RELIABILILTY OF SPATIAL DATA AND ITS ANALYSIS IN GIS

CHEUNG CHUI-KWAN

M. PHIL.

THE HONG KONG

POLYTECHNIC UNIVERSITY

2000

Abstract of thesis entitled 'Reliability of Spatial Data in GIS' submitted by Cheung Chui-kwan for the degree of Master of Philosophy at The Hong Kong Polytechnic University in April 2000

Geographical data in geographical information system (GIS) are not error-free. Accuracy of each object in the GIS should be attached with their data description. This is particularly important when the data is used for decision-making.

In this study, we focus on modeling positional error of spatial features in GIS. A reliability model of a spatial feature is proposed in this study. It is measured by discrepancies of the spatial feature. Since the measured spatial feature may contain positional errors, a simulation technique is adopted to simulate the positional errors of the spatial features. Most possible measured spatial features are generated based on the assumption of the nodal errors' distribution. For each measured spatial feature generated in the simulation, its discrepant area can be computed. An average of discrepant areas is an indicator of the reliability of the spatial features.

In this study, we describe three further developments on the reliability of line segment, which is a basic unit of the linear feature of GIS, to the previous studies. First, two possible statistical distributions, both uniform and bivariate normal distributions, of the errors of line segment's nodes are discussed. While in the previous studies, the uniform distribution was the only distribution case discussed. Second, an error ellipse model, instead of the error circle model, is used for describing the errors of the nodes. Third, an effect of error dependent relationship of two nodes on the reliability of line segment is further discussed. From our results, it is noticed that different combinations of correlated nodal errors yield different reliability of a line segment.

Apart from the simulation approaches, another reliability model is derived from a newly developed approach - the numerical integration technique - in order to validate the simulated results, mainly due to the fact that accuracy of the simulation

approaches has caused worry in some circles. After comparing these two methods, we notice that the simulated and the numerical results are approximately the same, but they have different computational time. In the reliability model of a line segment, we can achieve the numerical result in a shorter time. On the other hand, the simulated result can be obtained in a shorter time in other reliability models. This is due to the complexity of the numerical model depending on the amount of nodes of the spatial feature itself. The reliability model based on both methods is extended to the reliability of both 2D and 3D linear features, both 2D and 3D areal features, and 3D volumetric features in GIS. It also concludes that error ellipse parameters affect the reliability of a spatial feature.

Furthermore, an error propagation model in buffer spatial analysis is derived based on both the simulation and the numerical analysis approaches. It is observed that the size of a buffer affects the reliability of the buffer. The reliability model proposed in this project is thus applicable to all features of GIS and buffer GIS operations for error description.

ACKNOWLEDGEMENTS

I would like to thank Dr. Wenzhong Shi, my supervisor of the degree of Master of Philosophy for his valuable suggestions and comments on my research. I also wish to thank the supervisor of my first degree - Bachelor of Science in applied mathematics, Wan-kai Pang for providing some basic concepts of simulation and introducing Dr. Shi as the supervisor of my M. Phil. program.

I gratefully acknowledge the financial support and administration from the Departmental Research Committee and the Research and Postgraduate Studies Office of the Hong Kong Polytechnic University Grants.

I would also like to thank Angela Kwong, Dr. Ping-kei Leung, Dr. Chiu-lai Lin and Dr. Matthew Pang for their research assistance in an early phase of this study. I wish to thank Stanley Leung and Kenneth Yau for their technical support. Moreover, I would like express my thanks to my colleagues in the department of Land Surveying and Geo-Informatics for their support. Finally, I thank my family for supporting my M. Phil. study.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

During the last two decades, the handling of spatial data has undergone a great change. Maps are now not only drawn on a paper sheet, but are also stored in computers. Users can view maps via the worldwide web (WWW) and even download them through the file transfer protocol (FTP).

A geographical information system (GIS) can be defined as a software package, which provides users with a tool to input, store, analyze, retrieve and transform geographical data (Cassettari, 1993). GIS is now widely applied in many different areas including military applications, environmental studies and geological exploration. However, geographical data in GIS is not error-free (Heuvelink, 1998). Due to the complexity of the geographical world, it is virtually impossible to represent the world completely. Some of the man-made utilities such as water pipes and road networks can be represented by point, line and polygons but most natural phenomena cannot (Burrough, 1986). The geographical representation limits its scope. Differences between the database contents and the phenomena they represent depend on the characteristic of the phenomena. These differences may be referred to as the 'quality' of the representation.

Although ignoring these differences is the principal method of dealing with them, to a certain extent the data quality affects decisions made with the geographical data. Goodchild (1991) stated that an accuracy of each object in the database would be attached depending on the type of data and source of errors. However, there are still no standard models to calculate the errors.

Accuracy is defined by the relationship between the measurement and the reality. This relationship can be described by errors. There exist different standards

of data quality in different organizations. A National Committee on Digital Cartographic Data Standards (NCDCDS) was established under the auspices of the American Congress of Surveying and Mapping (ASCM) in 1982 (Aalders, 1999). The reports of this committee mentioned that data quality was composed of five components (Moellering, 1985, 1987). Moreover, Guptill and Morrison (1995) stated that the ICA Commission on Spatial Data Quality added two additional components. Overall, errors are most likely to be grouped into five categories: lineage errors, positional errors, attribute errors, logical inconsistency, and completeness. The first errors refer to an existence of the history of the geographical data (including source material, dates and processing applied). They are difficult to examine because little geographical data carries its history. The positional errors are due to a missing entity, an incorrectly place entity or a disordered entity. Existing research in the positional errors is going to be elaborated later on. Attribute errors occur because of placing the wrong code for an attribute. Fisher (1992) computed the error standard deviation of an existing digital elevation model from the root mean square error (RMSE). Goodchild et al (1992) pointed out that a stochastic error model that could estimate the uncertainty associated with outputs in GIS. Veregin (1995) summarized classification error matrix, which stores error level for each class of attribute value, by using proportion correctly classified (PCC) index while Næsset (1996) applied weighted Kappa coefficient for indices of error. Maybe the most common method used for an attribute classification is the classification error matrix. Placing a code in a wrong location causes the logical inconsistency. Users should check logical consistency after inputting the geographical data to make sure no blunders (or careless mistakes) in the geographical data. Completeness is related to whether a given data set contains all information it claims to. Among these errors, the positional errors are the focus of this study.

1.1 Positional Errors in GIS

There are two types of data structures used for representing geographical data in GIS: raster and vector. In this study, the positional errors in a vector-based GIS are studied. They are considered to be of three types: blunders or mistakes, systematic errors and random errors (Wolf and Brinker, 1994). Mistakes are due to carelessness on the part of the observer mainly. For example, the observer may misinterpret the target. Blunders or mistakes may occur through failure in technique or failure of the equipment but can be detected and eliminated. Systematic errors occur according to a system (such as digitization), which are always expressed by mathematical formulation. If an "error" is removed from a measurement, the value of that measurement should be improved. This error, known as a systematic error can be identified and corrected. After eliminating blunders or mistakes and correcting systematic errors, users may notice the existence of the so-called random errors. Apart from blunders or mistakes, systematic errors and random errors, modeling errors are introduced during the transfer of both the reality to the digital database and the source map to the digital database (Bolstad et al.; 1990; Dunn et al., 1990; Keefer et al., 1991; Maffini et al., 1989). Their existence is due to map generalization. Dutton (1999) stated that some map generalization studies (Cromley and Campbell, 1990, 1992; Zhan and Buttenfield, 1996) concerned scale-changing operations but failed to relate either techniques or results to specific map scales while some (Muller, 1987; Topfer and Pillewizer, 1966) restricted themselves to relative scale change but their graphics do not resize appropriately. Hence, further studies on modeling errors are necessary. In this study positional random errors are modeled.

In addition to the positional errors inherent in the input data, other errors are introduced during GIS operations, perhaps due to scales, dates and map projections. As a result, the derived spatial data probably accumulates more errors and has different error characteristics from the input spatial data. In other words, errors of the source spatial data will be transferred to the derived data via GIS operations. A process of error transference from source to derived spatial data is called error

propagation. Error propagation is modeled mathematically in order to describe the error mechanism of a particular GIS operation. Its model should be derived for each GIS operation based on an empirical relation among different source data. Most existing error propagation models concerned attribute errors in raster-based GIS (Arbia et al, 1998; Haining and Arbia, 1993; Heuvelink and Burrough, 1993; Heuvelink et al, 1989; Newcomer and Szajgin, 1984; Shi and Ehlers, 1996; Veregin, 1995) and little attention has been paid on positional error propagation in vector-based GIS (Stanislawski et al, 1996; Zhang et al, 1998). This is due to the complexity of vector-based data. Buffer spatial analysis is a basic GIS data transformation function in which a zone of some specified width is delineated around the spatial feature, and its error propagation model will be derived.

1.2 A Review of Previous Studies

Within a vector-based GIS, elementary spatial features are point, linear and areal features. An error model for a point has been studied for a long time in the fields of geodesy, surveying and mapping. The positional errors of the point are usually distributed in the proximity of an error ellipse centered at the true location of the point while existing studies of the positional errors model of a line segment can be mainly classified into three approaches: (a) error-band models derived from simulation techniques and error propagation law; (b) confidence region models based on rigorous statistical approaches; and (c) reliability models of a line segment based on simulation and integration techniques.

An error-band model is a band around the expected line segment. The epsilon band model (Perkal, 1966) was created by rolling a circle along the line segment and this band is similar to a buffer around the line segment. Chapman et al (1997) stated that the width of the band determined by a function of different uncertainties accumulated these uncertainties into the final stage. This model can be applied

during the execution of many spatial operations easily but it seems to be odd that the true line segment is definitely located within the band.

Some researches have derived the error-band model using the error distribution of the points on a line segment. Dutton (1992) simulated the error distribution of the line segment using Monte Carlo simulation technique and the error distribution of the line segment was derived. Moreover, Caspary and Scheuring (1993) and also Shi (1994) derived error band models using errors of an arbitrary point on the line segment. Shi (1994) further developed a number of error indicators for line segments, as an extension to the point error indicators, based on the assumption of the line segments. These studies, however, are based on the assumption that the errors of the two nodes were independent. The shape of the error bands is, therefore, the minimum in the middle of the line segments and maximum at the two nodes. This result is due to the assumption of the independence (of the errors between two nodes) and the nonexistence of model errors. A more generic description on the positional errors of the line segments considers the case of interrelation between the two nodal errors and was discussed in Shi and Liu (2000).

Alesheikh (1998) proposed the rigorous uncertainty model of a line segment and stated that the existing model was a subset of the proposed model. However, Alesheikh's model may not be the final stage in the development of the error-band model. It will be difficult to determine the confidence coefficient of the confidence region for the line segment if the confidence intervals for the points on the line segment are used. Shi (1994) created the confidence region for a line segment based on integrating the simultaneous confidence intervals for the points on the line segment instead of the confidence intervals for the points.

A confidence region error model is a band surrounding the measured line segment, containing the true line segment with the probability larger than a predefined confidence coefficient. The confidence region error model (Shi, 1994) for a line segment in a two-dimensional (2D) GIS was developed by using rigorous statistical derivations. This model was extended to the confidence volume of a three-

dimensional (3D) GIS features. A generic model (Shi, 1998) was further developed for the confidence space for a N-dimensional features.

The reliability of a line segment was discussed by Stanfel and Stanfel (1994). The extent of a discrepancy, in which the boundary was the true line segment and the observed line segment, was defined by its area; and this discrepancy was a measure of the reliability of the line segment. Easa (1994) considered that the model of Stanfel and Stanfel oversimplified the discrepant area, and then Easa (1995) estimated the reliability of the line segment using the Monte Carlo simulation technique. It was concluded that the analytical solutions of the reliability of the line segment, proposed by Stanfel and Stanfel, might have some adjustments.

In modeling the positional errors, the previous reliability models assumed that the nodal errors followed a uniform distribution for the simplification and a normal distribution should be more appropriate according to the measurement technologies (Stanfel and Stanfel, 1993). Generally, they considered an error circle to be a feasible region that the measured nodes of the line segment lie inside. A further modification to the model would be an error ellipse model (Easa, 1995). What is more, the previous reliability models of a line segment are concerned with the independent nodal errors. This is, in fact, a simplification to the real world cases in which the nodal errors were dependent (Keefer, Smith and Gregoire, 1988).

The above error models of a line segment can be extended to error models of a linear feature while error models of an areal feature's boundary are similar to those of the line segment. Dutton (1992) generalized his error band to an error band for the boundary of an areal feature while Stanfel et al (1995) and Chapman et al (1997) applied their reliability models of a line segment into reliability models of an areal feature's boundary too. Although they further developed errors models of an areal feature based on integrating the positional errors of the areal feature's boundary straightforwardly, the errors models of the line segment may not be extended to those of an areal feature. The positional errors of an areal feature are different from those of the boundary of the areal feature. The interior part of the areal feature may not be

error-free and hence a positional error model of an areal feature should consider the interior part of the areal feature apart from its boundary. It is a fact that there is little research modeled the positional errors of an areal feature.

The research stated above is mainly concerned with the error models for studying the positional errors of spatial features in a 2D GIS. Very little research exists in the modeling of the positional errors in a higher-dimensional GIS. Shi (1998) derived the confidence region model of linear features from a strictly statistical approach. In this project, the reliability model of spatial features either in the 2D or 3D GIS will be taken into account.

In buffer spatial analysis for a vector-based GIS, an absolute accuracy has been researched (Zhang et al 1998). It was concluded that the absolute accuracy in buffer spatial analysis was inversely proportional to the width of the buffer. A weakness with this method is the need to choose an error-band model. Users have to determine which error-band model should be used to describe the error of the source linear feature. Error-band models include epsilon band (Blakemore, 1984; Chrisman, 1982; Perkal, 1966), E-band (Caspary and Scheuring, 1993), g-band (Shi and Liu, 2000) and so forth. Based on the error-band model, the error propagation model for the linear feature can be obtained. Until now, no standard methods can be implemented to determine the most feasible error-band model for the specified spatial feature.

1.3 Scope and Objectives

In this study, the reliability of a 2D line segment will be studied. First, two error ellipse models will be introduced to compute the reliability of the line segment, the uniform and normal. Second, an error ellipse model, instead of the error circle model, will be used for describing the nodal errors of the line segment. Third, the correlated nodal errors of the line segment will be discussed. The reliability model

will be further extended to the reliability model for spatial features in either 2D or 3D GIS. Furthermore, an error propagation model in buffer spatial analysis will be proposed.

In this project, positional error model and error propagation model in vector-based GIS will be proposed based on a simulation technique and a numerical integration approach. Three areas include the reliability model of spatial features in 2D GIS; the reliability model of spatial features in 3D GIS; and the error propagation model in buffer spatial analysis. Objectives of this study are

a) to model the reliability of both 2D and 3D spatial features in GIS;
b) to investigate effects of error models, nodal error distribution and correlation problem of nodal errors on the reliability of spatial features;
c) to model the reliability of a buffer around spatial features; and
d) to investigate an effect of buffer size on the reliability in buffer spatial analysis.

This dissertation is divided into seven chapters. Some statistical theories will be introduced briefly in Chapter 2. Chapter 3 will define how the reliability is measured and what assumptions are being made. The reliability model of spatial features will be elaborated in Chapter 4 based on simulation techniques, while in the next chapter, another reliability model will be derived from a newly developed approach - the numerical integration technique - and this model is used to validate the simulated results. The error propagation model in the buffer spatial analysis will be proposed in Chapter 6. Conclusions and recommendations will be given in the last chapter.

CHAPTER 2
A BRIEF REVIEW OF RELAVENT STATISTICAL THEORIES

Mathematical and statistical theories form the basis for analysis error in spatial data. In this chapter, some basic statistical theories used in this study will be stated. A study of positional errors is derived from an error distribution of a point and two possible error distributions will be shown in Session 2.1. The next session will introduce two methods to model the positional errors.

2.1 Error Distributions of a Point

A random variable associates a numerical value with each outcome of an experiment. The random variable can be classified into two types: a discrete random variable and a continuous random variable. The random variable is the discrete random variable if it has a finite number of values. It is said that a discrete random variable may assume a countable number of values. The random variable is the continuous one if it is in an interval or several intervals of real numbers. It has an infinite number of values. The discrete random variable has a discrete distribution while the continuous random variable has a continuous distribution. The following examples demonstrate the difference between the discrete distribution and the continuous distribution.

Suppose two coins are tossed simultaneous. Four possible outcomes are 'head-head', 'head-tail', 'tail-head' and 'tail-tail' where 'head-tail' means that the first coin is head and the second one is tail, and so on. Because the outcome for the first coin does not affect that for the second one, each outcome has the same probability ¼. If X is the number of heads obtained, each outcome of the experiment corresponds to a particular value of X. I call X a random variable. In this example,

the value of X belongs to a set {0, 1, 2}. Then, X has a discrete distribution. Another example is that a number X is chosen at random between 0 and 1. X is continuous random variable because the value of X is in the range of 0 and 1. I call the distribution of X is continuous.

Two continuous distributions, uniform distribution and normal distribution are considered in the following sessions.

2.1.1 Uniform Distribution

A continuous random variable X is said to have a uniform distribution on [a,b] if its probability density function (p.d.f.) is

$$f(x)=\begin{cases}\dfrac{1}{b-a} & \text{if } a \le x \le b,\\ 0 & \text{otherwise,}\end{cases} \tag{2.1}$$

where a and b are real numbers (see Figure 2.1).

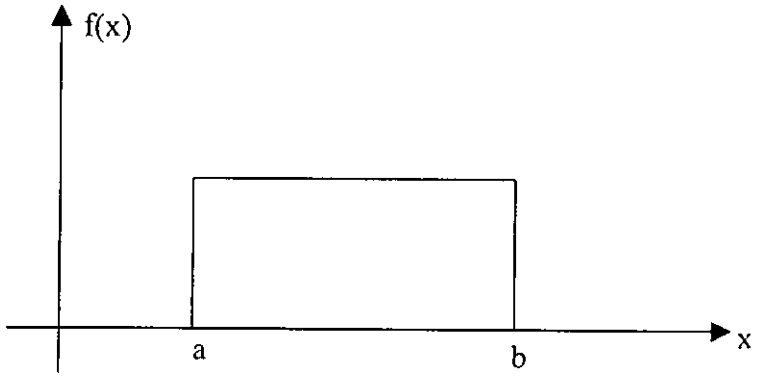


Figure 2.1. Probability density function of a uniform distribution.

A mean of X, E(X) and its variance, Var(X) are

$$E(X) = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{a+b}{2}, \text{ and}$$
$$Var(X) = \int_a^b x^2 \cdot \frac{1}{b-a} dx - \mu^2 = \frac{(b-a)^2}{12}. \quad (2.2)$$

2.1.2 Normal Distribution

A continuous random variable X is said to have a normal distribution if its p.d.f. is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \quad -\infty < x < +\infty, \quad (2.3)$$

where μ and σ are real numbers and $\sigma > 0$.

Figure 2.2 shows the special case of f(x) where μ is equal to 0 and σ is equal to 1. The standard normal distribution is the normal distribution having a mean equal to 0 and a standard deviation equal to 1. The letter Z is used to represent the standard normal random variable. Generally, f(x) is symmetrical about μ. A mean of X, E(X) and its variance Var(X) are

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx = \mu, \text{ and}$$
$$Var(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx - E(X)^2 = \sigma \quad (2.4)$$

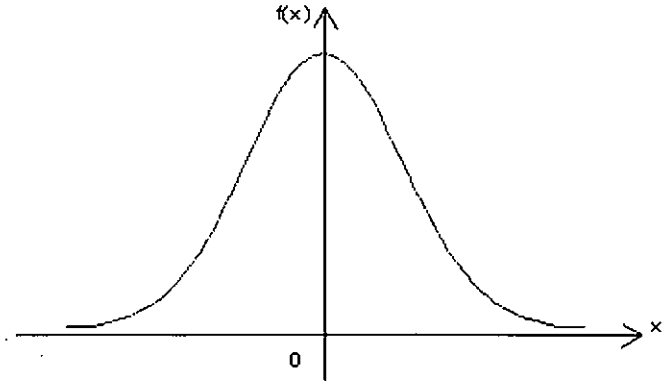


Figure 2.2. Probability density function of a standard normal distribution.

Equation (2.3) involves only one random variable X and so X has a univariate distribution. If two random variables are involved, a joint distribution of these two random variables is a bivariate distribution. Typically, a set of random variables has a multivariate distribution.

