, 1996.
89. Shi, W. and C. Cheung. "A Reliability of Spatial Objects in a 3D GIS." Proceedings of Urban and Regional Information Systems Association Annual Conference. 1999.
90. Stanfel and Stanfel. "A Generalization on Line Segment Reliability Measures." *Surveying and Land Information Systems,* Vol. 54, No. 1, 1994, pp. 49–52.
91. Thill, J. C. and Z. Yang. "A Model of Uncertainty for Integrating GPS Data and a Linear Reference System." Annual Meeting of the Association of American Geographers, Pittsburgh, April 2000.
92. McGranaghan, M. A. "Cartographic View of Spatial Data Quality." *Cartographica*. Vol. 30, No. 2-3, 1993.
93. Beard, K. and W. MacKaness, "Visual Access to Data Quality in Geographic Information Systems." *Cartographica*. Vol. 30., No. 2-3, 1993.
94. Burkholder, E. "Design of a Local Coordinate System for Surveying, Engineering, and LIS/GIS." *Surveying and Land Information Systems*. Vol. 53, No. 1, 1993, pp. 29–40.
95. Durgin, P. "Measurement Based Databases: One Approach to the Integration of Survey and GIS Cadastral Data." *Surveying and Land Information Systems*. Vol. 53, No. 1, 1993, pp. 41–47.
96. Ehlschlaeger, C. R. and M. F. Goodchild. "Uncertainty in Spatial Data: Defining, Visualizing, and Managing Data Errors." Proceedings of GIS/LIS, Phoenix AZ, http://everest.hunter.cuny.edu/~chuck/gislis/gislis.html. 1994.
97. Shi, W. Z. "A Generic Statistical Approach for Modelling Error of Geometric Features in GIS." *International Journal of*

Geographic Information Science, Vol. 12, No. 2, 1998, pp. 131–143.

98. Czerniak, R. J. “Collecting, Processing, and Integrating GPS Data into GIS.” *NCHRP Synthesis of Highway Practice 301*. Transportation Research Board, National Research Council, Washington, D.C., 2002.
99. Federal Geographic Data Committee, Ground Transportation Subcommittee Report 10/01/94 FGDC Ground Transportation Subcommittee. Position and Recommendations on Linear Referencing Systems.
100. Hallmark, S. “GPS to LRM: Integration of Spatial Point Features with Linear Referencing Methods.” Final Report. Iowa DOT. CTRE Management Project 00-68 2001.
101. Reis, T. and J. DuChateau. “A Workshop on Implementing Road Data Models: Coordinate Based and Cross Street Linear Referencing.” GIS-T Symposium. Arlington, VA, April 8, 2001.

ABBREVIATIONS AND ACRONYMS

ADOT—Arizona Department of Transportation
ATIS—Advanced Traveler Information Systems
ATSCS—Automatic Traffic Surveillance and Control System
AVL—Automatic Vehicle Location
BTS—Bureau of Transportation Statistics
CADD—Computer Aided Design and Drafting
CEP—Circular Error Probability
CIR—Color Infra–Red
CLs—Center Lines
COGO—Coordinate Geometry
CORS—Continuously Operating Reference Stations
CSDGM—Content Standard for Geospatial Data
CTA RR Network—Center for Transportation Analysis Rail Road Network
CVO—Commercial Vehicle Operations
DEM—Digital Elevation Model
DLG—Digital Line Graphs
DMI—Distance Measuring Instrument
DOQ—Digital Ortho Quad
DOQQ—Digital Ortho Quarter Quads
DOT– Department of Transportation
DQM—Data Quality Model
DX—Distance Expression
EDM—Electronic Distance Measurement
FGDC—Federal Geographic Data Committee
FHWA—Federal Highway Administration
FRA—Federal Railway Administration
GCP—Ground Control Point
GDT—Geographic Data Technologies
GIMS—Geographic Information Management System
GIS—Geographical Information Systems
GIS-T—Geographical Information System for Transportation
GNIS—Geographic Names Information System
GPS—Geographical Positioning System
GUI—Graphic User Interface
HARN—High Accuracy Reference Network
HDOM—Highway Design Optimization Model
HPMS—Highway Performance Monitoring System
IADOT—Iowa Department of Transportation
IHIS—Integrated Highway Information System
IMU—Inertial Measuring Unit
ISDMI—Integrated Surveillance and Data Management Infrastructure
ITMS—Integrated Traffic Monitoring System
ITS—Intelligent Transportation Systems
LE—Linear Element
LIDAR—Light Detection and Ranging
LLRM—Linear Location Referencing Model
LR—Linear Referenced
LRM—Linear Referencing Model
LRS—Linear Referencing System
LX—Location Expression
NAD—North American Datum
NCDOT—North Carolina Department of Transportation
NCHRP—National Cooperative Highway Research Program
NDDOT—North Dakota Department of Transportation
NGS—National Geodetic Survey
NHPN—National Highway Planning Network
NSDI—National Spatial Data Infrastructure
NSSDA—National Standard for Spatial Data Accuracy
NWN—National Waterway Network
OCTA—Orange County Transportation Authority
ODOT—Ohio Department of Transportation
ORNL—Oak Ridge National Laboratory
PCCP—Percentage of Correctly Classified Pixel
PLSS—Public Land Survey System
RIMS—Roadway Information Management System
RMSE—Root Mean Square Error
ROW—Right of Way
SEP—Spherical Error Probability
SRMP—State Route Mile Point
TS&W—Truck Size and Weight
USGS—United States Geological Survey
UTM—Universal Transverse Mercator
VaDOT—Virginia Department of Transportation
WisDOT—Wisconsin Department of Transportation

APPENDIX A

PROTOTYPE DATA ERROR MODEL (GISError) USER GUIDE

A1.0 INTRODUCTION

The Data Error Model (GISError) is a Microsoft® Windows application that allows the user to visualize the results of data errors and display how errors of different positional data sources affect each other. The program also provides an analysis of the results and shows the confidence intervals and buffers around data points and lines. The program enables the user to compute the probability of the intersection of two data sources to determine whether they are compatible and if they should be used together. The following data elements could be entered into the program by the user:

- Points: defined by a Cartesian (X, Y) Coordinate, such as the Easting and Northing in a State Plane system.
- Lines: defined by two points, i.e. the starting and end points, which are also defined in the Cartesian Coordinate frame.
- Points on a Line: defined by a distance from the starting point. These correspond to the Linear Reference System used for transportation applications, where the line would represent the road centerline and the point is an event on the centerline, such as an accident. The distance is the mileage value of the event from the beginning of the route or the nearest intersection.
- Error Values: The user can also enter an Error Value for both points and lines. This value represents the mean error (one sigma) of the data element. For example, a point whose position was determined with a GPS receiver may have a sigma of 3 feet. As is well known, the one sigma interval represents a probability of 67 percent. The error value of a line shows the uncertainty of the location of the whole line as a buffer in the specified coordinate frame.

This program uses a wizard-style interface that allows users to navigate through the program using the "Next" and the "Back" buttons. The window shown in Figure A-1 appears when the program starts.

A2.0 USER INPUTS

The Data Error Model Program starts with the Input screen, where the user enters the information required to perform the analysis. The data input screen of the program shows the information entered into the test program, such as the exact coordinate values for both data elements and the error values. The required inputs are described below.

Output Resolution

The output resolution controls the scale of the output image. The user enters the output resolution in the window area shown in Figure A-2. This user input can be an integer or a floating-point number. The program uses this value as the map scale of the output image. A smaller value for output resolution, therefore, produces a more detailed output image.

Features

The users must provide coordinates for at least one feature. The Data Error Model supports two types of features: point and line. For a point in 2D, users are required to provide X and Y coordinates. For a line, the user needs to input coordinates for the start and the end points of the line (see Figure A-3).

The program allows a 1D point on a line as an input. For this linearly referenced point, the user needs to specify a linear distance from the starting point of the line where the point event is located. The program can then calculate the point location on the specific line and show one-dimensional errors associated with this point.

Error Values

- When users choose a feature, they must also specify the error value associated with the feature. This error represents the mean error (one sigma) of the data element. For example, a point whose position was determined with a GPS receiver may have a sigma of 3 feet. The error value of a line shows the uncertainty of the location of the whole line as a buffer in the specified coordinate frame.

A3.0 VIEWING THE IMAGE

By pressing the "Next" button after entering all the required inputs, the program calculates the probability zones for each event entered and produces an output image that shows the probability zones around the measured event. This error typically displays as concentric circles around a point or a buffer that is rounded at the ends of a line segment. Both data elements are shown in the same window, so the user can make an empirical decision on whether the two features would actually intersect. For example, the image shown in Figure A-4 depicts the probability zones of a 2D point event and a line event and their intersection. The probability zones are shown at different levels of significance.

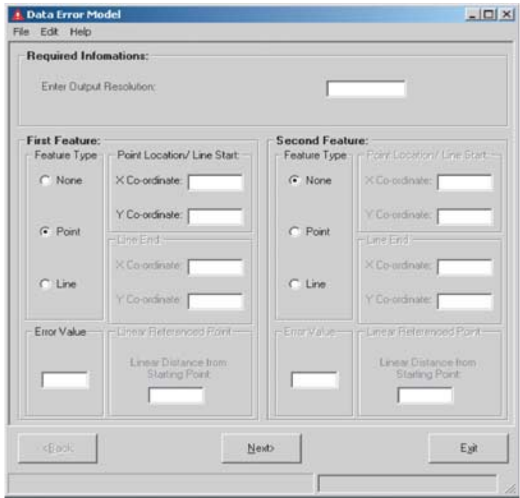


Figure A-1. Data error model input window.

Figure A-2. Output resolution.

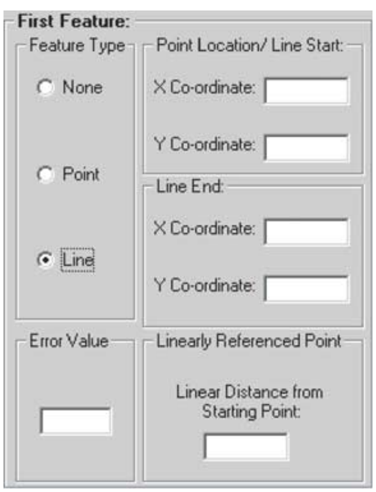


Figure A-3. Feature co-ordinates and associated error value.

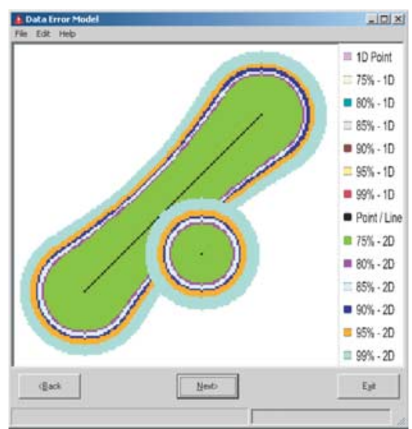


Figure A-4. The image window.

A4.0 VIEWING THE ERRORS

By pressing the "Next" button from the image screen, the program shows the error screen, depicted in Figure A-5. The results of the analysis are shown in this screen, which lists the errors of both the first and the second feature for different probabilities. This screen displays the maximum and the minimum error values associated with the feature(s) at various probability levels. These minimum and maximum values represent the radii of the probability zones for each significance level (i.e., 75%, 80%, 85%, 90%, 95%, and 99%). The program generates a graphical display of the situation showing error buffers of various probabilities at different significance levels and presented in different colors. Finally, the program displays the intersection probability of the two datasets. The intersection probability is a measure of the likelihood of the two data elements intersecting. It is important to note that the intersection probability is calculated relative to the first feature. The Intersection Probability shows the probability with which the second dataset will fall within the confidence buffer of the first dataset. For example, if the first dataset is inaccurate and has a large error, and the second dataset is very accurate and has a small error, and if the two data elements are separated by less than two times the error of the first dataset, then it is very likely that the second one will be within the first one's buffer. However, if the order of the datasets were reversed, which means that the accurate one comes first, then it is unlikely that the second dataset will be within the buffer of the first.

A5.0 OUTPUT RESULTS

The Data Error Model provides support for exporting the analysis in various formats. These are described below.

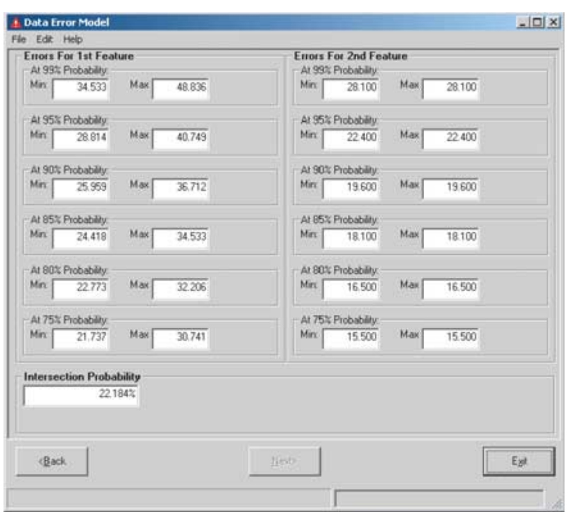


Figure A-5. Error analysis output window.

Save/Print Image

Users can print the image or save it from the "File" menu. When a user chooses the "Save Image" option, as shown in Figure A-6, the program saves the image in Windows Bitmap format.

Copy and Paste

Users can also perform a screen capture of the current image in the Windows system clipboard by clicking the "Copy Image" menu item of the "Edit" menu depicted in Figure A-7. The copied image can be pasted directly into any application that accepts bitmap images, such as Microsoft Word.

Exporting to GIS Applications

Users also can export the analysis result to ASCII grid format when they choose the menu item "Export To Grid" under the "File" menu (Figure A-8).

Third-party tools can convert the saved text file to a raster image. For example, the outputs can be exported to Grid Format in ArcToolbox from ESRI by choosing the ASCII to Grid option under Import to Raster tools from

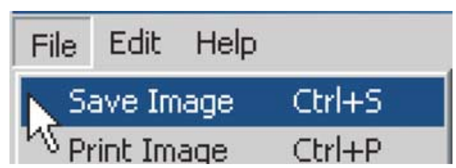


Figure A-6. Save/print options.

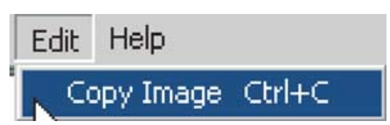


Figure A-7. Copy image option.

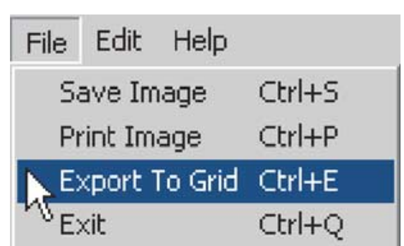


Figure A-8. Exporting to GIS applications.

TABLE A-1 Values used to represent features

	For 1st Feature	For 2nd Feature
For Line or Point Feature	150	250
99% Probability Zone	199	299
95% Probability Zone	195	295
90% Probability Zone	190	290
85% Probability Zone	185	285
80% Probability Zone	180	280
75% Probability Zone	175	275
For Linearly Referenced Point	100	200
99% Probability Zone for Linearly referenced point	149	249
95% Probability Zone for Linearly referenced point	145	245
90% Probability Zone for Linearly referenced point	140	290
85% Probability Zone for Linearly referenced point	135	235
80% Probability Zone for Linearly referenced point	130	230
75% Probability Zone for Linearly referenced point	125	225

ArcToolbox. The grid data can then be viewed using ArcMap from ESRI.

When exported to the grid format and viewed by GIS software such as ArcMap, the probability zones are formatted in the legend to show just the feature data values. The program uses the values between 100 and 199 and between 200 and 299 to represent probability zones for the first and second features, respectively. Users should refer to the look-up table shown in Table A-1 for relating probability zones with the feature data values when viewing the output in third-party GIS viewers.

