THE UNIVERSITY OF CALGARY

MODELING AND VISUALIZATION OF POSITIONAL UNCERTAINTY IN GIS AND ITS APPLICATION TO VISAT

by

Jianghong Zhou

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF GEOMATICS ENGINEERING

CALGARY, ALBERTA

DECEMBER, 1995

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "The Modeling and Visualization of Positional Uncertainty in GIS and Its Application to VISAT" submitted by Jianghong Zhou in partial fulfillment of the requirements for the degree of Master of Science.

M. A. Chapman

Supervisor, Dr. M. A. Chapman
(Department of Geomatics Engineering)

Ronxing Li

Dr. R. Li (Department of Geomatics Engineering)

Nigel Waters

Dr. N. Waters (Department of Geography)

8 May 1996

Date

ABSTRACT

Geographic Information Systems (GIS) are becoming more useful in a variety of disciplines especially in the context of decision-making. Errors involved in GIS could affect the data quality significantly. Therefore, there is a need to clarify and quantify the associated uncertainty so that the users can evaluate the accuracy level in a GIS database against their requirements. Positional data and attribute information are the two major components in GIS. This study focuses on the modeling and visualization of positional uncertainty due to random errors such as digitization error. The "Developed Error Band" model in 2-D is adapted for handling positional uncertainty and is extended to the 3-D situation. Modeling positional uncertainty of an object in 3-D GIS can be sub-divided into modeling the positional uncertainty of point, line segment, area and body object. The uncertainty of a line segment plays a key role since all other complex objects can be constructed from line segments. The confidence region at a certain probability level is used as an indicator of positional uncertainty. The true value of the object falls within this confidence region at the pre-defined probability level. Visualization of the positional uncertainty is implemented in three dimensions. Color is used as an indicator of the confidence level to show the concept of this type of modeling. Finally, the developed method of modeling and visualization is applied to the measurement in the VISAT project to test its feasibility.

ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisor, Dr. Mike A. Chapman, for his support, guidance and encouragement throughout my M. Sc. program, for introducing me to the topic of uncertainty management and providing invaluable suggestions in the writing of this thesis. I would also like to thank Dr. Chapman for giving me the opportunity to pursue graduate studies in the Department of Geomatics Engineering.

I would like to thank Dr. R. Li and Dr. Chapman for allowing me to join the VISAT project and obtain the data for this thesis study.

My greatest thanks go to my husband, James, whose support, encouragement and understanding have made the last two years both possible and a time that I have enjoyed.

Dedication

To my husband, James, and my parents

for their encouragement, understanding and love

TABLE OF CONTENTS

CHAPTER TWO

CHAPTER THREE

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 Data Quality in Geographic Information System

Geographic Information Systems (GIS) technology has been widely used in many fields and has been proven to be a valuable tool for spatial and environmental data analysis. Many new systems and procedures have been implemented to take advantage of GIS. A typical GIS consists of the data acquisition, storage, processing, management and analysis, as well as product generation functions. Each part of GIS deals with the data. But in fact, much of the data embedded in a GIS database is contaminated by systematic and random errors or a combination of the two. These errors are introduced during each stage of digital data development. Errors within GIS data, however, can limit the usage of GIS technology. So in the GIS community the users are sensitive to data quality issues.

Data quality (which is considered to be synonymous with certainty, reliability, and confidence) is an important attribute associated with all data (Laurini and Thompson, 1992). Data quality in GIS was initially discussed in the literature in the early 1980s (e.g. Chrisman, 1982). The concept of `fitness for use' (Chrisman, 1983) was accepted as the definition of data quality. The US National Committee Digital Cartographic Data Standards Tasks Force (DCDSTF, 1988) has adopted this definition formally for inclusion in a US national standard for exchange of spatial data. The standard requires a quality report that provides the basis for a user to make the final judgment -- the conversion to information by interpretation for a particular use (Chrisman, 1991a).

Quality information includes the following six aspects (Chrisman, 1983):

1. Lineage Records -- tracing the procedure of data collected and derived, information transformed and data updated.

2. Positional Accuracy -- evaluating the spatial information.

3. Attribute Accuracy -- estimating the accuracy of classification or fuzziness.

4. Logical Consistency -- describing the integrity of the data structure, a logical inconsistency error could result from the positional error, thematic error or both.

5. Completeness -- dealing with dependencies between data items.

6. Temporal Reference -- recording GIS evolution over time.

Space, time and attributes all interact. Quality information forms an additional dimension to tie these components together. All these aspects inform the users of the suitability of the data for their applications.

1.2 Error

The complex issues of quality are best evoked by the word `error'. Errors that are inherent in a spatial database have attracted close attention for many years. For instance, geodesists, surveyors and photogrammetrists expend a great amount of effort to reduce the error in their measurement products. For these disciplines, full attention is on the reduction of the standard deviation between a positional measurement and the true value. Although this may be a reasonable approach, it does not lead to the full exploitation of spatial information. It should be noted that error is not a completely bad thing. A measure of error helps to control error. The field of GIS should put significant effort into the development of methods to report and visualize the error in a database. An appropriate error estimation provides crucial information which must be preserved in judging the 'fitness for use' and its correct interpretation.

1.2.1 Error Sources

Errors could arise at any stage in a GIS and many errors could be subsequently propagated through the data sets. As such, they represent a major unknown quantity in assessing the accuracy of a spatial database. Hunter and Beard (1992) gave a detailed overview of the causes of error that may reside in spatial databases. Normally, errors come from the three procedures of a GIS development (i.e. data collection and compilation, data processing and data misuse).

- **Data collection and compilation**

  Errors introduced through collection and compilation are also referred to as "source error" or "inherent error". The data for a GIS may come from remotely sensed images; a field survey; GPS; cartography or direct measurement. Consequently, the error could be caused by a poor satellite platform or a low pixel resolution; by a geodetic base used or surveying equipment used (e.g. surveying error); by poor satellite precision or the positioning method used (e.g. GPS error); by map resolution or digitization error (e.g. cartographic error). Such errors as, instrument variability and environmental factors, will all contribute to error accumulation.

- **Data processing**

  Typical data processing errors occur during computational round-off, digital conversion, data editing functions, coordinate adjustment, and the most common spatial operations now available in many GIS packages -- Boolean operations (such as polygon overlay).

- **Data misuse**

  Beard (1989) cites several cases of misuse of maps due to factors such as lack of training or education, divergence from conventions or expectations, and the inappropriate use of maps. The major problem is that the users misuse the data and ignore what quality control is invoked during map production. So they do not adequately prevent this type of error.

1.2.2 Forms of Error Information

Goodchild (1988) suggests that the above errors occur in the spatial database in the form of positional error (also known as "spatial error", "locational error" or "cartographic error"), attribute description error (also known as "aspatial error", "attribute error", or "descriptive error") and modeling of spatial variation error. Although the terminology often varies, it is generally accepted that all these errors contribute to the overall cumulative error derived from the spatial database. This is what the users ultimately want to determine. Figure 1.1 summarizes the interactions between these errors and the typology of errors.

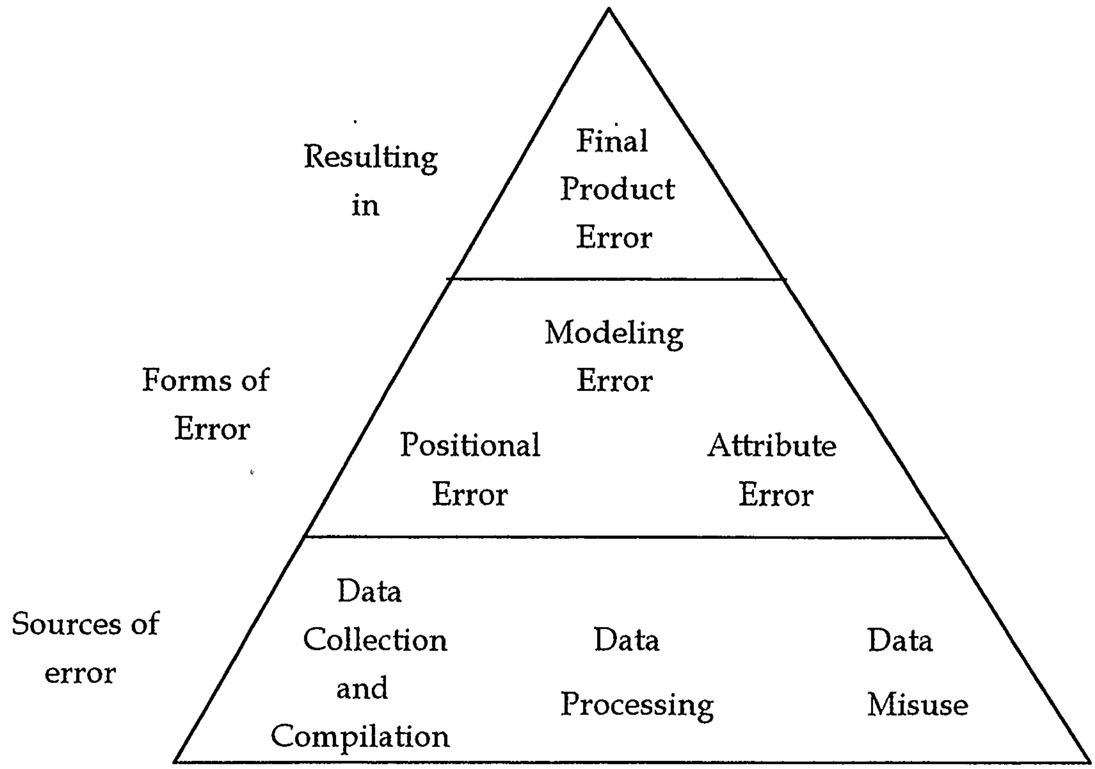

Figure 1.1 A Typology of Error in Spatial Databases (from Hunter, 1991)

1.3 Uncertainty

If error is referred to the deviation from a true value and no measurement is absolutely accurate, then uncertainty is the surrogate of error. In this sense, impreciseness, fuzziness, incompleteness and ambiguity of data are all considered to be forms of uncertainty (Gong, 1993). The problem with this definition is that, in reality, a decision maker is more likely to want to know "*how_likely* am I to find certain desired characteristics at a given location?". This is particularly obvious for a natural phenomena (e.g. boundaries are not widthless), or when attributes classification are not very definite. For example, the boundary between a swamp and a lake is extremely "fuzzy" and should be described as a buffer rather than a line; the difference between a sample with 80% sand and another one with 90% sand is minimal, they are described by different terms such as loamy or sandy. It is clear, that "uncertainty" is used to mean those simplifications that are introduced when a GIS represents parts of the real world in a database to identify land-related entities, to describe them and to locate them in space and time.

1.3.1 Classification of Uncertainty

No matter how good the conditions are, there are always uncertainties affecting the reliability of GIS databases. To understand better the consequences of those uncertainties, uncertainty can be classified into four categories (Bedard, 1987):

- **Conceptual uncertainty**

  Refers to the fuzziness in the identification of an observed reality, for example: an area is a forest or it is not?

- **Descriptive uncertainty**

  Refers to the uncertainty in the attribute values of an observed reality, for example: determining the density of a certain forest area according to a classification rule.

- **Locational / positional uncertainty**

  Refers to the uncertainty of an object's position with respect to the space, for example: error ellipses.

- **Meta-uncertainty**

  Refers to the magnitude of the first three types of uncertainty, for example: a standard error ellipse with a probability of 39.3%.

Uncertainty greatly affects people when they deal with planning and decision making. With the measurement of uncertainty, they can judge what kind of error is acceptable, or how much uncertainty they can tolerate in their products. Since a GIS is a computer-based database system for handling positional and attribute data, positional and attribute uncertainty are two essential components which should be included in a GIS database.

1.3.2 Approaches in Handling Uncertainty

Two approaches are considered to be appropriate for handling uncertainty in a GIS database:

- Since uncertainty originates in the source data, these sources of uncertainties are propagated through the model, leading to uncertainty in the model outputs. Uncertainty in model outputs must be displayed in an intuitive manner, to be used effectively by decision makers. This approach was summarized by Joy and Klinkenberg (1994) in Figure 1.2. This concept is used throughout this study.

- Another way of managing error in spatial databases lies in uncertainty reduction and absorption. Bedard (1987) recognized that the process of formalizing procedures and requirements helps reduce uncertainty between the model (as defined by the database) and the real world. For example, using better data processing methods; collecting more data; improving spatial or temporal resolution; and improving model calibration. However, regardless of the amount by which uncertainty is reduced, the uncertainty will never be removed completely. Thus, there will always be some residual uncertainty which users must decide either to absorb if they wish to use the data, or to find other source data if they reject the data. The amount of uncertainty absorbed can be considered as the risk associated with using the data or product (Hunter and Goodchild, 1993). The users make their own evaluation of the suitability of the data for their purposes according to the data quality statement supplied by the producers.

Assess Uncertainty in Source Data

↓

Propagate Uncertainty through the Model

↓

Visualization of Uncertainty

Figure 1.2 An Approach for Handling Uncertainty

(Revised from Joy and Klinkenberg, 1994)

1.4 The Objectives of the Uncertainty Research

Uncertainty in GIS is important and has attracted more and more attention in recent years. Three significant research objectives have been identified in the literature (Klinkenberg and Xiao, 1990):

1. The development of error models which can be used to compute measures of uncertainty. In other words, to know the accuracy and precision of GIS-products. This information can be described as the data quality dimension of a database, which serves to estimate whether the data is accurate and

precise enough for a specific application. The detail and quality of the data required will vary with the individual application.

2. The minimization of errors. By investigating various sources of errors and their propagation, it can be determined where and how to improve the quality of GIS products, and how to reduce the errors.

3. The determination of the best trade-off between cost and quality. Theoretically, the products should be as free from error as possible. Practically, however, as the accuracy or precision of a product is improved, so does the cost concomitantly increase. A GIS product is often expected to be generated at a relatively low cost but with acceptable accuracy and precision. Otherwise, the product might be either wasteful or useless.

1.5 Scope and Objectives of This Study

The approach of handling uncertainty including uncertainty assessment, modeling and visualization of uncertainty is used in this study. Imprecision introduced by random error at any point or as a result of any process cannot be removed later (Walsh et. al. 1987). Thus, random imprecision caused by any factor will limit overall precision. Random error in a GIS that is inherited from the original data source during data acquisition such as measurement, digitization, etc. will be dealt with. In this study, some well-developed theories for handling uncertainty will be presented. "Probability Theory" is considered to be the most suitable theory to handle "positional uncertainty"; while "Fuzzy Sets" have been specially developed to handle "attribute uncertainty" although they could handle boundary uncertainty as well. Since positional data are fundamental information in a GIS, more effort must be focused on the "positional uncertainty". After comparing the "epsilon band" model, "error band" model and "Developed Error Band" model, the "Developed Error Band" model was selected to model and visualize the positional uncertainty in a GIS as it provides a quantitative method measuring the uncertainty information. This method handles the random error caused by imprecision, vagueness and low resolution during data collection. This model is developed according to the available two dimensional spatial information. To best describe the three dimensional world, this model will be extended to the 3-D situation. Consequently, 3-D positional uncertainty modeling and visualization are the objectives of this study. This model will be applied to the error analysis in the VISAT project (Schwarz et. al., 1993) to test its feasibility.

CHAPTER TWO

THEORIES OF HANDLING UNCERTAINTY IN GIS

2.1 Introduction

Many methods and models have been developed for handling uncertainties in GIS. However, a complete theoretical system which can handle "geo-uncertainty" has not been developed because there are so many different kinds of uncertainties in GIS. Several existing theories, such as probability theory, the mathematical theory of evidence and fuzzy set theory can be adopted and applied to handle uncertainty problems in GIS. Each theory makes its own assumptions, has its own advantages and disadvantages, and can, so far, only handle certain kinds of uncertainty efficiently.

2.1.1 Probability Theory

Probability theory can be used to handle uncertainty caused by random errors. It describes the likelihood of some random events and it is possible to obtain a precise statement of the uncertainty involved. A random event is one whose relative frequency of occurrence approaches a stable limit as the number of observations or repetitions of an experiment is increased to infinity (Mikhail, et., 1981). The limit of the relative frequency of occurrence of a random event is known as the probability of an event. The probability of a random event is a number which lies somewhere between 0 and 1. In a mathematical representation, the probability P(E) of the event E occurring is $0 \leq P(E) \leq 1$. For example, let X represent the error of distance measurements. The random event X that takes on the specific numerical value x is represented mathematically by the expression X = x, and the probability of this random event occurring, represented mathematically by P(X = x), is given by the function P(x), i.e.,

$$P(x) = P(X = x) \tag{2.1}$$

Here, X is a random variable, and P(x) is its probability function. There is another function with comparable importance

$$F(x) = P(X \leq x) \qquad \text{for all x.} \tag{2.2}$$

F(x) is known as the cumulative probability distribution function of X. It is interpreted as the probability of the event, that the random variable takes on a value that is equal to or less than x.