Let Σ denote an $n \times n$ real symmetric matrix which is positive definite where n is a positive integer. Let μ denote the $n \times 1$ matrix such that μ^T (the transpose of μ) is equal to $[\mu_1, \mu_2, \ldots, \mu_n]$, where each μ_i is a real constant for $i = 1, \ldots, n$. Finally, let x denote the $n \times 1$ matrix such that $x^T = [x_1, x_2, \ldots, x_n]$. A joint p.d.f. of n random variables $X_1, X_2, \ldots, X_n$ is

$$\frac{1}{(2\pi)^{n/2}\sqrt{|\Sigma|}} \exp\left(-\frac{(x-\mu)^T \Sigma^{-1}(x-\mu)}{2}\right), \quad -\infty < x_i < \infty \tag{2.5}$$

The matrix μ is the matrix of means of the random variables $X_1, X_2, \ldots, X_n$. The matrix Σ, which is given by

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_{nn} \end{pmatrix}, \tag{2.6}$$

where σ_{ij} is the covariance of X_i and X_j.

This matrix is the covariance matrix of the multivariate normal distribution.

2.2 Simulation

Simulation is a method used to study a system that may be a group of units working in an interrelated manner. The purpose of system studies is to gain an understanding of the overall operation. Simulation provides a description of system behavior. The following definition is given by Shannon (1975, 2pp.):

"Simulation is the process of designing a model of a real system and conducting experiments with this model for the purpose either of understanding the behavior of the system or of evaluating various strategies (within the limits imposed by a criterion or set of criteria) for the operation of the system."

Simulation includes different processes (techniques): variance reduction, simulation validation and so forth. Each simulation process is specialized to special types of system but some general characteristics are common to and useful for a wide variety of practical problems. Simulation is good at answering specific "what if" questions. It requires a model be constructed that represents system behavior in terms of mathematical and logical relationships between variables. In the Monte Carlo simulation, which is a subset of variance reduction techniques, a population of interest is simulated. From the pseudo-population, repeated random samples are drawn. The statistic under study is computed in each pseudo-sample and its sample distribution is examined for insights into its behavior.

It is not essential to involve computers in simulation. However, applying the computers can enhance the efficiency of the simulation. Nowadays, simulation is highly computer intensive. And the sample data is not physical observed but represented by the set of computer commands used to generate the data.

Simulation has some advantages and disadvantages. Here are some of the advantages of simulation. A simulation model can be realistic because it captures the actual characteristics of the system being modeled. It can be completely controlled and completely observed. Moreover, it is possible to reproduce random events identically via sequences of pseudo-random numbers that can exhibit the characteristics of truly random numbers. It does not require a great level of mathematical sophistication. On the other hand, it is subject to important disadvantages. It cannot produce exact results because a system is composed of one or more elements that are subject to random behavior. It is time-consuming when solving some mathematical problems: solvable integration problems, parameter estimations for a population distribution, and so forth. In conclusion, simulation is not a panacea. It offers powerful advantages but suffers from significant disadvantages. It is a fact that analytical solution, where the system is expressed as a mathematical model, is more accurate and is usually more easily obtained than simulation result. Therefore, simulation cannot replace mathematical analysis. Nevertheless, it is a practical method for gaining an understanding of unsolvable mathematical models.

2.3 Numerical Integration: Gaussian Quadrature

A mathematical model of a complex situation for interpreting experimental results and predicting results is always constructed in some fields including surveying by describing the important features in mathematical terms. Occasionally, there may be a formal analytical solution procedure available but a great set of expressions may be involved. A numerical procedure leading to meaningful numerical results is available and preferable. Numerical analysis is a branch of mathematics in which such numerical procedures are studied. There are various main problems in numerical analysis: solving systems of linear equations, eigenvalue problems, interpolation, evaluating integrals, solving differential equations, and optimization.

Here numerical integration is implemented to study the reliability problem. The integrals are mainly of the form

$$I(f) = \int_a^b f(x)dx \tag{2.7}$$

where a and b are finite.

Very few integrals can be evaluated exactly by analytical methods. There is a desire for developing numerical methods to approximate the integral to be calculated I(f) that cannot be calculated analytically. Besides, it is often faster to integrate the integrable functions numerically rather than evaluating them exactly using a complicated antiderivative of f(x). The approximation of I(f) is usually referred to as numerical integration or quadrature.

Four well-known numerical methods for evaluating the integral in Equation (2.7) are the trapezoidal rule, Simpson's rule, the Newton-Cotes integration formula, and Gaussian quadrature. The first three methods are based on a lower-order polynomial approximation of the integrand f(x) on subintervals. And the length of each subinterval is unchanged over the whole interval. Gaussian quadrature uses polynomial approximations of f(x) of increasing degree. The length of each subinterval is varied by f(x) in order to minimize an approximation error. The resulting integration formula is extremely accurate in most cases. Therefore, Gaussian quadrature is implemented in this study to model the reliability problems of GIS spatial features and will be explored in Chapter 5.

2.4 Summary

In this chapter, some statistical theories relating to the assumption stated in Chapter 3 are briefly introduced and two methods used to evaluate positional error of spatial features. Under assumptions, two distributions were considered: uniform and

normal. In Session 2.1, their properties were demonstrated. You may notice that in the normal distribution (see Session 2.1.2), matrix Σ in Equation (2.5) must be positive definite. The definition of the term 'positive definite' is seen in Appendix. Moreover, this study will sample the positional error of spatial features and estimate the confidence interval for the mean of the positional error (see Appendix). From the definition of the confidence interval for the population mean, the random variable must be normally distributed. However, the distribution of the positional error of a spatial feature may not be normal. Therefore, the central limit theorem stated in Appendix must be applied. How these apply in this study will be elaborated in Chapter 4. Finally, simulation and numerical methods, which are used to model the positional error of GIS spatial features and the error propagation in buffer spatial analysis, are introduced in the last two sessions. The details of the implementation are shown in Chapter 4, Chapter 5 and Chapter 6.

CHAPTER 3
PROBLEM DEFINITION

Spatial features and buffers studied here will be introduced in Sessions 3.1 and 3.2. Then, the general definition of reliability of the spatial features and the buffers will be given. Assumptions will be stated in the last session.

3.1 Spatial Features

In the real world, geographical variation is infinitely complex. The world has to be represented in a discrete manner in a computer environment, GIS. A map is one of the methods used to describe geographic reality. Three fundamental units of the map are point, linear and areal features. They characterize features of various shapes and types.

A map can be treated as a graph in mathematics. The three basic units stated above can be defined using graph theory. Nodes are used to represent point features. They also represent the beginning or ending nodes of every linear feature and occur at the intersection of linear features. Arcs (order sets of points) are the elementary unit of any line and any curve to describe the location and the shape of a spatial feature. The linear and the areal features composed of line segments are considered. Two nodes form a line segment. Therefore, line segment is studied instead of arc, since an arc can band smoothly.

3.2 Buffer

Buffer operation is a basic GIS data transformation function in which a buffer of some specified width is delineated around a spatial feature. Figure 3.1 illustrates buffers around spatial features: point, linear and areal features. When a desired

distance is given, GIS builds the buffer outward from the selected features. Buffering allows GIS users to retrieve features that lie within the desired distance of the features such as 1 mile of a school.

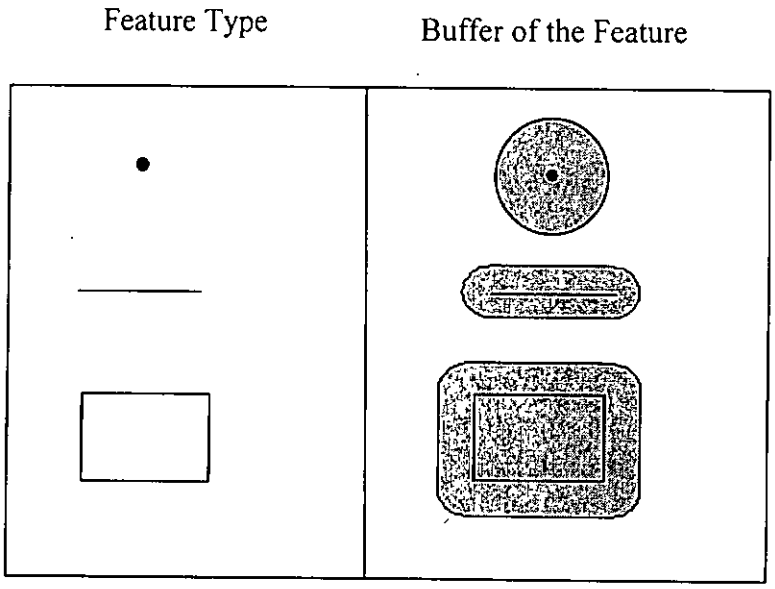


Figure 3.1. Buffer around features.

3.3 Reliability of Spatial Features

Reliability is used in surveying with different meanings. Reliability of a measurement is determined by three main factors: measurement instruments, measurement errors and the age of source data (McGrew, 1993). Uren and Price (1994) stated that the extent of detecting and eliminating gross errors was a measure of reliability, whereas Allan (1993) mentioned an indicator of reliability given by an estimate of the standard error of a residual. From the statistical point of view, the difference between a sample result and the result from a complete count taken under the same conditions is measured by what I would refer to as the reliability of the sample result (Hansen, 1993) where the result from a complete count should be referred to the actual result. From the statistical view, Stanfel and Stanfel (1993, 1994) and Easa (1995) considered an average area of discrepancy between the actual and the observed line segments as a measure of the reliability of the line segment.

This measure will be further developed for the reliability of spatial features in this study.

The difference between a sample feature and the expected feature in fact is an error in the sample feature and hence the reliability and the difference have an inverse relationship. Let R(x) denote the reliability function for a spatial feature x and Error(x) denote a function used for the error measurement in the measured feature. Their relationship can be expressed mathematically. That is,

$$R(x) = f\left(\frac{1}{\text{Error}(x)}\right) \tag{3.1}$$

In the reliability model for spatial feature, Equation (3.1) can be modified as Equation (3.2).

$$R(\text{spatial feature}) = f\left(\frac{1}{\text{Error}(\text{spatial feature})}\right) \tag{3.2}$$

Since Error(spatial feature) is a measure of the discrepancy of the spatial feature, it should be a function of the sample (measured) location and the actual location of the spatial feature. Furthermore, the actual location is unchanged and hence Error(spatial feature) is the function of the sample location of the spatial feature only. Let NP denote the number of nodes of a 2D spatial feature and (x_j, y_j) denote the coordinate of a node of the spatial feature for j =1, 2... NP. The mathematical expression of the reliability problem in 2D GIS is

$$R(x_1, y_1, x_2, y_2, \ldots, x_{NP}, y_{NP}) = f\left(\frac{1}{\text{Error}(x_1, y_1, x_2, y_2, \ldots, x_{NP}, y_{NP})}\right) \tag{3.3}$$

Since each node in 3D GIS can be represented by x, y and z where z records height of the node, the reliability problem in 3D GIS is expressed in Equation (3.4).

$$R(x_1, y_1, z_1, \ldots, x_{NP}, y_{NP}, z_{NP}) = f\left(\frac{1}{Error(x_1, y_1, z_1, \ldots, x_{NP}, y_{NP}, z_{NP})}\right) \quad (3.4)$$

In the buffer spatial analysis, the error in the source spatial feature will be propagated. The propagated error depends on the source error and so it should be in terms of the source error. Then let a function g denote the propagated error in terms of the source error. Thus the reliability problem can be expressed in Equation (3.5).

$$R(x_1, y_1, x_2, y_2, \ldots, x_{NP}, y_{NP}) = f\left(\frac{1}{g(Error(x_1, y_1, x_2, y_2, \ldots, x_{NP}, y_{NP}))}\right) \quad (3.5)$$

In this study, an area of the discrepancy of a spatial feature measures the reliability of the spatial feature. This discrepancy is a difference between the measured and the actual spatial features. The geometrical representation of this discrepancy in different spatial features will be elaborated in Chapter 4. Also, the discrepancy in buffer spatial analysis is a difference between the measured and the actual buffers (see Chapter 6). In order to study possible measured location of a spatial feature, an assumption of nodal errors of the spatial feature is made in the following session.

3.4 Assumption

In 2D problems, an error ellipse may be established around a point to indicate precision regions of different probabilities. The orientation of the ellipse (relative to the Cartesian coordinate system) of which the axes are called x- and y-axes normally, depends on the correlation between values of x and y on the ellipse (see the dash lines with arrows in Figure 3.2). If x and y are uncorrelated, the two axes of the ellipse will be parallel to axis x and axis y (see Figure 3.3). If the semi-major axis

and the semi-minor axis have the same length, the ellipse becomes a circle. Similarly, in the case of 3D problems, error ellipsoids may be established around a point to indicate precision regions of different probabilities.

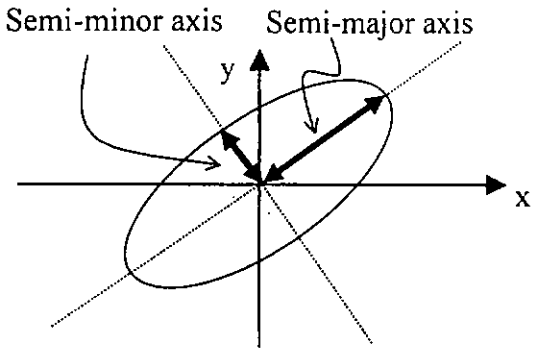


Figure 3.2. Error ellipse.

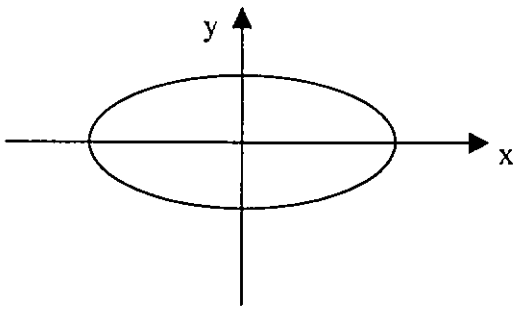


Figure 3.3. Error ellipse with orientation parallel to the axes.

The reliability of a spatial feature is determined by the discrepancy of the spatial feature in which the measured spatial feature includes neither gross errors nor systematic errors. Due to random errors, it is assumed that positional errors of the nodes of the spatial feature are mostly distributed "within" a feasible region (either the error ellipse in 2D GIS or error ellipsoid in 3D GIS) centered at the actual location of the node. And the probability of the positional error lying in the feasible region is larger than or equal to a pre-defined confidence coefficient. Though the actual location of the node of the spatial feature is not known, from the statistical point of views, a mean of any variable X is close to its actual value. As a result, the actual location of the node on the spatial feature should be referred to as its corresponding expected (mean) node.

Now that linear features, areal features and 3D volumetric features consist of nodes, the above assumption about a node is significant for the reliability of spatial features.

3.5 Summary

Positional error in 2D linear and areal features and 3D linear, areal and volumetric features will be investigated. Also, error propagation in buffer spatial analysis will be focused on, too. To explain my proposed model easily, the definition of these features was stated in Session 3.1 and Session 3.2. Second, reliability does not have a unique definition in GIS. In this study, the reliability is defined from the statistical view. The difference between a sample result and the expected result under the same conditions is used to measure the reliability of the sample result. This difference in fact is an error in the sample result and hence the reliability and the difference have an inverse relationship. The reliability model will be derived by sampling under the assumption of the nodal error. This assumption was given in Session 3.4. In the following chapters, the proposed reliability model will be elaborated.

CHAPTER 4
MODELING RELIABILITY OF SPATIAL FEATURES BY SIMULATION

The reliability of a spatial feature is estimated by the discrepancy of the spatial feature. In this chapter, geometrical definitions of the discrepancy of spatial features in both 2D and 3D GIS are given. The expected and the measured spatial features will shape the discrepancy of a spatial feature. In order to investigate random errors of the spatial feature, samples of the measured spatial feature are generated based on the spatial feature's distribution of the spatial feature. This distribution can be derived from nodal error distributions of the spatial feature because a node is a fundamental unit of any spatial feature. Here the sample measured spatial features will be generated using simulation techniques.

This chapter will propose a simulation model on the reliability of spatial features. The first section will study the reliability problem in a 2D GIS and the second will study the reliability problem in a 3D GIS. In these two sections, the discrepancy of spatial features is defined and some examples of the reliability of the spatial features will be given.

4.1 Reliability of Spatial Features in 2D GIS

Spatial features in 2D GIS considered in this study are linear features and areal features but not point features, mainly due to the fact that positional error of a point is commonly assumed to be distributed within an error ellipse. A line segment is an element of either a linear or areal feature and its reliability is discussed first. Reliability of a linear feature and an areal feature will be investigated. It is a fact that the reliability of the areal feature is distinct from that of the areal feature's boundary.

4.1.1 Line Segments

The discrepancy of a line segment is defined by the difference between the expected location and the measured location of the line segment; this difference is caused by measurement errors, which cannot be eliminated. The shaded area in Figure 4.1 shows the discrepancy of the line segment. The solid line segment represents the expected line segment composed of the two expected nodes (μ_{x_1}, μ_{y_1}) and (μ_{x_2}, μ_{y_2}); the dash line segment represents the measured line segment composed of the two measured nodes (x_1, y_1) and (x_2, y_2).

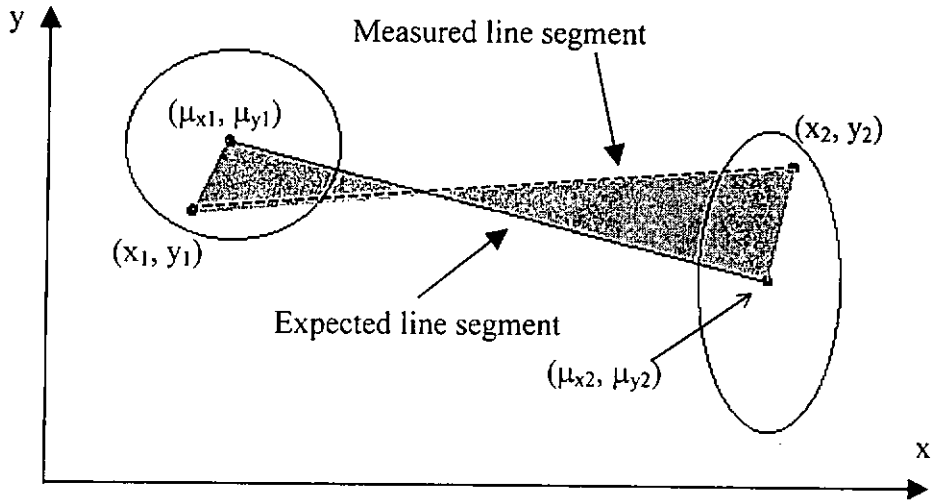


Figure 4.1. Discrepancy of a line segment in the two-triangle case.

For each node of the line segment, its error ellipse centered at its expected location can be seen in Figure 4.1 whereby its measured location is inside its error ellipse. Let (x_{12}, y_{12}) denote the intersecting point of the measured and the expected line segments. Then, I have

$$\begin{cases} x_{12}(x_1, y_1, x_2, y_2) = \dfrac{\mu_{y_1} - k_2\mu_{x_1} + k_1 x_1 - y_1}{k_1 - k_2} \\ y_{12}(x_1, y_1, x_2, y_2) = y_1 + k_1(x_{12} - x_1) \end{cases} \quad \text{if } \mu_{x_1} \neq \mu_{x_2}$$

and

$$\begin{cases} x_{12}(x_1, y_1, x_2, y_2) = \mu_{x_1} \\ y_{12}(x_1, y_1, x_2, y_2) = y_1 + k_1(x_{12} - x_1) \end{cases} \quad \text{if } \mu_{x_1} = \mu_{x_2} \tag{4.1}$$

where $k_1 = \dfrac{y_2 - y_1}{x_2 - x_1}$ and $k_2 = \dfrac{\mu_{y_2} - \mu_{y_1}}{\mu_{x_2} - \mu_{x_1}}$; and

k_1 must not equal to k_2 if the intersecting point exists.