APPENDIX B

RESULTS OF CASE STUDIES

Each case study represents a combination of two different data elements provided by TRANSMAP or ODOT as described below:

- A: GDT Road Centerline—ODOT GPS Point from video logging van
- B: TRANSMAP Centerline—ODOT Centerline
- C: TRANSMAP Centerline—ODOT Centerline
- D: TRANSMAP Centerline—ODOT Crash Data
- E: TRANSMAP Centerline—ODOT GPS Point from video logging van
- F: TRANSMAP Feature Points—ODOT GPS Point from video logging van
- G: TRANSMAP GPS Image Location—ODOT Centerlines
- H: TRANSMAP Feature Points—ODOT Centerlines
- I: TRANSMAP GPS Image Location—ODOT Crash Data
- J: TRANSMAP GPS Image Location—ODOT GPS Point from video logging van
- K: TRANSMAP Feature Points—ODOT Crash Data
- L: GDT Road Centerline—ODOT Crash Data
- M: ODOT Road Centerline—ODOT Crash Data
- N: ODOT Road Centerline—ODOT Crash Data.

The following pages contain the graphics for each case study showing the following:

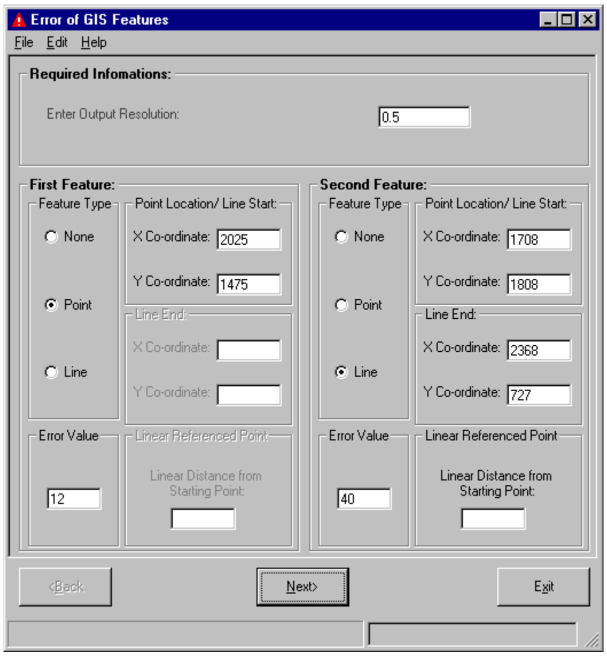
- The data input screen of the GISError program—showing the information entered into the test program, such as the exact coordinate values for both data elements and the error values.
- The analytical results of the analysis are shown in the Error of GIS Feature screen. It lists the Errors of both the first and the second feature for different probabilities. The results represent the minimum and maximum error radius for a certain probability. This window also shows the Intersection Probability of the two data elements.
- Graphic of the Error of the GIS Feature—This error typically displays as concentric circles around a point or a buffer that is rounded at the ends of a line segment. Both data elements are shown in the same window, so the user can make an empirical decision on whether the two features would actually intersect.
- The last window displays the corresponding datasets in a GIS format. The two datasets are displayed in the ArcView GIS and a screen shot of the area used for the analysis is shown. This gives the user a good understanding on how these datasets and data elements look like in the real world.

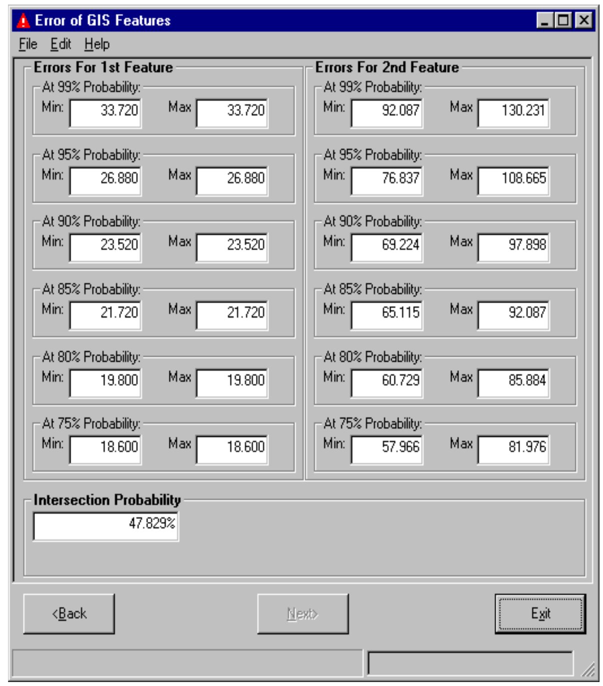
Example A-1

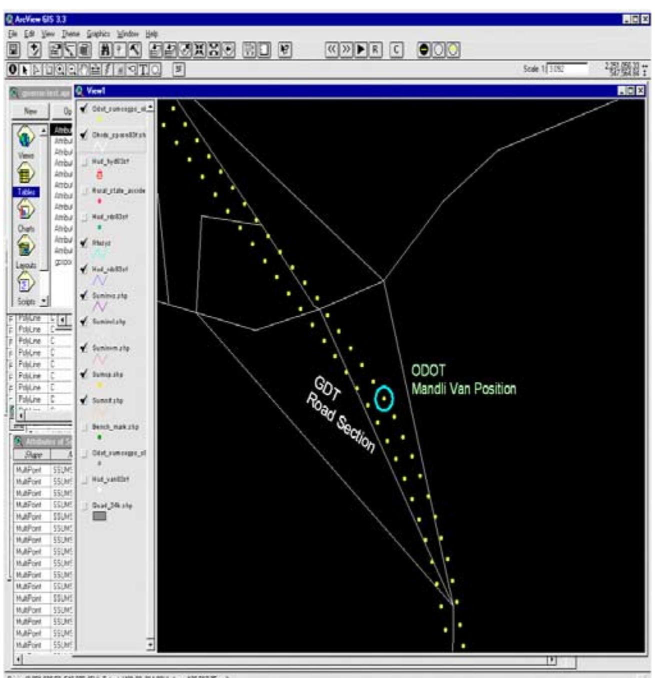
Point Feature
ODOT Mandli Van GPS Point

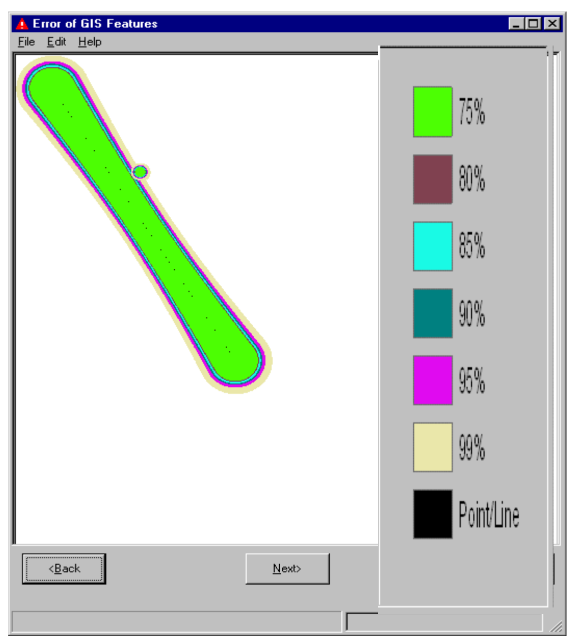
Line Feature
GDT Road Section

Coordinate System
State Plane, Ohio North, NAD 1983, US feet
x-shift: -2,250,000
y-shift: -545,000







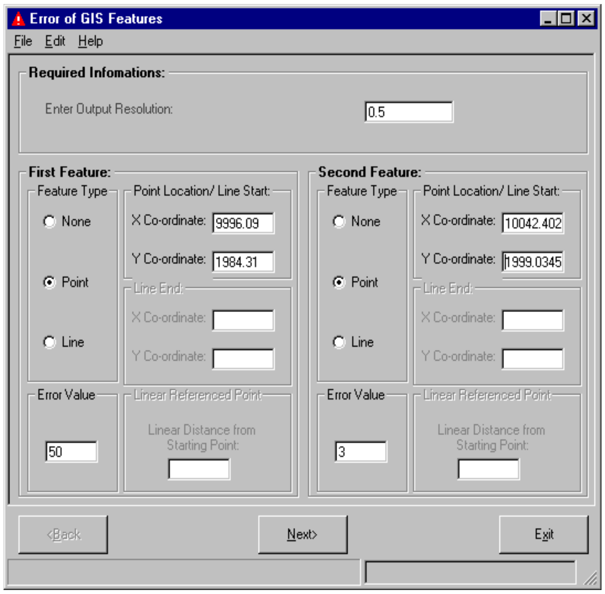


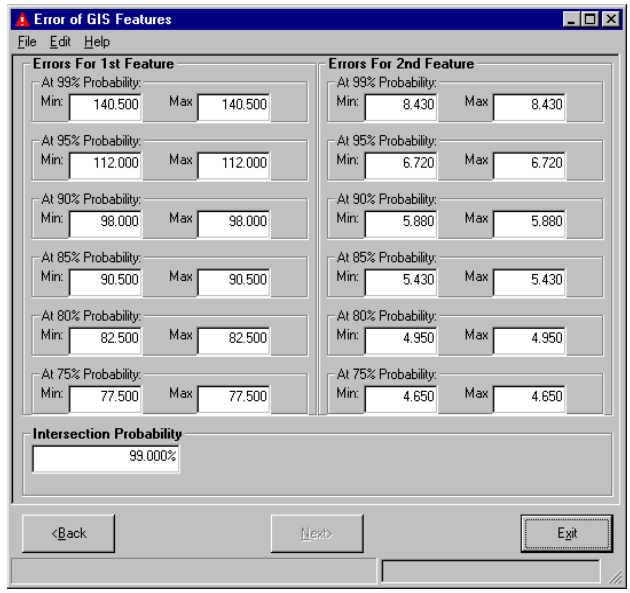
Example B-1

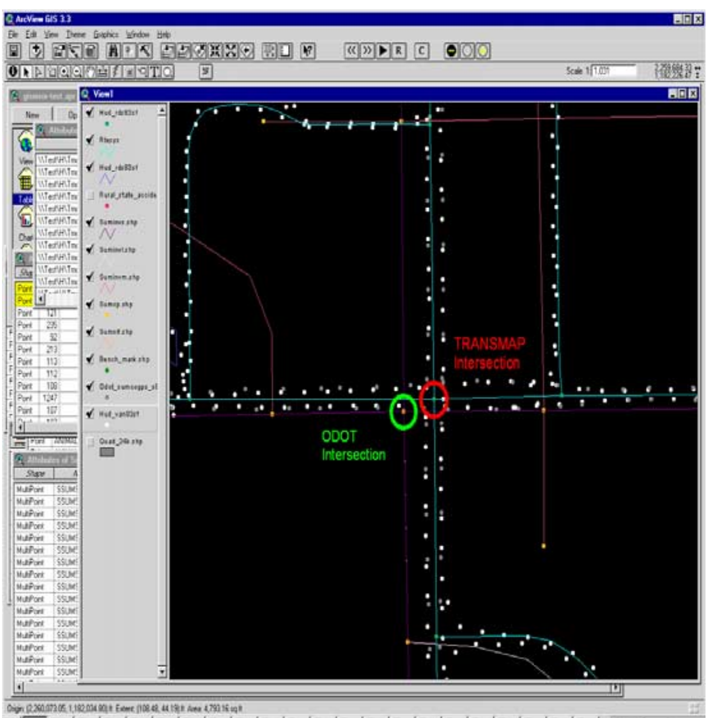
Point Feature 1
ODOT Intersection

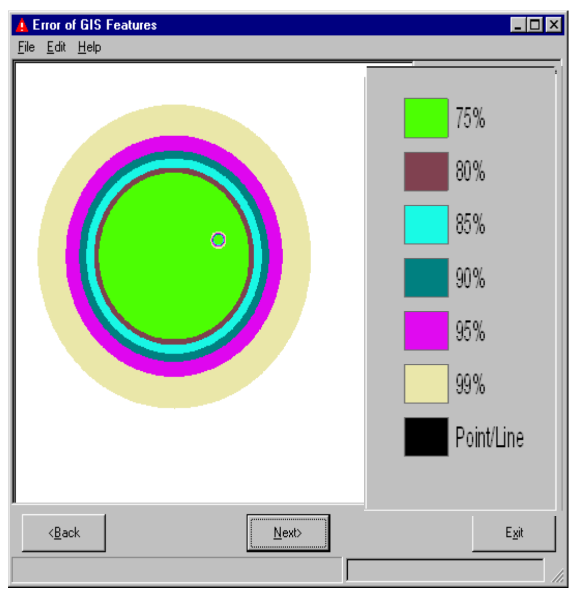
Point Feature 2
TRANSMAP Intersection

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,250,000
y-shift: -1,180,000