2.1.2 Fuzzy Sets

The fuzzy set theory was introduced by Zadeh (1965) to handle imprecise information. The fuzzy set concept is used to determine the degree to which an object is a member of a set. The degree of "belonging" of each element to a fuzzy set is indicated by a membership grade, which is usually denoted by a real number ranging from 0 to 1. The higher the membership grade, the higher the degree that an element belongs to the set. The mathematical definition of a fuzzy set, A, is as follows:

$$\{x_i / \mu_A(x_i)\} \forall x_i \in U \tag{2.3}$$

where

"$\forall x_i$" denotes "for all x",

"/" is a separator to separate a set element x_i and its membership grade $\mu_A(x_i)$,

"$\mu_A(x_i)$" is a membership function of x_i, which is often viewed as a characteristic function in a classical set A of U,

"U" is the universe of discourse, whose generic elements are denoted x_i.

A fuzzy set is made up of a set of ordered pairs. Each pair consists of an element from the universe of discourse and its membership grade.

For instance, a fuzzy set for the linguistic term of observer's uncertainty, can be expressed as Table 2.1.

linguistic	probability p	grade
certain	100% ~ 90%	1.0 ~ 0.9
reasonably certain	90% ~ 70%	0.9 ~ 0.7
moderately certain	70% ~ 60%	0.7 ~ 0.6
not certain	< 60%	< 0.6

Table 2.1 Defining Uncertainty Using Probability

If the probability level is more than 90%, the uncertainty is best described as "certain", corresponding to a membership grade between 1.0 and 0.9; if the probability level is less than 60%, the uncertainty can be described as "uncertain", corresponding to a membership grade less than 0.6; otherwise, the membership grade will be between 0.9 and 0.6 showing the observer's lack of confidence.

The standard operations in fuzzy sets theory include the complement, union and intersection, which are defined as:

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x) \qquad \text{(complement)} \qquad (2.4)$$

$$\mu_{A \cup B}(x) = \max[\mu_A(x), \mu_B(x)] \qquad \text{(union)} \qquad (2.5)$$

$$\mu_{A \cap B}(x) = \min[\mu_A(x), \mu_B(x)] \qquad \text{(intersection)} \qquad (2.6)$$

Fuzzy sets theory can provide a good representation for geographic information. In a fuzzy representation, classes can be defined as fuzzy sets and spatial entities as set elements. Each spatial entity is associated with a group of membership grades to indicate the extent to which the entity belongs to certain classes. Entities with a class mixture or intermediate conditions can be described by membership grades. For example, if a ground area contains a mixture of two cover types, 'soil' and 'vegetation', it may have two membership grades indicating the extent to which it is associated with the two classes (Wang, et al. 1990). Fuzzy sets tell us how distinct any object might be, deal with possibility of membership and handle the imprecision in terms of uncertainty.

2.1.3 Evidential Theory

Mathematical theory of evidence proposed by Shafer (1976), also called evidential theory, is a generalization of the classical probability theory. The interval between a belief and a plausibility (low and high probabilities) represents the uncertainty of the knowledge about the event. This theory provides a mathematical framework for the description of incomplete knowledge.

A set of mutually exclusive and exhaustive hypotheses is called a frame of discernment Y.

$$Y = \{P(H_1), P(H_2), \ldots, P(H_j), \ldots, P(H_n), \Theta\} \qquad (2.7)$$

where

$P(H_j)$ is the probability that hypothesis H_j is true,

Θ is the uncommitted or distributed support.

$$\sum_{j=1}^{n} P(H_j) + \Theta = 1 \qquad (2.8)$$

The belief function is defined as

$$Bel(H_i) = P(H_i) \qquad (2.9)$$

The plausibility function is defined as

$$Pl(H_i) = 1 - \sum_{j=1}^{n} P(H_j) = Bel(H_i) + \Theta \qquad (\mathrm{i} \neq \mathrm{j}) \qquad (2.10)$$

The interval $[Bel(H_i), Pl(H_i)]$ represents the uncertainty because of the incompleteness of evidence for hypothesis H_j due to uncommitted support Θ.

The evidential theory has been applied in the classification of remote sensing images. It can be used to solve the uncertainty problem in judging whether an object belongs to a subset A. For example, if a classified pixel of image belongs to forest or other classes.

2.2 The Uncertainty Dealt with in This Study

Point errors in GIS are mostly inherited from the original data source, such as field data or digitized map, usually from positional error and attribute error. Since the positional error is a significant error component, we will only deal with positional uncertainty in this study. When people make measurements, they usually want high accuracy and precision. So it is assumed that all input to a GIS is carefully checked (i.e. no gross errors or blunders are involved, systematic errors such as map deformation, survey error and mathematical model error are corrected). Thus, we will assume that the positional point errors in this study are random in nature. Based on the point errors, and since the point is the basic constructive geometrical type, we can further derive the uncertainty of a line segment, an area object, and finally the uncertainty of a solid boundary.

2.3 Adopted Theory of Handling Positional Uncertainty

Three existing theories can solve different aspects of the data quality. Due to the characteristics of the uncertainty dealt with in this study, "probability theory" is adopted to handle the positional uncertainty. We will start the investigation of uncertainty from a point perspective, employ the Law of Error Propagation and uncertainty indicators (error ellipses or error ellipsoids with confidence levels), use the analytical model of "Developed Error Band" to construct the line segment uncertainty, and conclude with the uncertainty of a solid boundary.

CHAPTER THREE

MODEL DEVELOPMENT OF POSITIONAL UNCERTAINTY

3.1 Systematic Effects and Random Effects

Measurements in GIS are carried out for a variety of purposes. No matter for what purpose, it is necessary that the measurements possess a certain degree of reliability or precision, and employ a suitable mathematical model representing the physical situation since computations are often involved with the results of the measurements. These two aspects need to be considered to ensure successful results. There are two factors which may affect the reliability: systematic effects and random effects (Richardus, 1966).

Systematic effects

> "When the effect of an incorrect choice of the mathematical model, deviations of standardlization and identification is of a regular and consistent character it will be called systematic."

The systematic effect can be reduced or eliminated by several methods :

- Using better data inputs -- It may be reduced to a magnitude which is negligible by a better choice of mathematical model or by the use of a better instrument. For example, using the high precision technology of INS (Inertial Navigation System) instead of GPS technology alone.

- Calibrating data -- By a determination of the error independently from the measurements and the application of a corresponding correction to the observation. For example, using the fixed numbers provided by the manufacturer to calculate the corrections.

- Applying special methods -- The systematic influence may be derived by special methods considering the processing of the data provided by the measurements. For example, lens distortion errors can be corrected by using the function of lens distortion parameters.

Random effects

When a sufficiently large number of measurements is made under the same set of circumstances, the observations may follow a set pattern, which can be described by the following characteristics:

1. All observations fluctuate about a central value which can be represented by a point or a line in a plot.

2. Positive and negative deviations from this central value are equally frequent.

3. Small deviations are more frequent than large ones.

It is generally accepted that, if the measurements follow a pattern and have the above characteristics, the behavior of the measurements can be described by a theoretical continuous curve (See Figure 3.1) -- the curve of the density function (Equation 3.1), given by the law of the Normal or Laplace-Gaussian Probability Distribution. Random effects can be reduced by increasing the number of measurements.

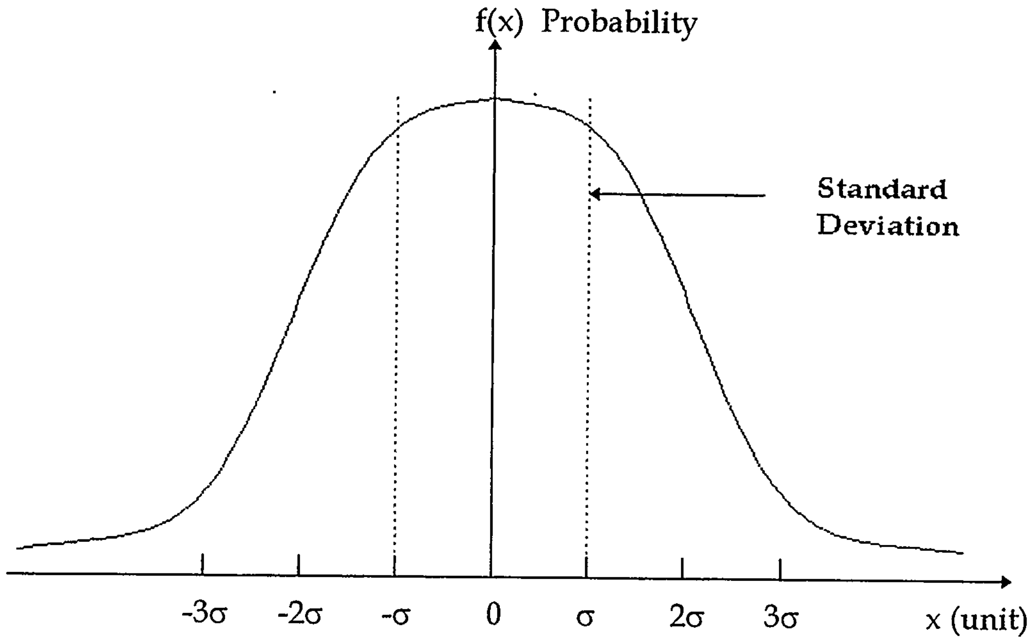

Figure 3.1 Normal Distribution Curve

We can see from Figure 3.1, if the measurements cluster closely about the mean, the standard deviation will be small; conversely if the scatter is large, the standard deviation will be large. High precision measurements belong to a normal distribution with a small standard deviation. Low precision measurements belong to a normal distribution with a large standard deviation.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\xi)^2/2\sigma^2} \qquad (-\infty < x < +\infty) \qquad (3.1)$$

The density of the observation f(x) is expressed in terms of deviations $(x-\xi)$ from the mean ξ of a hypothetical infinite number of the observations. The variance is σ^2. The positive square root of the variances is defined as the standard deviation. The deviations $(x-\xi)$ are usually called the random errors. Many people in the field of surveying indicate the precision of an observation by its standard deviation. This gives the values $x_1 = -\sigma$ and $x_2 = \sigma$ as limits in between which an observation may fall with a probability level of 68.3% as expressed by:

$$prob\{x_1 < x < x_2\} = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-x^2/2\sigma^2} dx = 68.3\% \tag{3.2}$$

3.2 Accuracy and Precision

GIS employs quantitative techniques for a very good reason. Nothing is wrong with a qualitative statement, but it will carry more weight if it is possible to make a statement quantitatively. The advantage of a quantitative statement is its precision. A quantitative statement is superior only if the following conditions are met: validity and accuracy (Matthews, 1981). The results of shooting at a target provides a good analogy of accuracy and precision and illustrates the differences between a qualitative and quantitative statement in terms of these two concepts (Figure 3.2).

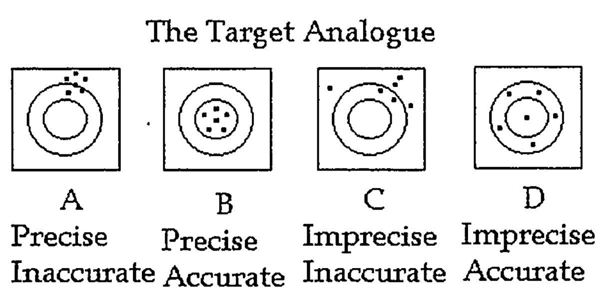

Figure 3.2 Accuracy and Precision

We can see that case B is accurate and precise and, therefore, the best of the four cases which is the case we always look for. Both A and C are seriously in error. D is accurate but its imprecision allows a range of possible interpretations and could cause the confusion. Even if both B and D are accurate, we also want to have high precision. Both the concepts of accuracy and precision are useful in describing the positional uncertainty of GIS coordinate data.

Bolstad et. (1990) gave the following definitions for accuracy and precision in GIS:

Accuracy (average error) measures the nearness of quantities to their true values, the error is the difference between the GIS coordinates and the true value, so the discrepancy between the mean value and the true value determines the accuracy of the coordinates.

Precision (mean error) measures the degree of conformity of measurements among themselves, the error is the deviation of the position of points from their mean location, so the deviation reflects the precision of the coordinates.

Total error for a particular measurement can be considered as the sum of the mean (systematic) and deviation (random) errors. For a particular GIS database, the accuracy can be considered to reflect the average error associated with the database, the precision reflects the variation or distribution of errors. Since the precision and accuracy are related to the statistical distribution of the errors, they can be modeled by a probability density function (Hogg and Craig 1965). For errors that follow a normal distribution, the mean error would characterize the accuracy and the variance or standard deviation could be used to characterize the precision.

3.3 Positional Uncertainty of Point

A "point" is an elementary component for describing geometry in a GIS. A point is generally represented by its 3-D coordinates (x, y, z). In order to understand the essence of positional uncertainty of a point, we must model the error of the point coordinates. In GIS, coordinates of a digital point are usually collected through measurement or computation from measurements with a certain accuracy and precision. In reality, it is impossible to obtain the true value from the measurements. That means no measurement is absolutely accurate, and we are not sure about the measurements. Thus, point precision is often used as a quality measurement for point coordinates instead of using point accuracy, although they are different theoretically. In the case of high accuracy, people could only use precision to judge the quality.

3.3.1 Positional Uncertainty Indicators

If the errors in the (x, y, z) coordinates of a point are random and normally distributed, the precision of point coordinates can be represented by the standard deviation. The uncertainty of a point can then be indicated by a 3-D error ellipsoid. From a statistical viewpoint, an error ellipsoid may be defined as a surface of equal probabilities of the location of a point in 3-dimensional space according to the law of Normal Distribution. By analysis of the error ellipsoid, people can determine the weakest direction of the established point and the uncertainty of the position of the point along this direction. It may be further interesting to determine the standard deviation of the point location in any arbitrary direction. Furthermore, an overall understanding of the precision can be obtained from the mean-radius of the error ellipsoid, which is the square root of the average of the variances of the three spatial coordinates of the point (Malhotra, 1969). The probability density function in 3-dimensional space is given by the following:

$$p = prob(x, y, z)dx, dy, dz \tag{3.3}$$

The equi-probability surface determines the uncertainty of the position of the point. Therefore, we can use the confidence level to measure the uncertainty. For example, when we say the error of a point has a 75% probability of being within the error ellipsoid, it means we are 75% certain that the point is within the error ellipsoid. The confidence level can be considered as the fourth dimension information about the point.

3.3.2 The Error Ellipse and Confidence Region

If we are only concerned with the 2-D uncertainty of a point position, the probability density function in two dimensions may be expressed as:

$$prob(x - dx \le \xi \le x + dx, y - dy \le \eta \le y + dy) = f(x, y)dxdy \quad (3.4)$$

when x and y are independent, the family of equations for error ellipses can be expressed as:

$$\frac{(x-\xi)^2}{\sigma_x^2} + \frac{(y-\eta)^2}{\sigma_y^2} = c^2 \quad (3.5)$$

where ξ and η are the mean values of coordinates (x, y), σ_x and σ_y are the standard deviations. When c is equal to 1, we can obtain the equation of the standard ellipse, whose semi major axes are equal to σ_x and σ_y along the x and y axes, respectively (Figure 3.3). The error ellipse has a statistical significance. Since x and y are independent and have random errors with variance σ_x^2 and σ_y^2 respectively, they both have normal distributions, and $\frac{(x-\xi)^2}{\sigma_x^2} + \frac{(y-\eta)^2}{\sigma_y^2}$ has a χ^2 distribution with two degrees of freedom.