The discrepant area of the line segment is given by

$$A_1(x_1, y_1, x_2, y_2) = \frac{1}{2}\left(\left|\mu_{x_1} y_1 + x_1 y_{12} + x_{12}\mu_{y_1} - x_1\mu_{y_1} - x_{12} y_1 - y_{12}\mu_{x_1}\right| + \left|x_2 y_{12} + x_{12}\mu_{y_2} + \mu_{x_2} y_2 - y_2 x_{12} - y_{12}\mu_{x_2} - x_2\mu_{y_2}\right|\right) \tag{4.2}$$

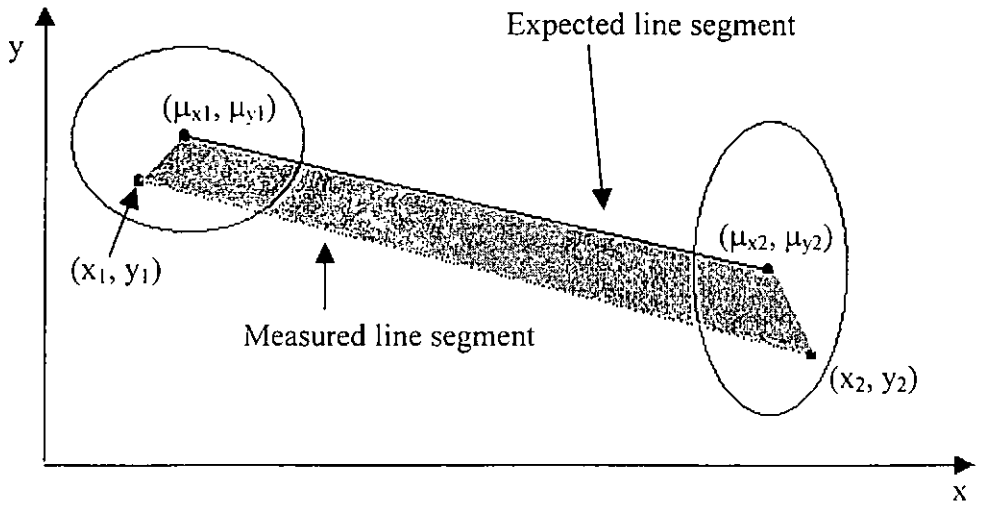


Figure 4.2. Discrepancy of a line segment in the quadrangle case.

Figure 4.1 shows the measured line segment intersecting the expected line segment. This case is called a two-triangle case. In another case, the measured and the expected line segments do not intersect. I call this a quadrangle case (see Figure 4.2). The discrepant area is given by

$$A_2(x_1, y_1, x_2, y_2) = \frac{1}{2}\left|x_1 y_2 + x_2 \mu_{y_2} + \mu_{x_2}\mu_{y_1} + \mu_{x_1} y_1 - x_2 y_1 - \mu_{x_2} y_2 - \mu_{x_1}\mu_{y_2} - x_1 \mu_{y_1}\right|. \quad (4.3)$$

Therefore, the discrepant area of the line segment will be calculated depending on its case. In order to study an effect of the random errors, the simulation is implemented under the assumption. The nodal errors of the line segment are uniformly (or normally) distributed within an error ellipse.

4.1.1.1 Algorithm for uniform distribution

Let us consider uniformly distributed nodal errors inside the error ellipse. The mathematical expression of the error ellipse is given by

$$\left(\frac{x_i - \mu_{x_i}}{a_i}\right)^2 + \left(\frac{y_i - \mu_{y_i}}{b_i}\right)^2 - c_i\left(\frac{x_i - \mu_{x_i}}{d_i}\right)\left(\frac{y_i - \mu_{y_i}}{e_i}\right) = 1 , \quad (4.4)$$

where $i = 1$ or 2;

a_i, b_i, d_i and e_i are non-zero real numbers; and

c_i is a real number.

If c_i is zero, a_i and b_i will be equal to the length of two semi-axes of the error ellipse and the two semi-axes will be parallel to the x- and the y-axes of the coordinate system. Given the error ellipse parameters a_i, b_i, c_i, d_i and e_i, the discrepant area of the line segment can be determined. According to the assumption, I generate two nodes for the two expected nodes of the line segment and then calculate the discrepant area. This is the first iteration. After the simulation is repeated N times where N is a positive integer, a mean (or average) of the discrepant

area is obtained. The algorithm used to compute the mean of the discrepant area is shown below.

Algorithm 4.1

Step 1 Input the total number of replications for the simulation N, the two expected nodes of the line segment, and parameters of the two error ellipses: $a_1, b_1, c_1, d_1, e_1, a_2, b_2, c_2, d_2$ and e_2.

Step 2 Set j = 0 where j is used to count the number of replications for the simulation.

Step 3 Generate four random numbers: u_1, u_2, u_3 and u_4 in [0,1]; and increase j by 1.

Step 4 Transform these random numbers to two generated nodes for the two nodes of the line segment from the following equations

$$\begin{aligned} x_1 &= 2(u_1 - 0.5)a_1 + \mu_{x_1} \\ y_1 &= 2(u_2 - 0.5)b_1 + \mu_{y_1} \\ x_2 &= 2(u_3 - 0.5)a_2 + \mu_{x_2} \\ y_2 &= 2(u_4 - 0.5)b_2 + \mu_{y_2} \end{aligned} \quad . \tag{4.5}$$

Step 5 Decrease j by 1 and go to step 3 if neither (x_1, y_1) or (x_2, y_2) lies in its error ellipse.

Step 6 Compute an area of the discrepancy from Equation (4.2) or (4.3).

Step 7 Go to step 3 if j is less than N.

Step 8 Compute the average of the discrepant area.

4.1.1.2 Algorithm for normal distribution

Another feasible distribution of the nodal errors is normal distribution. Under this assumption, the error ellipse refers to a (1-α)% confidence region for an expected node of the line segment. Thus, for i =1 or 2, let $\rho_{x_i y_i}$ denote a correlation coefficient of x_i's error and y_i's error; and σ_{x_i} and σ_{y_i} denote sample standard derivations of x_i's error and of y_i's error respectively. The key properties of a correlation coefficient

are well known. When x_i's and y_i's errors are linearly independent, $\rho_{x_i y_i} = 0$. When error of x_i = error of y_i, I have $\rho_{x_i y_i} = 1$; and when error of x_i = -error of y_i, I have $\rho_{x_i y_i} = -1$. However, it is only defined for distributions having marginal probability density function with finite variance. Hence, the correlation coefficient does not exist in the uniform case. However, it does not mean that x_i's and y_i's errors are independent in the uniform case. In this study, I only consider correlated errors (or linearly dependent errors). The mathematical expression of the error ellipse for the node of the line segment in Equation (4.4) is modified as Equation (4.6).

$$\left(\frac{x_i - \mu_{x_i}}{a_i}\right)^2 + \left(\frac{y_i - \mu_{y_i}}{b_i}\right)^2 - 2\rho_{x_i y_i}\left(\frac{x_i - \mu_{x_i}}{\sigma_{x_i}}\right)\left(\frac{y_i - \mu_{y_i}}{\sigma_{y_i}}\right) = 1 \text{ , } i = 1 \text{ or } 2 \qquad (4.6)$$

where $a_i = \frac{\sigma_{x_i}}{\sqrt{-2\log(\alpha)}}$ and $b_i = \frac{\sigma_{y_i}}{\sqrt{-2\log(\alpha)}}$.

The algorithm in the normal case is similar to that in the uniform case but the only exception is how to generate nodes of the line segment.

A standard normally distributed node is generated by the Box-Muller method. This method is an exact method that uses two independent random variables v_1 and v_2 in [0, 1] to sample two independent standard normal variables u_1 and u_2:

$$\begin{cases} u_1 = \cos(2\pi v_2)\{-2\ln(v_1)\}^{0.5} \\ u_2 = \sin(2\pi v_2)\{-2\ln(v_1)\}^{0.5} \end{cases} . \qquad (4.7)$$

According to Algorithm 4.2 in the following, the average of the discrepant area of the linear feature can be estimated.

Algorithm 4.2

Step 1 Input the total number of replications for the simulation N, the two expected nodes on the line segment, and parameters of the two error ellipses: $a_1, b_1, \sigma_{x_1}, \sigma_{y_1}, \rho_{x_1 y_1}, a_2, b_2, \sigma_{x_2}, \sigma_{y_2}, \rho_{x_2 y_2}$, and correlation coefficients among the nodal errors.

Step 2 Set j = 0 where j is used to count the number of replications for the simulation.

Step 3 Generate four standard normal random numbers u_1, u_2, u_3 and u_4; and increase j by 1.

Step 4 Transform these random numbers to two generated nodes for the two nodes of the line segment from the following equations

$$
\begin{aligned}
x_1 &= \sigma_{x_1} u_1 + \mu_{x_1} \\
y_1 &= \rho_{x_1 y_1} \sigma_{y_1} u_1 + \sqrt{1-\rho_{x_1 y_1}^2}\,\sigma_{y_1} u_2 + \mu_{y_1} \\
x_2 &= \rho_{x_1 x_2} \sigma_{x_2} u_1 + \rho_{x_2 y_1} \sigma_{x_2} u_2 + \sqrt{1-\rho_{x_1 x_2}^2-\rho_{x_2 y_1}^2}\,\sigma_{x2} u_3 + \mu_{x_2} . \\
y_2 &= \rho_{x_1 y_2} \sigma_{y_2} u_1 + \rho_{y_1 y_2} \sigma_{y_2} u_2 + \rho_{x_2 y_2} \sigma_{y_2} u_3 \\
&\quad + \sqrt{1-\rho_{x_1 y_2}^2-\rho_{y_1 y_2}^2-\rho_{x_2 y_2}^2}\,\sigma_{y_2} u_4 + \mu_{y_2}
\end{aligned}
\tag{4.8}
$$

Step 5 Decrease j by 1 and go to step 3 if neither (x_1, y_1) or (x_2, y_2) lies in its error ellipse.

Step 6 Compute an area of the discrepancy from Equation (4.2) or (4.3).

Step 7 Go to step 3 if j is less than N.

Step 8 Compute the average of the discrepant area.

4.1.2 Linear Features - Polylines

The discrepancy of a linear feature is defined in the same manner. An example of the discrepancy of a linear feature is illustrated in Figure 4.3 whereby the linear feature consists of three nodes and two line segments. An area of the discrepancy (the shaded area in Figure 4.3) is the measure of the reliability of the linear feature.

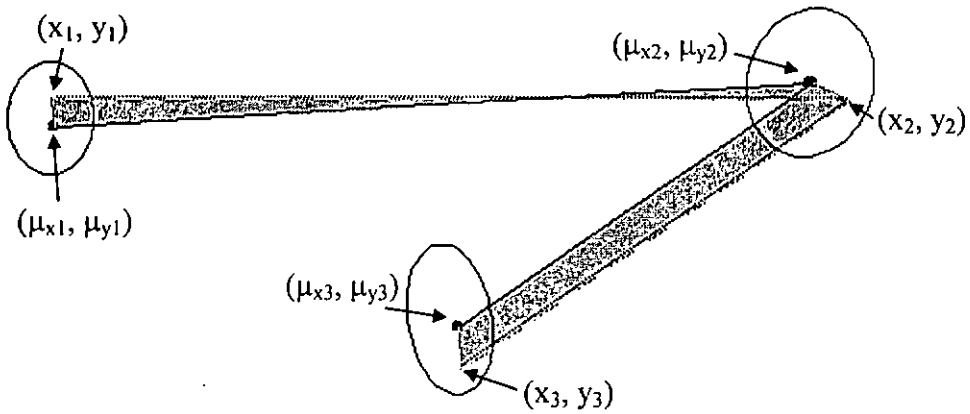


Figure 4.3. Discrepancy of a linear feature.

I call this discrepancy a triangle-quadrangle case. In this example, the discrepancy has four possible cases: triangle-triangle, triangle-quadrangle, quadrangle-triangle, and quadrangle-quadrangle cases. A linear feature consists of NP nodes where NP is a positive integer and then there are 2^{NP-1} possibilities of its discrepancy. It may be impossible to derive one and only one equation used to calculate the discrepant area among these four cases. As a result, which case the discrepancy is should be determined, and then compute the discrepant area from the equation in the corresponding case.

According to Figure 4.3, the discrepant area of the linear feature cannot be computed by summing the discrepant areas of the line segments of the linear feature. Otherwise, the area of an extra region, which occurs if either the first measured line segment intersects the latter expected line segment or the first expected line segment intersects the latter measured line segment, will be computed twice during calculating the discrepant area of the linear feature.

Let (x_{120}, y_{120}) denote the intersecting point of the measured line segment connected by (x_1, y_1) and (x_2, y_2) and the expected line segment connected by (μ_{x_2}, μ_{y_2}) and (μ_{x_1}, μ_{y_3}). The subscript of x_{120} is '120', where the first character represents the measured line segment connected by (x_1, y_1) and (x_2, y_2); the

remaining two characters represent the expected line segment connected by (μ_{x_2},μ_{y_2}) and (μ_{x_3},μ_{y_3}). Then, I have

$$\begin{cases} x_{120} = \mu_{x_1} \\ y_{120} = y_1 + \dfrac{y_2 - y_1}{x_2 - x_1}(x_{120} - x_1) \end{cases} \quad \text{if} \begin{cases} x_1 \neq x_2 \\ \mu_{x_1} = \mu_{x_2} \end{cases}$$

$$\begin{cases} x_{120} = x_1 \\ y_{120} = \mu_{y_1} + \dfrac{\mu_{y_2} - \mu_{y_1}}{\mu_{x_2} - \mu_{x_1}}(x_{120} - \mu_{x_1}) \end{cases} \quad \text{if} \begin{cases} x_1 = x_2 \\ \mu_{x_1} \neq \mu_{x_2} \end{cases} \quad (4.9)$$

$$\begin{cases} x_{120} = \dfrac{(\mu_{x_1} - \mu_{x_2})(y_2 x_1 - x_2 y_1) + (x_2 - x_1)(\mu_{y_2}\mu_{x_1} - \mu_{x_2}\mu_{y_1})}{(y_2 - y_1)(\mu_{x_1} - \mu_{x_2}) + (x_2 - x_1)(\mu_{y_2} - \mu_{y_1})} \\ y_{120} = y_1 + \dfrac{y_2 - y_1}{x_2 - x_1}(x_{120} - x_1) \end{cases} \quad \text{if} \begin{cases} x_1 \neq x_2 \\ \mu_{x_1} \neq \mu_{x_2} \end{cases}$$

The area of the so-called extra region is

$$A_3(x_1, y_1, x_2, y_2, x_3, y_3) = \frac{1}{2}\left|\mu_{x_2} y_2 + x_2 y_{120} + x_{120}\mu_{y_2} - x_2\mu_{y_2} - x_{120} y_2 - \mu_{x_2} y_{120}\right|. \quad (4.10)$$

The expected discrepant area of the linear feature in the triangle-triangle case is

$$A_4(x_1, y_1, x_2, y_2, x_3, y_3) = A_1(x_1, y_1, x_2, y_2) + A_2(x_2, y_2, x_3, y_3) + \delta_{120} A_3(x_1, y_1, x_2, y_2, x_3, y_3) \quad (4.11)$$

where $\delta_{120} = \begin{cases} 1 & \text{if either the measure line segment composed of } (x_1, y_1) \text{ and } (x_2, y_2) \text{ intersects the expected line segment composed of } (\mu_{x_2},\mu_{y_2}) \text{ and } (\mu_{x_3},\mu_{y_3}) \text{ or the expected line segment composed of } (\mu_{x_1},\mu_{y_1}) \text{ and } (\mu_{x_2},\mu_{y_2}) \text{ intersects the measured line segment composed of } (x_2, y_2) \text{ and } (x_3, y_3); \\ 0 & \text{otherwise.} \end{cases}$

The expected discrepant areas of the linear feature in the triangle-triangle, the quadrangle-triangle and the quadrangle-quadrangle cases A_5, A_6 and A_7 are given by Equations (4.12), (4.13) and (4.14) respectively.

$$A_5(x_1,y_1,x_2,y_2,x_3,y_3) = A_1(x_1,y_1,x_2,y_2) + A_1(x_2,y_2,x_3,y_3) + \delta_{120}A_3(x_1,y_1,x_2,y_2,x_3,y_3) \quad (4.12)$$

$$A_6(x_1,y_1,x_2,y_2,x_3,y_3) = A_2(x_1,y_1,x_2,y_2) + A_1(x_2,y_2,x_3,y_3) + \delta_{120}A_3(x_1,y_1,x_2,y_2,x_3,y_3) \quad (4.13)$$

$$A_7(x_1,y_1,x_2,y_2,x_3,y_3) = A_2(x_1,y_1,x_2,y_2) + A_2(x_2,y_2,x_3,y_3) + \delta_{120}A_3(x_1,y_1,x_2,y_2,x_3,y_3) \quad (4.14)$$

Procedures used to estimate the discrepant area of the linear feature are similar to Algorithms 4.1 and 4.2; with respect to the uniformly or normally distributed nodal errors of the linear feature. Two modified algorithms are given below respectively.

Algorithm 4.3

Step 1 Input the total number of replications for the simulation N; the total number of nodes on the linear feature NP; the NP expected nodes; and parameters of NP error ellipses.

Step 2 Set j = 0 where j is used to count the number of replications for the simulation.

Step 3 Generate 2 x NP random numbers u_i (where i = 1, 2, ..., NP) in [0,1]; and increase j by 1.

Step 4 Transform these random numbers to NP generated nodes for NP nodes of the line segment from the following equations

$$\begin{aligned} x_i &= 2(u_{2i-1} - 0.5)a_i + \mu_{x_i} \\ y_i &= 2(u_{2i} - 0.5)b_i + \mu_{y_i} \\ &\text{for } i = 1,\ldots, NP. \end{aligned} \quad (4.15)$$

Step 5 Decrease j by 1 and go to step 3 if at least one of $(x_1, y_1), \ldots, (x_{NP}, y_{NP})$ does not lie in its error ellipse.

Step 6 Compute an area of the discrepancy.

Step 7 Go to step 3 if j is less than N.

Step 8 Compute the average of the discrepant area.

Algorithm 4.4

Step 1 Input the total number of replications for the simulation N; the total number of nodes on the linear feature NP; the NP expected nodes; parameters of the NP error ellipses; and correlation coefficients among the nodal errors.

Step 2 Set j = 0 where j is used to count the number of replications for the simulation.

Step 3 Generate 2 x NP standard normal random numbers u_i (where i = 1, 2, ..., NP) in [0,1]; and

increase j by 1.

Step 4 Transform these random numbers to NP generated nodes for NP nodes of the line segment from the following equations

$$x_i = \sum_{k=1}^{i-1} \rho_{x_k x_i} u_{2k-1} \sigma_{x_i} + \sum_{k=1}^{i-1} \rho_{x_i y_k} u_{2k} \sigma_{x_i} + \sqrt{1 - \sum_{k=1}^{i-1} \rho_{x_k x_i}^2 - \sum_{k=1}^{i-1} \rho_{x_i y_k}^2} \, u_{2i-1} \sigma_{x_i} + \mu_{x_i}$$

$$y_i = \sum_{k=1}^{i} \rho_{x_k y_i} u_{2k-1} \sigma_{y_i} + \sum_{k=1}^{i-1} \rho_{y_k y_i} u_{2k} \sigma_{y_i} + \sqrt{1 - \sum_{k=1}^{i} \rho_{x_k y_i}^2 - \sum_{k=1}^{i-1} \rho_{y_k y_i}^2} \, u_{2i} \sigma_{y_i} + \mu_{y_i}$$

, for i = 1,..., NP. (4.16)

Step 5 Decrease j by 1 and go to step 3 if at least one of $(x_1, y_1), \ldots, (x_{NP}, y_{NP})$ does not lie in its error ellipse.

Step 6 Compute an area of the discrepancy.

Step 7 Go to step 3 if j is less than N.

Step 8 Compute the average value of the discrepant area.

Algorithms 4.1, 4.2, 4.3 and 4.4 are similar. Algorithms 4.1 and 4.3 are used to model the discrepancy in the uniform case while algorithms 4.2 and 4.4 are used to model the normal case. As a result, in step 3 algorithms 4.1 and 4.3 generate uniform random numbers and algorithms 4.2 and 4.4 generate normal random numbers.

Second, in the uniform case there is no correlation relation of the nodal error and hence step 4 in the uniform case is different from that in the normal case. Also, algorithms 4.3 and 4.4 are the general algorithms. When NP = 2, algorithms 4.1 and 4.2 are achieved.

4.1.3 Areal Features

The reliability of the boundary of an areal feature is probably defined by a combination of reliabilities of line segments circumscribing the areal feature. It is a fact that the reliability of the boundary of the areal feature is different from that of the areal feature for the interior of the areal feature may have errors. Consequently, the reliability of the areal feature is discussed here.

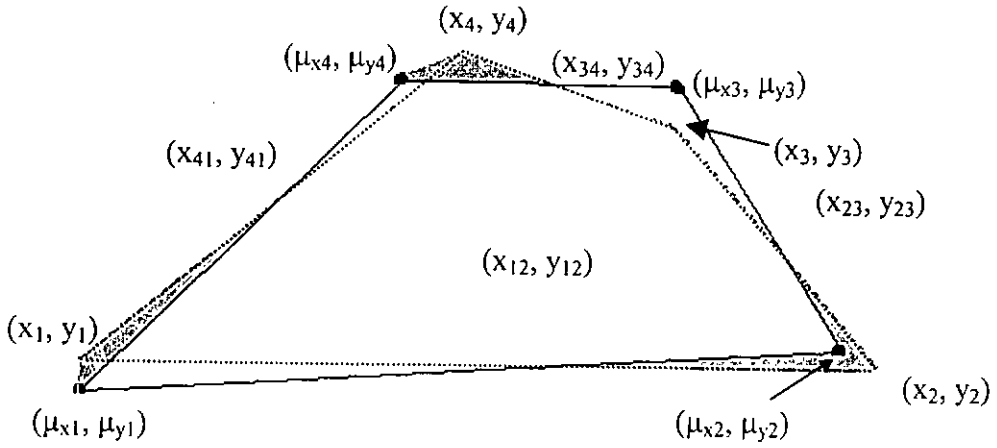


Figure 4.4. Discrepancy of an areal feature.