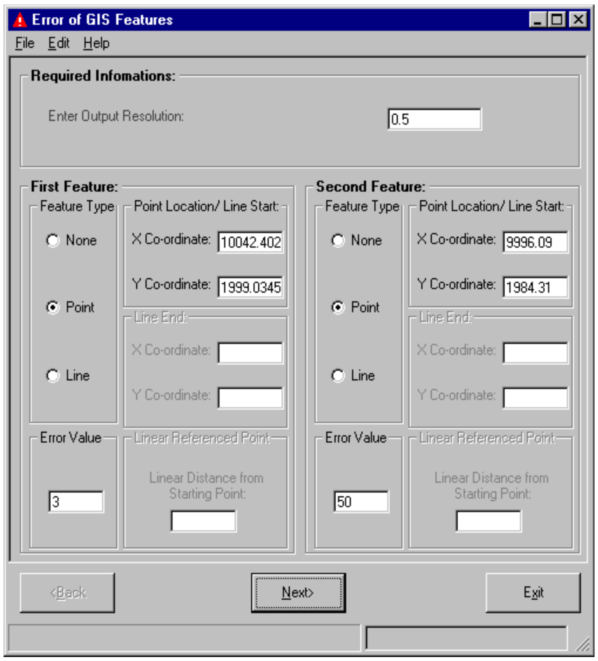


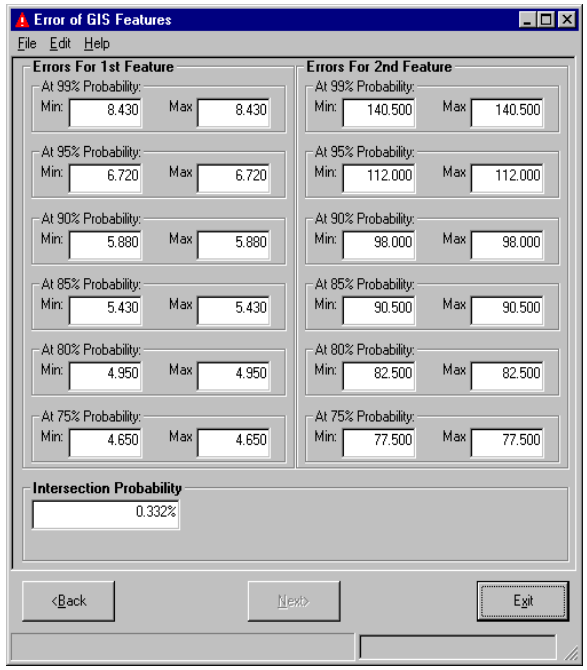
Example B-2

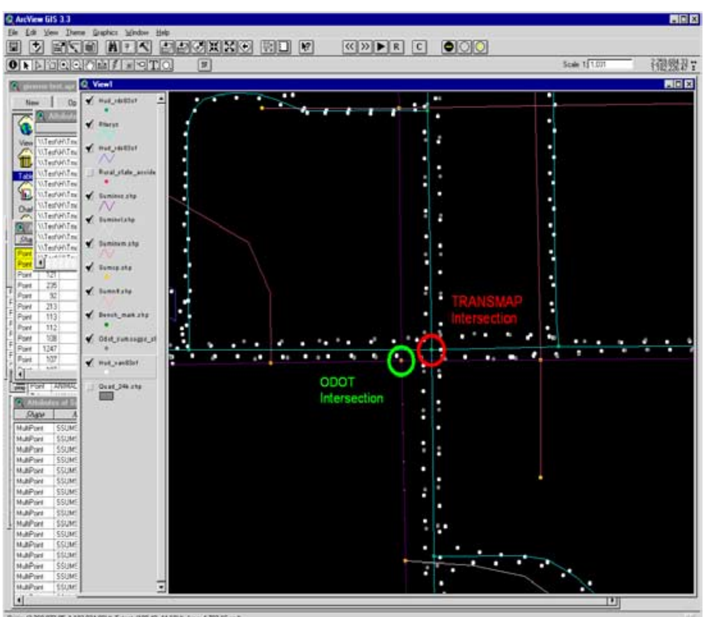
Point Feature 1
TRANSMAP Intersection

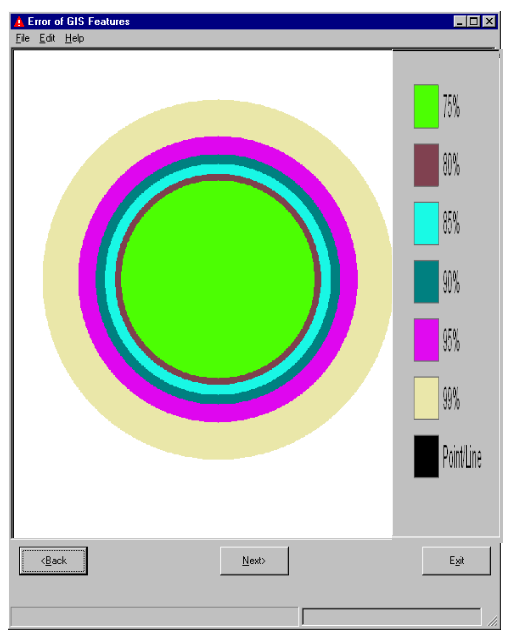
Point Feature 2
ODOT Intersection

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,250,000
y-shift: -1,180,000







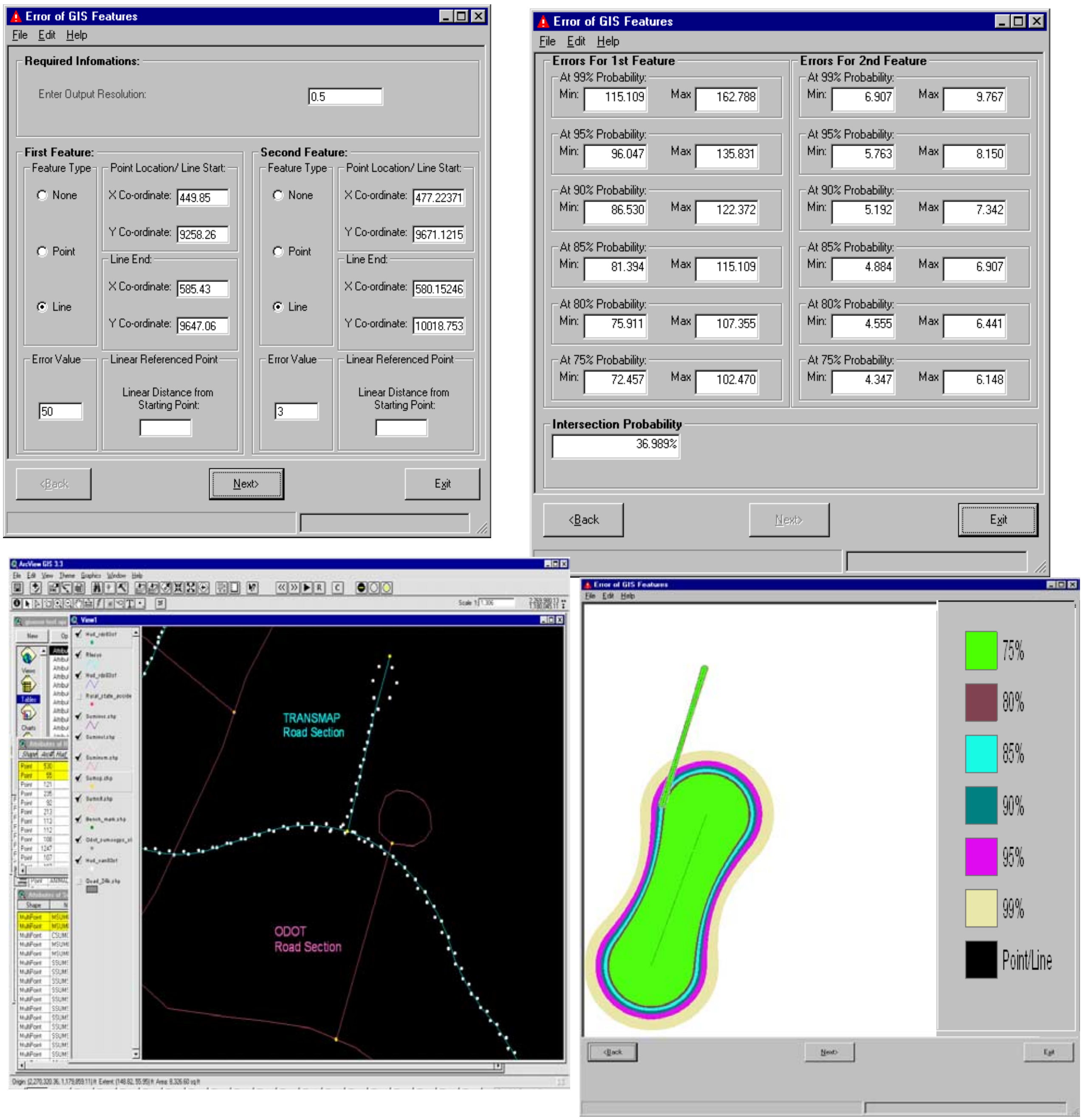


Example C-1

Feature 1
ODOT Road Section

Feature 2
TRANSMAP Road Section

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,270,000
y-shift: -1,170,000

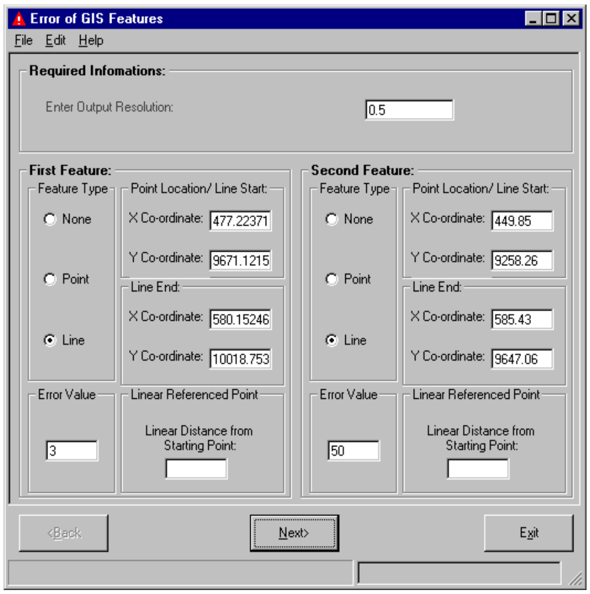


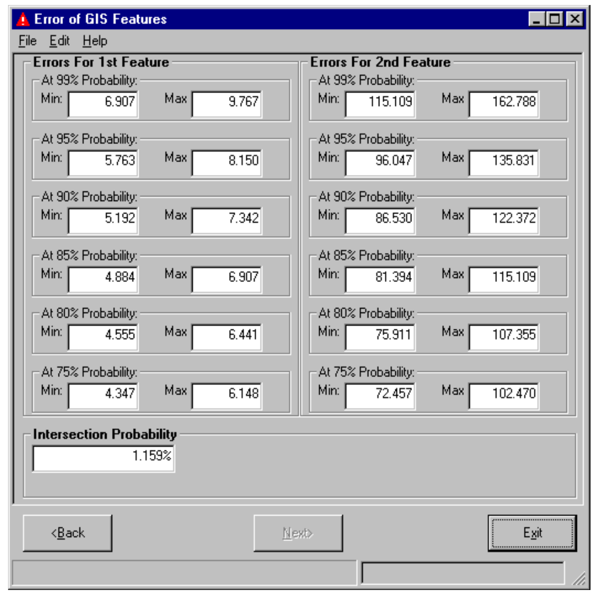
Example C-2

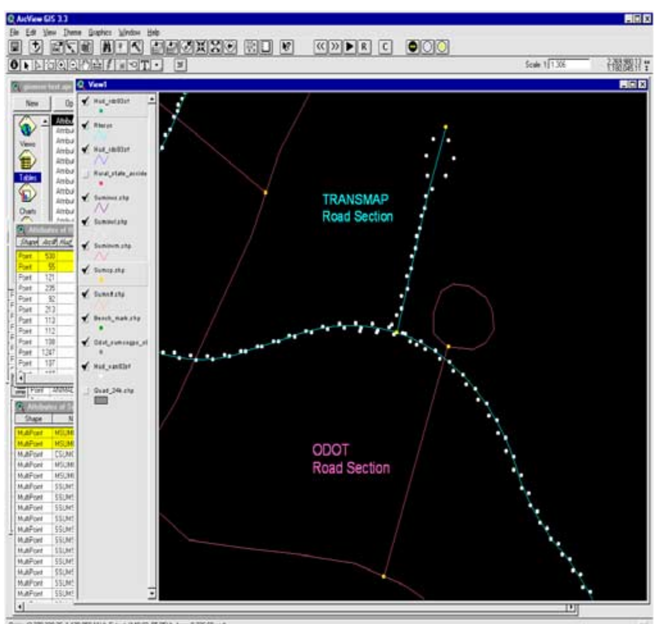
Feature 1
TRANSMAP Road Section

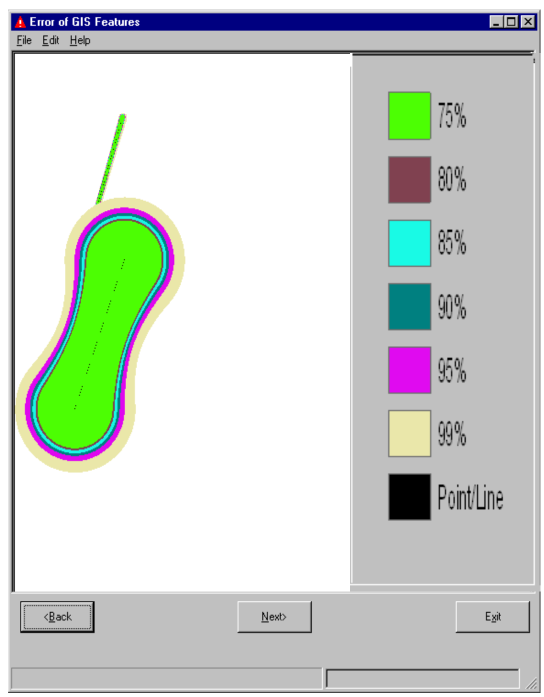
Feature 2
ODOT Road Section

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,270,000
y-shift: -1,170,000







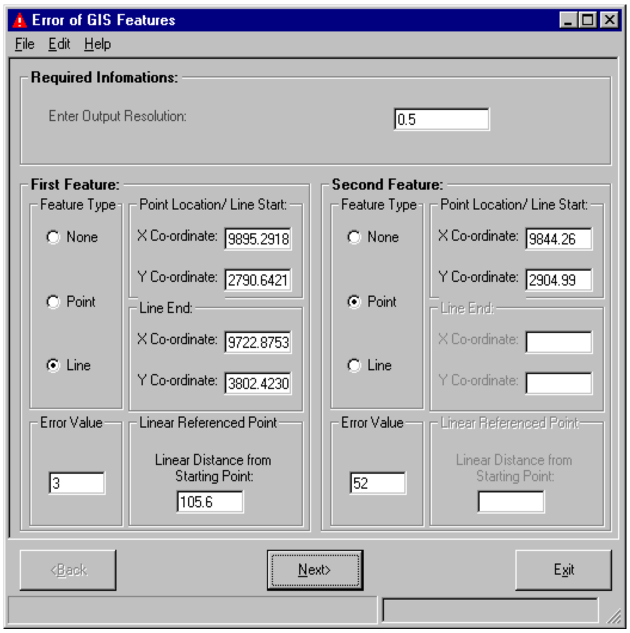


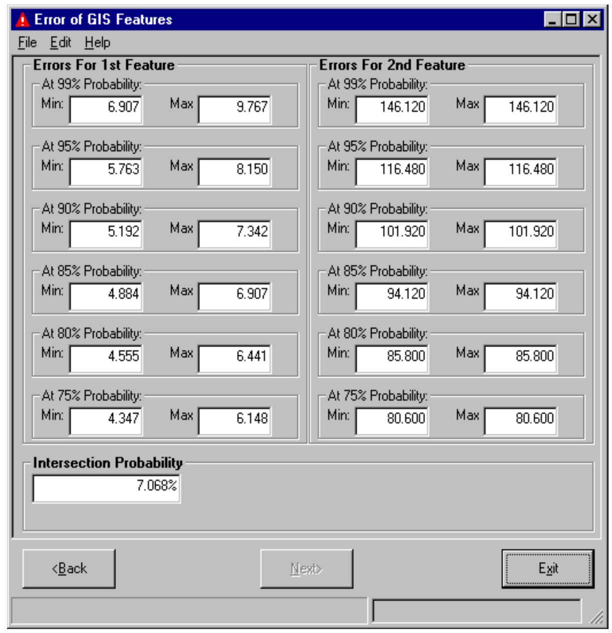
Example D-1

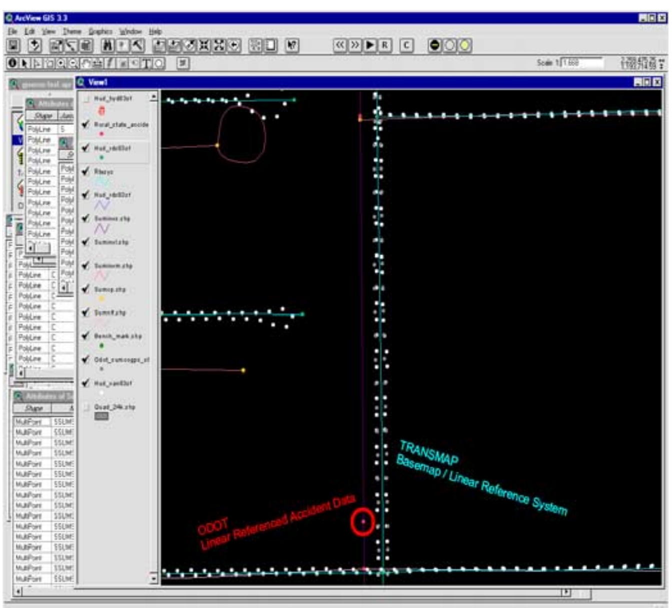
Linear Referenced Point Feature 1
TRANSMAP Road Section
& ODOT Accident Linear Distance