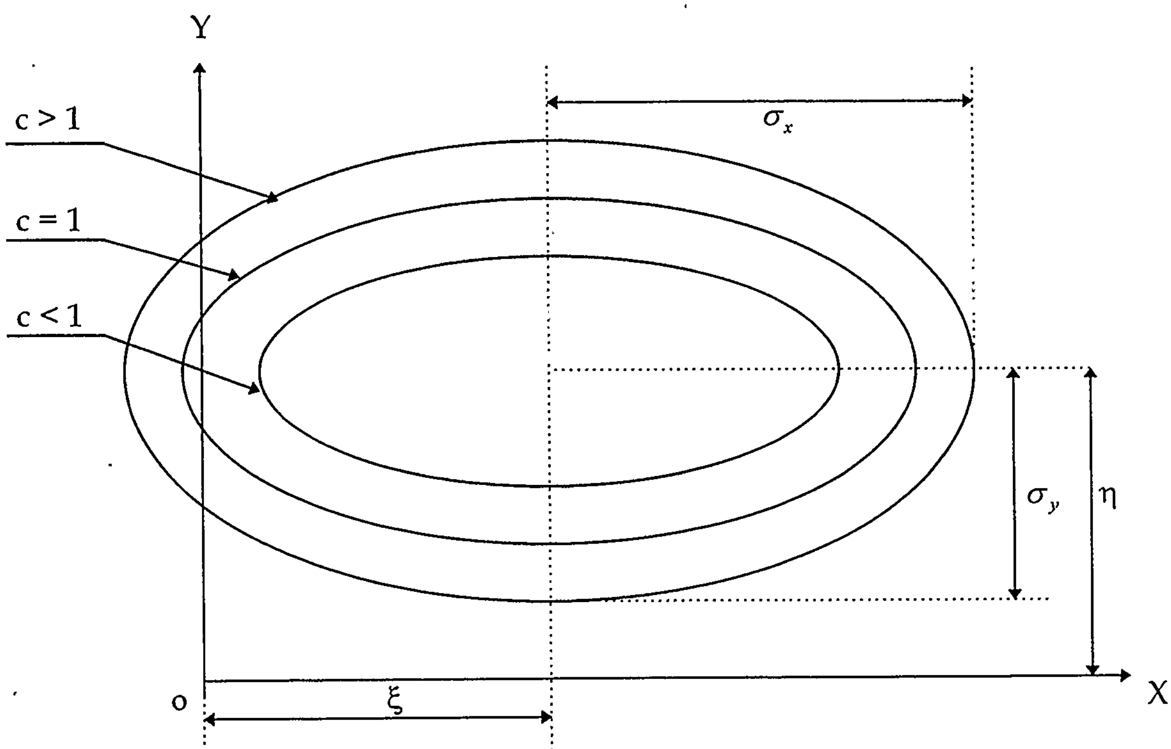

Figure 3.3 The Standard Error Ellipses at Different Probability Level

From the χ^2 distribution table, we can find the probability that the point lies within the standard ellipse is 38.5%. We can show various values of the expansion factor c corresponding to different probability percentages as shown in Table 3.1:

c	Probability (%)
1.0	38.5
2.0	85.6
3.0	98.9
4.0	99.8

Table 3.1 Probability Percentage of an Error within an Error Ellipse

3.3.3 The Error Ellipsoid and Confidence Region

Similarly, the probability density function in three dimensions is given by:

$$p = prob(x,y,z)dx,dy,dz \tag{3.6}$$

The equation of the equal probability surface is:

$$\frac{(x-\xi)^2}{\sigma_x^2}+\frac{(y-\eta)^2}{\sigma_y^2}+\frac{(z-\varsigma)^2}{\sigma_z^2}=c^2 \tag{3.7}$$

This equation represents a family of error ellipsoids for every value of c. Here, ξ, η and ς are the mean values of coordinates (x, y, z), σ_x, σ_y and σ_z are the standard deviations. If c = 1, the resulting equation is generally known as the Standard Error Ellipsoid, whose semi-major axes are equal to σ_x, σ_y and σ_z along the x, y and z axes, respectively.

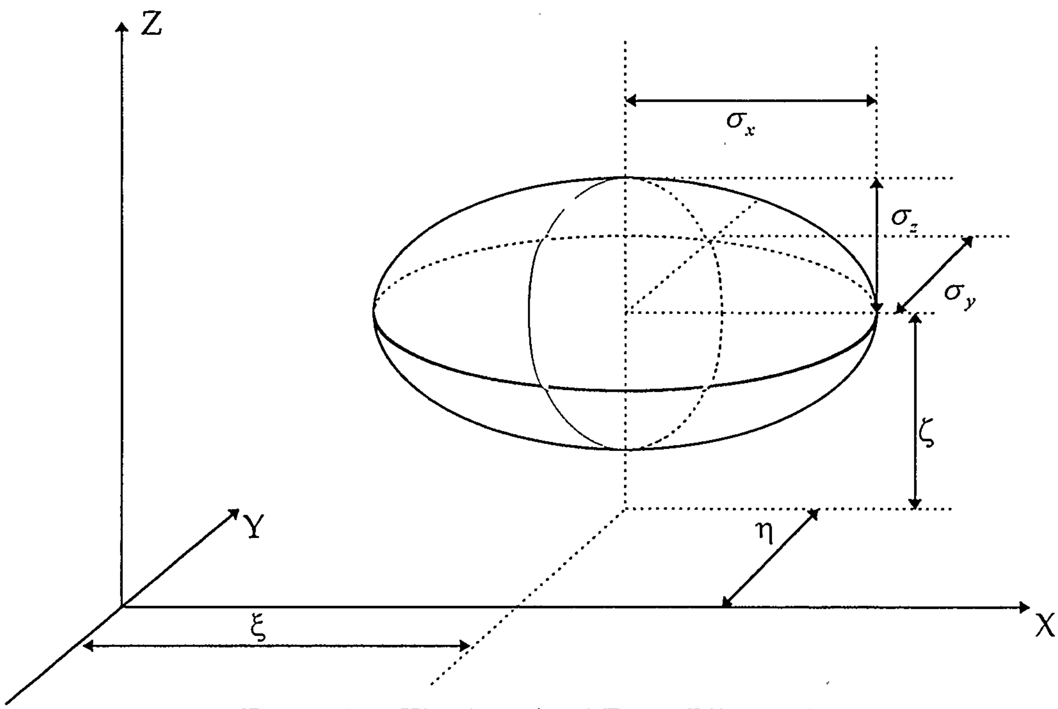

Figure 3.4 The Standard Error Ellipsoid

Since $\frac{x-\xi}{\sigma_x}$, $\frac{y-\eta}{\sigma_y}$ and $\frac{z-\varsigma}{\sigma_z}$ have a normal distribution N(0, 1), then $\frac{(x-\xi)^2}{\sigma_x^2}+\frac{(y-\eta)^2}{\sigma_y^2}+\frac{(z-\varsigma)^2}{\sigma_z^2}$ has a χ^2 distribution with three degrees of freedom. From the χ^2 distribution tables, the probability that $\frac{(x-\xi)^2}{\sigma_x^2}+\frac{(y-\eta)^2}{\sigma_y^2}+\frac{(z-\varsigma)^2}{\sigma_z^2}$ is less than or equal to unity is 0.199. In other words, the probability of locating a point within the Standard Error Ellipsoid is 19.9 percent.

Table 3.2 gives the probability percentage of an error to be within an error ellipsoid with semi-major axes equal to $c*\sigma_x$, $c*\sigma_y$ and $c*\sigma_z$, along the x, y and z axes of the error ellipsoid. The error ellipsoids represent confidence regions. Different confidence regions corresponding to different values of c can be obtained to indicate the data quality (Malhotra, 1969).

c	Probability (%)
1.0	19.9
2.0	73.4
3.0	96.8
4.0	99.9

Table 3.2 Probability Percentage of an Error within an Error Ellipsoid

3.4 Positional Uncertainty of Line Segments

In many cases, there is a certain tolerance level in the selection of the mathematical model. In general, the simplest one is selected in order to avoid unnecessary complications. In this study, the error due to the mathematical model is negligible.

A line is another primitive element in GIS. The fundamental idea of this study is to handle the uncertainty of a line segment derived from the uncertainty of a point. There are several geographic features represented as lines in GIS (Mark, 1989). Table 3.3 gives a typology of geographic lines in GIS.

Among those lines in Table 3.3, different lines are related to different errors or uncertainty. For example: There are no specific points in the real world to define forest boundaries or soil boundaries due to imprecision in information. People have to deal with the error / uncertainty caused by line interpretation, measuring or manipulation. Since the error sources involved are from human interpretation, people often use fuzzy sets theory to qualify them rather than giving a quantitative statement. For political boundaries or road lines, there are specific points to define them in reality. People have to deal with the measuring or manipulation errors. Because of the randomness of the errors, we could use probability theory to give a statistical analysis. So this study will focus on the determination, modeling and visualization of the line uncertainty which is composed of specific points with known random errors.

Basically, there are only two models for describing the line uncertainty: the epsilon band model and the error band model.

Real World Features	Example
1. Mathematical line	Latitude, longitude, azimuth distance
2. Legislated line	Political boundary
3. Line feature	
a) line feature	Road, railway
b) line-like feature (variable width)	Stream, river
4. Zero set of feature	
a) well defined surface	Contour
b) complex / dynamic	Shoreline
5. Area-class boundary	Forest boundary, soil boundary

Table 3.3 Typology of Geographic Lines (adapted from Mark and Csillag, 1989)

3.4.1 The Epsilon Band Model

The original epsilon band model was elaborated by Chrisman (1982) based on the theory of epsilon distance (Perkal, 1956) as a representation of the positional uncertainty of a digitized line. The epsilon band is an area of constant width ε on both sides of a line or polygon boundary. The model proposes that the true position of a line will lie within a band at some displacement from the measured position or digitized position (Figure 3.5).

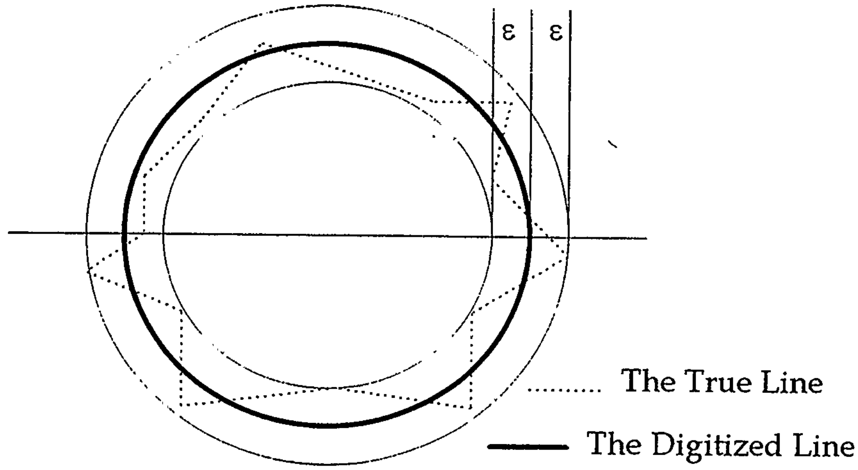

Figure 3.5 Epsilon Band Boundary

The width of the band may be set to the standard deviation of the uncertainty of the line but should not be used to define error in the strict sense of a "buffer" corridor used in GIS applications (Crain, Gong and Chapman, 1993). In Figure 3.5, the epsilon band model shows a dashed line representing the true location of the line to be digitized; the central solid line is the measured line; the two parallel lines describe the region of an epsilon band; and the true line falls within the epsilon band region of the digitized line with a certain probability.

The probability of the true value in relation to a measured value can be modeled as that shown in Figure 3.6.

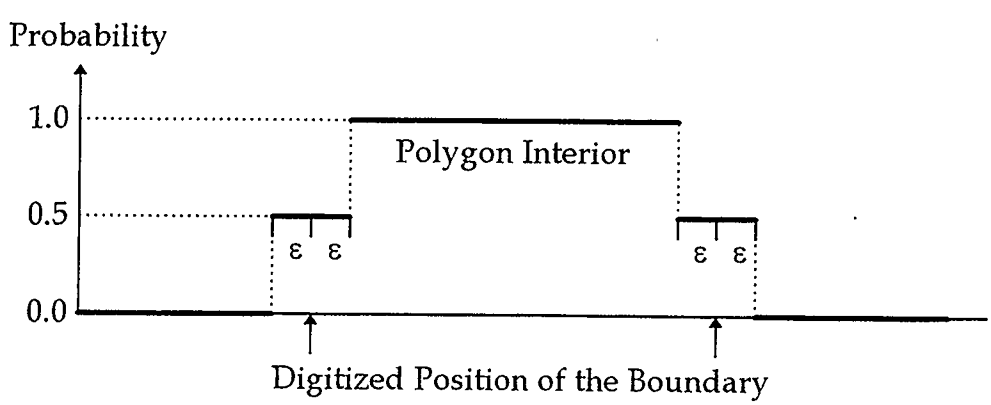

Figure 3.6 The Probability of a True Value Lying within a Band (Cross Section)

In Figure 3.6, the model describes the probability of a point falling within the boundary. The probability of a point inside the boundary is 1.0, that means the point is definitely inside the boundary; the probability of a point within the epsilon band is 0.5; the probability of a point outside of the boundary is 0.0. The epsilon band model can be viewed as a discrete approximation to this continuous, symmetric epsilon model. Some people describe the epsilon band as the volume occupied by rolling a ball along the line.

3.4.1.1 The Application of the Epsilon Band

The epsilon band can be used as an indicator to measure the error / uncertainty associated with arcs, or boundaries of the features represented on the map. When the width of the epsilon band increases, the band area also increases. Chrisman (1982) applied the epsilon measurement program to data obtained from GIRAS digital file - the U. S. Geological Survey's Land Use/Land Cover series. He quantified three types of error effects: line width drafting error might have an average deviation of 12.5 meters; digitizing error had a deviation of 8.3 meters with a map scale of 1:10000; and round-off contributed 2.9 meters average deviation. These error effects combined to give an average deviation of 15.2 meters by using the Law of Error Propagation based on the assumption that these errors were independent and random. Therefore, it was decided that an epsilon band of 20 meters would be quite conservative. So the defined band width of 20 meters was applied to all lines. About 7% of the total area of the map was covered by the epsilon band. The area covered by the epsilon band represented an uncertainty of the land use / land cover class.

The epsilon band of uncertainty can also be visualized by using graphical methods. In ARC/INFO, the system creates a look-up table with different widths of bands and corresponding symbols to display the arcs with different shades, patterns and thicknesses. The uncertainty of arcs can be displayed in different colors.

3.4.1.2 The Limitation of the Epsilon Band

The epsilon band model is a simple and useful concept used to describe the uncertainty of the data. However, there are some limitations in the application of this model. Goodchild and Dubuc (1987) pointed out that the epsilon band model is not completely satisfactory as a cartographic error model.

First, although the model proposes that every true line lies entirely within the epsilon band with a probability of 1.0, we would expect intuitively that no such deterministic upper limit to error exists. Instead, it would seem that larger errors are simply less likely. Error models of simple measurements, such as a Gaussian distribution, places no upper limit on the size of the errors.

Second, the model provides no distribution of error within the epsilon band. Although intuition might suggest that the most likely position for the real line (the true position) is the center of the epsilon band.

Third, while the epsilon band provides a model of deviation for a point on the line, it does not model the line itself. In fact, GIS users are interested in measures for accuracy of objects such as lines, boundaries and polygons as well.

From the above observations, we can see that the epsilon band model needs reviewing. A more suitable model would have a continuous distribution with asymptotic tails centered on the true line and the deterministic epsilon distance would be replaced by a standard deviation parameter. The most suitable one may be a Gaussian Distribution.

3.4.2 The Error Band Model

Objects in a spatial information system are points, lines and polygons. A straight line is captured by digitizing both end points. From this, all points along the line are determined. Boundaries and polygons are defined by lines which border a certain area. The accuracy of these objects is a function of the positional accuracy of the points involved. The uncertainty of the location of a line can be represented by an epsilon band. Because of the limitation of the epsilon band model, a few researchers (Caspary and Scheuring, 1992; Dutton, 1992) have developed the error band model which is based on the epsilon band model. The error band model follows the assumptions that the coordinate errors of the end points are independent and equal $\sigma = \sigma_x = \sigma_y$; the error propagation law yields the positional error of points along the line, the error at the midpoint of the line is the smallest, and the error at two end points is the largest. The error band is not a rectangle but it is a band bordered by sagging lines rather than by straight lines like the epsilon band. Further investigations have been carried out to determine the shape of the band and any points within the band. The results enable the computation and visualization of the correct error band containing the true straight line with a fixed probability.

Under these assumptions, with two end points $p_1(x_1, y_1)$ and $p_2(x_2, y_2)$ having the same standard deviation $\sigma = \sigma_x = \sigma_y$, any arbitrary point $p_i(x_i, y_i)$ on the line segment p_1p_2 can be derived from the end points according to the Law of Error Propagation:

$$x_i = (1 - \frac{l_i}{l})x_1 + \frac{l_i}{l}x_2 \tag{3.8}$$

$$y_i = (1 - \frac{l_i}{l})y_1 + \frac{l_i}{l}y_2 \tag{3.9}$$

$$\sigma_{x_i}^2 = \sigma_{y_i}^2 = (1 - 2 \bullet \frac{l_i}{l} + 2 \bullet \frac{l_i^2}{l^2})\sigma^2 \tag{3.10}$$

where $0 \le l_i \le l$. When $l_i = \frac{l}{2}$, we have a minimum $\sigma_{x_i} = \sigma_{y_i} = \frac{1}{\sqrt{2}}\sigma = 0.707\sigma$. That means the standard deviation in the middle of the line is smaller than at the end points by the factor of $\frac{1}{\sqrt{2}}\sigma$. Because of this, the error band is not a rectangle but a band bordered by sagging lines according to the equations above. In Figure 3.7, the circles at the ends of the line show the positional errors of points p_1 and p_2, respectively.