The shaded area in Figure 4.4 for instance is the discrepancy of the areal feature. The solid line represents the expected boundary of the areal feature, and the dash line represents the measured boundary. The discrepancy of the areal feature shown in this example is a triangle-triangle-triangle-triangle case, which is one of 16 possible cases. When the areal feature consists of NP nodes (a natural number NP is larger than 2) the quantity of all possibilities of the discrepancy of the areal feature is 2^{NP}. The discrepant area of the areal feature may not be defined from one equation only, as if the area of the discrepancy of the linear feature may not. This discrepant

area of the areal feature can be computed by summing up the discrepant areas of four line segments being boundaries of the areal feature; and the area of the interior part of the region which is bounded by the shaded region, the expected and the measured areal features; and then subtracting the area of the expected areal feature. In order to determine the interior part of the region that is bounded by the shaded region, the expected and the measured areal features (called region I), intersecting points of the NP measured line segments and the NP expected line segments should be computed in the first step. In addition to these intersecting points being some nodes of region I, I have to decide whether the measured node (x_i, y_i) is inside the expected areal feature for i = 1, 2, ..., NP. If the measured node is a point in the expected areal feature, it will be a node of region I. The area of region I can be estimated after its nodes are determined. The discrepant area of the areal feature in the remaining cases can be derived in the same manner. This computation considers the relationship between the discrepant area of a linear feature and that of an areal feature.

The algorithm for the discrepancy of the areal feature is the same as that for a discrepancy of a linear feature except for the mathematical expression of the discrepant area.

4.1.4 Examples

Easa (1995) considered a uniformly distributed error of nodes of a line segment, and the researcher's data values are used for comparison with the simulation model proposed here. Two expected nodes of the line segment (μ_{x_1}, μ_{y_1}) and (μ_{x_2}, μ_{y_2}) are (0, 0) and (1000, 0). In Equation (4.4) parameters a_1, b_1, c_1, a_2, b_2 and c_2 are 148, 148, 0, 54, 54 and 0. Also parameters d_1, e_1, d_2 and e_2 are any nonzero real numbers. A meter is a unit of length in this study. The total number of replications for the simulation N is 1000 because Easa found that the simulated result will be stable when N = 1000.

According to Algorithm 4.1, the average discrepant area of the line segment is 35987.2m^2 nearly to 1 decimal place. Running this simulation model again, I get another value of 35100.4m^2. Both values tend to be 35626.0m^2, which is Easa's result. The unstable average discrepant area in the simulation model is due to random variables. However, the result will be highly accurate if the total number of iterations for the simulation N is large. Meanwhile, increasing N will lead the algorithm to be time inefficient during execution. The value of N should be considered based on an acceptant level or a tolerance.

A (1-α)% confidence interval for the mean of the simulation result is evaluated where α is a positive real number smaller than 1. Algorithm 4.1 should be repeated NR times where NR is a positive integer. A set of these NR average values of the discrepant area is further divided into NB even subgroups where NB is a positive integer. Since each subgroup has NR/NB average values of the discrepant area, a mean of these NR/NB average values called a 'sub-mean' of the discrepant area in each subgroup is computed. Then, from the central limit theorem, a (1-α)% confidence interval for mean of this 'sub-mean' is given by

$$\overline{A} \pm \frac{t_{\alpha/2}\sigma_A}{\sqrt{NB}}, \tag{4.17}$$

where variable A is the mean of the discrepant area in each subgroup, i.e. the 'sub-mean' of the discrepant area; $\overline{A}$ is a mean of A; σ_A is the sample standard derivation of A. In the simulation model, I set N = 1000, NR = 50 and NB = 10 (same values as Easa's model).

Table 4.1 shows mean $\overline{A}$ and the 95% confidence interval for mean $\overline{A}$ of the 'sub-mean' of the discrepant area. The expected line segment is connected by (0, 0) and (1000, 0). The first four columns in Table 4.1 record some parameters of two error ellipses. Parameters c_1 and c_2 in Equation (4.4) are zero; parameters $\rho_{x_1y_1}$ and $\rho_{x_2y_2}$ in Equation (4.6) are zero. Also, the remaining parameters in Equation (4.4) or Equation (4.6) are any non-zero real number because in the left-hand side of these two equations, the third part becomes zeros. The last two columns

in Table 4.1 give the mean value $\overline{A}$ of the 'sub-mean' of the discrepant area and its 95% confidence interval for a mean of A under two different assumptions of the error of the nodes on the line segment. The first is uniformly distributed and the second normally distributed. In these two columns, the first value is the mean of the 'sub-mean' of the discrepant area; the value in the brackets is the 95% confidence interval for the mean of the 'sub-mean' of the discrepant area. In the following, I name the mean of the 'sub-mean' *the mean of the discrepant area* for simplification. Figure 4.5 shows a distribution of the line segment in the uniform case, given that parameters of the two error ellipses are exactly equal to that in the second row of Table 4.1; i.e. $a_1 = 100$, $b_1 = 196$, $a_2 = 30$ and $b_2 = 78$ except NR = NB = 1. A set of line segments in this illustration involves the expected line segment and a thousand generated line segments; an error ellipse of each node is figured. For data values in Table 4.1, distribution of these line segments is shown in Figure 4.6 if NR = NB =1. Both Table 4.1 and Figure 4.6 are used to compare the line segments under various error ellipses.

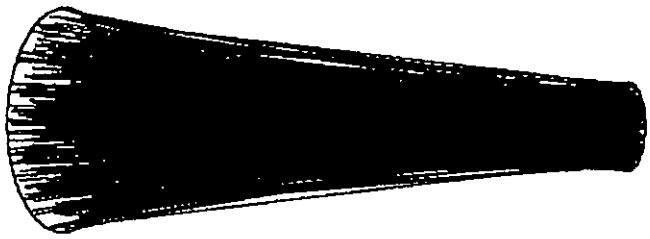
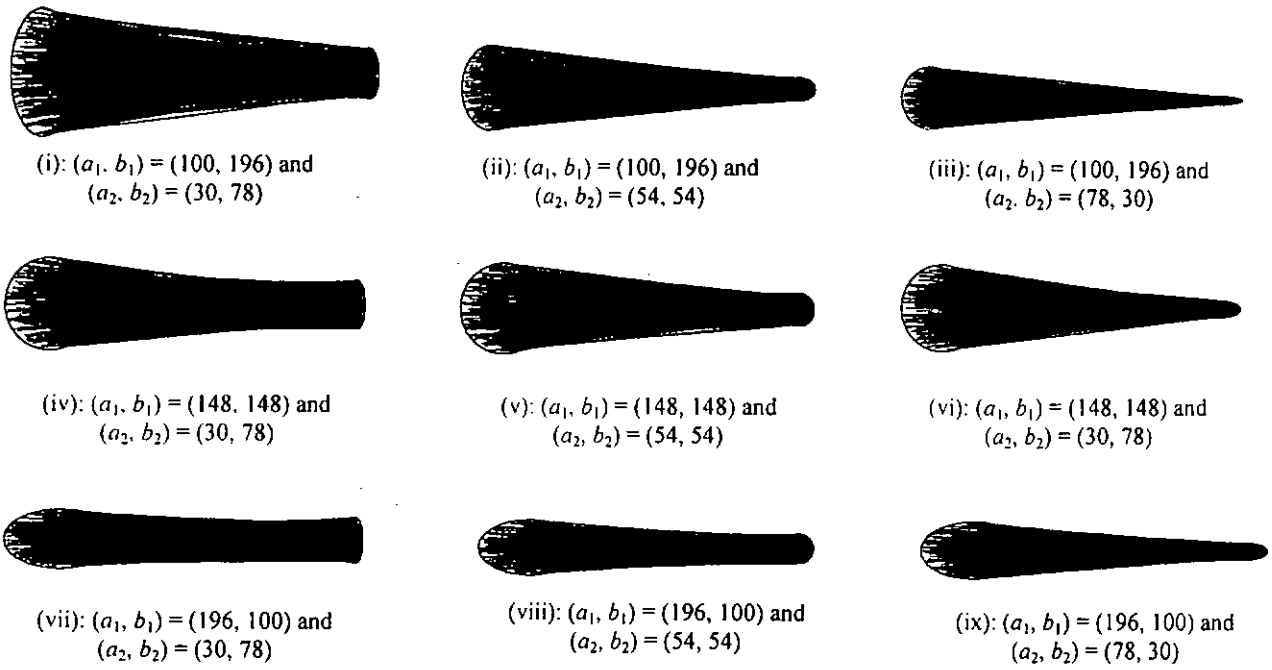


Figure 4.5. Distribution of a line segment in a uniform independent case.

Table 4.1. Area of discrepancy of line segments.

a_1 (m)	b_1 (m)	a_2 (m)	b_2 (m)	Average area * (m^2)	
				Uniform Random Variable	Normal Random Variable
100	196	30	78	48442.7 (48226.0, 48659.4)	34977.7 (34776.8, 35178.5)
100	196	54	54	45900.1 (45669.7, 46130.4)	32857.7 (32719.0, 32996.4)
100	196	78	30	43248.2 (42932.5, 43564.0)	30902.0 (30758.1, 31046.0)
148	148	30	78	39260.7 (39058.2, 39463.3)	28513.2 (28274.5, 28751.9)
148	148	54	54	35987.2 (35905.9, 36068.5)	25950.2 (25739.4, 26161.0)
148	148	78	30	33145.8 (33000.3, 33291.2)	23835.9 (23730.0, 23941.8)
196	100	30	78	30379.6 (30255.8, 30503.4)	22006.6 (21910.8, 22102.5)
196	100	54	54	26669.4 (26556.9, 26781.9)	19265.8 (19173.6, 19358.1)
196	100	78	30	23480.7 (23411.6, 23549.8)	16998.1 (16883.9, 17112.4)



(i): $(a_1, b_1) = (100, 196)$ and $(a_2, b_2) = (30, 78)$

(ii): $(a_1, b_1) = (100, 196)$ and $(a_2, b_2) = (54, 54)$

(iii): $(a_1, b_1) = (100, 196)$ and $(a_2, b_2) = (78, 30)$

(iv): $(a_1, b_1) = (148, 148)$ and $(a_2, b_2) = (30, 78)$

(v): $(a_1, b_1) = (148, 148)$ and $(a_2, b_2) = (54, 54)$

(vi): $(a_1, b_1) = (148, 148)$ and $(a_2, b_2) = (30, 78)$

(vii): $(a_1, b_1) = (196, 100)$ and $(a_2, b_2) = (30, 78)$

(viii): $(a_1, b_1) = (196, 100)$ and $(a_2, b_2) = (54, 54)$

(ix): $(a_1, b_1) = (196, 100)$ and $(a_2, b_2) = (78, 30)$

(a) Uniform random case

(i): $(a_1, b_1) = (100, 196)$ and $(a_2, b_2) = (30, 78)$

(ii): $(a_1, b_1) = (100, 196)$ and $(a_2, b_2) = (54, 54)$

(iii): $(a_1, b_1) = (100, 196)$ and $(a_2, b_2) = (78, 30)$

(iv): $(a_1, b_1) = (148, 148)$ and $(a_2, b_2) = (30, 78)$

(v): $(a_1, b_1) = (148, 148)$ and $(a_2, b_2) = (54, 54)$

(vi): $(a_1, b_1) = (148, 148)$ and $(a_2, b_2) = (78, 30)$

(vii): $(a_1, b_1) = (196, 100)$ and $(a_2, b_2) = (30, 78)$

(viii): $(a_1, b_1) = (196, 100)$ and $(a_2, b_2) = (54, 54)$

(ix): $(a_1, b_1) = (196, 100)$ and $(a_2, b_2) = (78, 30)$

(b) Normal random case

Figure 4.6. Distribution of line segments in an independent case.

In Figure 4.6, if either a increases or b decreases, the measured line segment has a decreasing variation on its distribution. From Figure 4.6(a)(i) to Figure 4.6(a)(iii), the line segment for instance has a decreasing variation on its distribution when b_2 reduces from 78 m to 30 m. Similarly, if b_1 decreases from 196 m to 100 m, the variation of the distribution of the line segment also drops (see Figures 4.6(a)(i), 4.6(a)(iv) and 4.6(a)(vii)).

Besides, the mean of the discrepant area (the mean of the 'sub-mean' of the discrepant area) in the uniform case is larger than that in the normal case as shown in Table 4.1 and the same result is indicated in Figure 4.6. It is because uniform random variables can be viewed as normal random variables with infinite standard derivation.

In the above examples, line segments are generated under the assumption of independent random variables. Correlated random variables will be considered in the following to generate measured line segments because nodal errors of a spatial feature may be linearly dependent. Figure 4.7 shows an expected line segment of nodes (0,0) and (1000,0) under different error ellipse parameters. If parameters a_i, b_i and c_i in Equation (4.4) are given, parameters $\sigma_{x_i}, \sigma_{y_i}$ and $\rho_{x_i y_i}$ in Equation (4.6) will be obtained and vice versa. Figure 4.7 shows the expected line segment with different correlation coefficients when $a_1 = 100$, $b_1 = 196$, $a_2 = 30$, $b_2 = 78$, $\rho_{x_1 y_1} = -1.0$, $\rho_{x_2 y_1} = \rho_{x_1 y_2} = \rho_{y_1 y_2} = 0.0$ and the remaining error ellipse parameters are changing in the normal case.

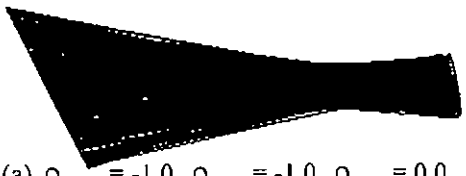
In Figures 4.7(a) - (e), when $\rho_{x_1 x_2} = -1.0$ and $\rho_{x_2 y_2}$ increases from –1.0 to 1.0, there is an increasing variation of the distribution of the line segment. In Figures 4.7(f) - (j), the distribution of the line segment does not have any significant changes in its variation when $\rho_{x_1 x_2} = 0.0$ and $\rho_{x_2 y_2}$ changes from –1.0 to 1.0. Besides, if $\rho_{x_1 x_2} = 1.0$ and $\rho_{x_2 y_2}$ is increases by 0.5, the variation of the distribution of the line

segment will be reduced. In Figures 4.8(a) - (c), the variation of the line segment's distribution decreases when $\rho_{y_1y_2}$ is reduced from –1.0 to 1.0, $\rho_{x_1y_1} = 0.0, \rho_{x_1x_2} = -1.0$ and $\rho_{x_2y_1} = \rho_{x_1y_2} = \rho_{x_2y_2} = 0.0$. In Figures 4.8(d) - (f), there are no significant changes in the variation of the distribution of the line segment when $\rho_{x_2y_1}$ changes from –1.0 to 1.0, $\rho_{x_1y_1} = -0.5, \rho_{x_1x_2} = 0.0$, $\rho_{x_1y_2} = -1.0$ and $\rho_{y_1y_2} = \rho_{x_2y_2} = 0.0$.

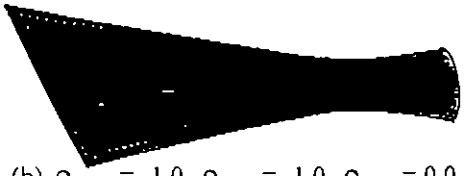
When only one of the error ellipse correlation coefficients increases and the remaining are zero, the variation of a line segment's distribution will be discussed. If either $\rho_{x_1y_1}$ or $\rho_{x_2y_2}$ increases from –1.0, the variation decreases. During a change of either $\rho_{x_1y_1}$ or $\rho_{x_2y_2}$, the variation has a minimum value at varying the correlation coefficient to zero. In another example, the variation will increase if $\rho_{y_1y_2}$ increases. However, the variation has no significant changes when I vary $\rho_{x_1x_2}$. This is due to the fact that the expected line segment in these examples is parallel to the x-axis and so both y_1 and y_2 determine which case (two-triangle or quadrangle) the line segment belongs to regardless of x_1 and x_2. x_1, y_1 and x_2 are considered to be the same values in Figures 4.1 and 4.2. If y_2 in Figure 4.1 has identical magnitude with this in Figure 4.2 but they are in an opposite sign, the discrepant area in a quadrangle case is larger than that in a two-triangle case. This implies that $\rho_{y_1y_2}$ affects the discrepant area of a line segment mainly given that the expected line segment is parallel to the x-axis of the coordinate system.

For the example of a linear feature, the three expected nodes are (0, 0), (1000, 0) and (1500, 866). The parameters of the three error ellipses a_1, b_1, a_2, b_2, a_3 and b_3 are 100m, 196m, 30m, 78m, 100m and 196m respectively. In the uniform case, the mean of the 'sub-mean' of the discrepant area and the 95% confidence interval for mean of the 'sub-mean' of the discrepant area are 80222.1m^2 and (79972.1m^2, 80472.1m^2). In the normal case, the mean and the confidence interval are 57730.9m^2 and (57493.9m^2, 57967.9m^2). Using the three expected nodes of the linear feature for

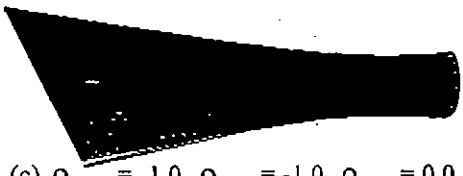
the areal feature results in the mean and the 95% confidence interval 184109.3m^2 and (183484.0m^2, 184734.5m^2) respectively in the uniform case. In the normal case, the mean and the confidence interval are 133342.2m^2 and (132795.9m^2, 133888.4m^2).