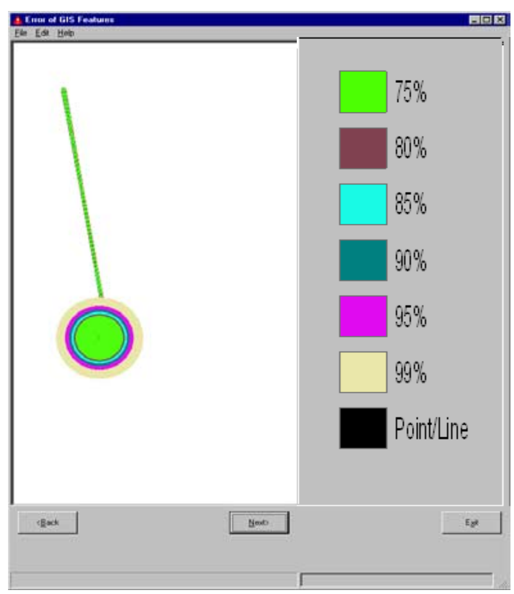
Point Feature 2
ODOT Accident Point

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,250,000
y-shift: -1,190,000







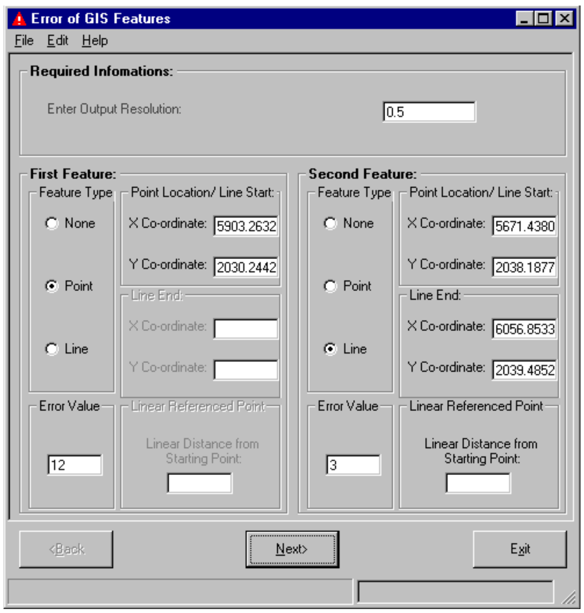


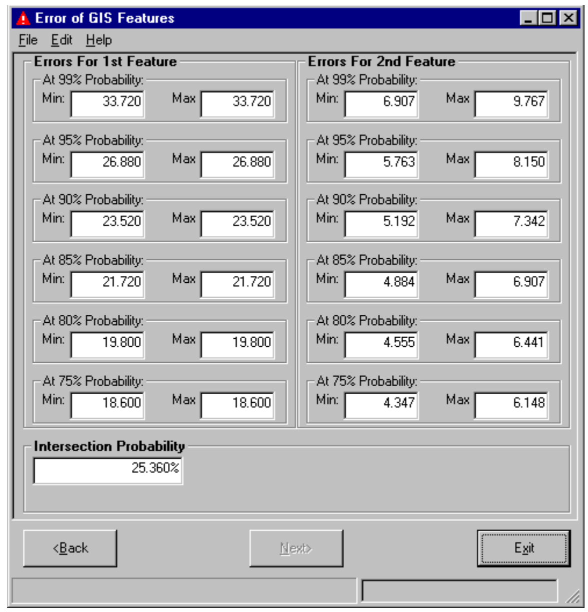
Example E-1

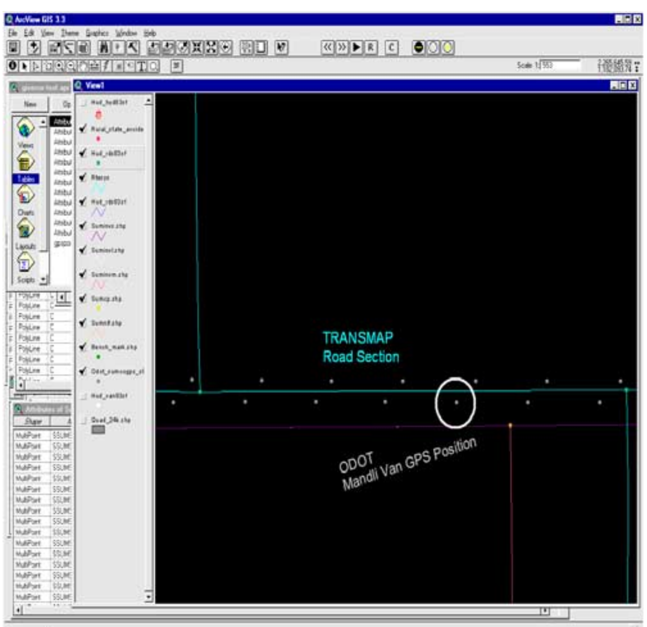
Point Feature 1
ODOT Mandli Van GPS Point

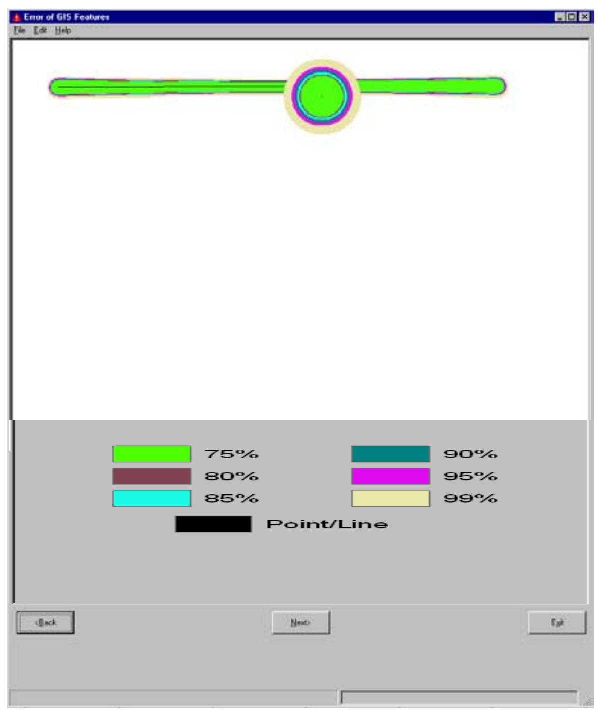
Line Feature 2
TRANSMAP Road Section

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,260,000
y-shift: -1,180,000







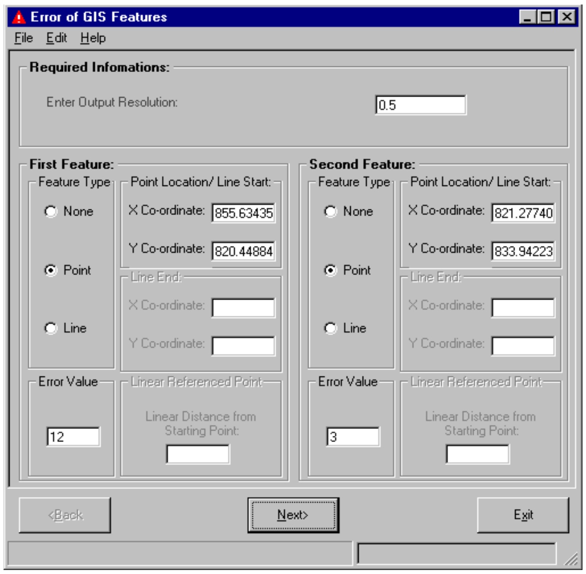


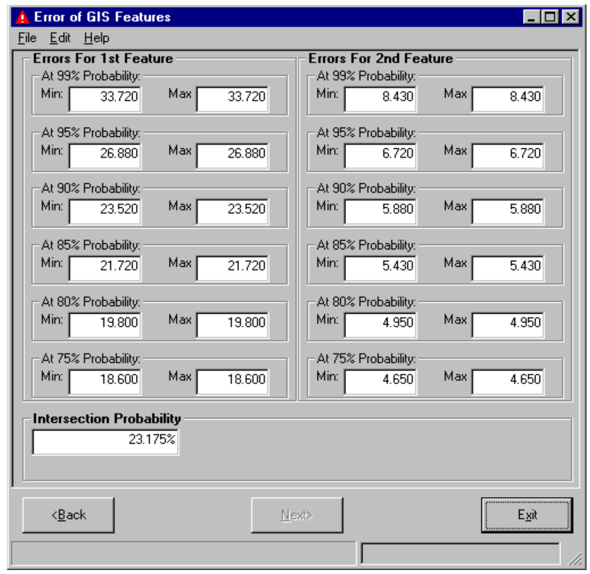
Example F-1

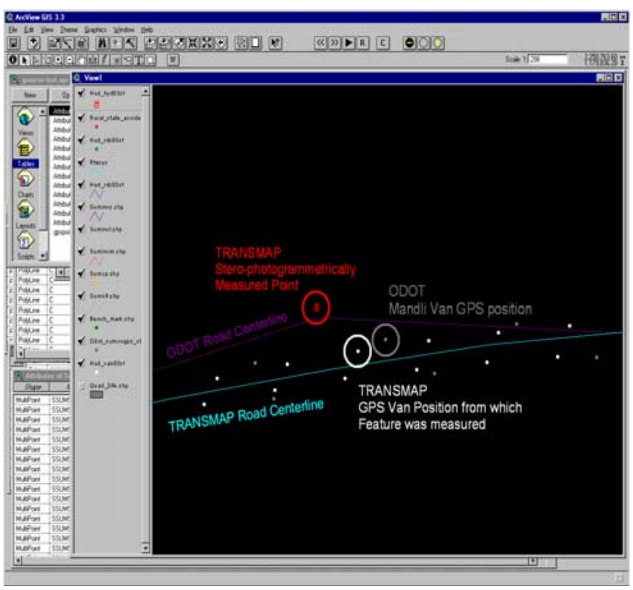
Point Feature 1
ODOT Mandli Van GPS Point

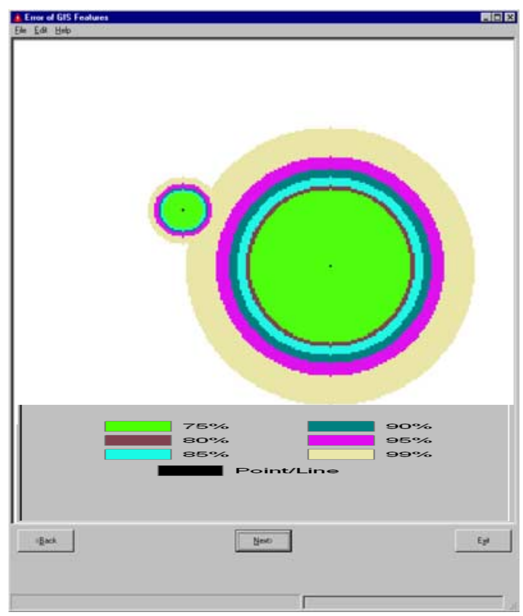
Point Feature 2
TRANSMAP Measured Point

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,250,000
y-shift: -1,179,000







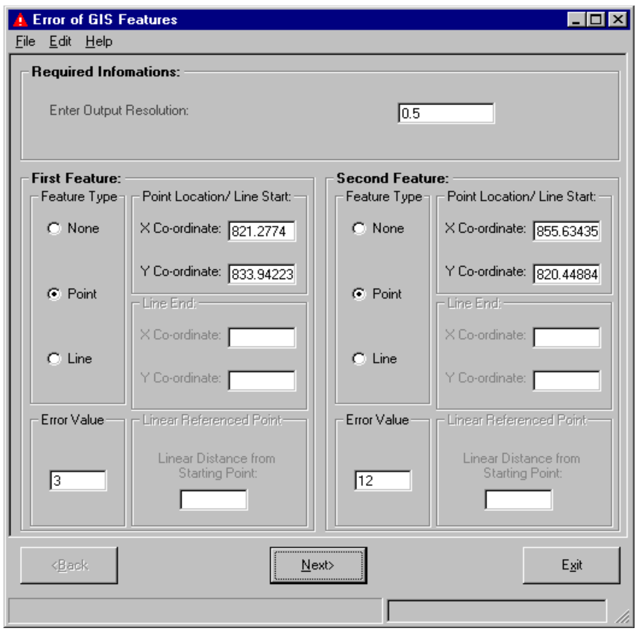


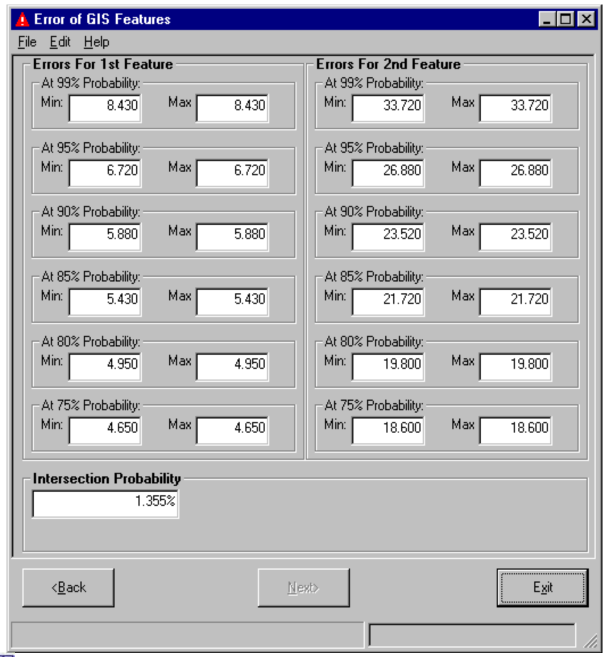
Example F-2

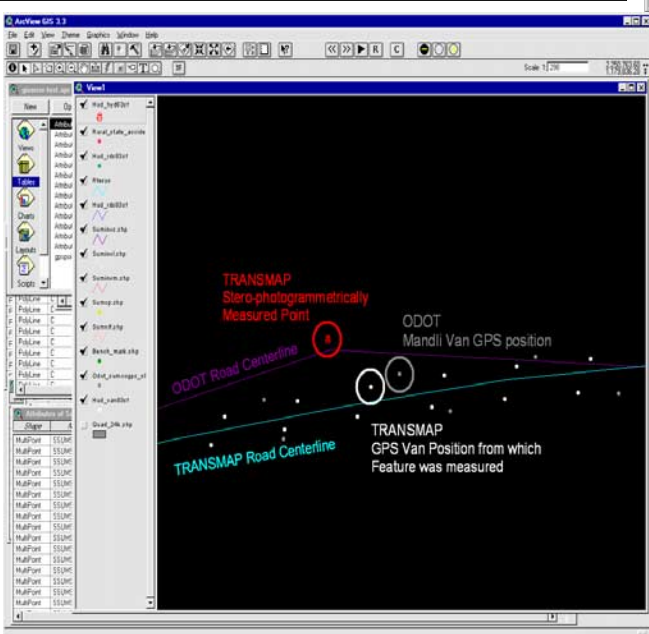
Point Feature 1
TRANSMAP Measured Point

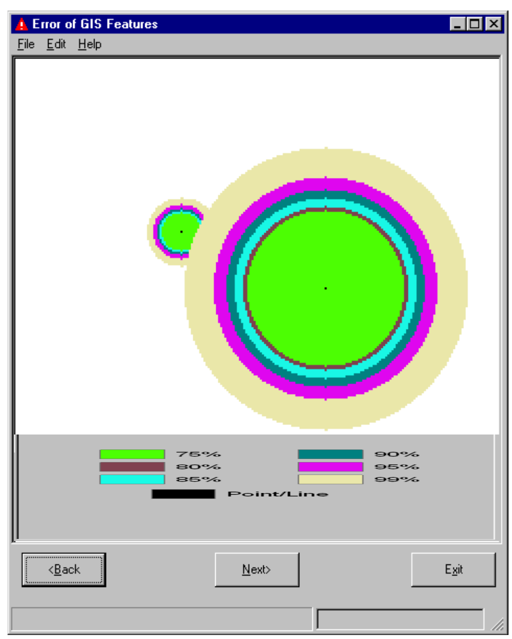
Point Feature 2
ODOT Mandli Van GPS Point