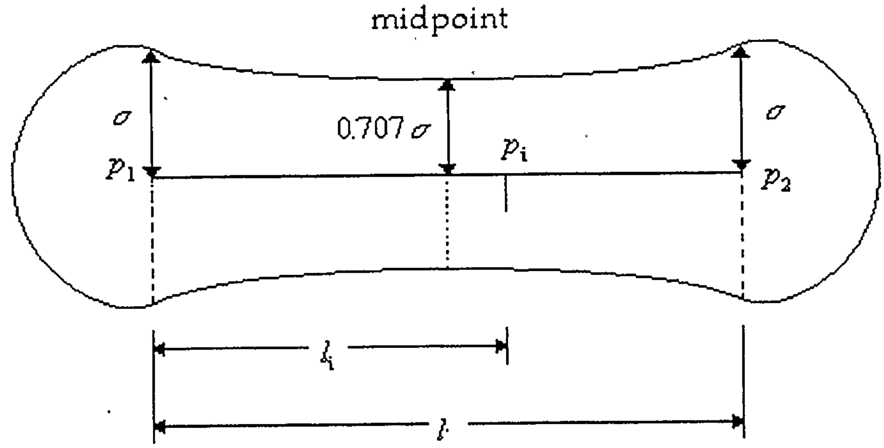

Figure 3.7 The Error Band

The area of this error band can be computed by

$$A_2 = \pi\sigma^2 + 2l\sigma\int_0^1 \sqrt{1 - 2x + 2x^2}\,dx \approx \pi\sigma^2 + 1.62l\sigma \tag{3.11}$$

Compared to the area of the epsilon band $A_1 = \pi\sigma^2 + 2l\sigma$, and if the standard deviation is much smaller than the length ($\sigma << l$), we can approximate

$A_2 \approx 0.81A_1$. The uncertainty area of the error band is smaller than the area of a conventional epsilon band by the factor, 0.81. The error band can be interpreted as an area bordered by the envelope of the error circles of all points along the line. For example, how far a point may wander from the center of a stream and still be in the water. It is very difficult to describe the shape of such an error band analytically, although the error band can be approximated by four straight lines in Figure 3.8.

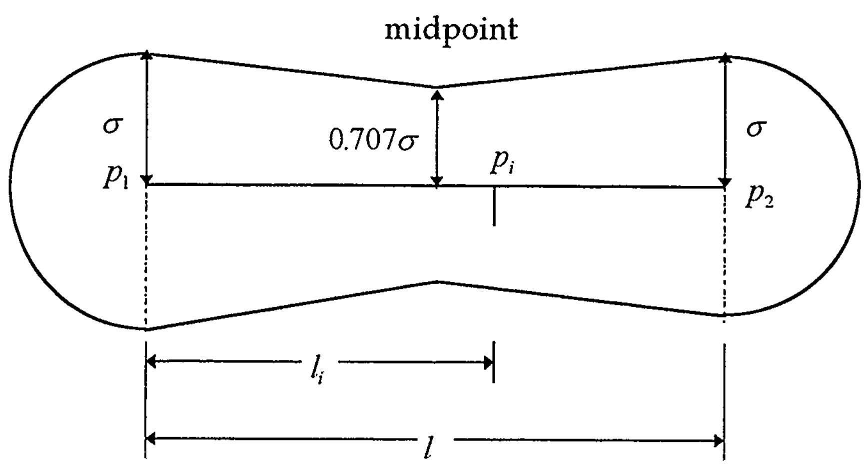

Figure 3.8 Approximation of the Error Band

The area can be calculated by

$$A_3 = \pi\sigma^2 + 4\frac{(1+\frac{1}{\sqrt{2}})\frac{l}{2}\sigma}{2} = \pi\sigma^2 + 1.707l\sigma \tag{3.12}$$

$$A_3 \approx 0.81A_1 \tag{3.13}$$

The error bands discussed so far are simplified two-dimensional representations of lines. A nearly correct map of the area, where the random lines are located with a certain probability, can be derived by a Monte-Carlo approach. The coordinates of the end points of the line are generated according to a desired distribution of the coordinate errors, its expectation and standard deviation. By this, a random position of the straight line is simulated. Figure 3.9 illustrates the results of 100 such random trials.

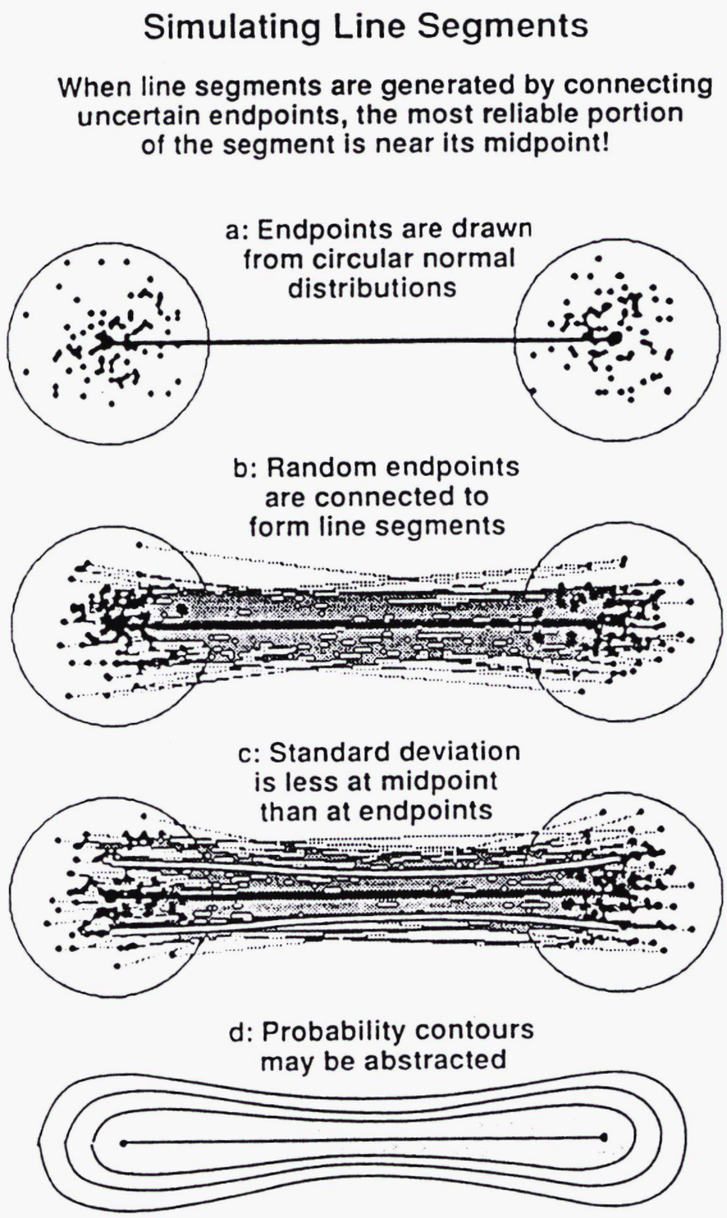

Figure 3.9 Simulating Line Segments

(Adapted from Dutton, 1992)

The area with a certain probability of containing the random line is described as a probability contour. The fact that dispersion is greatest at the endpoints and least around the midpoint may seem odd at first. This indicates that when linear features are digitized, the displacement error tends to be greatest near the measured points and least midway between them, where there are no explicit coordinates! No matter how much care is taken in positioning its endpoints, a segment's center point will prove to be a more reliable location, even though it is more fictional (Dutton, 1992).

3.5 The "Developed Error Band" Model

According to the above sections, either for the epsilon band model or the error band model, the confidence region of a line segment can only describe five different relationships between a line segment and the location of a point: definitely in; possibly in; definitely out; possibly out and ambiguous. Obviously, we need a model to describe the positional relationships continuously and analytically instead of by using simulation methods. Shi (1993) developed the error band model and provided quantitative uncertainty data by using the probability distribution of a line segment within the range [0,1]. This model includes two newly developed features: the confidence region of a line segment in the form of a statistical formula and the probability distribution perpendicular to the line segment. This model is called the "Developed Error Band" model. The uncertainty of any object is based on the uncertainty of the line segment.

3.5.1 Probability Distribution

A line segment is defined by two end points p_1 and p_2 and an arbitrary point p is on the line segment, where $r = l_r / l$ and $r \in [0,1]$.

In this study, we only consider the error of a line segment caused by the random error of the two end points. It is assumed that the two end points are independent and the two measured stochastic vectors $p_1 = (x_1, y_1)^T$ and $p_2 = (x_2, y_2)^T$ follow the two-dimensional normal distribution according to the multi-variate normal distribution:

$$p_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \rightarrow N_2\left(\begin{pmatrix} \xi_1 \\ \eta_1 \end{pmatrix}, \begin{pmatrix} \sigma_{x_1}^2 & \sigma_{x_1 y_1} \\ \sigma_{y_1 x_1} & \sigma_{y_1}^2 \end{pmatrix}\right) \tag{3.14}$$

$$p_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \rightarrow N_2\left(\begin{pmatrix} \xi_2 \\ \eta_2 \end{pmatrix}, \begin{pmatrix} \sigma_{x_2}^2 & \sigma_{x_2 y_2} \\ \sigma_{y_2 x_2} & \sigma_{y_2}^2 \end{pmatrix}\right) \tag{3.15}$$

where $\xi_1, \eta_1, \xi_2, \eta_2$ are the mean value of coordinates of two end points while the two end points have the same variance $\sigma_{x_1}^2 = \sigma_{x_2}^2 = \sigma_x^2$, $\sigma_{y_1}^2 = \sigma_{y_2}^2 = \sigma_y^2$ and covariance $\sigma_{x_1 y_1} = \sigma_{y_1 x_1} = \sigma_{x_2 y_2} = \sigma_{y_2 x_1} = \sigma_{xy} = \sigma_{yx}$. Since the variance and covariance can be estimated, by using the linear interpolation, an arbitrary point on a measured line segment between p_1 and p_2 can be represented by

$$p_r = \begin{pmatrix} x_r \\ y_r \end{pmatrix} = (1-r)p_1 + rp_2 = \begin{pmatrix} (1-r)x_1 + rx_2 \\ (1-r)y_1 + ry_2 \end{pmatrix} \quad \text{where } r \in [0,1] \tag{3.16}$$

The expectation of coordinate x_r

$$E(x_r) = E[(1-r)x_1 + rx_2] = (1-r)E(x_1) + rE(x_2) = (1-r)\xi_1 + r\xi_2 = \xi_r \tag{3.17}$$

Similarly, we can derive the expectation of coordinate y_r

$$E(y_r) = (1-r)\eta_1 + r\eta_2 = \eta_r \tag{3.18}$$

According to the Law of Error Propagation, the variance and covariance of x_r and y_r can be derived also:

$$\sigma_{x_r}^2 = (1-r)^2\,\sigma_{x_1}^2 + r^2\,\sigma_{x_2}^2 = (1-2r+2r^2)\,\sigma_x^2 \tag{3.19}$$

$$\sigma_{y_r}^2 = (1-r)^2\,\sigma_{y_1}^2 + r^2\,\sigma_{y_2}^2 = (1-2r+2r^2)\,\sigma_y^2 \tag{3.20}$$

$$\sigma_{x_r y_r} = (1-2r+2r^2)\sigma_{xy} = \sigma_{y_r x_r} \tag{3.21}$$

With above equations, we found

$$p_r = \begin{pmatrix} x_r \\ y_r \end{pmatrix} \rightarrow N_2\left(\begin{pmatrix} (1-r)\xi_1 + r\xi_2 \\ (1-r)\eta_1 + r\eta_2 \end{pmatrix}, (1-2r+2r^2)\begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{pmatrix}\right) \tag{3.22}$$

Based on these assumptions, we can state that an arbitrary point on the measured line segment is represented as a stochastic vector following the bi-variate normal distribution defined by Equation 3.22.

Shi (1993) found that the "perpendicular distribution" of a line segment for all $r \in [0,1]$ also follows a two dimensional normal distribution. The density function in the perpendicular direction is

$$f'(y') = \int_{-\infty}^{\infty} f'(x',y')dx' = \frac{1}{(2\pi)^{1/2}\sigma_{y'}} \exp[-(y'-\eta_{y'})^2 / 2\sigma_{y'}^2] \tag{3.23}$$

where

$$y' = -\sin\theta[(1-r)x_1 + rx_2] + \cos\theta[(1-r)y_1 + ry_2] \tag{3.24}$$

$$\eta_{y'} = -\sin\theta[(1-r)\xi_1 + r\xi_2] + \cos\theta[(1-r)\eta_1 + r\eta_2] \tag{3.25}$$

$$A = \cos\theta\sigma_{yx} - \sin\theta\sigma_x^2 \tag{3.26}$$

$$B = \cos\theta\sigma_y^2 - \sin\theta\sigma_{xy} \tag{3.27}$$

$$\sigma_{y'}^2 = [A(-\sin\theta) + B\cos\theta](1-2r+2r^2) \tag{3.28}$$

θ is the angle between the line segment and x axis.

when θ = 0, we have

$$y' = (1-r)y_1 + ry_2 \tag{3.29}$$

$$\eta_{y'} = (1-r)\eta_1 + r\eta_2 \tag{3.30}$$

$$A = \sigma_{yx} \tag{3.31}$$

$$B = \sigma_y^2 \tag{3.32}$$

$$\sigma_{y'}^2 = \sigma_y^2(1-2r+2r^2) \tag{3.33}$$

Since $r \in [0,1]$, obviously, when r = 0 or 1, $\sigma_{y'}^2$ has the maximum value σ_y^2. Take the derivative of $\sigma_{y'}^2$

$$\frac{\partial \sigma_{y'}^2}{\partial r} = \sigma_y^2(-2+4r) = 0 \tag{3.34}$$

Thus r = 1/2. That means when r is equal to 1/2, $\sigma_{y'}^2$ has the minimum value. The conclusion is that the line segment has the maximum variance at the two end points and minimum variance at the midpoint, if the two end points have the same variance σ_y^2.

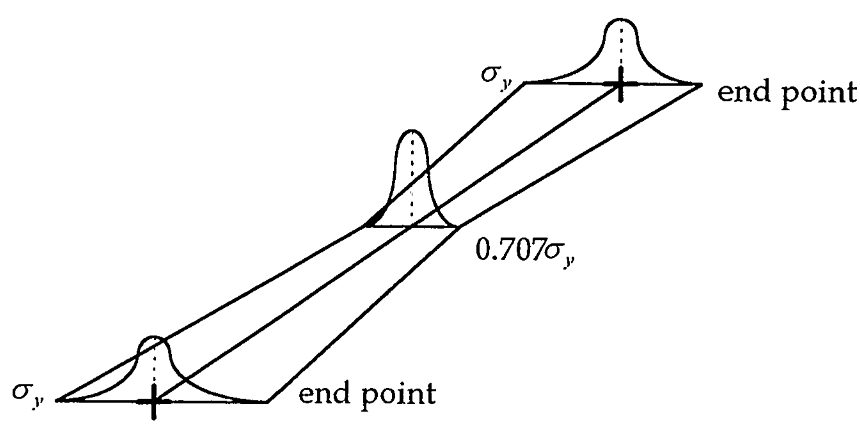

Figure 3.10 Probability Distribution of Line Segment

The probability distribution of a line segment is an analytical expression and was used instead of the Monte-Carlo simulation techniques determined by Dutton. We can use this result to quantify the uncertainty such as how close a point is to the boundary.

3.5.2 Confidence Region

According to the two dimensional normal distribution, we can construct a confidence region such that all points (corresponding to $r \in [0,1]$) will fall into the region at the same time with a probability greater than a pre-defined confidence level. For a fixed r, a confidence interval for $(1-r)\xi_1 + r\xi_2$ can be defined by standardizing the stochastic variable $(1-r)x_1 + rx_2$, we have

$$\frac{(1-r)x_1 + rx_2 - (1-r)\xi_1 - r\xi_2}{(1-2r+2r^2)^{1/2}\sigma_x} \to N(0,1) \tag{3.35}$$

The confidence interval can be defined by an expansion factor *c*. If the *c* is greater than 1, the probability level is increased.

3.6 Further Development of the "Developed Error Band" Model in 3-D

Those models of positional uncertainty, e.g. epsilon band model and error band model, deal with two dimensional information. Sometimes people use the metadata of uncertainty indicators as the third dimension other than the data itself. In reality, objects possess 3-dimensional positional information (x, y, z). There are z values in most GIS packages. People often ignore the third dimension either because of the limitation of the models or it is not necessary for the specific geographical application. For some GIS applications such as geological information, people need to know the distribution of the error for the subsurface in the vertical direction. Definitely, we need to consider the third dimension while modeling the positional uncertainty.

Because of the advantages of the "Developed Error Band" model, we can easily extend the positional uncertainty in three dimensions by using the developed error band model and error ellipsoids. With similar assumptions:

1. Coordinate errors of end points are independent and follow a three dimensional normal distribution.

2. A straight line segment connects two end points spatially.

we can derive any arbitrary point on the line segment between end points by using the Law of Error Propagation. If we simplify, and let all the covariance of the end points equal to zero and allow the two end points to have different variances, we have

$$p_1 = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} \rightarrow N_3\left(\begin{pmatrix} \xi_1 \\ \eta_1 \\ \varsigma_1 \end{pmatrix}, \begin{pmatrix} \sigma_{x_1}^2 & 0 & 0 \\ 0 & \sigma_{y_1}^2 & 0 \\ 0 & 0 & \sigma_{z_1}^2 \end{pmatrix}\right) \tag{3.36}$$

$$p_2 = \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} \rightarrow N_3\left(\begin{pmatrix} \xi_2 \\ \eta_2 \\ \varsigma_2 \end{pmatrix}, \begin{pmatrix} \sigma_{x_2}^2 & 0 & 0 \\ 0 & \sigma_{y_2}^2 & 0 \\ 0 & 0 & \sigma_{z_2}^2 \end{pmatrix}\right) \tag{3.37}$$

For any arbitrary point p, $r \in [0,1]$, we have

$$p_r = \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} \rightarrow N_3\left[\begin{bmatrix} (1-r)\xi_1 + r\xi_2 \\ (1-r)\eta_1 + r\eta_2 \\ (1-r)\varsigma_1 + r\varsigma_2 \end{bmatrix}, \begin{bmatrix} (1-r)^2\sigma_{x_1}^2 + r^2\sigma_{x_2}^2 & 0 & 0 \\ 0 & (1-r)^2\sigma_{y_1}^2 + r^2\sigma_{y_2}^2 & 0 \\ 0 & 0 & (1-r)^2\sigma_{z_1}^2 + r^2\sigma_{z_2}^2 \end{bmatrix}\right] \tag{3.38}$$

That means the point also follows the three dimensional distribution and falls into the region with a probability level greater than the pre-defined confidence level. For example, with the standard deviation of the two end points, a point falls within the confidence region defined by the end points with a probability of 19.9%.