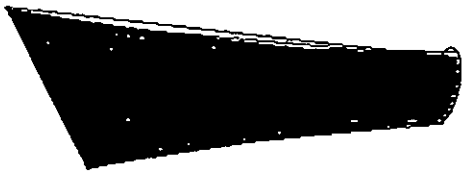
(a) $\rho_{x_1y_1} = -1.0$, $\rho_{x_1x_2} = -1.0$, $\rho_{x_2y_1} = 0.0$, $\rho_{x_1y_2} = 0.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = -1.0$



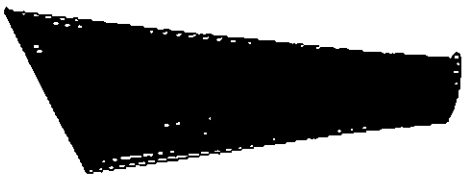
(b) $\rho_{x_1y_1} = -1.0$, $\rho_{x_1x_2} = -1.0$, $\rho_{x_2y_1} = 0.0$, $\rho_{x_1y_2} = 0.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = -0.5$



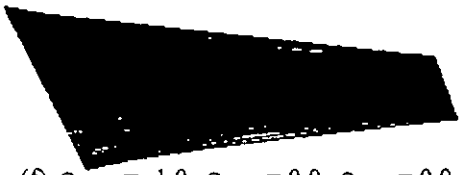
(c) $\rho_{x_1y_1} = -1.0$, $\rho_{x_1x_2} = -1.0$, $\rho_{x_2y_1} = 0.0$, $\rho_{x_1y_2} = 0.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = 0.0$



(d) $\rho_{x_1y_1} = -1.0$, $\rho_{x_1x_2} = -1.0$, $\rho_{x_2y_1} = 0.0$, $\rho_{x_1y_2} = 0.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = 0.5$

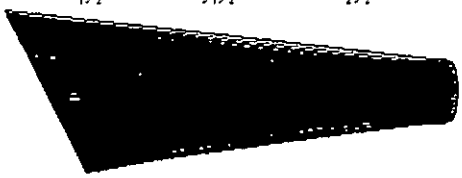


(e) $\rho_{x_1y_1} = -1.0$, $\rho_{x_1x_2} = -1.0$, $\rho_{x_2y_1} = 0.0$, $\rho_{x_1y_2} = 0.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = 1.0$

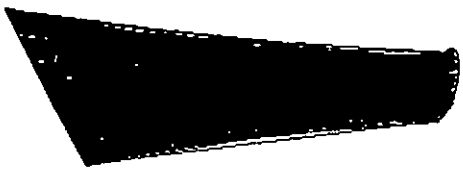


(f) $\rho_{x_1y_1} = -1.0$, $\rho_{x_1x_2} = 0.0$, $\rho_{x_2y_1} = 0.0$, $\rho_{x_1y_2} = 0.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = -1.0$

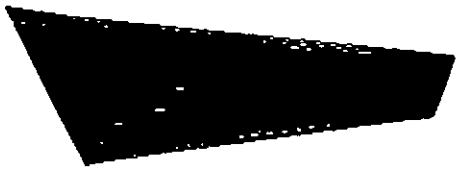
(g) $\rho_{x_1y_1} = -1.0$, $\rho_{x_1x_2} = 0.0$, $\rho_{x_2y_1} = 0.0$, $\rho_{x_1y_2} = 0.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = -0.5$



(h) $\rho_{x_1y_1} = -1.0$, $\rho_{x_1x_2} = 0.0$, $\rho_{x_2y_1} = 0.0$, $\rho_{x_1y_2} = 0.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = 0.0$



(i) $\rho_{x_1y_1} = -1.0$, $\rho_{x_1x_2} = 0.0$, $\rho_{x_2y_1} = 0.0$, $\rho_{x_1y_2} = 0.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = 0.5$



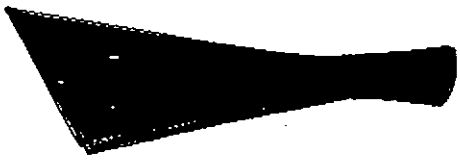
(j) $\rho_{x_1y_1} = -1.0$, $\rho_{x_1x_2} = 0.0$, $\rho_{x_2y_1} = 0.0$, $\rho_{x_1y_2} = 0.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = 1.0$

(k) $\rho_{x_1y_1} = -1.0$, $\rho_{x_1x_2} = 1.0$, $\rho_{x_2y_1} = 0.0$, $\rho_{x_1y_2} = 0.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = -1.0$

(l) $\rho_{x_1y_1} = -1.0$, $\rho_{x_1x_2} = 1.0$, $\rho_{x_2y_1} = 0.0$, $\rho_{x_1y_2} = 0.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = -0.5$

(m) $\rho_{x_1y_1} = -1.0$, $\rho_{x_1x_2} = 1.0$, $\rho_{x_2y_1} = 0.0$,
$\rho_{x_1y_2} = 0.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = 0.0$

(n) $\rho_{x_1y_1} = -1.0$, $\rho_{x_1x_2} = 1.0$, $\rho_{x_2y_1} = 0.0$,
$\rho_{x_1y_2} = 0.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = 0.5$



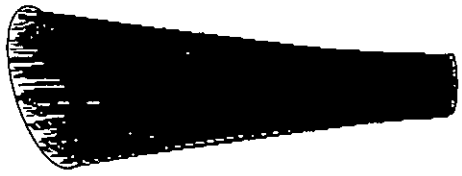
(o) $\rho_{x_1y_1} = -1.0$, $\rho_{x_1x_2} = 1.0$, $\rho_{x_2y_1} = 0.0$,
$\rho_{x_1y_2} = 0.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = 1.0$

Figure 4.7. Distribution of line segments with $\rho_{x_1y_1} = -1.0$.

(a) $\rho_{x_1y_1} = 0.0$, $\rho_{x_1x_2} = -1.0$, $\rho_{x_2y_1} = 0.0$,
$\rho_{x_1y_2} = 0.0$, $\rho_{y_1y_2} = -1.0$, $\rho_{x_2y_2} = 0.0$

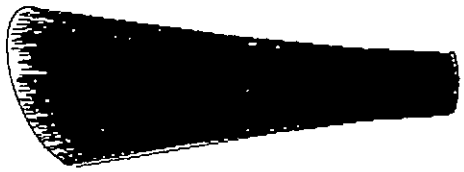
(b) $\rho_{x_1y_1} = 0.0$, $\rho_{x_1x_2} = -1.0$, $\rho_{x_2y_1} = 0.0$,
$\rho_{x_1y_2} = 0.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = 0.0$

(c) $\rho_{x_1y_1} = 0.0$, $\rho_{x_1x_2} = -1.0$, $\rho_{x_2y_1} = 0.0$,
$\rho_{x_1y_2} = 0.0$, $\rho_{y_1y_2} = 1.0$, $\rho_{x_2y_2} = 0.0$



(d) $\rho_{x_1y_1} = -0.5$, $\rho_{x_1x_2} = 0.0$, $\rho_{x_2y_1} = -1.0$,
$\rho_{x_1y_2} = -1.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = 0.0$

(e) $\rho_{x_1y_1} = -0.5$, $\rho_{x_1x_2} = 0.0$, $\rho_{x_2y_1} = 0.0$,
$\rho_{x_1y_2} = -1.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = 0.0$



(f) $\rho_{x_1y_1} = -0.5$, $\rho_{x_1x_2} = 0.0$, $\rho_{x_2y_1} = 1.0$,
$\rho_{x_1y_2} = -1.0$, $\rho_{y_1y_2} = 0.0$, $\rho_{x_2y_2} = 0.0$

Figure 4.8. Distribution of line segments with $\rho_{x_2y_2} = 0.0$.

4.2 Reliability of Spatial Features in 3D GIS

The reliability of spatial features including linear features, areal features and volumetric features in 3D GIS will be studied. Since the algorithm used to calculate the area of the discrepancy of a spatial feature (a measure of the reliability of the spatial feature) is similar to that in 2D GIS, this section will concentrate on defining the discrepancy of 3D spatial features.

The computation will assume that an ellipsoid is centered on a true node of a spatial feature. A mathematical expression of the error ellipse is given by the following:

For i = 1, ..., NP, I have

$$\left(\frac{x_i-\mu_{x_i}}{a_i}\right)^2+\left(\frac{y_i-\mu_{y_i}}{b_i}\right)^2+\left(\frac{z_i-\mu_{z_i}}{c_i}\right)^2+d_{1,i}\left(\frac{x_i-\mu_{x_i}}{d_{2,i}}\right)\left(\frac{y_i-\mu_{y_i}}{d_{3,i}}\right)$$
$$+e_{1,i}\left(\frac{x_1-\mu_{x_1}}{e_{2,i}}\right)\left(\frac{z_1-\mu_{z_1}}{e_{3,i}}\right)+f_{1,i}\left(\frac{y_1-\mu_{y_1}}{f_{2,i}}\right)\left(\frac{z_1-\mu_{z_1}}{f_{3,i}}\right)=1 \qquad (4.18)$$

where a_i, b_i, c_i, $d_{2,i}$, $d_{3,i}$, $e_{2,i}$, $e_{3,i}$ $f_{2,i}$ and $f_{3,i}$ are non-zero real numbers; and

$d_{1,i}$, $e_{1,i}$ and $f_{1,i}$ are any real numbers.

If $d_{1,i}$, $e_{1,i}$ and $f_{1,i}$ are zero, a_i, b_i and c_i will be the semi-axes of the error ellipsoid which are parallel to the x-, to the y- and to the z-axes of the coordinate system.

In the normal case, the error ellipsoid refers to a $(1-\alpha)$% confidence region for mean of a node of the line segment, and then Equation (4.18) is modified as shown below.

$$\left(\frac{x_i-\mu_{x_i}}{a_i}\right)^2+2\left(\frac{\rho_{x_iz_i}\rho_{y_iz_i}-\rho_{x_iy_i}}{k}\right)\left(\frac{x_i-\mu_{x_i}}{\sigma_{x_i}}\right)\left(\frac{y_i-\mu_{y_i}}{\sigma_{y_i}}\right)$$
$$+\left(\frac{y_i-\mu_{y_i}}{b_i}\right)^2+2\left(\frac{\rho_{x_iy_i}\rho_{y_iz_i}-\rho_{x_iz_i}}{k}\right)\left(\frac{x_i-\mu_{x_i}}{\sigma_{x_i}}\right)\left(\frac{z_i-\mu_{z_i}}{\sigma_{z_i}}\right) \qquad (4.19)$$
$$+\left(\frac{z_i-\mu_{z_i}}{c_i}\right)^2+2\left(\frac{\rho_{x_iy_i}\rho_{x_iz_i}-\rho_{y_iz_i}}{k}\right)\left(\frac{y_i-\mu_{y_i}}{\sigma_{y_i}}\right)\left(\frac{z_i-\mu_{z_i}}{\sigma_{z_i}}\right)$$
$$=1$$

where $i = 1, 2, \ldots, NP$;

$$k=-2(\ln(1-\alpha))(1-\rho_{x_iy_i}^2-\rho_{x_iz_i}^2-\rho_{y_iz_i}^2+2\rho_{x_iy_i}\rho_{x_iz_i}\rho_{y_iz_i});$$

$$a_i=\frac{\sigma_{x_i}}{\sqrt{-2\log(\alpha)}};$$

$$b_i=\frac{\sigma_{y_i}}{\sqrt{-2\log(\alpha)}};\text{and}$$

$$c_i=\frac{\sigma_{z_i}}{\sqrt{-2\log(\alpha)}}$$

4.2.1 Line Segments

The discrepancy of a linear feature in 3D GIS is defined by the difference between the expected linear feature and the measured linear features; and this difference is due to measurement error. Similar to the 2D problem, the discrepancy of a line segment is studied first. Figure 4.9 shows the discrepancy of a line segment in 3D GIS. The solid line segment represents the expected line segment composed of the two expected nodes, $(\mu_{x_1},\mu_{y_1},\mu_{z_1})$ and $(\mu_{x_2},\mu_{y_2},\mu_{z_2})$. And the dash line segment represents the measured line segment composed of the two measured nodes, (x_1, y_1, z_1) and (x_2, y_2, z_2). The discrepancy of the line segment is the shaded plane, either flat or twisted. For each node of the line segment, its error ellipsoid centered on its expected node can be seen and its measured node is mostly in the error ellipsoid.

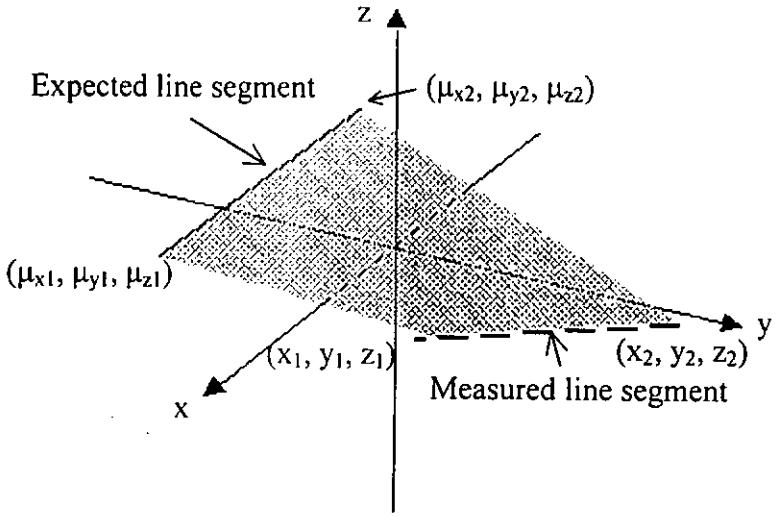


Figure 4.9. Discrepancy of a line segment in 3D GIS.

The discrepant area of the line segment (the shaded plane) is determined differently depending on its case. Three possible cases exist: (a) the measured and the expected line segments do not intersect and they are on a flat plane; (b) they intersect; (c) cases neither (a) nor (b) is a possibility.

It is assumed that the measured locations and the expected locations of all nodes should be on a flat plane and their two corresponding line segments should not intersect. In such a situation, the discrepant area can be computed as follows.

Let 1×3 vectors A, B, C and D denote $(x_1 - \mu_{x_1}, y_1 - \mu_{y_1}, z_1 - \mu_{z_1})$, $(\mu_{x_2} - \mu_{x_1}, \mu_{y_2} - \mu_{y_1}, \mu_{z_2} - \mu_{z_1})$, $(x_1 - \mu_{x_2}, y_1 - \mu_{y_2}, z_1 - \mu_{z_2})$, and $(x_2 - \mu_{x_2}, y_2 - \mu_{y_2}, z_2 - \mu_{z_2})$ respectively. Then, I have

$$A_s = \frac{1}{2}|A \times B| + \frac{1}{2}|C \times D|, \tag{4.20}$$

where A × B is a vector cross product of A and B and so forth;

$|(c_1, c_2, c_3)|$ is the length or magnitude of vector (c_1, c_2, c_3).

In the second case, the measured and the expected line segments intersect at a point (x_{12}, y_{12}, z_{12}). Let 1×3 vectors A', B', C' and D' denote $(x_1-\mu_{x_1}, y_1-\mu_{y_1}, z_1-\mu_{z_1})$, $(x_{12}-\mu_{x_1}, y_{12}-\mu_{y_1}, z_{12}-\mu_{z_1})$, $(x_{12}-\mu_{x_2}, y_{12}-\mu_{y_2}, z_{12}-\mu_{z_2})$, and $(x_2-\mu_{x_2}, y_2-\mu_{y_2}, z_2-\mu_{z_2})$ respectively. The area of the discrepancy is expressed in Equation (4.21).

$$A_9 = \frac{1}{2}|A'\times B'| + \frac{1}{2}|C'\times D'|. \tag{4.21}$$

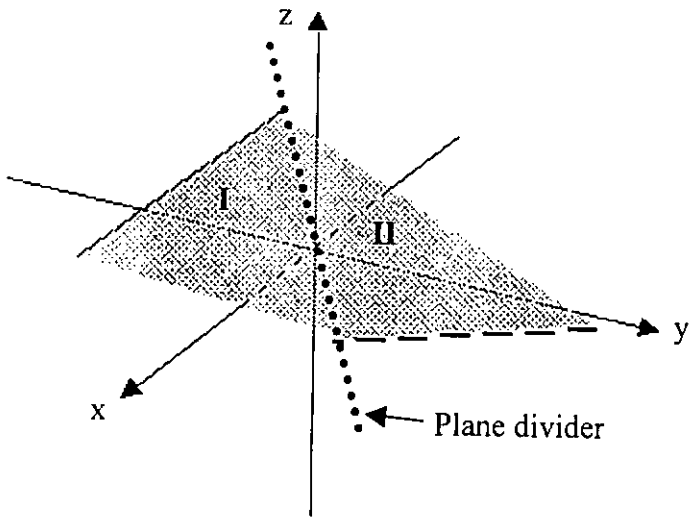


Figure 4.10. Two sub-planes for discrepancy of a line segment in 3D GIS.

It is also possible that both the measured and the expected line segments are neither on a 'flat' plane nor intersect. Under these circumstances, the area of the shaded plane cannot be computed exactly. The obscurity of the equation formed by the measured and the true nodes affects the discrepancy; the discrepancy cannot be readily calculated. To simplify and quantify such a case, the approximate area of the shaded plane will be obtained. Supposing the shaded plane shown in Figure 4.9 is twisted, the plane is divided into two sub-planes (see Figure 4.10). The dotted line called plane divider is used to divide the shaded plane into the two sub-planes I and II. Nodes on plane I are $(\mu_{x_1}, \mu_{y_1}, \mu_{z_1})$, $(\mu_{x_2}, \mu_{y_2}, \mu_{z_2})$ and (x_1, y_1, z_1); nodes on plane II are $(\mu_{x_2}, \mu_{y_2}, \mu_{z_2})$, (x_1, y_1, z_1) and (x_2, y_2, z_2). The total area of the shaded planes I and II is presented in Equation (4.22).

$$A_{10} = \frac{1}{2}\left|\left(\mu_{x_2} - \mu_{x_1}, \mu_{y_2} - \mu_{y_1}, \mu_{z_2} - \mu_{z_1}\right) \times \left(x_1 - \mu_{x_1}, y_1 - \mu_{y_1}, z_1 - \mu_{z_1}\right)\right| + \frac{1}{2}\left|\left(\mu_{x_2} - x_2, \mu_{y_2} - y_2, \mu_{z_2} - z_2\right) \times \left(x_1 - x_2, y_1 - y_2, z_1 - z_2\right)\right| \quad (4.22)$$

Another plane divider passing via $(\mu_{x_1}, \mu_{y_1}, \mu_{z_1})$ and (x_2, y_2, z_2) can be chosen, then the two sub-planes III and IV contain $(\mu_{x_1}, \mu_{y_1}, \mu_{z_1})$, (x_1, y_1, z_1) and (x_2, y_2, z_2), and $(\mu_{x_1}, \mu_{y_1}, \mu_{z_1})$, $(\mu_{x_2}, \mu_{y_2}, \mu_{z_2})$ and (x_2, y_2, z_2) respectively. The total area of the shaded planes III and IV is shown below

$$A_{11} = \frac{1}{2}\left|\left(\mu_{x_1} - x_1, \mu_{y_1} - y_1, \mu_{z_1} - z_1\right) \times \left(x_2 - x_1, y_2 - y_1, z_2 - z_1\right)\right| + \frac{1}{2}\left|\left(\mu_{x_1} - \mu_{x_2}, \mu_{y_1} - \mu_{y_2}, \mu_{z_1} - \mu_{z_2}\right) \times \left(x_2 - \mu_{x_2}, y_2 - \mu_{y_2}, z_2 - \mu_{z_2}\right)\right| \quad (4.23)$$

As a result, the area of the discrepancy of the line segment is approximated by the average of A_{10} and A_{11}.

4.2.2 Linear Features

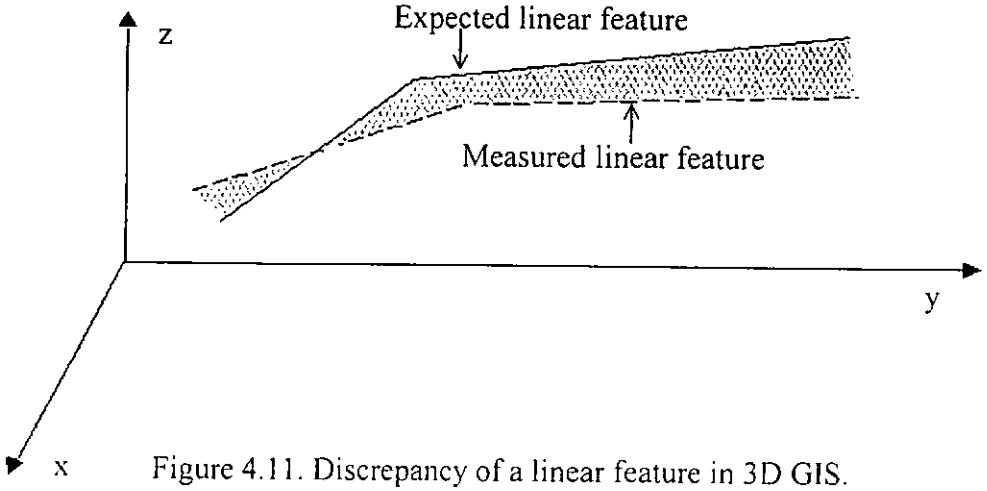


Figure 4.11. Discrepancy of a linear feature in 3D GIS.

Joining several line segments together yields a polyline, which is a broad line feature. In this instance, the linear feature only refers to an unclosed polyline. For instance, a linear feature passes via three nodes and its discrepancy is shown in Figure 4.11. The solid polyline represents the expected location and the dash polyline

represents the measured location. The area of the discrepancy of the polyline can be viewed as union of discrepancies of two line segments.

4.2.3 Areal Features

An areal feature as discussed in this paper refers to a polygon in the digital database sense. The discrepancy of the areal feature is distinct from the discrepancy of the boundary of the areal feature, which is in fact the discrepancy of the closed linear feature. The discrepancy of the areal feature should refer to a volume of the discrepant object. In Figure 4.12, the solid and the dash polylines represent boundaries of the expected areal feature and of the measured areal feature respectively; the shaded object is the discrepancy of the areal feature. A volume of the shaded object measures the reliability of the areal feature.