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,250,000
y-shift: -1,179,000







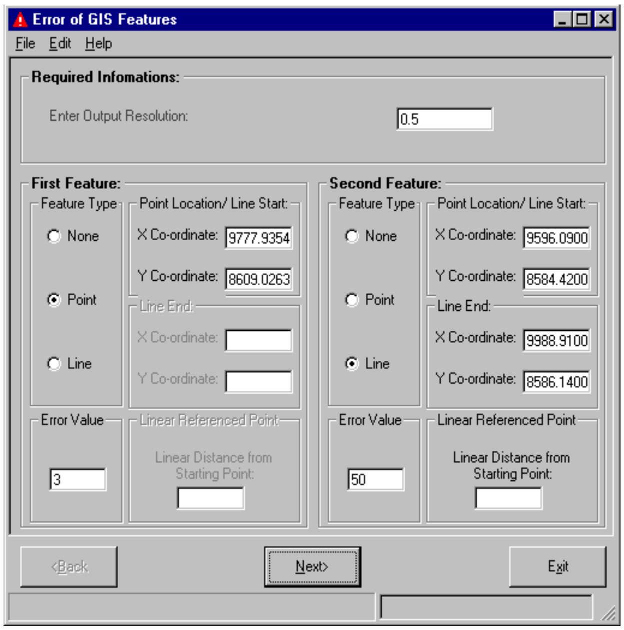


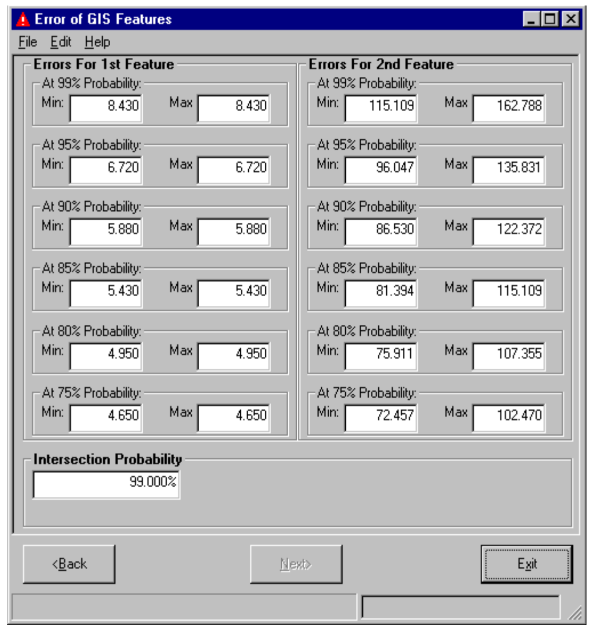
Example G-1

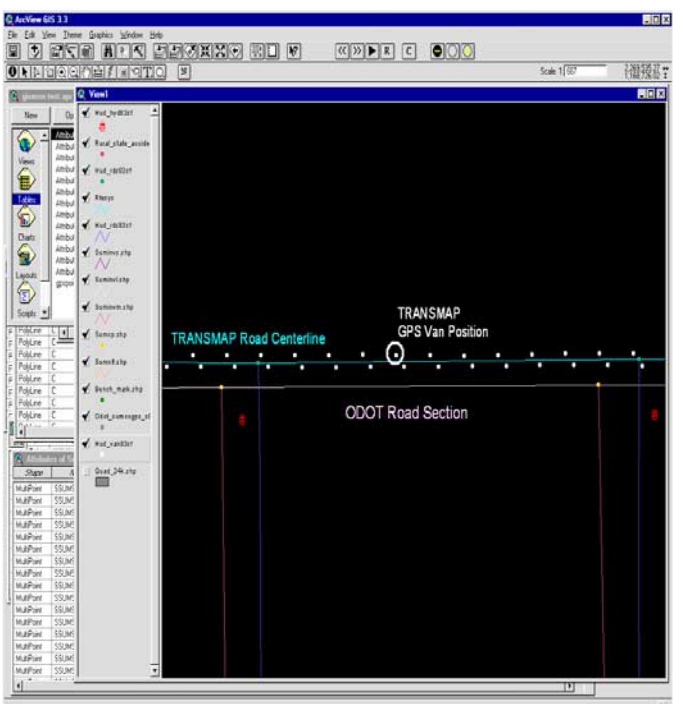
Point Feature 1
TRANSMAP Van Position

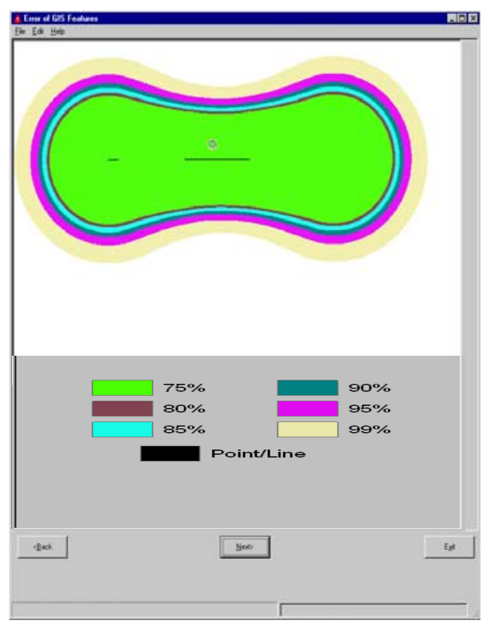
Line Feature 2
ODOT Road Section

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,260,000
y-shift: -1,160,000







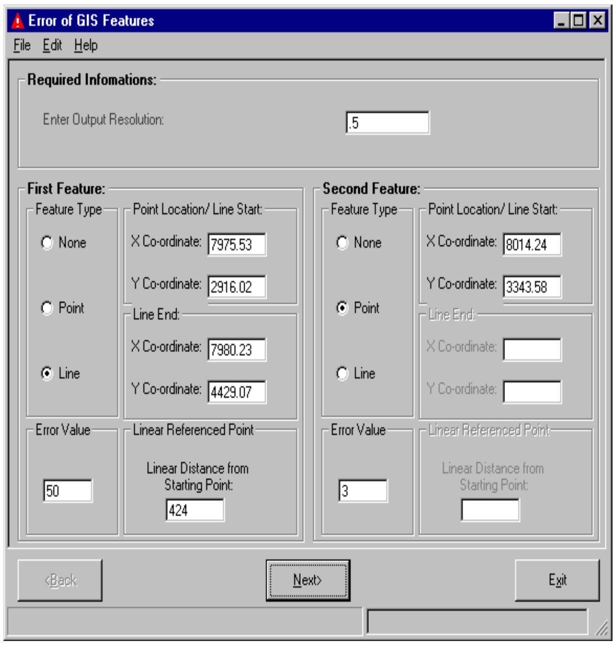


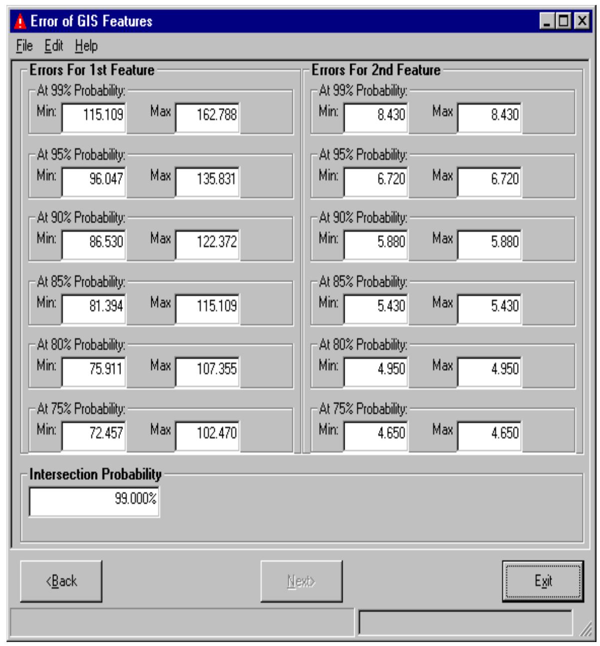
Example H-1

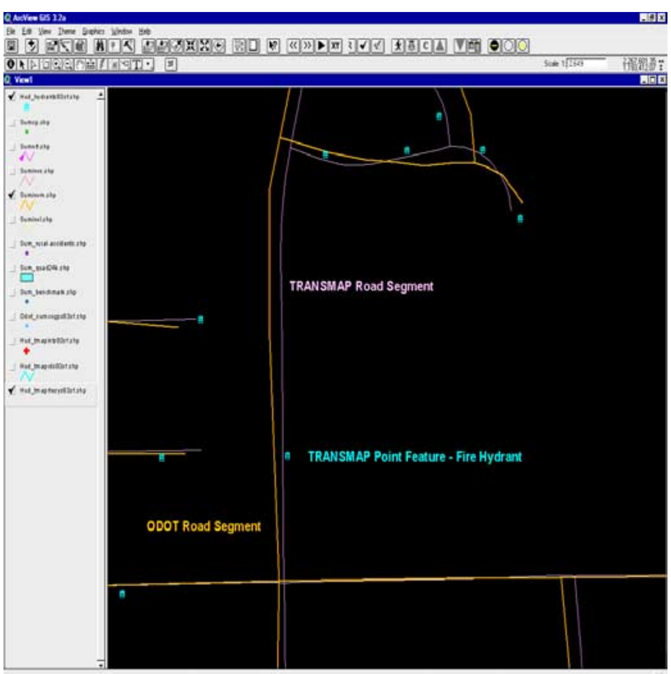
Line Feature
ODOT Road Section

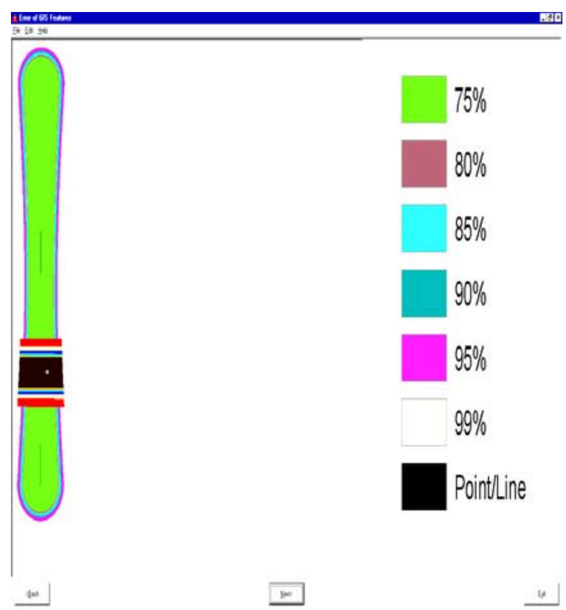
Point Feature
TRANSMAP Feature – Fire Hydrant

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,260,000
y-shift: -1,190,000







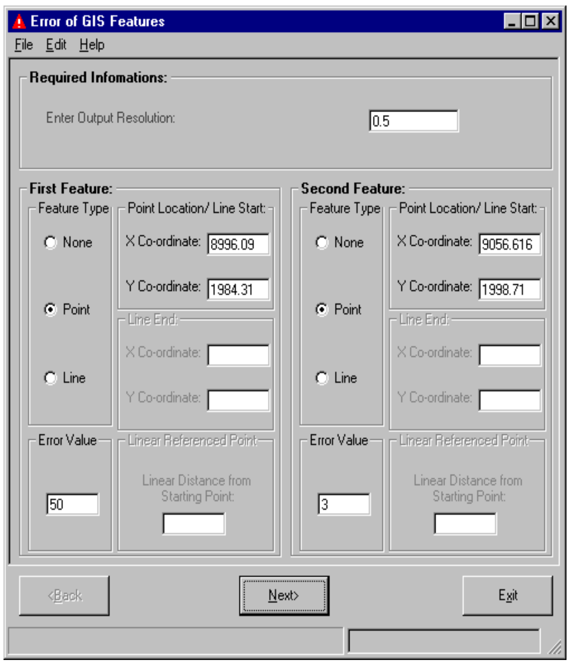


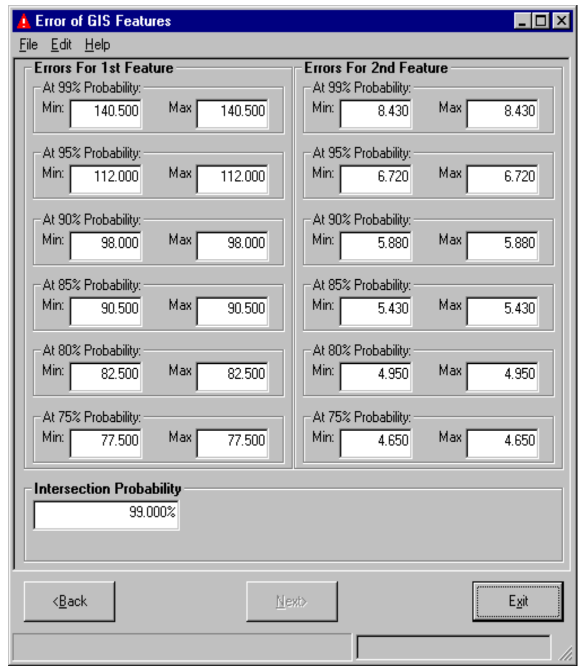
Example I-1

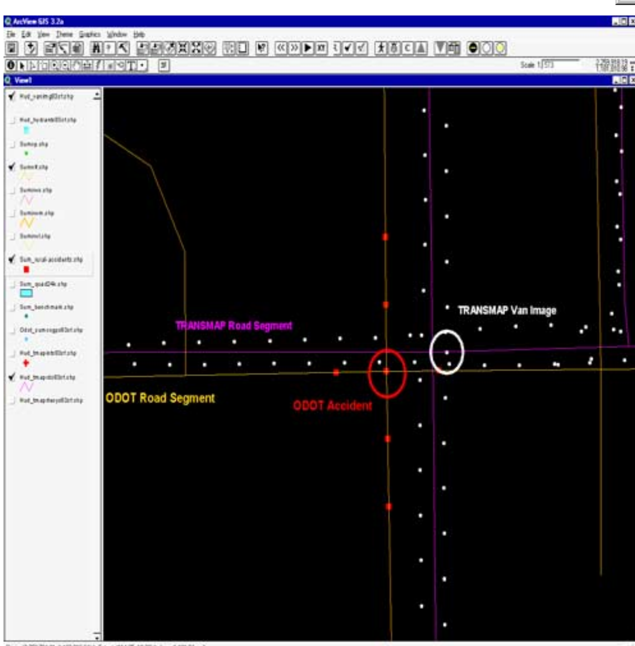
Linear Referenced
Point Feature 1
ODOT Accident

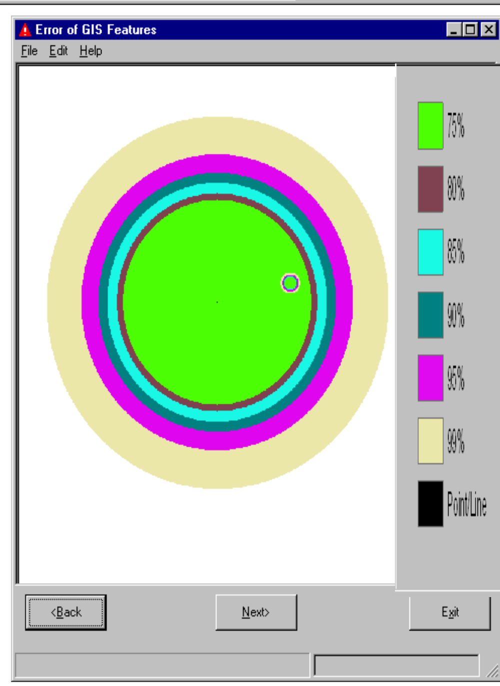
Point Feature 2
TRANSMAP GPS
Van Image

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,251,000
y-shift: -1,180,000