So, we have the same results:

- The error at the midpoint of a line segment is the smallest with the same probability level if the errors of the end points are the same. That means we are more certain about the midpoint.
- The errors at the two end points are the largest and have the same probability level. That means we are more uncertain about the end points.

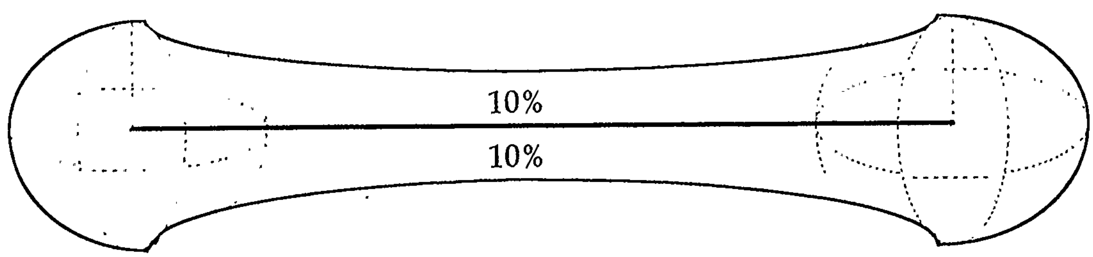

Figure 3.11 The "Developed Error Band" in 3-D

3.7 Positional Uncertainty of Polygon and Solid Boundaries

Once we have the uncertainty model of the line segment, we can use a line segment component to construct the area object and solid boundary.

3.7.1 Area Object

A line feature is defined as a feature composed of more than one line segment. If a line feature constitutes the boundary of an area object, it is a boundary line feature. An area object or polygon is defined as an area closed by a boundary line feature. The polygon uncertainty is described by the probability that a point belongs to the area object. For example, when an arbitrary point moves from the outside to the core area of the area object, the probability changes from 0 to 1. So we can build the polygon confidence region by using line segments as primitives.

3.7.2 Solid Boundary

Utilizing the same concept as that of the area object, we can use an area object to construct solid object because each side of the solid object is composed of an area element. The uncertainty of the solid boundary is described by the probability level that a point falls into the volume of the solid boundary. When an arbitrary point moves from the outside to the center of the solid boundary, the probability changes from 0 to 1.

CHAPTER FOUR

3-D VISUALIZATION OF POSITIONAL UNCERTAINTY

4.1 Introduction

The purpose of calculating uncertainty information is to allow GIS users (e.g. decision-makers) to judge the confidence level attached to a GIS. Once estimated, the uncertainty information must be presented in conjunction with the data to which it applies. Therefore, we need a method to communicate the uncertainty information to the users. How to convey the meta-data of uncertainty in a GIS is very important not only because users should be able to compare the accuracy of database outputs against the accuracy required for their tasks in order to make quality assessments, but also to protect the integrity of the past, present and future decisions that may utilize such information.

Obviously, the technique of graphics and imagery have already shown their power to the users seeking to understand physical phenomena represented numerically. The use of graphics has been proven to be the right trend in communicating uncertainty. DiBiase (1992) said that "The demand for effective graphic methods for data analysis and presentation has increased concomitantly. Collectively, these methods have come to be known as **visualization**". Visualization provides an ability to organize abstract concepts into intuitive pictures. It is a tool to see and explore complex relationships by the manipulation of geometry, color and motion.

What is the information to be visualized in GIS? Sinton (1978) suggests that all spatial data should be observed with respect to location, theme and time. The meta-data of uncertainty definitely has to follow this rule. In this study, we concentrate on the positional uncertainty, so the spatial dimension would be the most important part. Spatial dimension could refer to the number of dimensions supported by a data model or by the system software. This could be one dimensional as in distances and angles, two dimensional as a coordinate location on a plane, or three dimensional — x, y, z coordinates in 3-D space (Beard, 1991). For geologists, oceanographers, etc., the third dimension could be an important quality aspect of the data since they are more concerned about the z values.

The expression of positional uncertainty varies for each data type. In GIS, the point, line and polygon are the three basic elements. The positional uncertainty of a point is usually visualized as an error ellipse while the line is visualized as a epsilon band. Traditional GIS software package such as ARC/INFO uses the epsilon band as a measure of spatial uncertainty. It creates a look-up table with all the possible values of epsilon and assigns a symbol to each value of epsilon. All the symbols for lines correspond to different shades, patterns, colors and thickness with which each arc could be displayed at a different epsilon value. So the user may query the database and display the uncertainty information contained in the application such as sliver polygons caused by overlay operation or epsilon band uncertainty of the road arcs, river arcs or forest arcs.

Shi (1993) uses the analytical method of the confidence regions of a line segment to visualize the uncertainty according to the variances of the end points.

Figure 4.1 shows the confidence regions of a line segment for different probability levels. The variances for the end points are 1. The width of the confidence regions is exaggerated with respect to the line itself in order to show the shape of the confidence region clearly. The two white dots represent the locations of the two "measured" end points of the line segment. Figure 4.1 (a), (b), (c) and (d) shows the different confidence regions containing the true location of the line segment with the probabilities of 50%, 90%, 95% and 99.8%, respectively. When the probability level increases, the area of the confidence region increases as well. For example, when we have a high probability level, we are more certain about the true line falling within the confidence region. Figure 4.1 also demonstrates that when the variances of the end points are increasing, the confidence region is also increasing, which means the users are more uncertain about the measurements .

Figure 4.2 is the representation of the uncertainty of a polygon. The gray values represent the probability that a point belongs to the object. The darker that the gray value is, the higher the probability that the point belongs to the object. There is a higher certainty that the point at a dark region belongs to the object. In the figure, the interior region is black meaning that the probabilities of these points belonging to the object is equal to 1.

Most of the current GIS systems are based on a two dimensional data model. It is difficult for them to handle three dimensional objects and, therefore, it is impossible to directly edit and render 3-D objects with current 2-D GIS systems. Therefore, most of the visualization of the positional uncertainty is based on two dimensional information because of the limitation of the application or the available display tools. People can only view the variation of x and y coordinates without the z value. But we all know the z value has the same significance as that of x and y, so the information of vertical information must be included in the modeling and visualization.

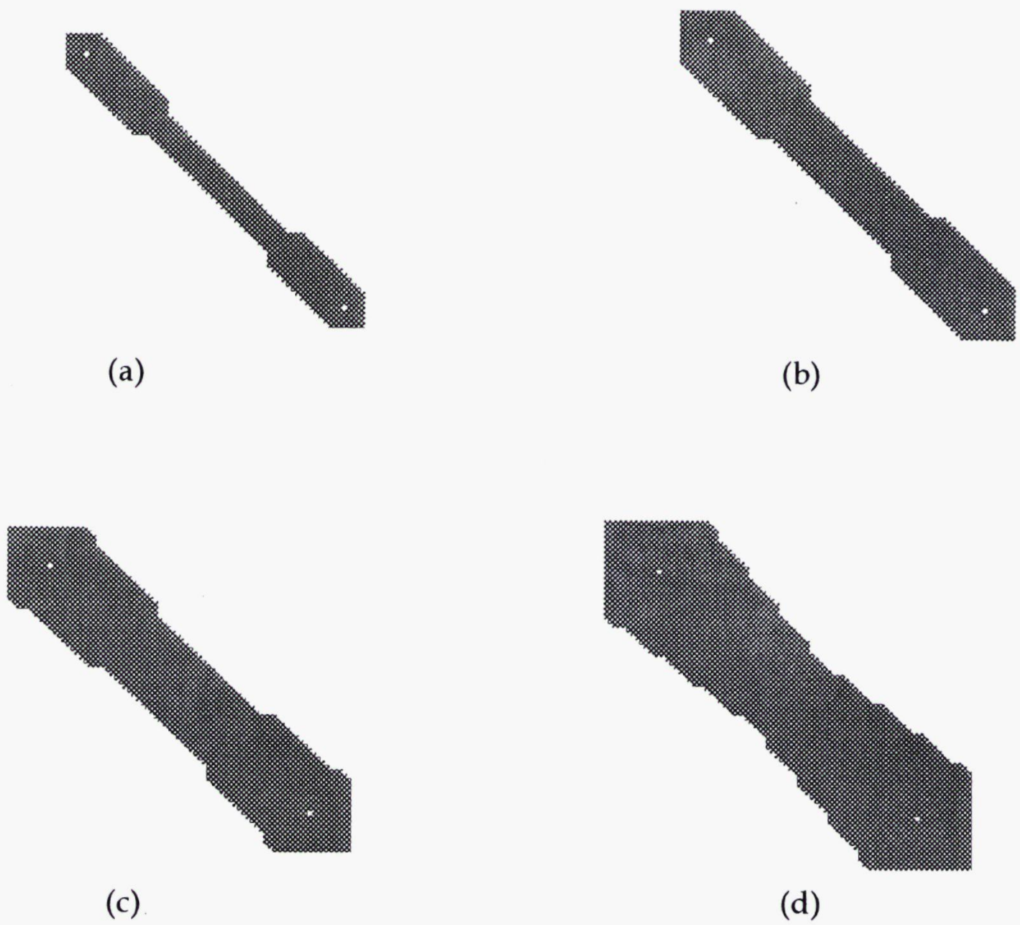

Figure 4.1 Confidence Regions of a Line Segment Varying According to Probability Level (from Shi, 1993)

(a) 50% probability level (b) 90% probability level

(c) 95% probability level (b) 99.8% probability level

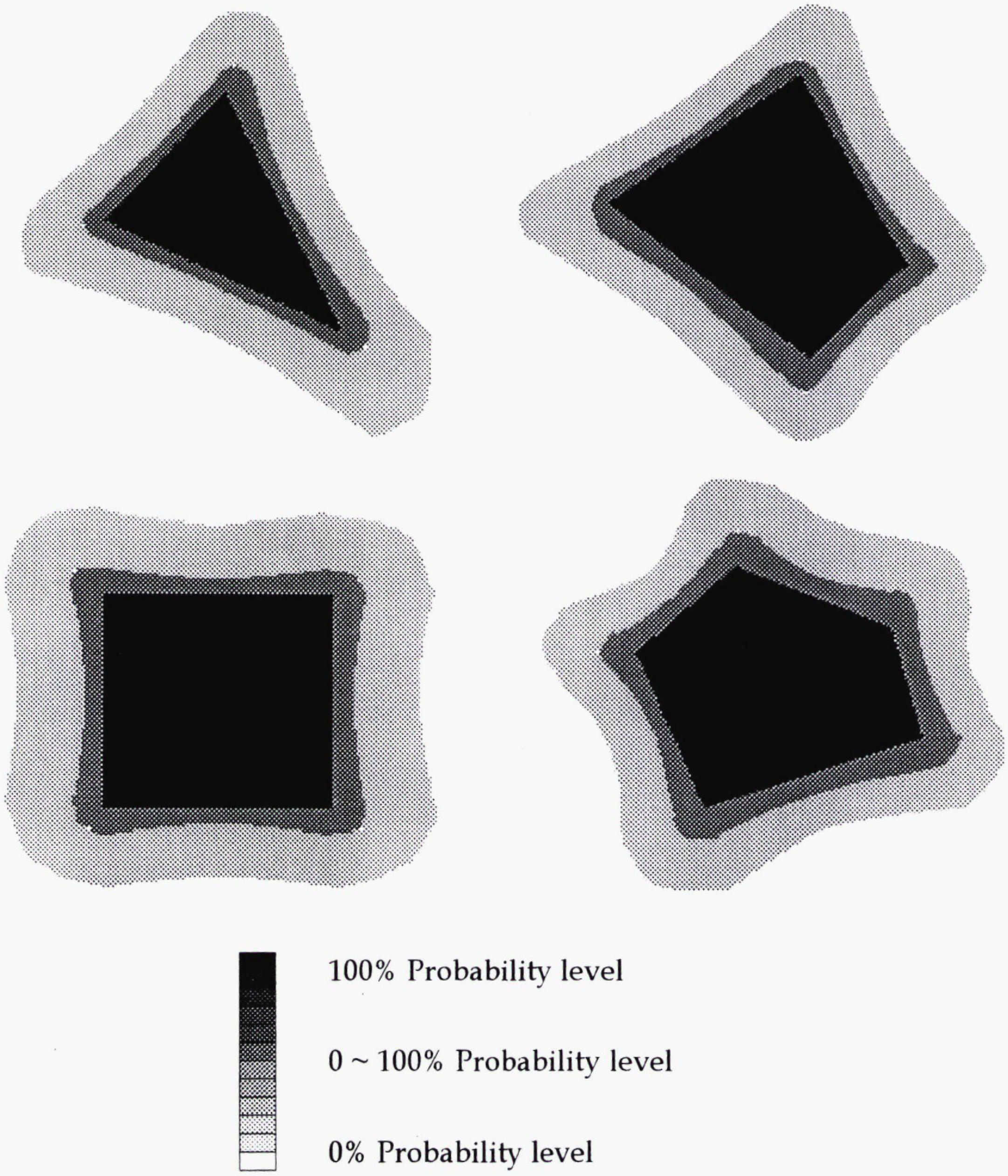

Figure 4.2 Positional Uncertainty of Area Objects (from Shi, 1993)

In Chapter three, we have the theoretical model for the positional uncertainty in three dimensions. The approach adopted in this study is to find a tool which has the capability of three dimensional visualization and to use that tool to implement the 3-D positional uncertainty model.

4.2 Tools for Visualization

The Graphic Library (GL) from a Silicon Graphic workstation was chosen to be the tool for visualizing 3-D positional uncertainty. GL is a library of subroutines for creating 2-D and 3-D color graphics and animation. It has the full functionality and flexibility to build an application using GL commands within the framework of the programming language. The C programming language provides the logical structure for our program, and GL commands provide the interface to the graphics software and hardware.

4.3 Data Structure of 3-D Objects

Because of the characteristics of the positional uncertainty, each object will have meta-data of uncertainty in addition to the spatial information of x, y and z values. The confidence regions at different probability levels are the indicators of positional uncertainty as determined by the "Developed Error Band" model.

Four different geometric object types are proposed to be visualized in 3-D (Figure 4.3):

- Point objects: zero-dimensional objects which have positions (mean values ξ, η, ζ) and corresponding standard deviations $(\sigma_x, \sigma_y, \sigma_z)$ but no spatial extension;
- Line objects: one-dimensional objects with length as the only measurable spatial extension, shape and position, which means that Line Objects are built up of connected line segments, while line segment is composed of two end points;

- Surface objects (polygon or area): two-dimensional objects with area and perimeter as measurable spatial extensions. They can have a 3-dimensional shape. Surface Objects are composed of surface segments or faces;

- Body objects (solid boundary): three-dimensional objects with volume and surface area as measurable spatial extensions. They are bordered by a surface, and built up from faces (Rikkers, et. al., 1994).

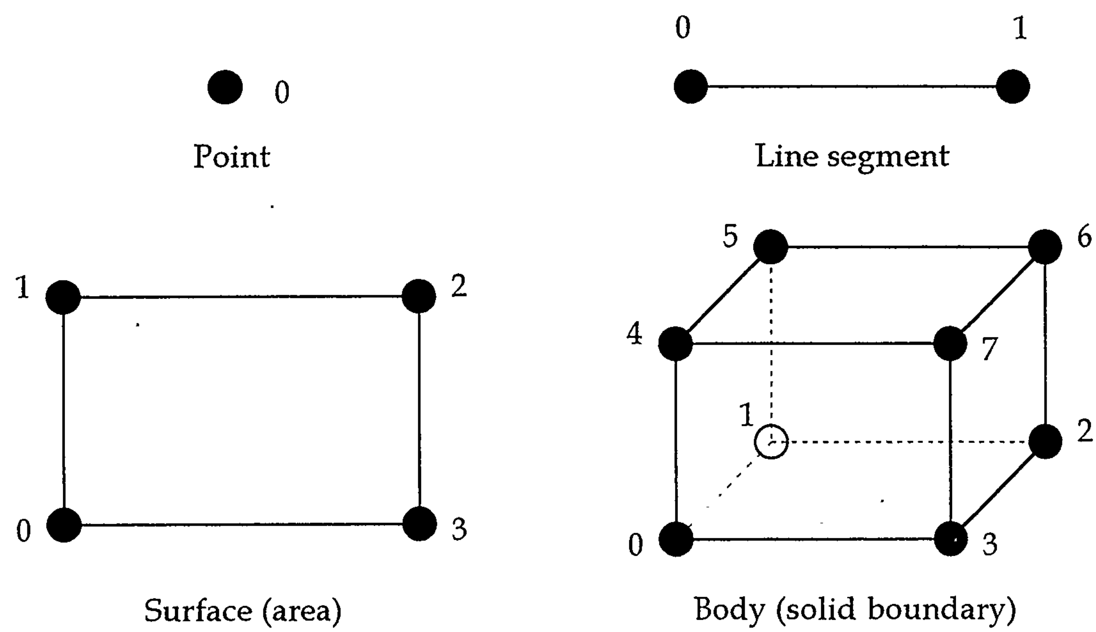

Figure 4.3 Geometric Object Types

Since the objective of this study is to investigate the possibility of modeling and visualization of positional uncertainty in 3-D, the management of 3-D positional uncertainty is an issue here. One method is to create a few tables containing all of the positional information by using the concept of a relational database structure.