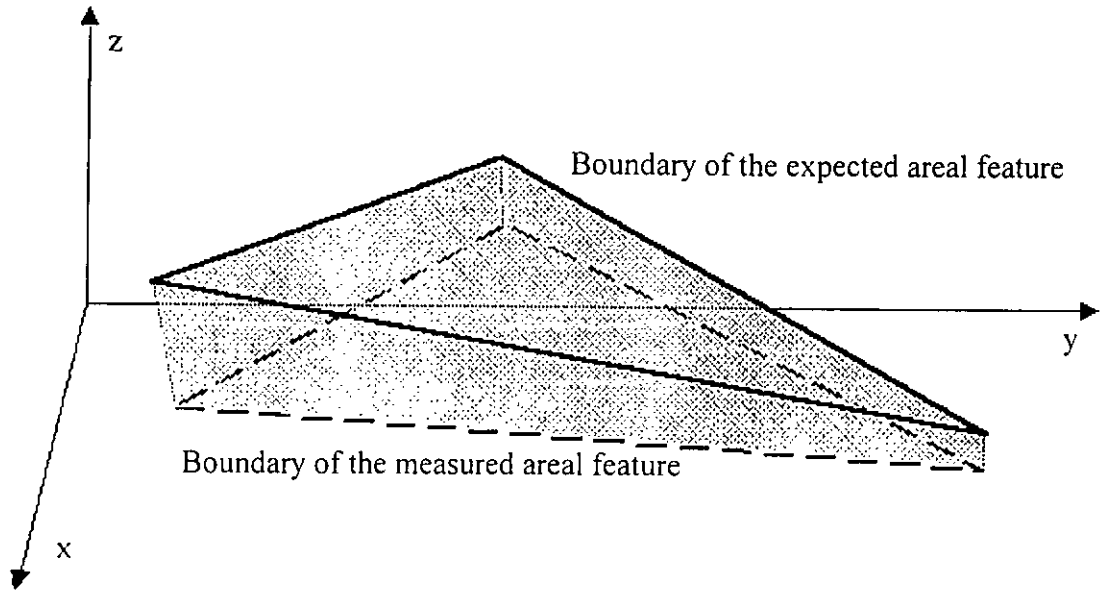


Figure 4.12. Discrepancy of an areal feature in 3D GIS.

4.2.4 Volumetric Features

In a 3D GIS, another important feature of spatial data is a volumetric feature. The discrepancy of the volumetric feature can be determined with Figure 4.14. This

illustration is an example of the volumetric feature whereby five surfaces S_1, S_2, S_3, S_4 and S_5 form the volumetric feature (see Figure 4.13). The difference between the measured and the expected volumetric features is the discrepancy of the volumetric feature. For gaining more information on this discrepancy, let us consider a difference between the expected volumetric feature and one of the measured surfaces such as the measured surface of S_4. This discrepancy (or difference) is the shaded object involving two sub-objects (see Figure 4.15). One contains nodes P_1, P_2, P_3 and P_4 in which P_1 is on the measured surface of S_1; P_2 and P_3 are intersecting points of S_1's expected surface and S_4's measured surface; P_4 is on the expected surface of S_1. The other contains nodes P_5, P_6, P_7, P_8, P_9 and P_{10} while P_5 and P_6 are intersecting points of S_4's measured surface and S_4's expected surface; P_7 and P_{10} are on the expected surface of S_4; P_8 and P_9 are on the measured surface of S_4. The shaded object in Figure 4.15 is related to the discrepancy between surface S_4 and the volumetric feature. Similarly, the discrepant volume between the remaining four surfaces and the volumetric feature is computed. The sum of the five discrepant volumes is a measure of the reliability of the volumetric feature.

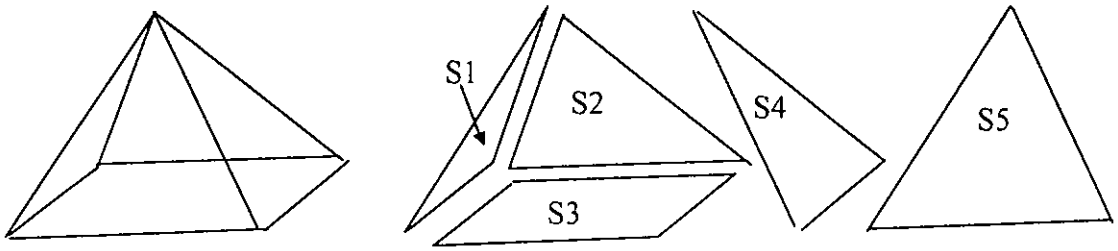


Figure 4.13. A volumetric feature in 3D GIS.

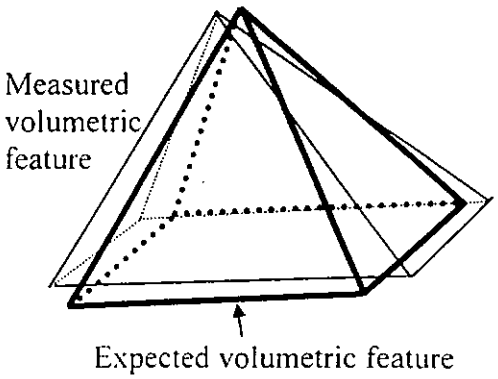


Figure 4.14. Discrepancy of a volumetric feature in 3D GIS.

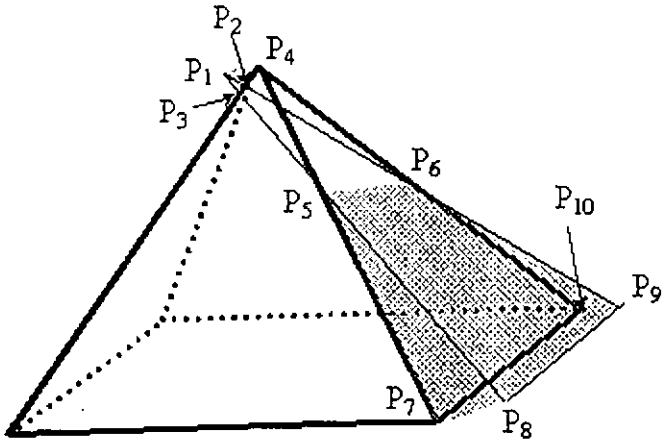


Figure 4.15. Discrepancy between a surface of the measured volumetric feature and the expected volumetric feature.

4.2.5 Examples

An expected line segment is connected by two expected nodes (0, 0, 0) and (1000, 0, 0). The parameters of the two error ellipsoids, a_1, b_1, c_1, a_2, b_2 and c_2 in both Equations (4.14) and (4.15) are 148, 148, 148, 66, 66 and 66 respectively; parameters $d_{1,1}$, $e_{1,1}$, $f_{1,1}$, $d_{1,2}$, $e_{1,2}$ and $f_{1,2}$ are zeros. In the uniform case, the mean of the 'sub-mean' of the discrepant area is 62627.0m^2 and the 95% confidence interval for mean of the 'sub-mean' of the discrepant area is in the range of 46903.3m^2 and 78350.7m^2. In the normal case, the mean is 49275.4m^2 and the 95% confidence interval is [36903.8m^2, 61647.0m^2]. As the result obtained in the 2D problem, the discrepant area of the line segment in the normal case is relatively smaller than that in the uniform case.

In the previous 2D model, the parameters of the nodal error ellipses may affect the reliability of a line segment. Table 4.2 shows the mean of the 'sub-mean' of the discrepant area and the 95% confidence interval for mean of the 'sub-mean' of the discrepant area under different values for the parameters of the error ellipsoids in both the uniform case and the normal case. The first six columns in Table 4.2 record values for the parameters, a_1, b_1, c_1, a_2, b_2 and c_2. The next two columns tabulate the

mean and the 95% confidence interval under the assumption of uniformly distributed nodal error and the last two columns tabulate the mean and the confidence interval under the assumption of normally distributed nodal error. Similar to the previous 2D study, the values of parameter a_1, b_1, c_1, a_2, b_2 and c_2 have the identical average in all rows. Also, the average of the first three columns is 148m and this of the next three columns 66m. As a result, for a positive integer i, parameter c_i will increase if parameter a_i increases and parameter b_i is a constant; parameter b_i will decrease if parameter a_i increases and parameter c_i are fixed, and so on.

It is noticed that when only one of parameters b_i and c_i keeps constant and a_i increases, the average discrepant area decreases. While parameter a_i is fixed and either parameter b_i or c_i varies the average discrepant area does not change significantly. The expected line segment and the measured line segment generated by the simulation estimate the average discrepant area. Due to the unchanged expected line segment, the average discrepant area depends on the measured line segment. According to the parameters of the two error ellipsoids the measured line segment will be generated. Then, parameters a_i, b_i and c_i may affect the average discrepant area of the line segment. Figure 4.16 illustrates how the measured line segment affects the average discrepant area.

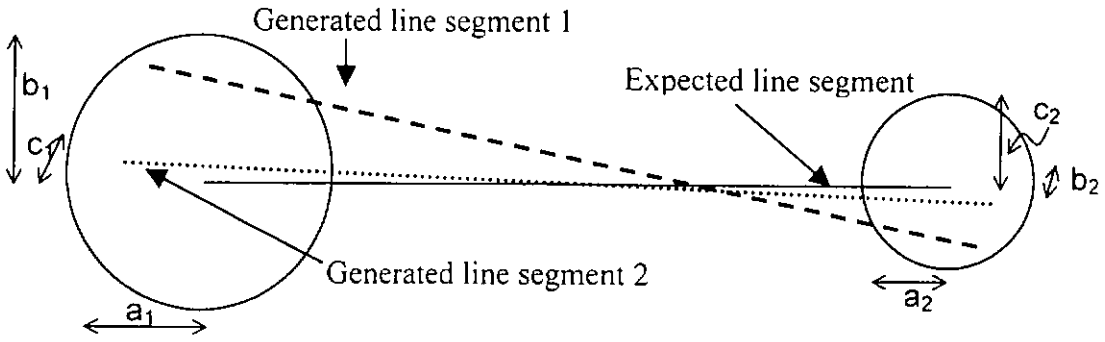


Figure 4.16. The effect of the measured line segment on the discrepancy.

Table 4.2. Discrepant area of a line segment under different parameters of error ellipsoids in a uniform case and in a normal case.

Parameters of two error ellipsoids in m						Discrepancy in a uniform case		Discrepancy in a normal case	
a_1	b_1	c_1	a_2	b_2	c_2	Mean	95% confidence interval	Mean	95% confidence interval
148	148	148	66	66	66	62627.0	[46903.3, 78350.7]	49275.4	[36903.8, 61647.0]
100	196	148	66	66	66	69564.5	[52098.9, 87030.1]	54619.5	[40906.1, 68332.9]
196	100	148	66	66	66	56424.7	[42257.9, 70591.5]	44379.2	[33236.6, 55521.8]
100	148	196	66	66	66	69539.2	[52080.1, 86998.4]	54663.8	[40939.3, 68388.3]
148	196	100	66	66	66	63844.0	[47814.5, 79873.6]	50232.1	[37619.9, 62844.3]
148	100	196	66	66	66	63888.0	[47847.6, 79928.4]	50193.2	[37591.0, 62795.5]
196	148	100	66	66	66	56396.3	[42236.5, 70556.2]	44347.8	[33213.2, 55482.4]
148	148	148	30	78	90	68789.9	[51516.4, 86063.4]	53847.8	[40327.8, 67367.8]
148	148	148	30	90	78	68702.0	[51452.6, 85951.4]	53887.4	[40357.9, 67417.0]
148	148	148	78	30	90	62154.1	[46548.8, 77759.4]	48806.7	[36552.5, 61060.9]
148	148	148	90	30	78	59741.4	[44741.8, 74740.9]	46956.6	[35167.2, 58746.0]
148	148	148	78	90	30	62164.3	[46555.9, 77772.7]	48875.6	[36603.8, 61147.4]
148	148	148	90	78	30	59866.2	[44835.4, 74897.0]	46894.3	[35120.5, 58668.1]

The solid line in Figure 4.16 represents the true location of the line segment. The dotted and the dash line segments represent the two generated line segments under the same assumption of the nodal errors: the generated line segment 1 and the generated line segment 2. Obviously, the discrepancy between the generated line segment 1 and the expected line segment is larger than that between the generated line segment 2 and the expected line segment. Moreover, the larger parameter a_i, the smaller the discrepant area, due to the fact that more generated line segments are near the expected line segment when parameter a_i is large.

For the example of a polyline (linear feature), the three expected nodes are (0, 0, 0), (50, 500, 707.1) and (1500, 500, 707.1). The parameters of nodal error ellipsoids a_1, b_1, c_1, a_2, b_2, c_2, a_3, b_3 and c_3 are 100m, 196m, 148m, 30m, 78m, 90m, 100m, 196m and 148m respectively. In the uniform case, the mean of the 'sub-mean' of the discrepant area and the 95% confidence interval for mean of the 'sub-mean' of the discrepant area is 113475.0m^2 and [84985.1m^2, 141964.9m^2] respectively. In the normal case, the mean and the 95% confidence interval are 89036.1m^2 and [66681.9m^2, 111390.2m^2] respectively.

The three nodes of an areal feature are chosen as that in the example of the linear feature. The mean and the 95% confidence interval are 2.3×10^7 m^3 and [1.8×10^7 m^3, 3.0×10^7 m^3] in the uniform case. In the normal case, the mean and the 95% confidence interval are 1.9×10^7 m^3 and [1.4×10^7 m^3, 2.3×10^7 m^3].

For the example of a volumetric feature, an additional node (to the existing three) is considered. This addition now specifies a volumetric feature. This additional node is (500, 707, 500) and its error ellipsoid has parameters a_4 = 30m, b_4 = 78m and c_4 = 90m. The mean is 5.7×10^7 m^3 and the 95% confidence interval is [4.2×10^7 m^3, 7.1×10^7 m^3] in the uniform case; the mean is 4.4×10^7 m^3 and the confidence interval is [3.3×10^7 m^3, 5.5×10^7 m^3] in the normal case.

4.3 Summary

The simulation model has been proposed to model the reliability of 2D and 3D spatial features. Since this model is a further development of the existing simulation-based models, a comparison between the proposed model and the existing models are made. It is concluded that the existing models may require some adjustments. In the 2D and 3D reliability models, some similar findings are obtained. It has been observed that the error ellipse model is required instead of the error circle model. The size and shape of the error ellipse of a node affects the discrepancy of a spatial feature. Furthermore, the correlated nodal error is significant in the reliability model. Correlation combination of the nodal error varies the discrepancy of the spatial feature. Also, the distribution of the nodal error affects the reliability in a certain extent. Usually, the discrepancy in the uniform case is greater than that in the normal case.

CHAPTER 5
RELIABILITY OF SPATIAL FEATURES BY ANALYTICAL METHOD

A simulation model has been developed in the previous chapter to investigate the reliability of a spatial feature in either 2D or 3D GIS. However, one of the weaknesses of the simulation techniques is that they are time-consuming. Stanfel (1996) suggested that a stochastic method could be used to approximate the discrepant area of a spatial feature. It was also pointed out that accuracy and speed of convergence should be further considered.

An alternative model will be proposed to describe the reliability of spatial features using numerical analysis in this chapter, while the reliability and the discrepancy of the spatial features are defined as in the previous chapter. This numerical model improves upon the past body of work by investigating the analytical method with a numerical solution. In the following, the numerical model of the reliability of spatial features including linear and areal features in 2D GIS will be demonstrated. Next, the numerical model of the reliability of spatial features including linear, areal and volumetric features in 3D GIS will be discussed. Finally, numerical results will be compared with simulated results in describing reliability of spatial feature in GIS.

5.1 Spatial Features in 2D GIS

5.1.1 Linear Features

It is mentioned that a line segment is the fundamental unit of spatial features in the previous chapter. The numerical model for the reliability of a line segment is derived, and then that of a linear feature and an areal feature is discussed later on.

The discrepancy of a line segment is in either the two-triangle or quadrangle case (see Figure 4.1 and Figure 4.2). The nodes of a spatial feature have been generated in simulation approaches and then joining the nodes together has created the so-called generated measured spatial feature, which has been further used to determine the case of the discrepancy. This simulation model considers the expected discrepant area of the line segment in a converse approach. The first step of the numerical model concerns the case of the discrepancy, mainly due to the fact that the expected discrepant area is expressed as integral.

In order to describe all of these possible cases under the assumption, integration is implemented to calculate the discrepant area of a line segment. From statistical theory, the expected discrepant area in the two-triangle case is as follows

$$E(A_1) = \iint_{D_1 \cup D_2} \iint f(x_1, y_1, x_2, y_2) \times A_1(x_1, y_1, x_2, y_2) dx_1 dy_1 dx_2 dy_2 \qquad (5.1)$$

where D_1 and D_2 are domains of (x_1, y_1) and (x_2, y_2) such that the two-triangle case occurs;

$f(x_1, y_1, x_2, y_2)$ is a joint probability density function of four random variables X_1, Y_1, X_2 and Y_2; and

$A_1(x_1, y_1, x_2, y_2)$ is the discrepant area in the two-triangle case (see Equation (4.2)).

Similarly, in the quadrangle case, the expected discrepant area can be calculated as follows

$$E(A_2) = \iint_{D_1 \cup D_2} \iint f(x_1, y_1, x_2, y_2) \times A_2(x_1, y_1, x_2, y_2) dx_1 dy_1 dx_2 dy_2 \qquad (5.2)$$

where D_1 and D_2 are domains of (x_1, y_1) and (x_2, y_2) such that the quadrangle case occurs;

$f(x_1, y_1, x_2, y_2)$ is a joint probability density function of four random variables X_1, Y_1, X_2 and Y_2; and

$A_2(x_1, y_1, x_2, y_2)$ is the discrepant area in the quadrangle case (see Equation (4.3)).

The domain for the above two integrals is elaborated here. The solid line segment in Figure 5.1 can be expressed as $x - \mu_{x_1} = \frac{\mu_{x_2} - \mu_{x_1}}{\mu_{y_2} - \mu_{y_1}}(y - \mu_{y_1})$ if $\mu_{y_1} \neq \mu_{y_2}$ and $y = \mu_{y_1}$ if $\mu_{y_1} = \mu_{y_2}$ where (x, y) is an arbitrary point on the solid line segment. The domains in the two-triangle case and in the quadrangle case can be modified. In Figure 5.1, the two nodal error ellipses are divided into two parts by the expected line segment. If the measured location of the left-hand sided node of the line segment lies insides region A and the measured location of the right-hand sided node lies insides region D, these two measured nodes will form two triangles with the expected line segment. This represents the two-triangle case. If these two measured endpoints lie inside region B and region C, it remains the two-triangle case. Thus, there are two possible domains of the integral for the two-triangle case. Similarly, two possible domains of the integral also exist for the quadrangle case. The first case occurs if the two measured nodes are inside regions A and C. The second case occurs if the two measured nodes are inside regions B and D.

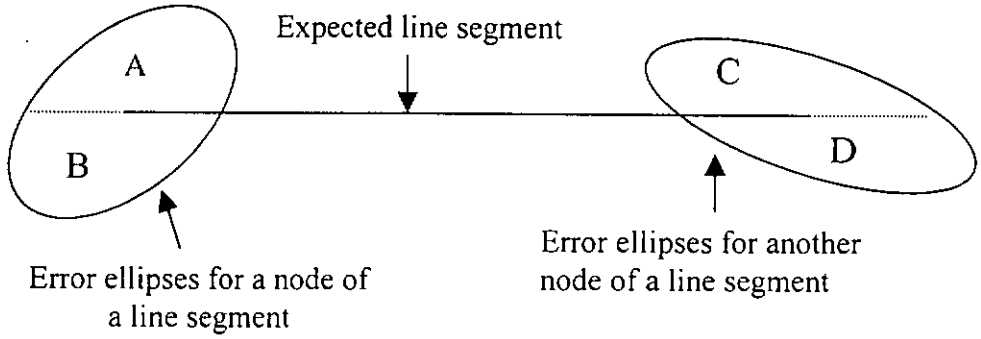


Figure 5.1. Domain in the integral of a line segment.

For the two-triangle case, Equation (5.1) can be modified as

$$E(A_1) = \iint_{A \cup D} \iint f(x_1, y_1, x_2, y_2) \times A_1(x_1, y_1, x_2, y_2) dx_1 dy_1 dx_2 dy_2 + \iint_{B \cup C} \iint f(x_1, y_1, x_2, y_2) \times A_1(x_1, y_1, x_2, y_2) dx_1 dy_1 dx_2 dy_2 \quad (5.3)$$

For the quadrangle case, Equation (5.2) can be modified as

$$E(A_2) = \iint_{A\cup C}\iint f(x_1,y_1,x_2,y_2)\times A_2(x_1,y_1,x_2,y_2)dx_1dy_1dx_2dy_2 + \iint_{B\cup D}\iint f(x_1,y_1,x_2,y_2)\times A_2(x_1,y_1,x_2,y_2)dx_1dy_1dx_2dy_2 \tag{5.4}$$

The overall expected discrepant area E is the sum of the expected discrepant area in both the two-triangle case and the quadrangle case. That is,

$$E = E(A_1) + E(A_2) \tag{5.5}$$

Therefore, given the joint probability density function f, E can be computed. The function f is represented by distributions of the nodal errors (of the line segment). Two feasible assumptions are made. The first is uniformly distributed nodal error; the second is normally distributed nodal error.