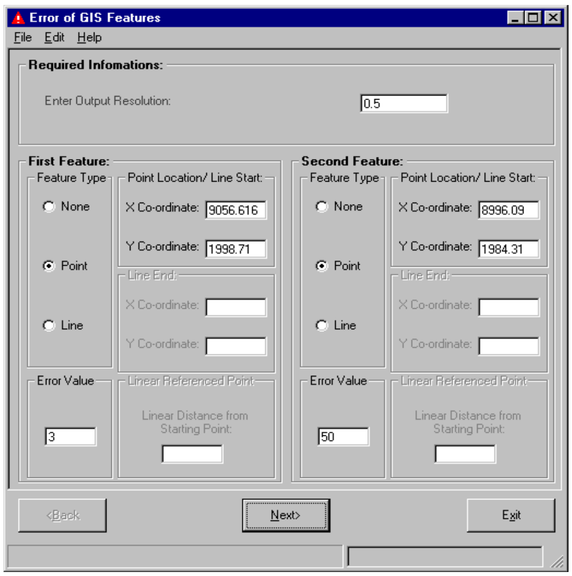


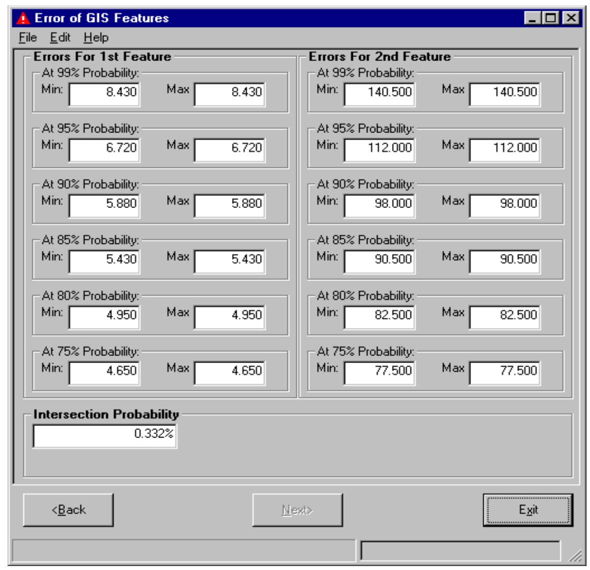
Example I-2

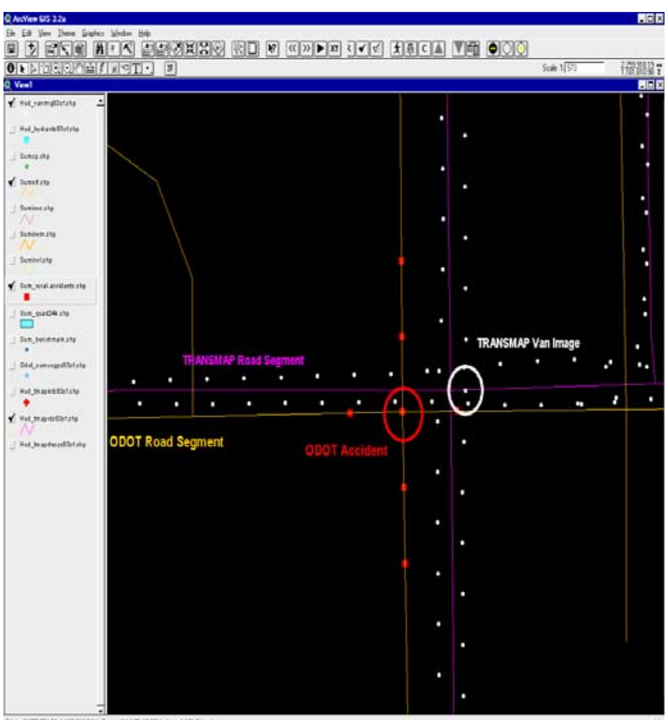
Point Feature
TRANSMAP GPS Van Image

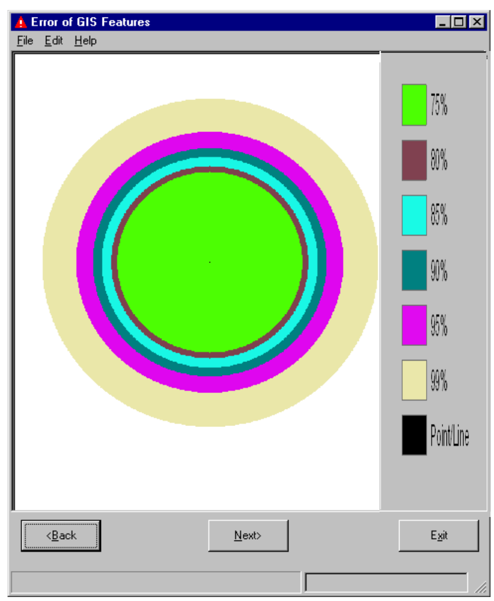
Linear Referenced Point Feature 1
ODOT Accident

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,251,000
y-shift: -1,180,000







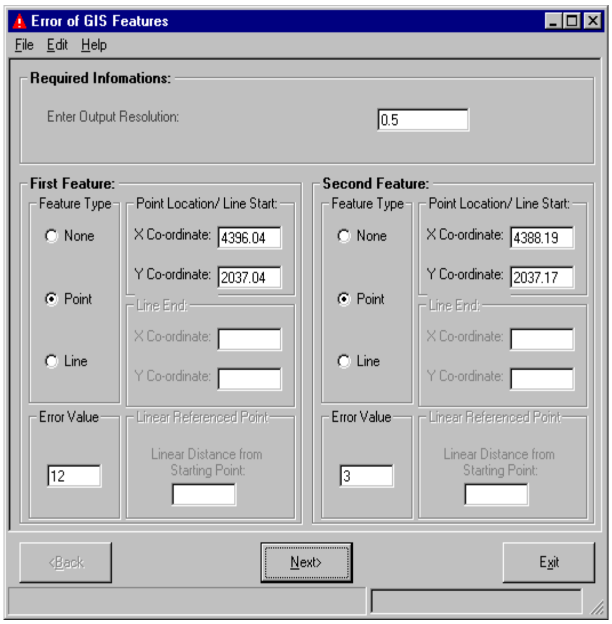


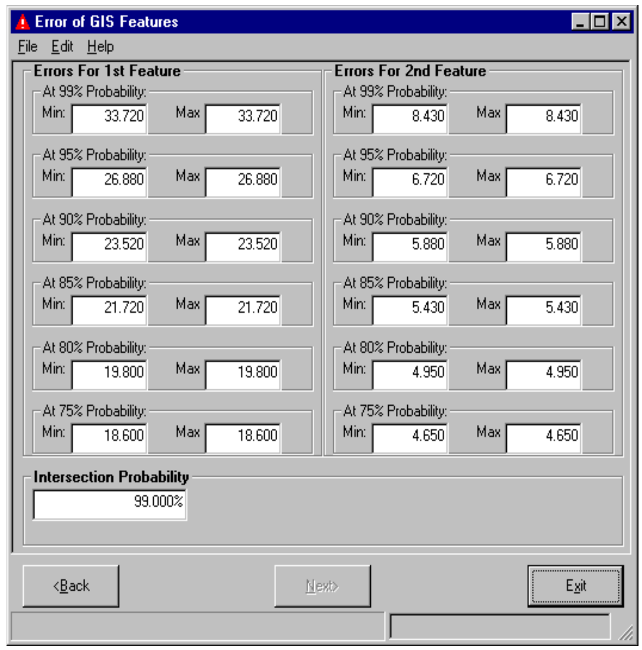
Example J-1

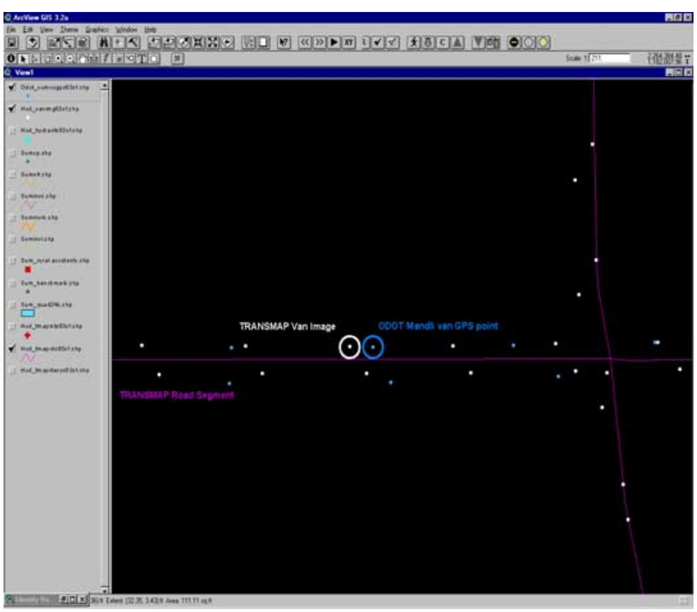
Point Feature 1
ODOT Mandli Van GPS Point

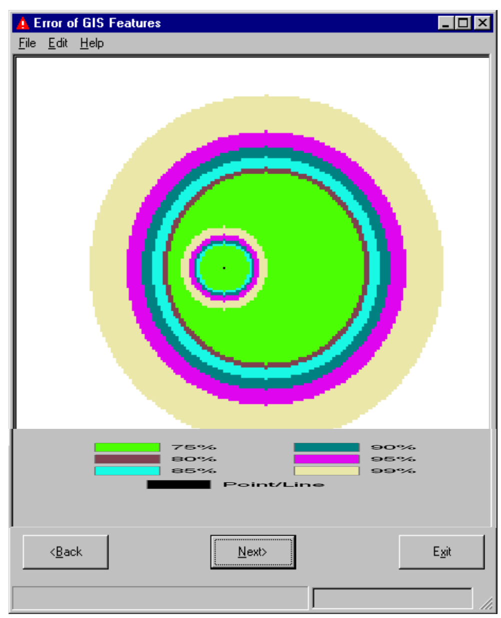
Point Feature 2
TRANSMAP GPS Van Image

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,260,000
y-shift: -1,180,000







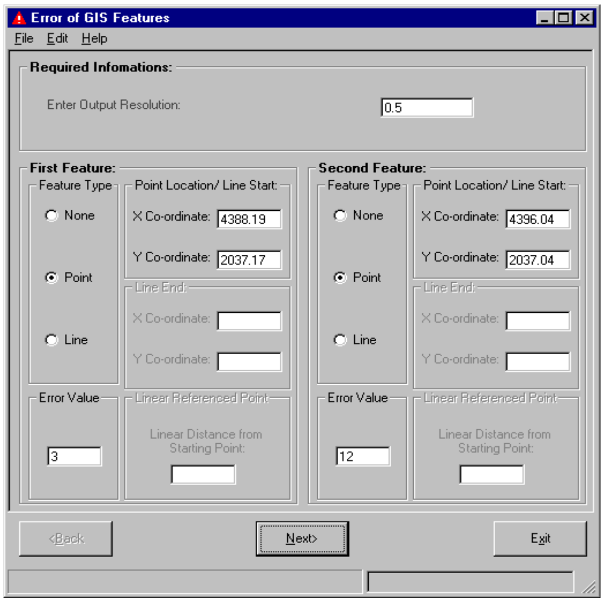


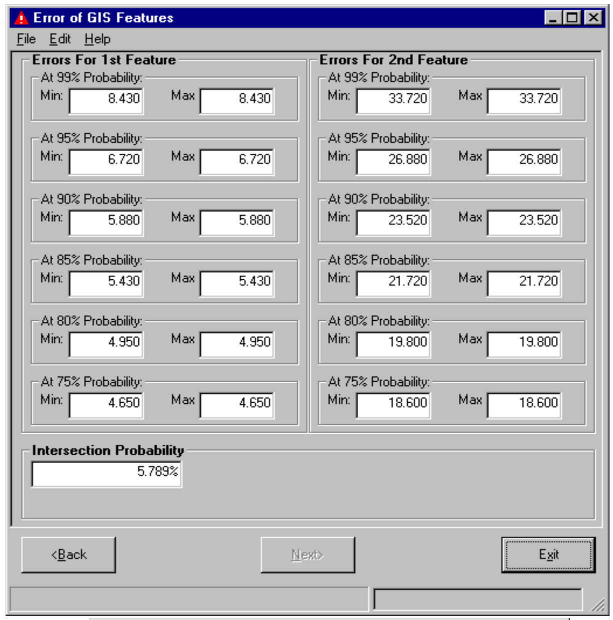
Example J-2

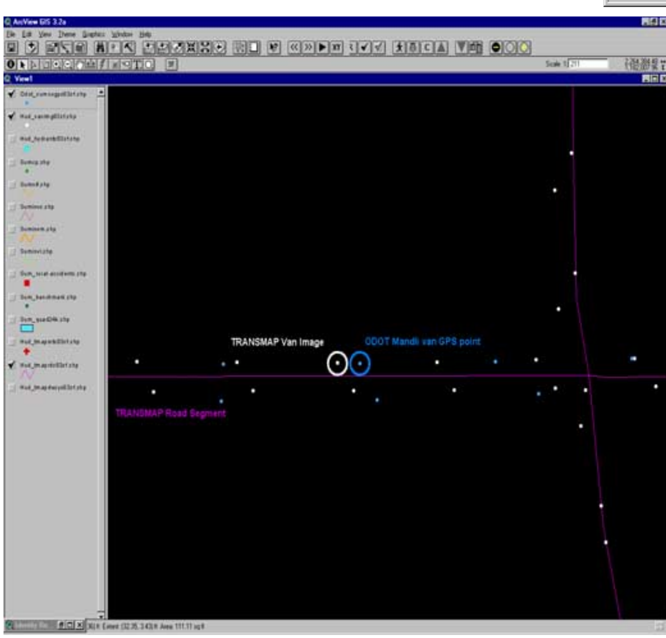
Point Feature 1
TRANSMAP GPS Van Image

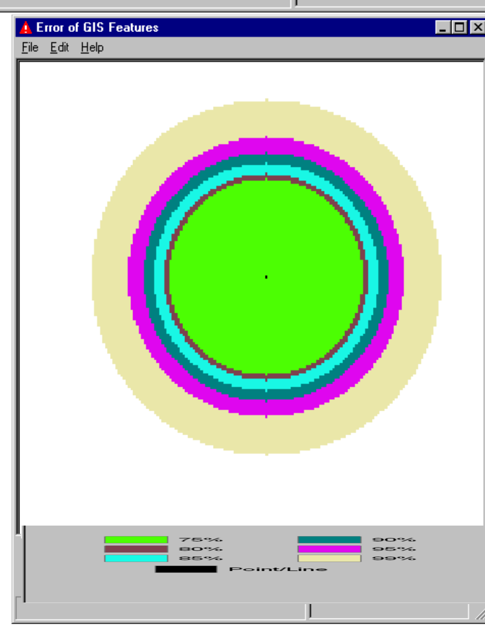
Point Feature 2
ODOT Mandli Van GPS Point

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,260,000
y-shift: -1,180,000