The relational database structure in its simplest form stores no pointers and has no hierarchy. Instead, the data are stored in simple records grouped in two-dimensional tables. Each table is usually a separate file. Data are extracted from a relational database through a procedure in which the user defines the relation that is appropriate for the query. This relation is not necessarily already present in the existing files, so the controlling program uses the methods of relational algebra to construct the new tables. Tables 4.1, 4.2, 4.3 and 4.4 give examples of data structures for an object point, line, area and solid boundary, respectively, which are employed in the implementation of visualization.

Point ID	ξ	η	ζ	σ_x	σ_y	σ_z
0	0	0	0	0.6	0.6	0.6
1	0	6	0	0.6	0.6	0.6
2	6	0	0	0.6	0.6	0.6
:	:	:	:	:	:	:
7	8	0	8	0.8	0.8	0.8

Table 4.1 Data Structure for Points

Each end point has a record of ID number, mean values (ξ, η, ζ) of the position (x, y, z) measured and corresponding standard deviations in Table 4.1.

Line ID	From (Point ID)	To (Point ID)
0	0	1
1	1	2
2	2	3
:	:	:
12	3	7

Table 4.2 Data Structure for Lines

Each line segment is constructed by the two ends of starting point and ending point in Table 4.2.

Area ID	Line ID			
0	0	1	2	3
1	7	6	5	4
2	0	8	4	1
:	:	:	:	:
5	3	11	7	4

Table 4.3 Data Structure for Surfaces (Areas)

Each area is composed of the line segments which are the boundary of the area in Table 4.3.

Body ID	Area ID					
0	0	1	2	3	4	5
:	:	:	:	:	:	:

Table 4.4 Data Structure for Body Objects (Solid Boundaries)

Each solid boundary is composed of the areas in Table 4.4.

In order to reduce the time that the system takes to draw the object on the computer screen, the data for each object are stored in its own table after calculating the error at each point. For complex objects, such as irregular polygons, since they are composed of elementary data types of the geometrical parts, linear interpolation and the Error Propagation law can be used to construct them.

4.4 Visualizations

In order to visualize the positional uncertainty, positional information is simulated in Figure 4.4. There are 8 points having the mean values of (ξ, η, ζ) and corresponding standard deviations ($\sigma_x, \sigma_y, \sigma_z$). The standard deviations are designed to be large compared to the lengths of the lines. This is make the shape of confidence regions more obvious. Figures 4.5(a), 4.5(b), 4.6(a), 4.6(b), 4.7(a) and 4.7(b) are the visualizations of the four elementary geometric objects. Color is used as the indicator of the probability level.

Point ID	ξ	η	ζ	σ_x	σ_y	σ_z
0	0.0	0.0	0.0	0.8	0.8	0.8
1	0.0	6.0	0.0	0.8	0.8	0.8
2	8.0	6.0	0.0	0.8	0.8	0.8
3	8.0	0.0	0.0	0.8	0.8	0.8
4	0.0	0.0	6.0	0.6	0.6	0.6
5	0.0	6.0	6.0	0.6	0.6	0.6
6	8.0	6.0	6.0	0.6	0.6	0.6
7	8.0	0.0	6.0	0.6	0.6	0.6

Figure 4.4 Simulated Data for the End Points

In Figure 4.5(a), the confidence region of line segment is visualized at a probability level of 19.9% with the given standard deviations. The uncertainty of a line segment has a bone shape because the middle point has the least error when the two end points have the same variance. Figure 4.5(b) is the profile of confidence regions of line segment at different probabilities of 19.9%, 73.4% and

96.8% with standard errors of $(\sigma_x, \sigma_y, \sigma_z)$, errors of $2(\sigma_x, \sigma_y, \sigma_z)$ and near certainty errors of $3(\sigma_x, \sigma_y, \sigma_z)$, respectively. The straight line in the center is the measurement or estimation of the true line defined by the two end points.

In Figure 4.6(a), the confidence region of a surface object or an area object is visualized at a probability of 19.9%. It has a concave shape and the central point has the least error when the four end points have the same variances. Figure 4.6(b) is the profile of the uncertainty of the area at different confidence levels of 19.9%, 73.4% and 96.8%. The volume of the confidence region increases when the probability level increases. The gray plane at the center is the measurement of the surface constructed by the four end points.

Figure 4.7(a) is the frame-work of eight end points. When the true end points fall within their own confidence regions, the true object of solid boundary will fall within the confidence region (Figure 4.7(b)) with the same probability level of 19.9%. Each surface of the solid boundary is concave.

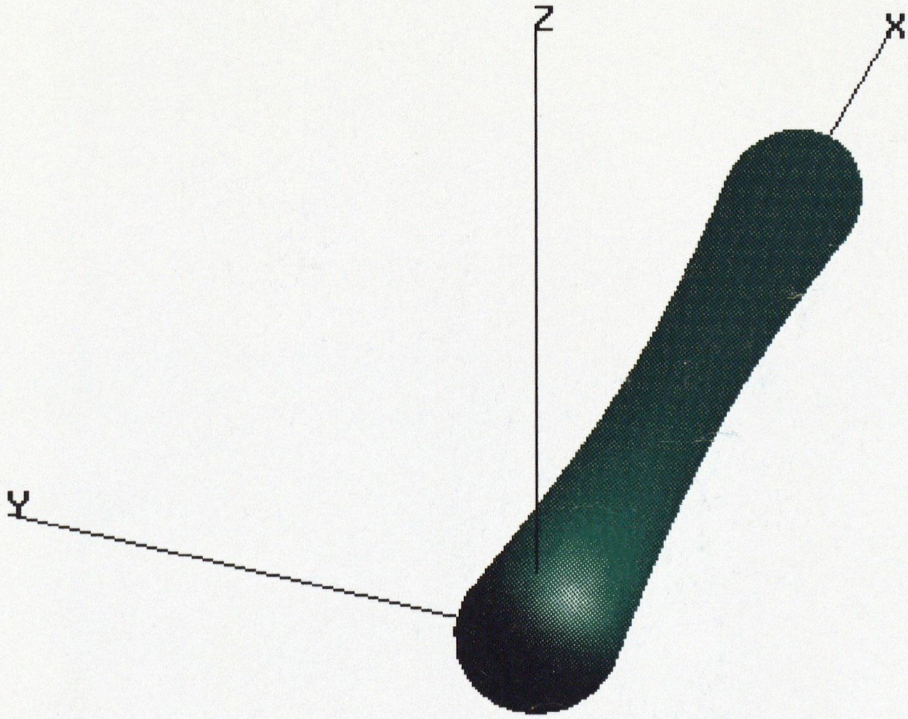

Figure 4.5 (a) Confidence Region For Line Segment

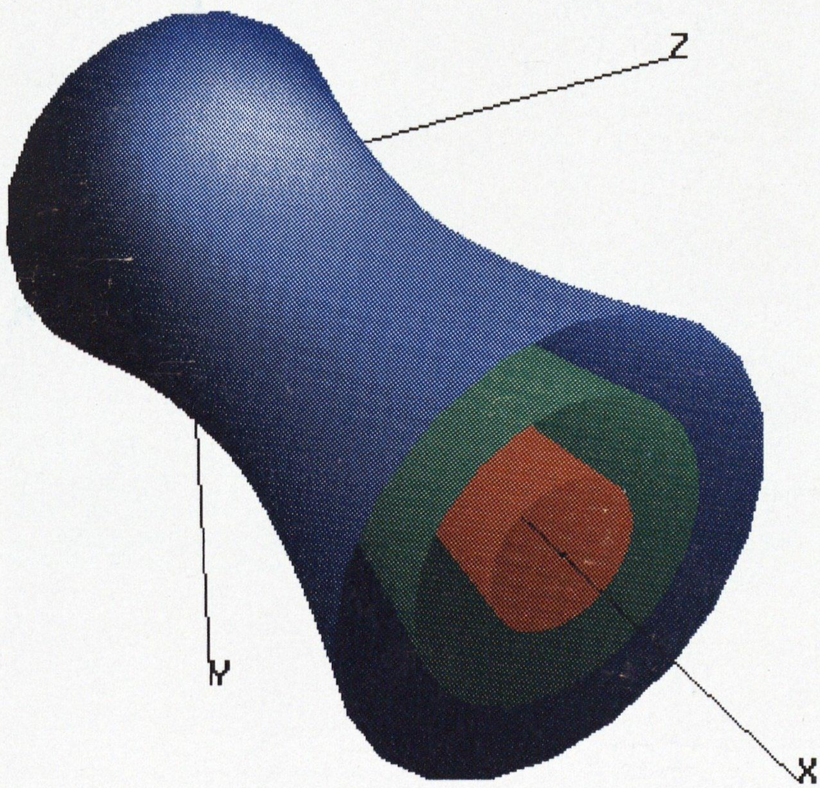

Figure 4.5 (a) Confidence Regions for Line Segment at Different Levels

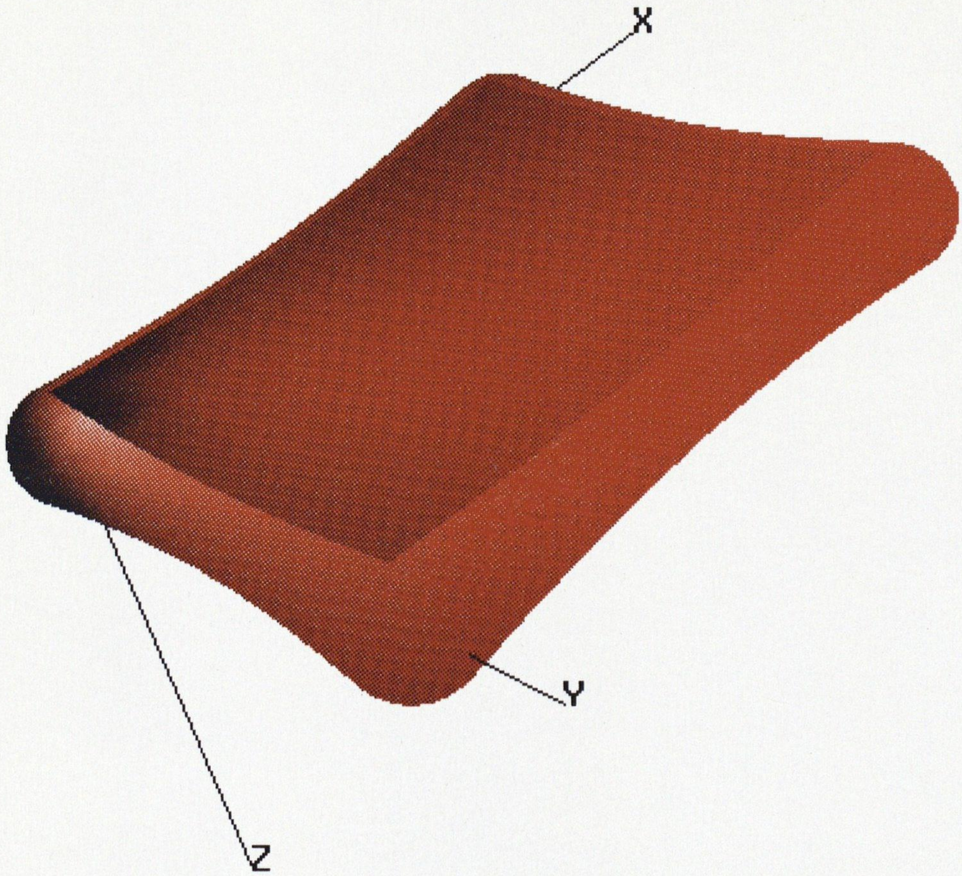

Figure 4.6 (a) Confidence Region for Area

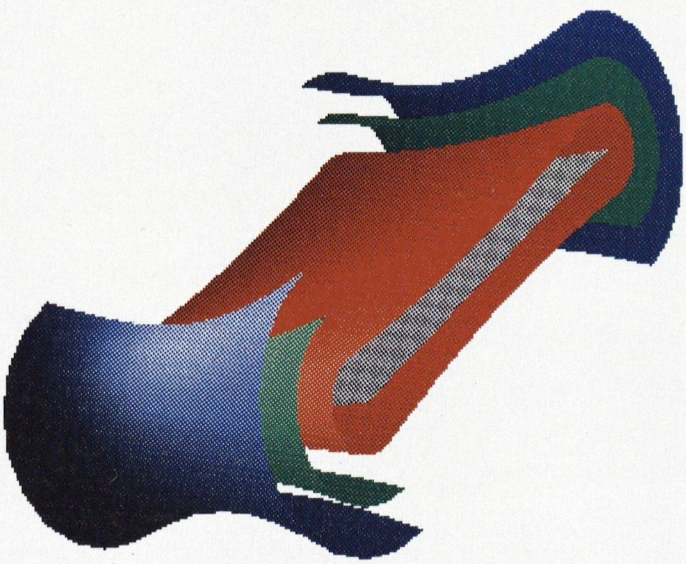

Figure 4.6 (b) Confidence Region for Area at Different Levels

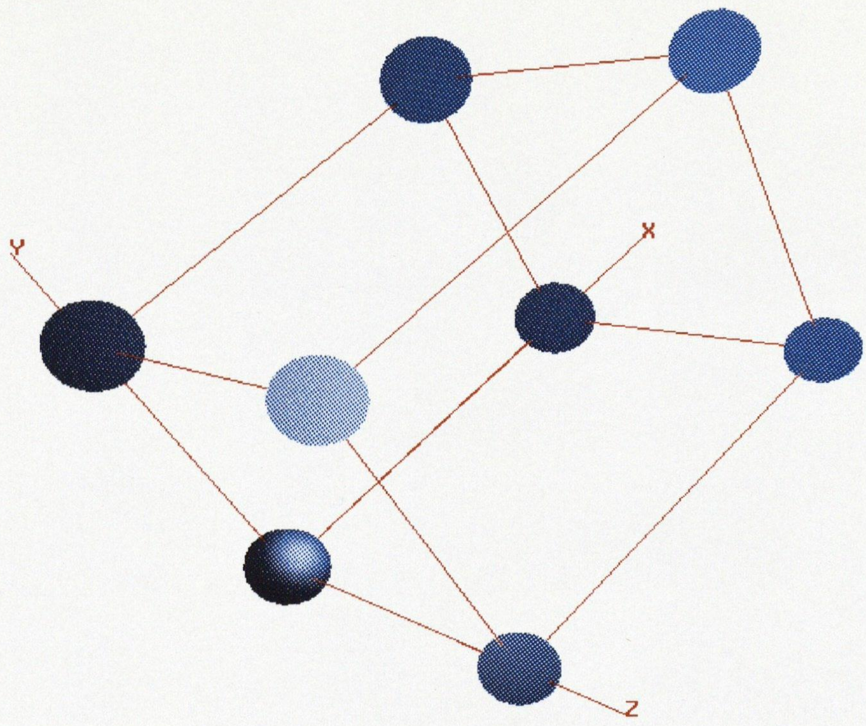

Figure 4.7 (a) Confidence Region for Points

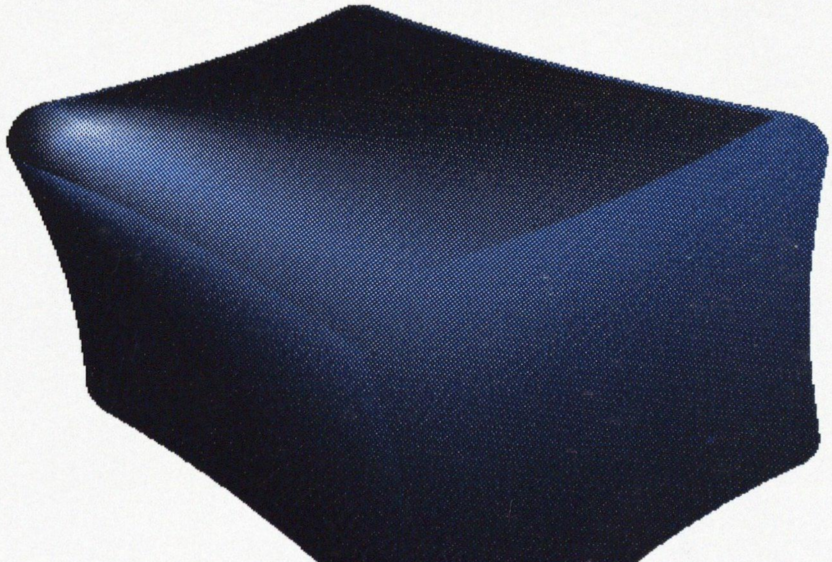

Figure 4.7 (b) Confidence Region for Solid Boundary

4.5 Reporting of Positional Uncertainty

Data analysis is an important part of GIS. There is a need for the reporting in certain cases for spatial database products. While the considerable time and effort investigating the modeling and visualization of uncertainty in spatial databases, the results will not be recognized until users start applying the techniques in operational situations. The reporting of the uncertainty will present the level of uncertainty which assists users in understanding the quality of output from the system (Hunter, et al., 1994). The reporting of positional uncertainty is included in this study.