Assume that both nodes of a line segment are bivariate uniformly distributed within their corresponding error ellipses whose equation is given by Equation (4.4) where two nodes are uncorrelated. A joint probability density function of X_1, Y_1, X_2, and Y_2 is

$$\begin{aligned} f(x_1,y_1,x_2,y_2) &= f(x_1,y_1)\times f(x_2,y_2) \\ &= \frac{1}{\text{area of the error ellipse for } (\mu_{x_1},\mu_{y_1})} \\ &\quad \times \frac{1}{\text{area of the error ellipse for } (\mu_{x_2},\mu_{y_2})} \\ &= \frac{2\pi a_1 b_1 d_1 e_1}{\sqrt{4d_1^2e_1^2 - a_1^2b_1^2c_1^2}} \times \frac{2\pi a_2 b_2 d_2 e_2}{\sqrt{4d_2^2e_2^2 - a_2^2b_2^2c_2^2}} \end{aligned} \tag{5.6}$$

if (x_1, y_1) and (x_2, y_2) are in their own error ellipses.

$f(x_1,y_1,x_2,y_2) = 0$, elsewhere.

Another feasible assumption is that both nodes of a line segment are bivariate normally distributed "within" their corresponding error ellipses whose equation is

given by Equation (4.6). A joint probability density function of four random variables X_1, Y_1, X_2, and Y_2 is

$$f(x_1, y_1, x_2, y_2) = \frac{1}{(2\pi)^2 |\Sigma|^{1/2}} \exp\left(\frac{-1}{2} \mu^T \Sigma^{-1} \mu\right) \tag{5.7}$$

where mean vector $\mu = \left(x_1 - \mu_{x_1}, y_1 - \mu_{y_1}, x_2 - \mu_{x_2}, y_2 - \mu_{y_2}\right)^T$;

covariance matrix $\Sigma = \begin{pmatrix} \sigma_{x_1x_1} & \sigma_{x_1y_1} & \sigma_{x_1x_2} & \sigma_{x_1y_2} \\ \sigma_{x_1y_1} & \sigma_{y_1y_1} & \sigma_{x_2y_1} & \sigma_{y_1y_2} \\ \sigma_{x_1x_2} & \sigma_{x_2y_1} & \sigma_{x_2x_2} & \sigma_{x_2y_2} \\ \sigma_{x_1y_2} & \sigma_{y_1y_2} & \sigma_{x_2y_2} & \sigma_{y_2y_2} \end{pmatrix}$; and

$\sigma_{x_ix_j}$, $\sigma_{x_iy_j}$, and $\sigma_{y_iy_j}$ are sample co-variances of X_i's and X_j's errors, X_i's and Y_j's errors, and Y_i and Y_j's errors respectively.

After function f is determined, the numerical integration will be implemented to calculate the multiple integral in Equations (5.3) and (5.4) because the integral may not be solved by finding antiderivatives. Multiple quadrature rules are the traditional approach to solve complex integration problems. For example, in an integral $\int_a^b g(x)dx$ where a and b are constants, the Gaussian quadrature approximates the integral by integrating the linear function that joins some of its points on the graph (Burden and Faires, 1993). In Figure 5.2, the solid curve is the graph of the function g. The first dash line segment is the line passing via $(x_1, g(x_1))$ parallel to the x-axis and the second one is the dash line segment passing via $(x_2, g(x_2))$ also parallel to the x-axis. The nodes $x_1, x_2, \ldots, x_n$ in the interval [a,b] and coefficients $c_1, c_2, \ldots, c_n$ are chosen to minimize the expected error obtained in performing the approximation $\int_a^b g(x)dx \approx \sum_{i=1}^{n} c_i g(x_i)$ for an arbitrary function g.

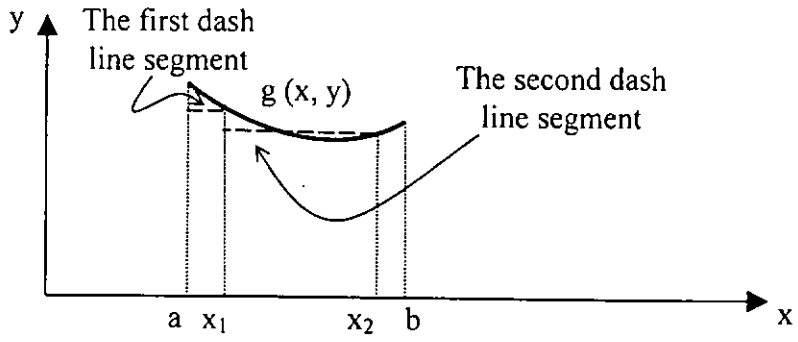


Figure 5.2. The Gaussian quadrature.

This technique can be modified in a straightforward manner for use in the approximation of multiple integrals. To apply the Gaussian quadrature to $\int_a^b \int_{c(x)}^{d(x)} g(x,y)\,dydx$, the integrals must first be translated. For each x in [a, b], the interval spans from [c(x), d(x)] to [-1, 1]. The results are expressed by the following formula

$$\int_a^b \int_{c(x)}^{d(x)} g(x,y)\,dydx \approx \int_a^b \left(\frac{d(x)-c(x)}{2} \sum_{j=1}^{n} c_{n,j}\, g\left(x, \frac{(d(x)-c(x))r_{n,j} + d(x) + c(x)}{2} \right) \right) dx \tag{5.8}$$

where the root $r_{n,j}$ and coefficients $c_{n,j}$ are constants (Stroud and Secrest, 1966).

In Figure 5.3, the domain of the double integral is divided into four parts. Ranges a and b in the x-direction are assumed to be divided into two parts. Similarly, ranges c(x) and d(x) in the y-direction are also divided into two parts (i.e. n = 2 in Equation (5.8)). As a result, the domain of the integral has four parts. The integral over each part can be calculated, then by summing up these values of the integral with their corresponding weights, the solution to the double integral g(x,y) is found.

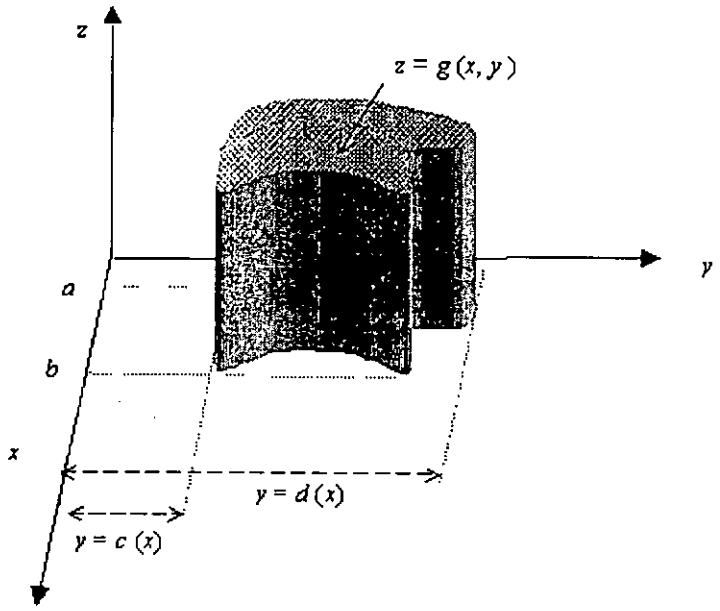


Figure 5.3. The domain of the double integral.

Algorithm 5.1 is used to compute:

$$\int_{a}^{b}\int_{c_1(x_1)}^{d_1(x_1)}\int_{c_2(x_1,y_1)}^{d_2(x_1,y_1)}\int_{c_3(x_1,y_1,x_2)}^{d_3(x_1,y_1,x_2)} g(x_1, y_1, x_2, y_2)\,dy_2\,dx_2\,dy_1\,dx_1 .$$

In the single integral, the x range is divided into n classes. In this multiple integral, ranges for x_1, y_1, x_2 and y_2 are divided into m, n, no and p classes respectively.

Algorithm 5.1

Step 1 Input ranges a and b; positive integers m, n, no, p (assume that the roots $r_{i,j}$ and coefficients $c_{i,j}$ are available for i equals m, n, no, p and for $1 \le j \le i$).

Step 2 Set $h_1 = (b - a) / 2$; $h_2 = (b + a) / 2$; $J = 0$.

Step 3 For i = 1, 2, ..., m do Steps 4-12.

Step 4 Set JX = 0;

$x = h_1 r_{m,i} + h_2$;

$d_{11} = d_1(x)$;

$c_{11} = c_1(x)$;

$k_1 = (d_{11} - c_{11}) / 2$;

$k_2 = (d_{11} + c_{11}) / 2$.

Step 5 For j = 1, 2, …, n do steps 6-11.

Step 6 Set JY = 0;

$y = k_1 r_{n,j} + k_2$;

$d_{21} = d_2(x, y)$;

$c_{21} = c_2(x, y)$;

$l_1 = (d_{21} - c_{21}) / 2$;

$l_2 = (d_{21} + c_{21}) / 2$.

Step 7 For kk = 1, 2, …, no do steps 8 – 10.

Step 8 Set JZ = 0;

$z = l_1 r_{no, kk} + l_2$;

$d_{31} = d_3 (x, y, z)$;

$c_{31} = c_3 (x, y, z)$;

$m_1 = (d_{31} - c_{31}) / 2$;

$m_2 = (d_{31} + c_{31}) / 2$.

Step 9 For k = 1, 2, …, p do

Set $w = m_1 r_{p, k} + m_2$;

$Q = g (x, y, z, w)$;

$JZ = JZ + c_{p, k}\ Q$.

Step 10 Set $JY = JY + c_{no, kk}\ JZ\ m_1$.

Step 11 Set $JX = JX + c_{n, j}\ JY\ l_1$.

Step 12 Set $J = J + c_{m, i}\ JX\ k_1$.

Step 13 Set $J = h_1\ J$.

Step 14 Output J.

Replacing x, y, z and w with x_1, y_1, x_2 and y_2 respectively, the overall expected discrepant area of a line segment can be computed.

In the general event of a linear feature in which a line segment is a special one, the average discrepant area of the linear feature composed of more than one line segment can be expressed as below.

$$E = \iint_{D_1 \cup \cdots \cup D_{NP}} \cdots \iint f(x_1, y_1, \cdots, x_{NP}, y_{NP}) \times A \, dx_1 dy_1 \cdots dx_{NP} dy_{NP} \tag{5.9}$$

where NP is the total number of the nodes of the linear feature;

$D_1, D_2, \ldots, D_{NP}$ are NP error ellipses for the NP expected nodes of the linear feature;

f is a joint probability density function of 2 x NP random variables X_1, Y_1, X_2, Y_2, ..., X_{NP} and Y_{NP}; and

A is the linear feature's discrepant area defined as Chapter 3.

This integral can be numerically solved by the Gaussian quadrature. Thus, the first step of the proposed numerical model is to define a domain for this multiple integral and express A in terms of co-ordinates of nodes.

For the line segment, there are two possible cases to calculate the discrepant area: the two-triangle case and the quadrangle case. However, for a linear feature, the discrepant area is more complex. It may be the two-triangle case, the quadrangle case or with both cases simultaneously. The following is an example to calculate the discrepant area of a linear feature of three expected nodes.

Suppose a linear feature has three expected nodes (μ_{x_1}, μ_{y_1}), (μ_{x_2}, μ_{y_2}) and (μ_{x_3}, μ_{y_3}). There are three error ellipses for the three nodes and these are centered at the corresponding expected nodes. Next, the domain of the integral is considered. Referring to Figure 5.4, the first error ellipse (on the left-hand side) is divided into two regions: A and B. Linking the left-hand sided node and the middle node forms the first line segment, and the second line segment is formed by linking the middle and the right-hand sided nodes. These two line segments and their extensions divide the first error ellipse into regions A and B; the second error ellipse (in the middle) into regions C and D; and the third error ellipse (on the right-hand side) into regions E and F.

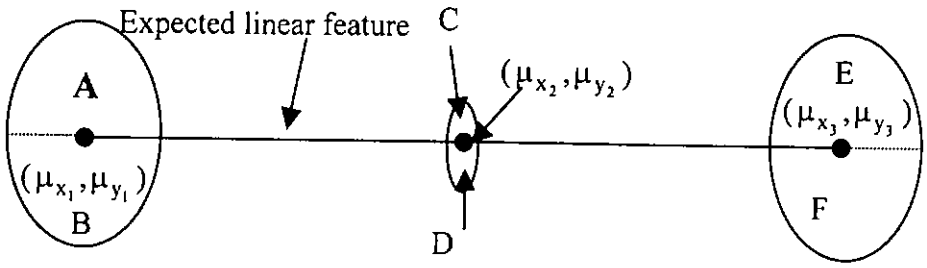


Figure 5.4. A linear feature of three nodes.

A measured location of the first node (x_1, y_1) may be within regions A or B. A measured location of the second node (x_2, y_2) may be within regions C and D. Similarly, the measured location of the third node (x_3, y_3) may be within regions E or F. Therefore, a measured linear feature can be within regions, for example, (A, C, E) or (A, C, F). There are eight combinations of the discrepancy and hence Equation (5.9) can be modified as follows

The expected discrepant area of the linear feature of three nodes E
$$= E_1 + E_2 + E_3 + E_4 + E_5 + E_6 + E_7 + E_8 \tag{5.10}$$

where $E_1 = \underset{A \cup C \cup E}{\int \cdots \int} f \times A_7 \, dx_1 dy_1 \cdots dx_{NP} dy_{NP}$;

$E_2 = \underset{B \cup D \cup F}{\int \cdots \int} f \times A_7 \, dx_1 dy_1 \cdots dx_{NP} dy_{NP}$;

$E_3 = \underset{A \cup D \cup F}{\int \cdots \int} f \times A_4 \, dx_1 dy_1 \cdots dx_{NP} dy_{NP}$;

$E_4 = \underset{B \cup C \cup E}{\int \cdots \int} f \times A_4 \, dx_1 dy_1 \cdots dx_{NP} dy_{NP}$;

$E_5 = \underset{A \cup C \cup F}{\int \cdots \int} f \times A_6 \, dx_1 dy_1 \cdots dx_{NP} dy_{NP}$;

$E_6 = \underset{B \cup D \cup E}{\int \cdots \int} f \times A_6 \, dx_1 dy_1 \cdots dx_{NP} dy_{NP}$;

$E_7 = \underset{A \cup D \cup E}{\int \cdots \int} f \times A_5 \, dx_1 dy_1 \cdots dx_{NP} dy_{NP}$; and

$E_8 = \underset{B \cup C \cup F}{\int \cdots \int} f \times A_5 \, dx_1 dy_1 \cdots dx_{NP} dy_{NP}$.

5.1.2 Areal Features

The discrepancy of an areal feature is defined in the previous chapter (see Figure 4.4) and the expected discrepant area of the areal feature can be expressed as

$$E = \underset{D_1 \cup \cdots \cup D_{NP}}{\iint\iint\cdots\iint} f(x_1, y_1, \cdots, x_{NP}, y_{NP}) \times A \, dx_1 dy_1 \cdots dx_{NP} dy_{NP} \tag{5.11}$$

where NP is the total number of nodes of the areal feature;

$D_1, D_2, \ldots, D_{NP}$ are NP error ellipse for the NP expected nodes of the areal feature;

f is a joint probability density function of 2 x NP random variables X_1, Y_1, X_2, Y_2, ..., X_{NP} and Y_{NP}; and

A is the areal feature's discrepant area defined as Chapter 3.

This equation seems to be same as Equation (5.9). In fact, f in both equations is multivariate joint probability function of X_1, Y_1, ..., X_{NP}, and Y_{NP}; but the only difference is the meaning of A. A refers to mathematical expression of the discrepant area of a linear feature in Equation (5.9) while it refers to mathematical expression of the discrepant area of an areal feature in Equation (5.11); these two mathematical expressions are distinct.

5.2 Spatial Features in 3D GIS

For the discrepancy of a line segment in 3D GIS, it is noticed that three possible cases exist in the previous chapter: (a) the measured locations of all nodes and their corresponding expected locations are on a flat plane given that the measured and the expected line segments do not intersect; (b) the measured and the expected line segment intersect; (c) neither case (a) or case (b) is a possibility. The discrepant area of a line segment in the first case is represented by A_8; that in the

second case is represented by A_9; the approximate discrepant area in the last case is represented by the average of A_{10} and A_{11}, mainly in order to simplify the complexity of the discrepant area in the third case. As a result, the expected discrepant area of the line segment is computed as per Equation (5.12).

$$E = \int_{D_1 \cup D_2} \cdots \int f \times A_8 dz_2 dy_2 dx_2 dz_1 dy_1 dx_1 + \int_{D_3 \cup D_4} \cdots \int f \times A_9 dz_2 dy_2 dx_2 dz_1 dy_1 dx_1 + \int_{D_5 \cup D_6} \cdots \int f \times \frac{(A_{10} + A_{11})}{2} dz_2 dy_2 dx_2 dz_1 dy_1 dx_1 \quad (5.12)$$

where D_1 and D_2 are the regions of the two ellipsoids given that case (a) occurs;
D_3 and D_4 are the regions of the two ellipsoids given that case (b) occurs;
D_5 and D_6 are the regions of the two ellipsoids given that case (c) occurs;
U is union of the two regions; and
f is a multivariate probability density function.

Practically, the three domains of the above three multiple integrals cannot be defined easily and thus a simplification is needed. The first and the last integrals on the right-hand side of Equation (5.12) should have the same integrand because the area of the quadrangle A_8 can be estimated by dividing the quadrangle into two triangles and then summing areas of these two triangles together. Moreover, the interval of the second integration on the right-hand side of Equation (5.12) is ambiguous. Its integrand is further approximated by 0.5 f x $(A_{10} + A_{11})$ and this approximation is larger than its exact value. Then, Equation (5.12) is modified as Equation (5.13).

$$E = \int_{D_1 \cup D_2} \cdots \int f \times \frac{A_{10} + A_{11}}{2} dz_2 dy_2 dx_2 dz_1 dy_1 dx_1 \quad (5.13)$$

where D_1 and D_2 are the regions of the two error ellipsoids.

In the uniform case, the mathematical expression of f is shown below

$$f(x_1, y_1, z_1, x_2, y_2, z_2) = f(x_1, y_1, z_1) \times f(x_2, y_2, z_2)$$
$$= \frac{1}{\textit{area of the error ellipsoid for } (\mu_{x_1}, \mu_{y_1}, \mu_{z_1})} \times \frac{1}{\textit{area of the error ellipsoid for } (\mu_{x_2}, \mu_{y_2}, \mu_{z_2})} \quad (5.14)$$

if (x_1, y_1, z_1) and (x_2, y_2, z_2) are in their error ellipsoids

$f(x_1, y_1, z_1, x_2, y_2, z_2) = 0$, otherwise

The joint probability density function of X_1, Y_1, Z_1, X_2, Y_2 and Z_2 in the normal case is

$$f(x_1, y_1, z_1, x_2, y_2, z_2) = \frac{1}{(2\pi)^{6/2}|\Sigma|^{1/2}} \exp\left(\frac{-1}{2}\mu^T \Sigma^{-1} \mu\right) \quad (5.15)$$

where $\mu = (x_1 - \mu_{x_1}, y_1 - \mu_{y_1}, z_1 - \mu_{z_1}, x_2 - \mu_{x_2}, y_2 - \mu_{y_2}, z_2 - \mu_{z_2})^T$;

$$\Sigma = \begin{pmatrix} \sigma_{x_1x_1} & \sigma_{x_1y_1} & \sigma_{x_1z_1} & \sigma_{x_1x_2} & \sigma_{x_1y_2} & \sigma_{x_1z_2} \\ \sigma_{x_1y_1} & \sigma_{y_1y_1} & \sigma_{y_1z_1} & \sigma_{x_2y_1} & \sigma_{y_1y_2} & \sigma_{y_1z_2} \\ \sigma_{x_1z_1} & \sigma_{y_1z_1} & \sigma_{z_1z_1} & \sigma_{x_2z_1} & \sigma_{y_2z_1} & \sigma_{z_1z_2} \\ \sigma_{x_1x_2} & \sigma_{x_2y_1} & \sigma_{x_2z_1} & \sigma_{x_2x_2} & \sigma_{x_2y_2} & \sigma_{x_2z_2} \\ \sigma_{x_1y_2} & \sigma_{y_1y_2} & \sigma_{y_2z_1} & \sigma_{x_2y_2} & \sigma_{y_2y_2} & \sigma_{y_2z_2} \\ \sigma_{x_1z_2} & \sigma_{y_1z_2} & \sigma_{z_1z_2} & \sigma_{x_2z_2} & \sigma_{y_2z_2} & \sigma_{z_2z_2} \end{pmatrix}$$; and

$\sigma_{x_ix_j}$, $\sigma_{x_iy_i}$ and $\sigma_{y_iy_j}$ are sample co-variances of x_i's and x_j's errors, x_i's and y_j's errors, and y_i's and y_j's errors respectively.