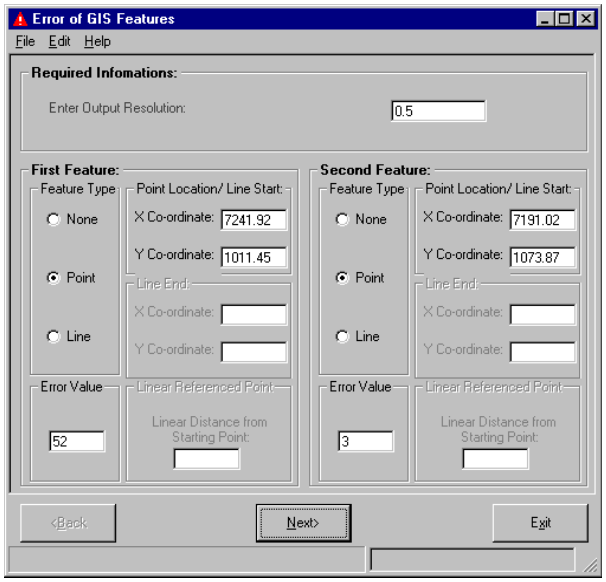


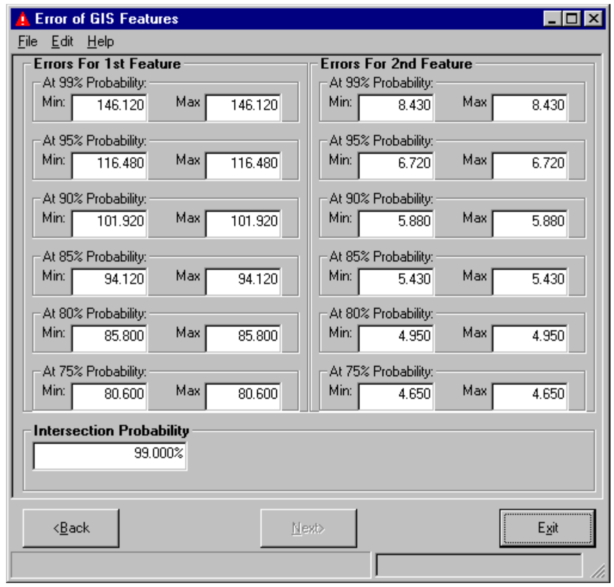
Example K-1

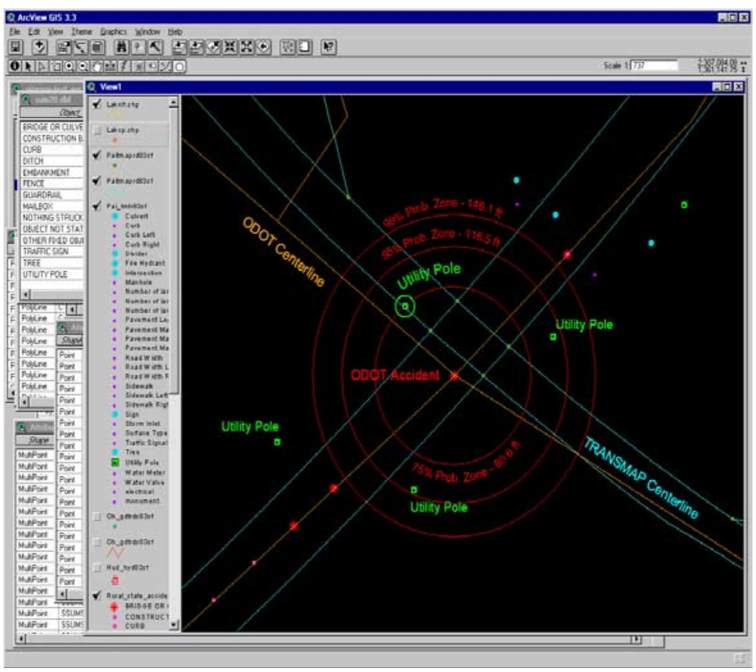
Point Feature 1
ODOT Accident

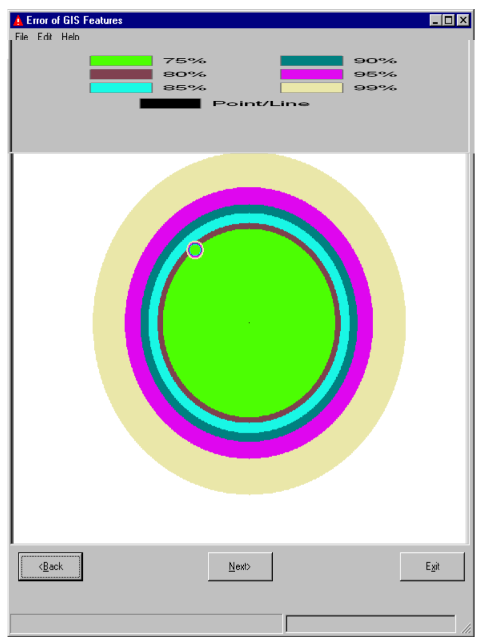
Point Feature 2
TRANSMAP Feature – Utility Pole

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,260,000
y-shift: -1,180,000







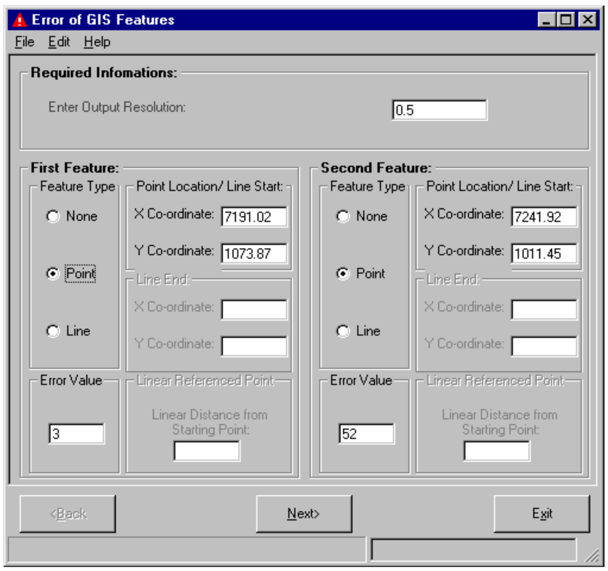


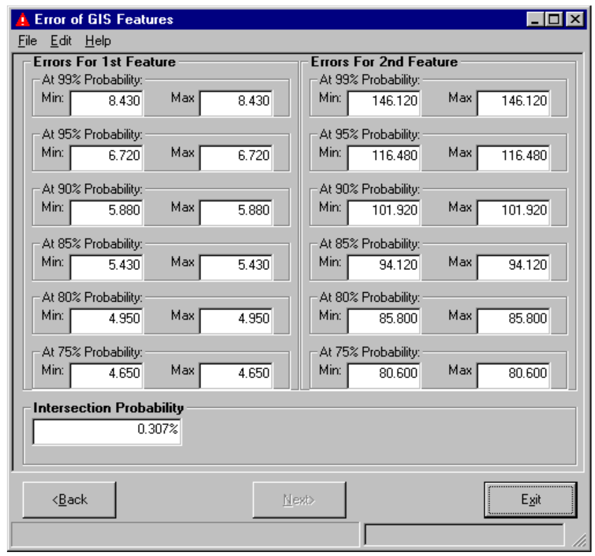
Example K-2

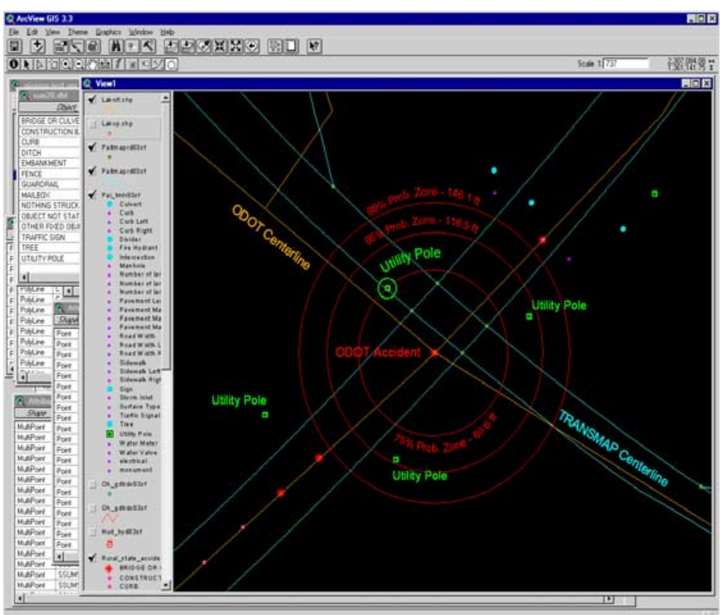
Point Feature 1
TRANSMAP Feature – Utility Pole

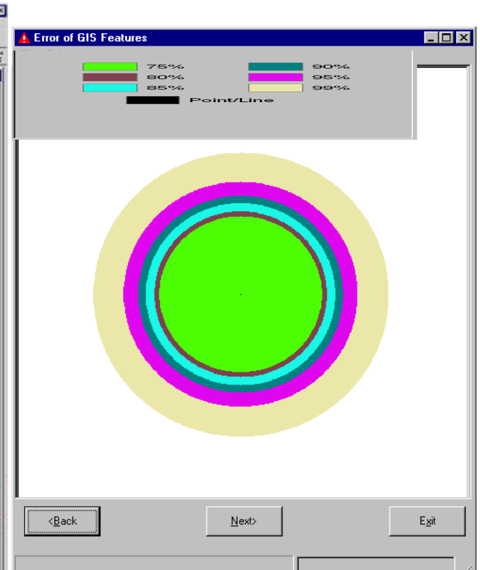
Point Feature 2
ODOT Accident

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,300,000
y-shift: -1,360,000







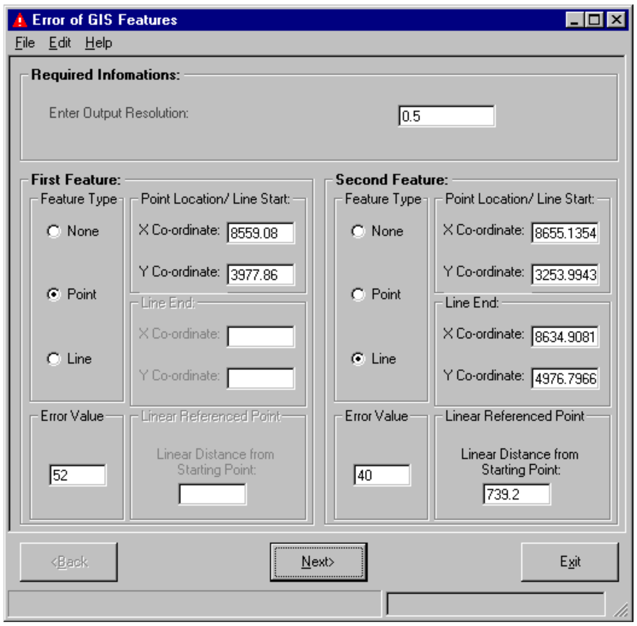


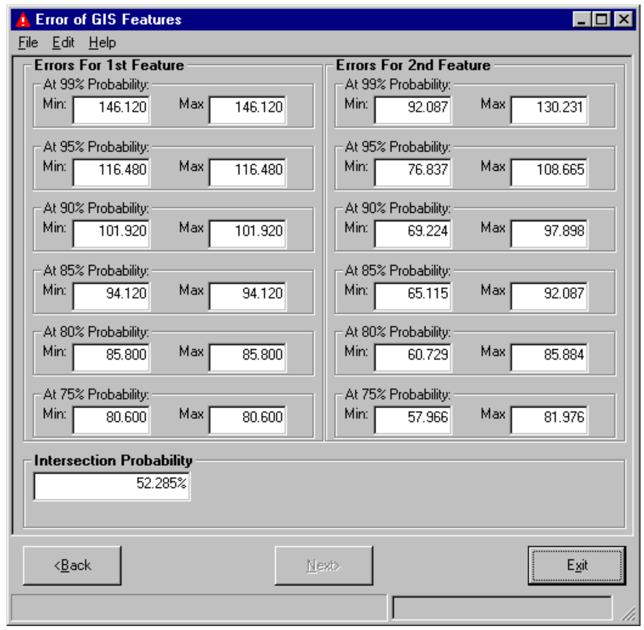
Example L-1

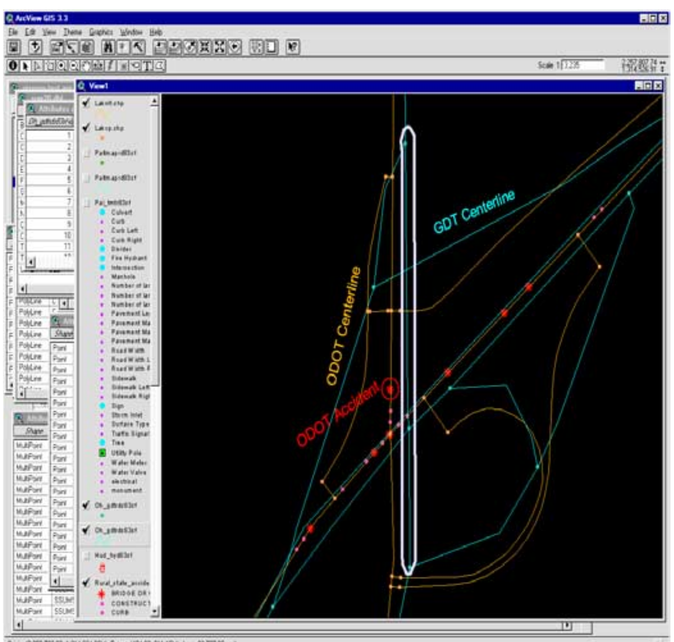
Point Feature 1
ODOT Accident

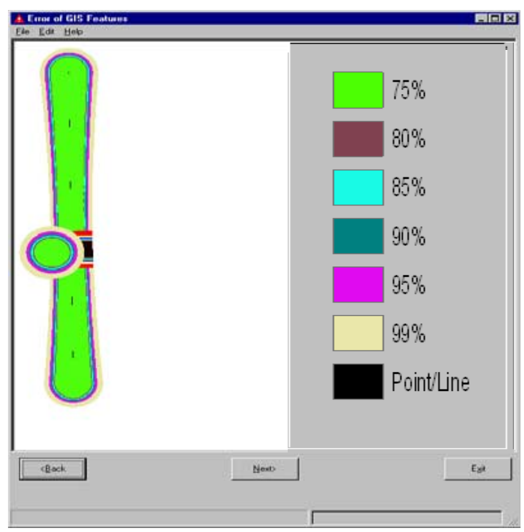
Line Feature 2
GDT Road Section and ODOT Accident Linear Distance

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,250,000
y-shift: -1,310,000