Regardless of whether one chooses to interpret the loci of positional uncertainty as a sausage or a bone, the "Developed Error Band" model provides a way to "query" spatial data at any desired level of confidence. The higher the confidence level users wish to have, the broader the band of error will be around any given feature (Dutton, 1991).

Relational databases have the great advantage that their structure is very flexible and can meet the demands of most queries that can be formulated using the rules of Boolean logic and many mathematical operations. They allow different kinds of data to be searched, combined, and compared. Two query functions were implemented in this study.

When variances corresponding to a certain confidence level are given, the confidence region will be displayed in order to compare the magnitude of the positional uncertainty with the spatial extension such as length, area or volume (Figure 4.8). The volume of the confidence region increases when the probability level increases. This allows the users or the decision-makers could intuitively know what errors they are prepared to tolerate, weigh the accuracy level against their GIS requirement and the cost effectiveness of collecting precise data.

On the other hand, the user could know the uncertainty information at a specific point. When a selected point on the computer screen is clicked, the coordinates, errors and confidence level at that point will be reported. This function will help people to understand their data quality. Figure 4.9 (a) and (b) are the reports of the query operations at different distances to the end points. When the points at the same confidence level are searched, the errors are smaller in the middle than those close to the end points. Figure 4.10 is the report of the query operations at different confidence levels. When the confidence level is high, the errors are large. These results proved that the query function works well and the results of the query operations are correct according to the principal of the "Developed Error Band" model.

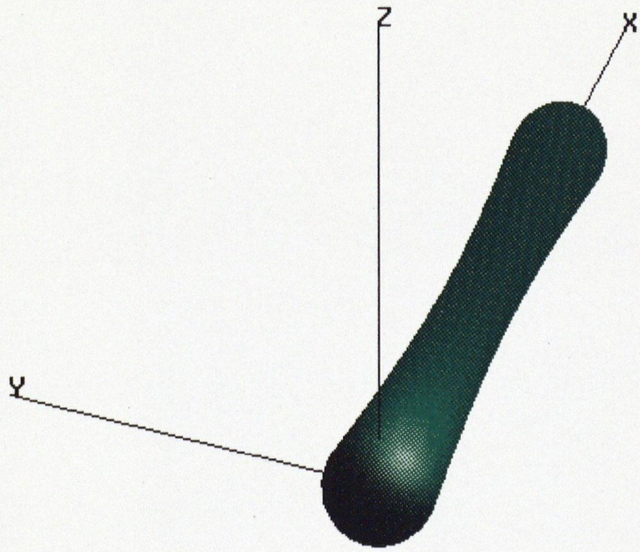

Figure 4.8 (a)

Confidence Region with Probability

Level of 19.9%

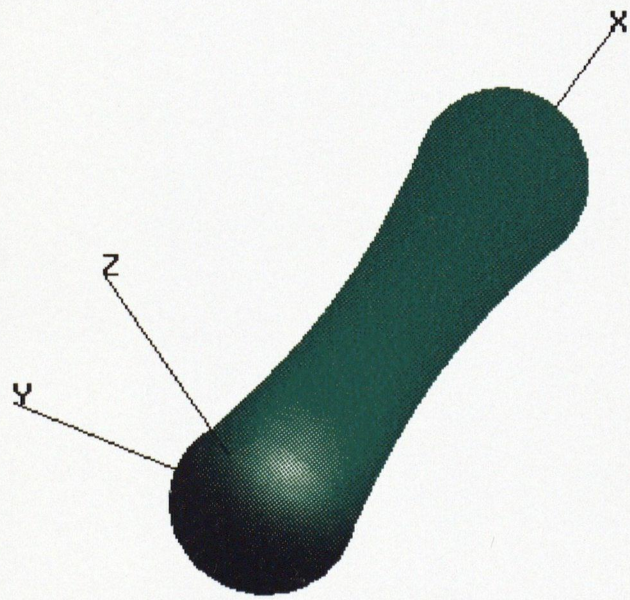

Figure 4.8 (b)

Confidence Region with Probability

Level of 73.4%

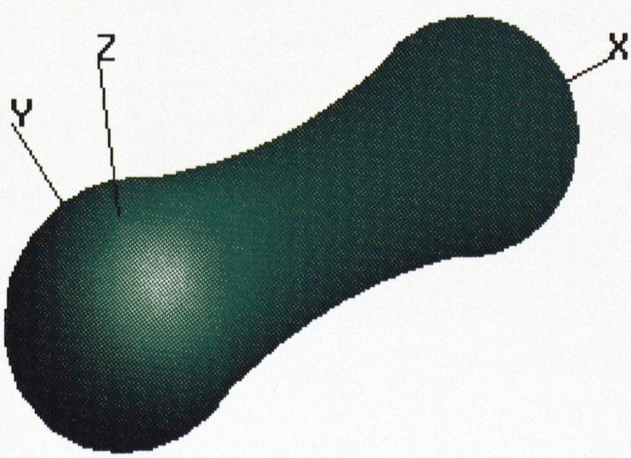

Figure 4.8 (c)

Confidence Region with Probability

Level of 96.8%

The coordinates of the selected point 1 are:
x= 0.00
y= 0.00
z= 0.00
The errors at this point are:
sx= 2.40
sy= 2.40
sz= 2.40
Within a confidence region of c=3 at a probability of 96.8 percent

The coordinates of the selected point 2 are:
x= 1.50
y= 0.00
z= 0.00
The errors at this point are:
sx= 2.00
sy= 2.00
sz= 2.00
Within a confidence region of c=3 at a probability of 96.8 percent

The coordinates of the selected point 3 are:
x= 3.00
y= 0.00
z= 0.00
The errors at this point are:
sx= 1.75
sy= 1.75
sz= 1.75
Within a confidence region of c=3 at a probability of 96.8 percent

The coordinates of the selected point 4 are:
x= 4.00
y= 0.00
z= 0.00
The errors at this point are:
sx= 1.70
sy= 1.70
sz= 1.70
Within a confidence region of c=3 at a probability of 96.8 percent

Figure 4.9(a) Report of Query at Different Distances

The coordinates of the selected point 5 are:
x= 4.50
y= 0.00
z= 0.00
The errors at this point are:
sx= 1.71
sy= 1.71
sz= 1.71
Within a confidence region of c=3 at a probability of 96.8 percent

The coordinates of the selected point 6 are:
x= 6.50
y= 0.00
z= 0.00
The errors at this point are:
sx= 2.00
sy= 2.00
sz= 2.00
Within a confidence region of c=3 at a probability of 96.8 percent

The coordinates of the selected point 7 are:
x= 7.00
y= 0.00
z= 0.00
The errors at this point are:
sx= 2.12
sy= 2.12
sz= 2.12
Within a confidence region of c=3 at a probability of 96.8 percent

The coordinates of the selected point 8 are:
x= 8.00
y= 0.00
z= 0.00
The errors at this point are:
sx= 2.40
sy= 2.40
sz= 2.40
Within a confidence region of c=3 at a probability of 96.8 percent

Figure 4.9 (b) Report of Query at Different Distances

The coordinates of the selected point 9 are:
x= 8.00
y= 0.00
z= 0.00
The errors at this point are:
sx= 2.40
sy= 2.40
sz= 2.40
Within a confidence region of c=3 at a probability of 96.8 percent

The coordinates of the selected point 10 are:
x= 8.00
y= 0.00
z= 0.00
The errors at this point are:
sx= 1.60
sy= 1.60
sz= 1.60
Within a confidence region of c=2 at a probability of 73.4 percent

The coordinates of the selected point 11 are:
x= 7.00
y= 0.00
z= 0.00
The errors at this point are:
sx= 1.41
sy= 1.41
sz= 1.41
Within a confidence region of c=2 at a probability of 73.4 percent

The coordinates of the selected point 12 are:
x= 8.00
y= 0.00
z= 0.00
The errors at this point are:
sx= 0.80
sy= 0.80
sz= 0.80
Within a confidence region of c=1 at a probability of 19.9 percent

Figure 4.10 Report of Query at Different Confidence Levels

CHAPTER FIVE

POSITIONAL UNCERTAINTY IN THE MEASUREMENTS OF VISAT

5.1 Introduction

Once the concept of modeling and visualization of positional uncertainty is investigated, the possibility of its application needs to be studied. In this chapter, an experiment using the application of positional uncertainty is reported with respect to the determination, representation, and display of the uncertainty during data acquisition (manual digitization) in the VISAT project.

5.2 VISAT Overview

VISAT (Video, Inertial and Satellite System) is a prototype of a real-time system for highway GIS acquisition that is currently being developed in the Department of Geomatics Engineering at The University of Calgary and Geofit Inc.. The primary purpose of this system is to provide precise spatial information with an absolute positioning accuracy of 0.3 m and a 0.1 m relative accuracy for objects seen within a 50 m radius of the cameras along a highway corridor (Schwarz et. al., 1993). The system employs a GPS (Global Positioning System) to update the absolute position information, an INS to provide orientation information, and an array of CCD (Charge-Coupled Device) cameras (from two up to six) to supply the digital images of the highway corridor. The vehicle carrying the INS, GPS receiver and the cameras travels at 50 - 60 km per hour, taking and recording images of objects every 0.7 seconds, while the GPS/INS system simultaneously records the time, positions and orientations of the cameras.

The on-board computer records the image data and navigation data on a mass storage unit. The deliverables of the field survey module of the VISAT system are binary digital image files with position and orientation parameters of the cameras and associated accuracy encoded in a reserved header area. In the post-processing procedure, a highway information system called GeoStation has been developed which processes digital video images controlled by GPS and INS from the VISAT system and generates a spatial database. This system was developed on a SUN workstation with two monitors using an X-windows and MOTIF environment. One monitor is used for the Overview system with the base map and P-lines (i.e. route of the vehicle) displayed; another monitor is used for a Multi-view system with multiple images and displaying the graphic user interface.

After the image data is loaded in the post-processing computer, the user can measure objects of the stereo images on the screen by using a mouse. Since all of the positioning and orientation parameters for each image are known from the header of the image, the 3-D ground coordinates of the desired point from at least two images can be measured by using digital photogrammetric triangulation functions.

5.3 Error Sources in VISAT

The purpose of developing the VISAT system is for the acquisition of fast, accurate and efficient object measurements from a moving vehicle. The measurements can be done in a comfortable office instead of in the field. This system is so successful that much has been written about the expected accuracy. Since this system is still under development, the achievable accuracy in varying circumstances under actual conditions should be investigated also.

In view of the complicated technologies incorporated in VISAT, it is easily understood that the errors are potentially numerous and varied. Table 5.1 is the summary of VISAT errors and their contribution to point positions. Under ideal conditions, the GPS/INS component of the system is the major contributor of the errors (almost half the total allocated system error budget). The GPS/INS error is involved in the phase of system calibration. Digitizing error is another significant contributor which occurs in the data extraction phase.

Error Source	Error Source	Contribution to Point Positioning Worst-Case (50m)
GPS/INS absolute position	10 - 20 cm	10 - 20 cm
INS absolute orientation	1 arcmin	1.5 cm
Time synchronization of system components	1 - 3 msec	5 cm
Differential offsets between system components	1 - 2 cm	1 - 2 cm
Rotational offsets between INS and CCD cameras	10 - 20 arcsec	1 cm
Digitizing error (no edge enhancement employed)	0.289 pixel	2.46 cm across track 140 cm along track
Digitizing error (employing edge enhancement techniques)	1/20 pixel	0.4 cm across track 25 cm along track

Table 5.1 Summary of VISAT Errors and Their Contribution to Point Positions

(Revised from Szarmes, 1994)

Although the errors are introduced at different stages in a time sequence, it is clear that the total error is not the sum of these errors at each individual stage. For instance: $\varepsilon_{total} \neq \varepsilon_1 + \varepsilon_2 + \cdots$, where ε_{total} is the total error at a particular point position, and ε_1, ε_2, etc., are the standard errors introduced by different processes. In this system, the fact is of critical importance that all the errors are considered independent. In order to obtain the best achievable accuracy, the errors from different phases should be reduced to an acceptable level. When an error analysis is carried out, all the errors should be considered and calculated separately. Since this study focuses on the positional error / uncertainty caused by random measurements, digitizing error will only be considered via the application of the "Developed Error Band" model providing every kind of error such as GPS/INS error and lens distortion which are considered independently and individually.

5.4 Digitization Error in Data Measurement

During the data extraction phase, the system operator will digitize the various objects in different images to obtain the measurements. Usually, the operator only measures once to get the point coordinates (x, y, z). If the same object within the same images is digitized repeatedly such as 100 times, the point position will be different each time. Therefore, the measurement repeatability or reliability needs to be studied. The reason is that a point accuracy of 0.289 pixel (Chapman, 1989) is the best that can be achieved during the digitization of a discrete digital image if sub-pixel edge detection techniques are not employed.

Assuming that the images have a size of 512 × 512 pixels; the CCD cameras have a horizontal field of view of 50°, resulting in an angular value of 0.0017 radians for one pixel. Over 50 m, an error of 0.289 pixel yields an average across-track error of 2.46 cm. However, this same 0.289 pixel yields an average along-track error of 140 cm. Figure 5.1 shows the plan view of the photogrammetric triangulation geometry in this situation. It is obvious that the short baseline length (1.7 m of baseline length is currently being used for the prototype system) between the two conjugate exposure stations limits the radius that can be precisely digitized. A significant improvement occurs if the distance to the object is only 25 m resulting in an along-track error of only 36 cm and an across-track error of only 1.2 cm. It becomes evident that if sub-pixel edge-detection techniques are employed, the digitizing errors will be reduced. Suppose an edge of the road line can be measured with an accuracy of 1/20 pixel (Szarmes, 1994), an across-track error for an object 50 m away will be reduced to 0.42 cm while an along-track error of 25 cm is committed. Table 5.2 lists the errors for different situations, it is obvious that the digitizing error is very sensitive in the direction of the image coordinate frame z-axis which is the direction of vehicle moving. This happens also, when the image enhancement techniques are applied. Therefore, positional uncertainty at different object distances will be emphasized and visualized.

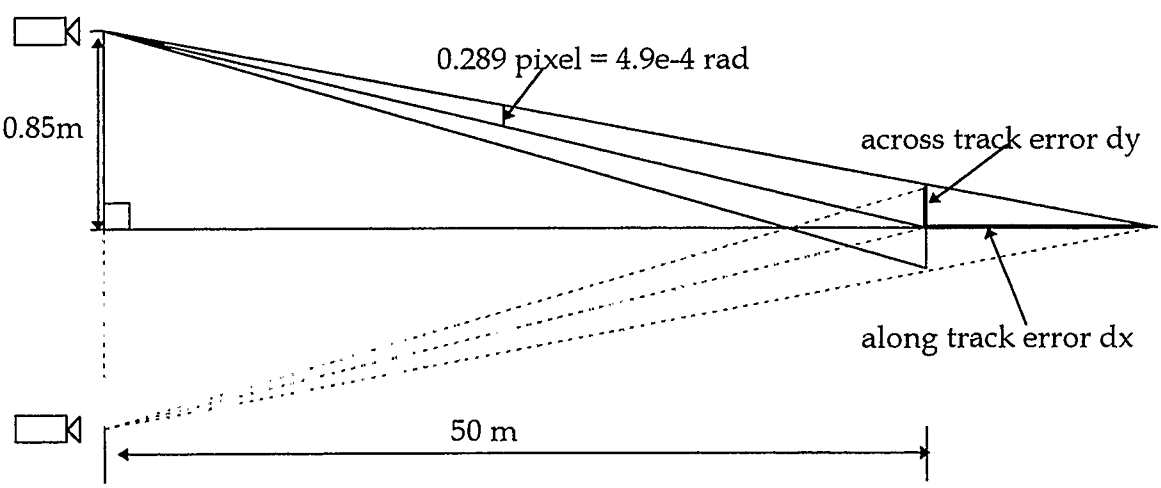

Figure 5.1 Plan View of Geometry of Photogrammetric Triangulation

(Revised from Szarmes, 1994)

Object distance(m)	No edge enhancement employed (0.289 pixel)		Edge enhancement techniques employed (1/20 pixel)	
	along-track dx	across-track dy	along-track dx	across-track dy
50	140 cm	2.46 cm	25 cm	0.42 cm
40	92 cm	1.95 cm	16 cm	0.34 cm
30	52 cm	1.48 cm	9 cm	0.25 cm
25	36 cm	1.22 cm	6 cm	0.20 cm

Table 5.2 Digitized Errors at Different Object Distances

5.5 Visualization of the Application to VISAT

Figure 5.2 are the visualizations of the uncertainty of digitized errors. Suppose a building boundary is digitized with each side being 10.0 m in length. Since the errors are so small compared to the object distance, the errors were exaggerated to show the shape of the boundary more clearly. The along-track errors are increased by a factor of 3 while the across-track errors are increased by a factor of 30. The results show that even with the difference of factors, the along-track error is still large compared to the across-track error. Because of the huge difference between the along-track errors and across-track errors, the shape at the end points seems sharp rather than round when the drawing interval is not small enough. Therefore, the additional image processing technique is necessary when an object is beyond 50 meters.