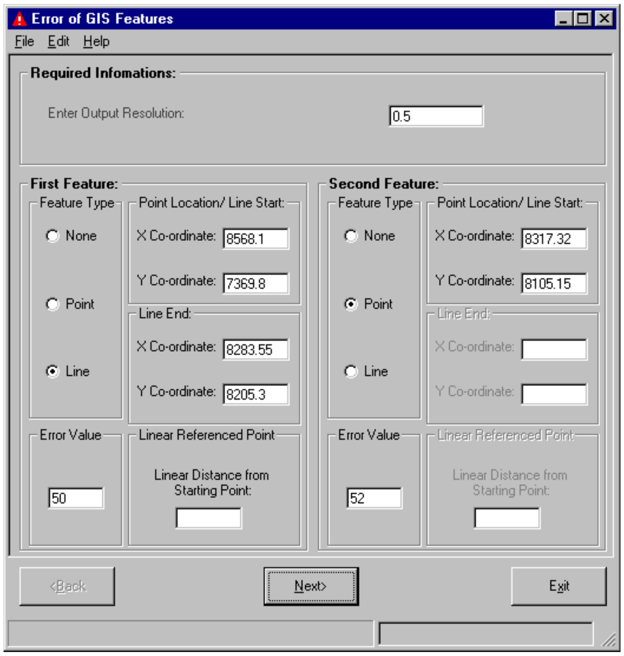


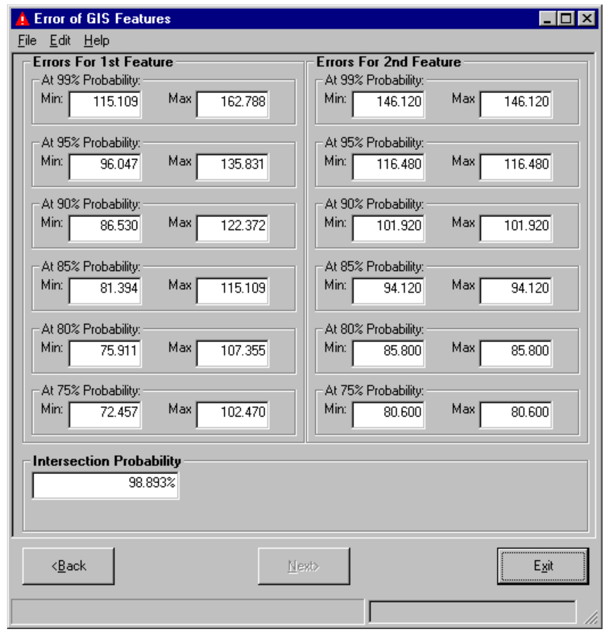


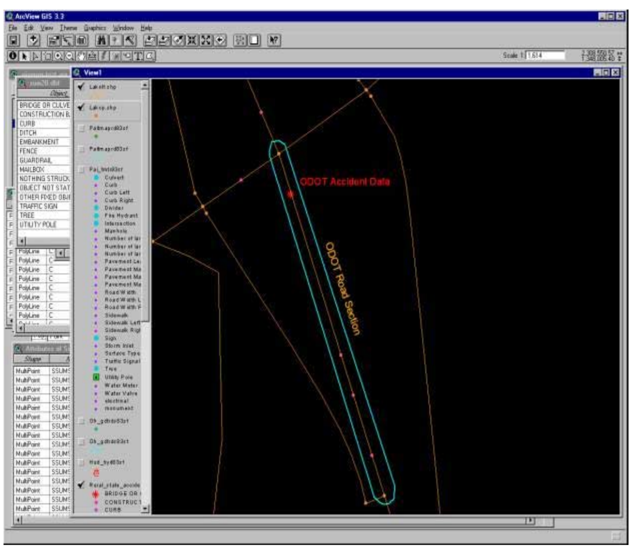
Example M-1

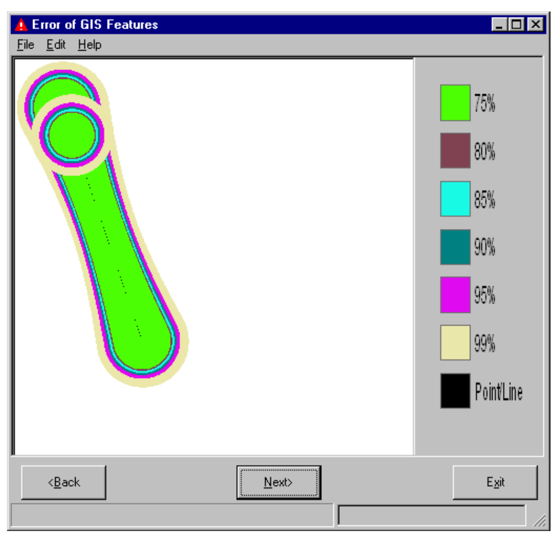
Line Feature1	**Point Feature2**
ODOT Road Section	ODOT Accident

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,300,000
y-shift: -1,340,000







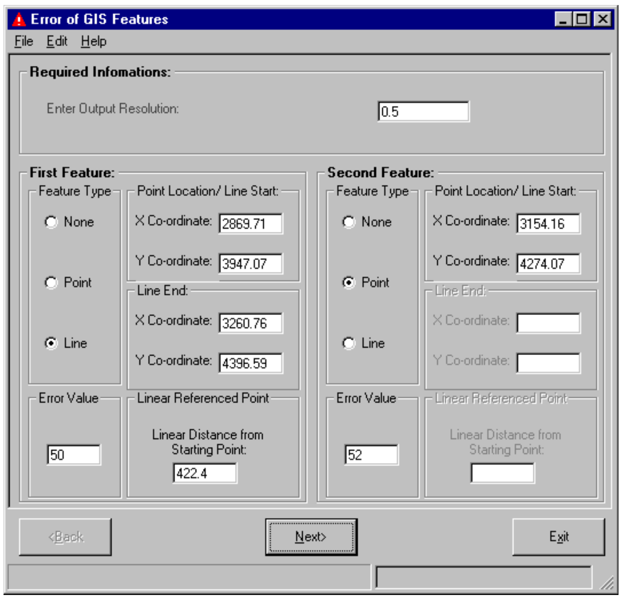


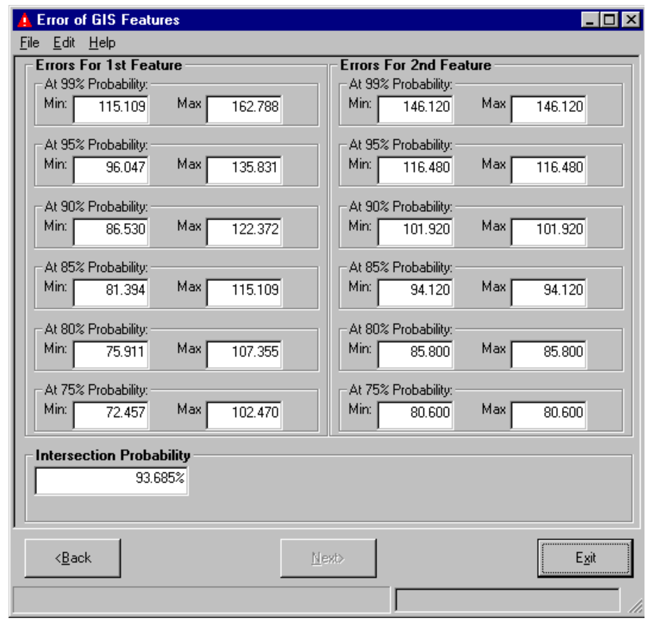
Example N-1

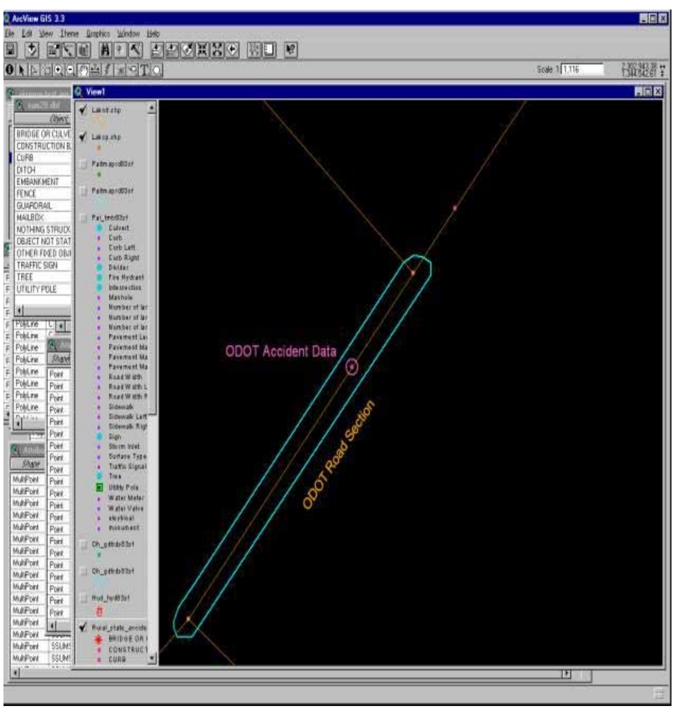
Line Feature1
ODOT Road Section

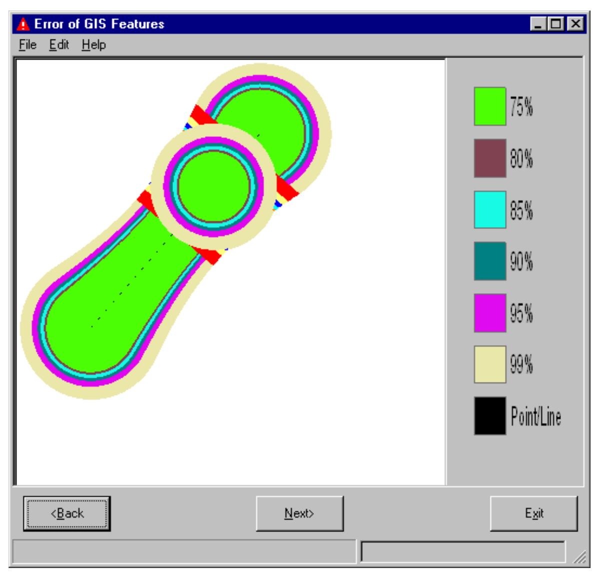
Point Feature2
ODOT Accident

Coordinate System
State Plane, Ohio South, NAD 1983, US feet
x-shift: -2,300,000
y-shift: -1,340,000









APPENDIX C

QUESTIONNAIRE FOR STATE DOT INTERVIEWS

BACKGROUND

This research study is sponsored by the National Cooperative Highway Research Program (NCHRP) to compile and develop information needed to address issues related to positional data accuracy and quality and to formulate methodologies to analyze their impacts or effects when considering trade-offs and/or transforming the positional data obtained from different measuring systems.

One of the critical components of this study is to assess the current and potential applications of spatial data and their sensitivity to positional and quality in various transportation applications. The mechanism for collection of this information is through structured interviews with selected state DOTs, MPOs, research and development organizations and other spatial data users. This interview guide is developed to facilitate the data collection effort and will be used to conduct in-person and telephone interviews. The specific objectives of the interview are to gather sufficient information on the following:

- Types of spatial data used and the types of transportation applications for which spatial data are used.
- Desired levels of accuracy and quality (including tolerance levels) of spatial data for the various applications.
- Sensitivity of transportation applications to accuracy and quality of spatial data.
- Effects of the quality of spatial data on the transportation applications.
- Potential and future applications of spatial data.

The questionnaire is divided into three parts.

Part I: General Organization Information

This section of the survey is intended to provide an overview of the organization and contacts for follow-up if clarification is required.

I-1 Organization Name ____

I-2 Type of organization

Federal ______, State _______, MPO _______, Other (specify) ___________________________

I-3 Who is the primary contact for spatial database products?

Name ____

Title ____

Branch/Division: ____

Address ____

Phone/Fax No ____

E-Mail ____

I-4 Who is the individual responsible for database standards within your organization (if different from I-3 above? Name, address, contact numbers, and e-mail

Part II: General Characteristics of Spatial Data

The following questions pertain to general characteristics of spatial data use

II-1 List the types of spatial data products used or maintained by your organization that is used for various transportation application. For each data product, provide the following (**You may use Table 1):**

- Name (and description) of product
- Source of data product
- Nominal scale
- Coverage area
- Applications or uses
- Desired level of resolution or precision
- Tolerance level
- Sensitivity of application to data errors

II-2 Describe the metadata records contained in the products described in Question II-1 (where applicable). Identify the metadata standard and the content for each data product.

____
____
____
____

II-3 What measuring techniques are used by your organization to collect or develop its own spatial data (GPS, Geodetic Survey, Air Photos)? For each data element (e.g., Line (street), Point (accident), Polygon Area (TAZ)) indicate (**You may use Table 2**):

- spatial data element
- measuring technique
- measuring instrument
- instrument error/precision
- resolution or precision for reporting.

II-4 Describe any decisions made based on the analysis of spatial data. Indicate whether the quality of data meets the minimum standards required by your organization?

____
____
____
____

II-5 Describe any adverse effects of using spatial data that does not meet your minimum quality standards or data with uncertain accuracy.

____
____
____
____

II-6 What are the plans by your organization for using the spatial data for emerging and future GIS-T applications. List the anticipated GIS-T applications

____
____
____
____

II-7 List the spatial data model and standard (National Standard for Spatial Data Accuracy—NSSDA) that your organization follows

____
____

Part III: Linear Referencing Methods and Data Standards

This section deals with the Linear Referencing Database Models and Standards that either have been developed or supported by the organization.

III-1 Describe the digital ground transportation network data models and standards that your organization developed or maintains and their capabilities.

- Name of data model or standard
- Capabilities e.g.,

1) Linear Referencing capabilities Yes _____ No ______
2) Dynamic segmentation capabilities Yes _____ No ______
3) Routing capabilities Yes _____ No ______
4) Address or location geo-coding capabilities Yes _____ No ______

III-2 Indicate the transportation network data models and standard that your organization has adopted. Also, indicate if these data models are based upon current or proposed NCHRP LRS data standard (NCHRP 20-27).

III-3 List the linear referencing method(s) (LRM) used by your organization to collect locational data for transportation applications. Identify typical attribute data collected associated with each method (**You may use Table 3**).

III-4 What measurement techniques are used by your organization to collect the LRS data? List the <u>resolution and precision</u> of each of the offset needed to report locations (e.g., 0.1, 0.01, 0.001) and <u>measurement position</u> (e.g., along the centerline or along the shoulder etc.) (**You may use Table 3).**

III-5 Describe any transformation techniques or algorithms that you use to transform between different linear referencing methods.

III-6 What are the typical applications of LRMs in your organization? What GIS-T applications do you currently have that would be enhanced/supported by the LRS? List them in order of priority.

III-7 What future applications do you feel should be supported by the LRS? List them with your top priority.

III-8 Is any of the application under III-6 and III-7 is sensitive to positional accuracy of the LRS network model? Please list the error range acceptable for the application.

Table 1.

	Name of data product	Description	Source of data	Nominal scale	Coverage	Application or uses	Desired resolution	Tolerance levels	Sensitivity of application to error
e.g.	NHPN	Highway	ONRL	1 : 100,000	U.S.	Planning	1-100m		

Table 2.

Spatial data element	Example	Measuring technique	Measuring instrument	Instrument precision (measurement error)	Precision	Remarks
line	street	Geodetic	chain			
Point	accident					
Polygon area						

Table 3.

LRM	Attribute	Measurement technique	Precision	Measurement Position	Application

Return completed questionnaire to:

Dr. Edward Fekpe
Transportation Division
Battelle
fekpee@battelle.org

Abbreviations used without definitions in TRB publications:

AASHO	American Association of State Highway Officials
AASHTO	American Association of State Highway and Transportation Officials
APTA	American Public Transportation Association
ASCE	American Society of Civil Engineers
ASME	American Society of Mechanical Engineers
ASTM	American Society for Testing and Materials
ATA	American Trucking Associations
CTAA	Community Transportation Association of America
CTBSSP	Commercial Truck and Bus Safety Synthesis Program
FAA	Federal Aviation Administration
FHWA	Federal Highway Administration
FMCSA	Federal Motor Carrier Safety Administration
FRA	Federal Railroad Administration
FTA	Federal Transit Administration
IEEE	Institute of Electrical and Electronics Engineers
ITE	Institute of Transportation Engineers
NCHRP	National Cooperative Highway Research Program
NCTRP	National Cooperative Transit Research and Development Program
NHTSA	National Highway Traffic Safety Administration
NTSB	National Transportation Safety Board
SAE	Society of Automotive Engineers
TCRP	Transit Cooperative Research Program
TRB	Transportation Research Board
U.S.DOT	United States Department of Transportation