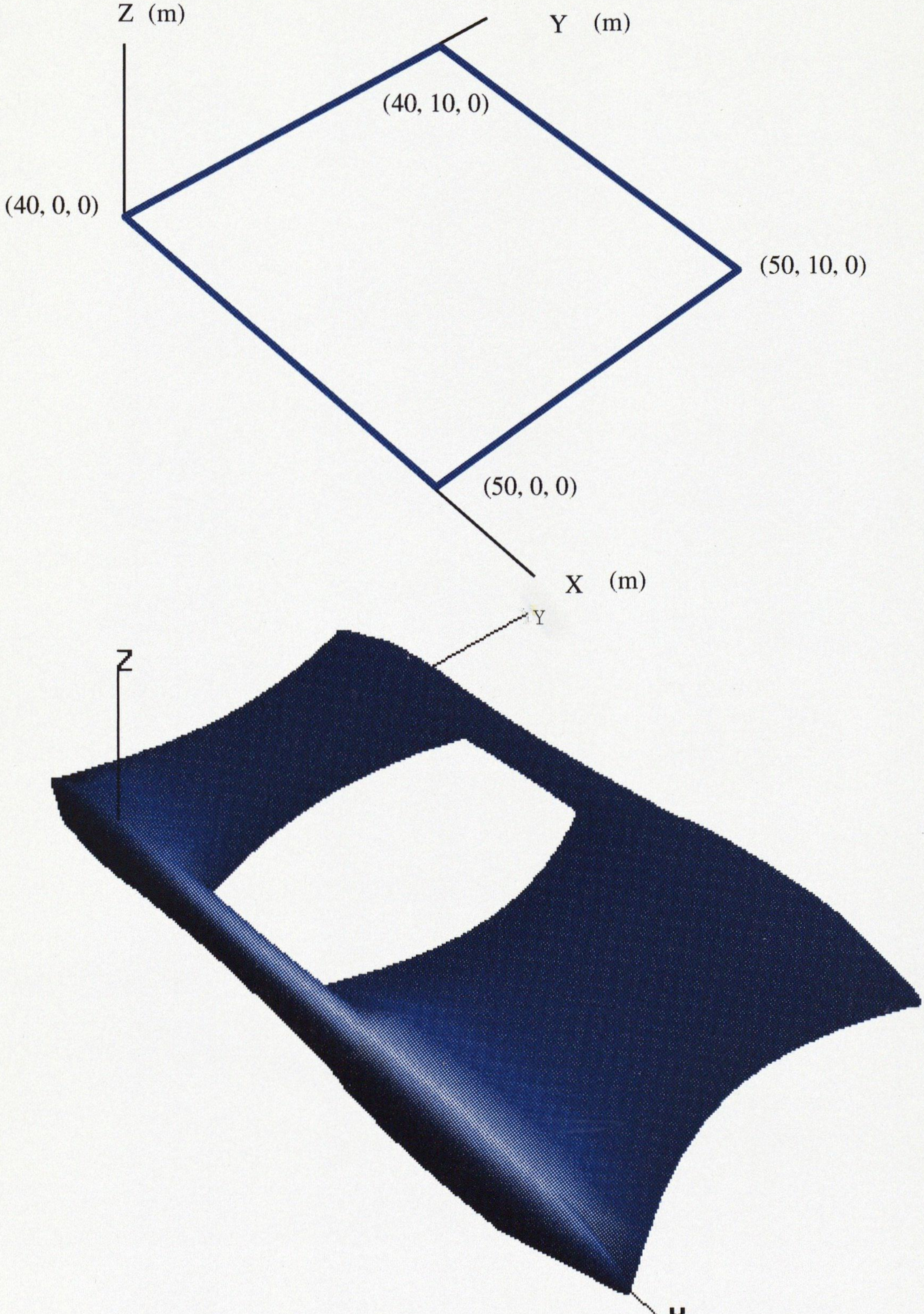

Figure 5.2 Confidence Region of Boundary in VISAT

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

Visualization of spatial data quality or positional uncertainty was identified as one of the most important research issues in GIS by NCGIA Research Initiative 7. Visualization should be explored as a method for capturing, interpreting and communicating quality information to users of GIS. Clearly, the quality of information varies spatially, and visual tools for the display of data quality will improve and facilitate the use of GIS, so positional uncertainty is a fundamental problem. At present, the visualization tools are either not available in existing GIS packages, error models are not well developed, or only developed in two dimensions. This thesis presents the theories which are used to handle the positional uncertainty due to random errors in GIS data. After the investigation of the different models, the "Developed Error Band" was adopted to model positional uncertainty, and this model has been extended to the situation of 3-dimensions. The 3-D graphic tools of the Graphic Library (GL) on the Silicon Graphic workstation was chosen to display the positional uncertainty containing three dimensional error information. Later, this method was applied to measurements in the VISAT project.

The sources of error / uncertainty in GIS have been discussed. The analysis of positional uncertainty has been emphasized. It is evident that error / uncertainty exists in all kinds of GIS applications, meta-data analysis and implementation should be part of the spatial information system to facilitate the understanding of data quality variations in digital data and to support decision-making.

Points, lines and polygons are three elementary geometric types used in spatial information systems. Error / uncertainty analysis is necessary for each type. The Law of Error Propagation plays a key role in the uncertainty analysis when only the point information is available. By using this law, the positional uncertainty of a line segment can be derived from points; the positional uncertainty of a boundary, an area, or a solid boundary can also be derived. A confidence region with a probability level is an indicator of the positional uncertainty. The "true" location falls into this confidence region with a probability higher than the pre-defined confidence level. Compared to the epsilon band model, the "Developed Error Band" model has a different shape of the confidence region. It is narrower in the middle of a line segment if the variances of the two end points are the same. The center points have smaller errors than other points on the line with the same confidence level and we are more confident about the center point. By using this model, the positional uncertainty of the boundary and area has been constructed.

With the same assumptions and simplifications that the end points are independent and follow a 3-D Normal Distribution, the "Developed Error Band" model has been extended to the 3-D situation. The error ellipsoid is used to indicate the point uncertainty with a specified confidence level. The positional uncertainty of line segment, area and solid boundary can be derived as well. Similarly, the middle point of the line segment has the smallest error on the line if the end points have the same variance. Compared to the 2-D case, if the confidence region is defined by the same standard deviation, the positional uncertainty in 3-D has a lower probability level of 19.9%, while the positional uncertainty in 2-D has a probability level of 38.5%. According to statistical theory, the probability level decreases when the degrees of freedom increases.

The "Developed Error Band" model has been implemented (coded in C) and visualized by using the Graphic Library tool provided by a Silicon Graphic workstation. A relational data structure, including the positional coordinates of the mean (x, y, z) and its meta-data ($\sigma_x, \sigma_y, \sigma_z$) has been built in order to manage the uncertainty information. Therefore, a realistic visualization for each geometric type has been displayed that was supported by GL. In addition, this implementation provides a method to query the spatial information at a certain probability level. Although this study was intended to investigate the positional uncertainty and visualization of the concept of this model, the results and its application to the VISAT project proved that the development of this method is appropriate. Since error / uncertainty exists, we have to fit them for use in a 3-dimensional space. The advantages of this approach can be summarized as follows:

- **Management**

The modeling and visualization of uncertainty will help users manage the information. This method describes the positional uncertainty not only qualitatively but also quantitatively. It helps in the understanding of the effect of uncertainty when users are in a decision-making process. They can determine the uncertainty that they are prepared to tolerate in their product.

- **Accuracy**

The users can look for other techniques to reduce the error if the uncertainty does not meet their requirements.

- **Efficiency**

In order to reduce the cost that may result from using incorrect information, people need to understand the nature and magnitude of uncertainty information in GIS.

- **Indicator**

From the point view of information theory, uncertainty is a measurement of information. In this study, confidence levels are used as indicators of the uncertainty. GIS users will have the choice of selecting the data which best fits their needs.

- **Prediction**

The uncertainty of any arbitrary point on the line segment can be derived from the two end points by using the "Developed Error Band" model. Other geometric type objects can also be derived from line segment. We can use this known information to predict unknown information.

6.2 Recommendations

The "Developed Error Band" model is based on certain assumptions and simplifications which only handle the positional uncertainty due to random errors. To understand the uncertainty in a more general case, further development can be considered from the following aspects.

6.2.1 Further Study of "Developed Error Band" Model

In GIS, the positional uncertainty is caused by random and systematic errors. How the final error is propagated from these errors and was not considered in this model. In a temporal GIS application, the fourth dimension of time is added, and the users need to know how much the uncertainty information has changed when the GIS is updated. When more than two spatial databases are combined, merging the uncertainty information and keeping them logically consistent are obviously problems to be solved.

6.2.2 Improvement of Visualization

The purpose of visualization and the query functions in this study was to demonstrate an approach to uncertainty management. The visualization and query functions are limited to the elementary geometric types and simple features. Improvement could be made to show even more complex features, and to provide more functions such as editing, reporting and searching, etc..

6.2.3 Integration of Positional and Attribute Uncertainty

This study did not address attribute uncertainty. In fact, attribute information is often related to the positional information. Fuzzy sets are suggested as a good technique to handle the attribute uncertainty. If the attribute uncertainty is assessed, the propagation of positional uncertainty and attribute uncertainty to the final stage is still not clear. A combined model to handle the integration of positional and attribute uncertainty needs to be developed and implemented.

6.2.4 Possibility of Implementing "Developed Error Band" Model in Commercial GIS Software

The feasibility of this modeling and visualization in a commercial GIS package needs to be studied. Most GIS packages on the market are based on two dimensional information or only have the capability of 2-D visualization. With three dimensional coordinates and their meta-data of uncertainty, the computational load is huge. These problems need to be addressed in the future.

REFERENCES

Altman, D. (1994). Fuzzy Set Theoretic Approaches for Handling Imprecision in Spatial Analysis, International Journal of Geographical Information Systems, Vol. 8, pp. 271-289.

Beard, K. (1991). Position Statement on Visualization of Data Quality, NCGIA Research Initiative 7: Visualization of Spatial Data Quality, Technical Paper 91-26, pp. c-11 - c-16.

Bedard, Y. (1987). Uncertainty in Land Information Systems Databases, Proceedings of Auto-Carto 8, pp. 175-184.

Bolstad, P. V., Gessler, P. and Lillesand, T. M. (1990). Positional Uncertainty in Manually Digitized Map Data, International Journal of Geographical Information Systems, Vol. 4, pp. 399-412.

Caspary, W. and Scheuring R. (1992). Error-bands as Measures of Geometrical Accuracy, Conference Proceedings of EGIS '92, Vol. 1, pp. 226-233.

Chapman, M. A. (1989). Ph.D. Thesis, The Department of Geodetic Science, Laval University, Quebec, Canada.

Chrisman, N. R., (1982). A Theory of Cartographic Error and Its Measurement in Digital Databases, Proceedings of Auto-Carto 5, pp. 159-168.

Chrisman, N. R., (1983). The Role of Quality Information in the Long-Term Functioning of a Geographic Information System, Proceedings of Auto-Carto 6, pp. 303-312.

Chrisman, N. R. (1991a). The Error Component in Spatial Data, Geographical Information Systems, edited by D. J. Maguire, M. F. Goodchild and D. W. Rhind, pp. 164-174

Crain, I. K., Gong, P. and Chapman, M. A. (1993). Implementation Consideration for Uncertainty Management in an Ecologically Oriented GIS, Symposium of GIS '93, Vancouver, B. C., Canada, pp. 167-172.

DCDSTF (1988). The Proposed Standard for Digital Cartographic Data, The American Cartographer 15, pp. 9-140.

DiBiase, D., MacEachren, A. M., Krygier, J. B. and Reeves, C. (1992). Animation and the Role of Map Design in Scientific Visualization, Cartography & geographic Information Systems, Vol. 19, no. 4, pp. 201-214.

Dunn, R., Harrison, A. R. and White, J. C. (1990). Positional Accuracy and Measurement Error in Digital Databases of Land Use: an Empirical Study, International Journal of Geographical Information Systems, Vol. 4, pp. 385-398.

Dutton, G. (1991). Handling Positional Uncertainty in Spatial Databases, Proceedings of 5th International Symposium on Spatial Data Handling, pp. 460-469.

Fisher, P. (1991). Spatial Data Sources and Data Problem, Geographical Information Systems, edited D. J. Maguire, M. F. Goodchild and D. W. Rhind, pp. 175-189.

Gong, P. (1993). Boundary Uncertainty in Digitized Maps I: Some Possible Determination Methods, The Department of Geomatics Engineering, The University of Calgary, Calgary, Canada, unpublished.

Goodchild, M. F. and Dubuc O. (1987). A Model of Error for Choropleth Maps with Applications to Geographic Information Systems, Proceedings of Auto-Carto 8, pp.165-174.

Heuvelink, G. B. M. and Burrough, P. A. (1993). Error Propagation in Cartographic Modeling Using Boolean Logic and Continuous Classification, International Journal of Geographical Information Systems, Vol. 7, pp. 231-246.

Hunter, G. J. (1991). Processing Error in Spatial Database: the Unknown Quantity, Proceeding of the Symposium on Spatial Database Accuracy, Melbourne, Australia, pp. 203-214.

Hunter, G. J. and Beard, K. (1992). Understanding Error in Spatial Databases, The Australian Surveyor, Vol. 37, pp. 108-119.

Hunter, G. J. and Goodchild, M. F. (1993). Managing Uncertainty in Spatial Databases: Putting Theory into Practice, The URISA Journal, Vol. 5, No. 2, pp. 55-63.

Hunter, G. J. and Goodchild, M. F. (1994). Design and Application of a Methodology for Reporting Uncertainty in Spatial Databases, Proceedings of the URISA '94, Vol. 1, pp. 771-785.

Joy, M. W. and Klinkenberg, B. (1994). Handling Uncertainty in Spatial Forest Model Integrated with GIS, Symposium of GIS '94, Vancouver, B. C., Canada, pp. 359-365.

Klinkenberg, B. and Xiao, Y. (1990). Some Conceptual Definitions in Error Analysis in GIS, Proceedings of GIS '90, pp. 1124-1130.

Laurini, R. and Thompson, D (1992). Fundamentals of Spatial Information Systems, Academic Press.

Mark, D. M. and Csillag, F. (1989). The Nature of Boundary on 'Area-Class' Maps, Cartographica, Vol. 26, pp. 65-77.

Malhotra, R. C. (1969). Matrix Treatment of the Error Ellipsoid, Civil Engineering Studies of Photogrammetry Series, No. 24, pp. 1-37.

Matthews, J. A. (1981). Quantitative and Statistical Approaches to Geography, Pergamon Press, pp. 1-11.

Mikhail, E. M. and Gracie, G. (1981). Analysis and Adjustment of Survey Measurements, Van Nostrand Reinhold Company, pp. 106-147.

Perkal, J. (1956). On Epsilon Length, Bulletin de Polonaise des Science, Vol. 4, pp. 399-403.

Richardus, P. (1974). Project Surveying, North-Holland Publishing / American Elsevier Publishing Company Inc., pp. 46-63.

Rikkers, R., Molenaar, M. and Stuiver, J. (1994). A Query Oriented Implementation of a Topological Data Structure for 3-Dimensional Vector Maps, International Journal of Geographical Information Systems, Vol. 8, No. 3, pp. 243-260.

Schwarz, K. P., El-Shiemy, N., Li, R. and Chapman, M. A. (1993). VISAT - A Mobile Highway Survey System of High Accuracy. Proceedings of VNIS'93, Ottawa, Canada, pp. 476-481.

Shafer, G. A. (1976). Mathematical Theory of Evidence. Princeton University Press.

Sinton, D. (1978). The Inherent Structure of Information as Constraint to Analysis, Harvard Papers on GIS.

Shi, W. (1993). Modeling Positional and Thematic Uncertainties in Integration of Remote Sensing and Geographic Information Systems, ITC Publication, No. 22.

Szarmes, M. (1994). Error and Accuracy Analysis of VISAT: A Video, Inertial and Satellite System for Cadastral-Related Data Collection, The Department of Geomatics Engineering, The University of Calgary, Calgary, Canada, unpublished.

Walsh, S. J., Lightfoot, D. R. and Butler, D. (1987). Recognition and Assessment of Error in Geographic Information Systems, Photogrammetric Engineering and Remote Sensing, Vol. 53, No. 10, pp. 1423-1430.

Wang, F. (1990). Fuzzy Information Representation and Processing in Conventional GIS Software: Database Design and Application, Taylor & Francis Ltd.

Zadeh, L A. (1965). Fuzzy Sets, Information and Control 8.