The integral in Equation (5.12) can be solved numerically using the Gaussian quadrature numerical integration. Consider the discrepant area of a linear feature. The expected discrepant area is presented in Equation (5.16).

$$E = \int_{D_1 \cup \cdots \cup D_{NP}} \cdots \int f(x_1, y_1, z_1, \ldots, x_{NP}, y_{NP}, z_{NP}) \times A \, dz_{NP} dy_{NP} dx_{NP} \cdots dz_1 dy_1 dx_1 \quad (5.16)$$

where NP is the total number of nodes on the linear feature;

$D_1, D_2, \ldots, D_{NP}$ are regions of the NP error ellipsoids for the NP nodes of the linear feature;

U is union of regions;

f is a multivariate joint probability density function; and

A is the discrepant area of the linear feature.

Similarly, the expected discrepant areas of an areal feature and of a volumetric feature are calculated by Equation (5.17).

$$E = \int_{D_1 \cup \cdots \cup D_{NP}} \cdots \int f(x_1, y_1, z_1, \cdots, x_{NP}, y_{NP}, z_{NP}) \times V \, dz_{NP} dy_{NP} dx_{NP} \ldots dz_1 dy_1 dx_1 \quad (5.17)$$

where NP is the total number of nodes on the linear feature;

$D_1, D_2, \ldots, D_{NP}$ are regions of the NP error ellipsoids for the NP nodes of the areal (or volumetric) feature;

U is union of domains;

f is a multivariate joint probability density function; and

V is the discrepant volume of the areal (or volumetric) feature.

5.3 Case Study

The following example calculates the expected discrepant area by the Gaussian quadrature in 2D GIS. Tables 5.1 and 5.2 illustrate the discrepant area of the line segment joined to two nodes (0,0) and (1000,0) of independent error in both the uniform and the normal cases respectively. The first four columns of both tables record values of the parameters of the error ellipse equations. The remaining parameters of the error ellipse equation are set to zero. The next three columns are the expected discrepant areas computed by the Gaussian quadrature with different

values of parameter m, where m = n = no = p. And the last column records the simulated results.

Table 5.1. Expected discrepant area of a line segment calculated from numerical integration and simulation in an independent uniform case.

a_1	b_1	a_2	b_2	Numerical result			Simulated result
				m = 5	m = 10	m = 15	
100	196	30	78	37469.9	55860.8	55846.1	48442.7
100	196	54	54	32360.4	48785.6	48761.7	45900.1
100	196	78	30	27250.9	41675.3	41643.2	43248.2
148	148	30	78	32360.4	47715.0	47712.9	39260.7
148	148	54	54	27250.9	40679.2	40667.1	35987.2
148	148	78	30	22141.3	33608.4	33587.3	33145.8
196	100	30	78	27250.9	39569.1	39597.6	30379.6
196	100	54	54	22141.3	32572.8	32572.6	26669.4
196	100	78	30	17031.8	25541.5	25531.5	23480.7

Table 5.2. Expected discrepant area of a line segment calculated from numerical integration and simulation in an independent normal case.

a_1	b_1	a_2	b_2	Numerical result			Simulated result
				m = 5	m = 10	m = 15	
100	196	30	78	31750.4	32296.7	32342.2	34977.7
100	196	54	54	29787.1	30410.4	30448.0	32857.7
100	196	78	30	28158.2	28847.9	28878.2	30902.0
148	148	30	78	25719.4	26091.6	26132.2	28513.2
148	148	54	54	23555.0	24021.2	24052.3	25950.2
148	148	78	30	21722.4	22254.6	22278.1	23835.9
196	100	30	78	19931.0	20064.2	20103.1	22006.6
196	100	54	54	17502.0	17770.5	17797.6	19265.8
196	100	78	30	15386.8	15746.1	15763.9	16998.1

From the above two tables, the numerical results converge as parameter m increases. It is a fact that a numerical result is an approximation of the expected discrepant area. The accuracy of the approximation is related to the number of the partitions in the domain. The domain of the integral has been divided into 5^4, 10^4 or

15^4 parts in the above examples. A difference between the numerical and the exact results can be minimized if the domain of the integral is divided into more parts. Theoretically, parameter m should be chosen as large as possible to obtain the convergent result but the larger m, the more computing time. Parameter m is set to be 10 after both aspects have been taken into account.

For the example of the linear feature composed of three nodes, the three expected nodes are (0, 0), (1000, 0) and (1500, 866). Error ellipse parameters a_1, b_1, a_2, b_2, a_3 and b_3 are 100m, 196m, 30m, 78m, 100m and 196m respectively. The covariance matrix, in the probability density function of the multivariate normal distribution f, is a 6x6 diagonal matrix with uncorrelated errors of nodes (x_1, y_1), (x_2, y_2), (x_3 and y_3) in the normal case while the confidence coefficient (1-α) 100% is 0.95. The expected discrepant areas of the line segment are 73871.0m^2 in the uniform case and 59536.5m^2 in the normal case, where m = n = no = p = 10

For the areal feature, the three expected nodes are chosen as that in the example of the linear feature. Its expected discrepant areas are $1.9x10^5$ m^2 in the uniform case and $1.3x10^5$ m^2 in the normal case.

The numerical model on the reliability of 3D spatial features is also applied to the example data of the proposed simulation model. The two expected nodes of the line segment are (0, 0, 0) and (1000, 0, 0). Error ellipsoid parameters a_1, b_1, c_1, a_2, b_2 and c_2 are 100m, 196m, 148m, 30m, 78m and 90m respectively. In the normal case, the covariance matrix, in the probability density function of the multivariate normal distribution f, is a 6x6 diagonal matrix with uncorrelated errors of nodes (x_1, y_1, z_1) and (x_2, y_2, z_2). The confidence coefficient (1-α) 100% is 0.95. The expected discrepant areas of the line segment are $9.2x10^4$ m^2 in the uniform case and $6.2x10^4$ m^2 in the normal case.

For the 3D linear feature composed of three nodes, the three expected nodes are (0, 0, 0), (500, 500, 707.1) and (1500, 500, 707.1). Error ellipsoid parameters a_1, b_1, c_1, a_2, b_2, c_2, a_3, b_3 and c_3 are 100m, 196m, 148m, 30m, 78m, 90m, 100m, 196m

and 148m respectively. The covariance matrix in the joint probability density function of the multivariate normal distribution f is a 9x9 diagonal matrix. The confidence coefficient (1-α) 100% is 0.95. The expected discrepant areas of the linear feature are $1.2x10^5$ m^2 in the uniform case and $8.6x10^4 m^2$ in the normal case. Using the three expected nodes of the 3D linear feature for the 3D areal feature results in expected discrepant volumes $2.3x10^7$ m^3 in the uniform case and $1.5x10^7$ m^3 in the normal case.

For the example of a 3D volumetric feature, an additional node (to the existing three) is considered. This addition now specifies a volumetric feature. This additional node is (500, 707.1, 500), and its error ellipsoid has parameters a_4 = 30m, b_4 = 78m and c_4 = 90m. The expected discrepant volume of the 3D volumetric feature in the uniform case is $6.4x10^7$ m^3 and that in the normal case is $3.8x10^7$ m^3.

5.4 Comparison between the numerical and the simulated results

Since the accuracy of the simulated result and the numerical result are unknown, their results are compared to check their accuracy. In the 2D examples, both their difference in results and the ratio of the numerical results to simulated results are tabulated in Table 5.3 and Table 5.4. In an ideal case (i.e. two methods provide the same value of the expected discrepant area), the ratio should be equal to one and the difference should be equal to zero.

Table 5.3. Comparison of the numerical results with the simulated results in an independent uniform case.

a_1	b_1	a_2	b_2	Difference			Ratio		
				m = 5	m = 10	m = 15	m = 5	m = 10	m = 15
100	196	30	78	1.1×10^4	-2.1×10^3	-2.1×10^3	0.77	1.15	1.15
100	196	54	54	1.4×10^4	-5.9×10^3	-5.9×10^3	0.71	1.06	1.06
100	196	78	30	1.6×10^4	-9.2×10^3	-9.2×10^3	0.63	0.96	0.96
148	148	30	78	6.9×10^3	-4.6×10^2	-4.4×10^2	0.82	1.22	1.22
148	148	54	54	8.7×10^3	-4.7×10^3	-4.7×10^3	0.77	1.13	1.13
148	148	78	30	1.1×10^4	-8.5×10^3	-8.5×10^3	0.67	1.01	1.01
196	100	30	78	3.1×10^3	1.6×10^3	1.6×10^3	0.90	1.30	1.30
196	100	54	54	4.5×10^3	-2.9×10^3	-2.9×10^3	0.83	1.22	1.22
196	100	78	30	6.4×10^3	-7.4×10^3	-7.4×10^3	0.73	1.09	1.09

Table 5.4. Comparison of the numerical results with the simulated results in an independent normal case.

a_1	b_1	a_2	b_2	Difference			Ratio		
				m = 5	m = 10	m = 15	m = 5	m = 10	m = 15
100	196	30	78	3.2×10^3	2.7×10^3	2.6×10^3	0.91	0.92	0.92
100	196	54	54	3.1×10^3	2.4×10^3	2.4×10^3	0.91	0.93	0.93
100	196	78	30	2.7×10^3	2.0×10^3	2.0×10^3	0.91	0.93	0.93
148	148	30	78	2.8×10^3	2.4×10^3	2.4×10^3	0.91	0.92	0.92
148	148	54	54	2.4×10^3	1.9×10^3	1.9×10^3	0.91	0.93	0.93
148	148	78	30	2.1×10^3	1.6×10^3	1.6×10^3	0.91	0.93	0.93
196	100	30	78	2.1×10^3	1.9×10^3	1.9×10^3	0.91	0.91	0.91
196	100	54	54	1.8×10^3	1.5×10^3	1.5×10^3	0.91	0.92	0.92
196	100	78	30	1.6×10^3	1.3×10^3	1.2×10^3	0.91	0.93	0.93

In Table 5.3, the difference between the numerical result and the simulated result is either positive or negative. This shows that the simulated result will be underestimated or overestimated if the numerical result is highly accurate. Hence the simulated results should have some adjustment. Besides, the ratio is in the range from 0.63 to 0.90 in the uniform case and is 0.91 in the normal case, where m = 5 in the numerical approach; these ratios are much further from 1. The ratios are in the range from 0.96 to 1.30 and the range from 0.91 to 0.93 in the uniform and the normal cases respectively, where m = 10 or m = 15. It is trivial that the simulation

model in the normal case is better than that in the uniform case. Let us consider the effects of error ellipse equations on the difference and the ratio.

Figures 5.5 and 5.6 plot the ratio and the difference against the error ellipse parameter a_i (i = 1 or 2) respectively from the results in Tables 5.3 and 5.4, where m = n = no = p =15. Under the assumption of the nodal error (either uniformly or normally distributed), three same symbols are displayed at a certain value of a_i (i =1 or 2). It is because for this value of a_i, other error ellipse parameters vary.

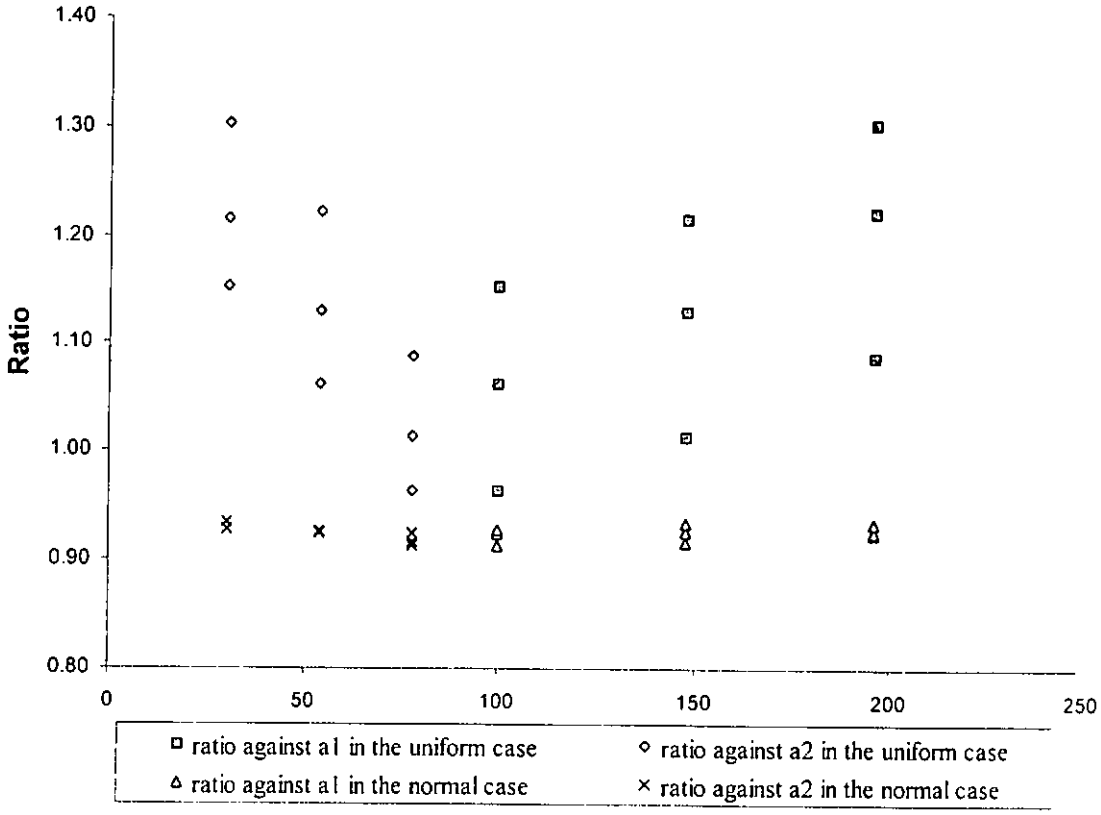


Figure 5.5. Ratio of the numerical results to the simulated results against the error ellipse parameter a_i, where i = 1 or 2.

In Figures 5.5 and 5.6, changing the error ellipse parameter a_i (either i = 1 or 2) does not affect the ratio and the difference significantly in the normal case. However, in the uniform case, both the ratio and the difference are controlled by the error ellipse parameter. For example, when a_i increases, the difference decreases but

the ratio increases. When a_2 increases, the difference increases but the ratio decreases.

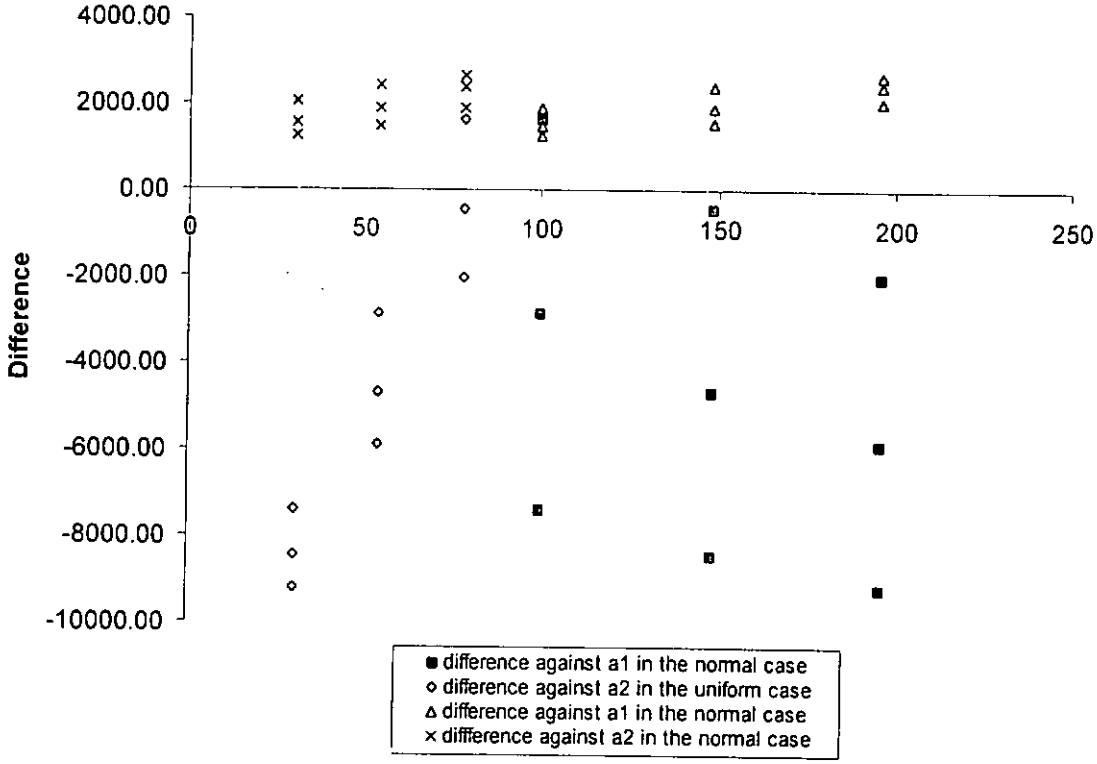


Figure 5.6. Difference between the numerical and the simulated results against the error ellipse parameter a_i, where i = 1 or 2.

Both the simulated and the numerical results are the approximation of the expected discrepant area of the spatial features. From the above figures, it is observed that the error ellipse affects the difference of the numerical results from the simulated results.

In the independent case, there is no correlation among the nodal errors. This may not always be true in the real world. I, therefore, further investigate the correlation cases. Table 5.5 shows the differences between the numerical and the simulated results (in the case of correlation) in the normal case.

Table 5.5. Discrepant area calculated from numerical integration and simulation under the assumption of the normally distributed correlated nodal errors.

ρ_{x1y1}	ρ_{x2y2}	ρ_{x1y2}	ρ_{x2y1}	ρ_{x1x2}	ρ_{y1y2}	Simulated result	Numerical result (m = 5)	Numerical result (m = 10)	Numerical result (m = 15)	Ratio (m=5)	Ratio (m=10)	Ratio (m=15)
-0.5	0.0	0.0	0.0	0.0	0.0	36207.1	30578.4	24573.3	24677.3	0.84	0.68	0.68
0.5	0.0	0.0	0.0	0.0	0.0	35030.9	30566.3	35008.1	35056.7	0.87	1.00	1.00
0.0	-0.5	0.0	0.0	0.0	0.0	35380.6	31377.6	28641.3	28466.8	0.89	0.81	0.80
0.0	0.5	0.0	0.0	0.0	0.0	35444.6	31084.4	30986.7	31293.7	0.88	0.87	0.88
0.0	0.0	-0.5	0.0	0.0	0.0	34944.4	31910.8	33069.2	32958.3	0.91	0.95	0.94
0.0	0.0	0.5	0.0	0.0	0.0	34904.9	31698.8	31873.8	32064.2	0.91	0.91	0.92
0.0	0.0	0.0	-0.5	0.0	0.0	34885.0	31381.4	34257.4	34394.9	0.90	0.98	0.99
0.0	0.0	0.0	0.5	0.0	0.0	34734.2	31644.8	35167.5	35119.1	0.91	1.01	1.01
0.0	0.0	0.0	0.0	-0.5	0.0	34939.5	31910.2	22361.3	23475.5	0.91	0.64	0.67
0.0	0.0	0.0	0.0	0.5	0.0	35160.5	31644.8	37551.5	36523.4	0.90	1.07	1.04
0.0	0.0	0.0	0.0	0.0	-0.5	30257.6	28168.2	29071.2	29090.1	0.93	1.03	0.96
0.0	0.0	0.0	0.0	0.0	0.5	37711.5	34334.6	35765.1	35774.0	0.91	0.95	0.95

The co-ordinates of the line segment used in Table 5.5 are (0,0) and (1000,0). The parameters for the error ellipse equations, a_1, b_1, a_2 and b_2 take the values 100m, 196m, 30m and 78m. Parameters m, n, no and p are set to be the same.

It is observed that the ratios of the numerical results to the simulated results ranges from 0.84 to 0.93, where m = n = no = p = 5; from 0.64 to 1.07, where m = no = p =10; from 0.67 to 1.04, where m = no = p = 15. These ratios are distinct from the ratio in the previous uncorrelated normal case (range from 0.91 to 0.93), mainly due to the effect of correlation coefficient. The ratios in the first four rows are a little bit smaller than that in the next eight rows, except the third one. The domains of the first four rows are very different to the domains of the next eight rows.

A domain of the integral is determined by two error ellipses. The correlation coefficients among the nodal errors cannot affect the error ellipse equations. However, the non-zero correlation coefficients $\rho_{x_1y_1}$ and $\rho_{x_2y_2}$ will rotate the error ellipses and change its shape, so that the pattern of the domain is different from that in the independence case. As a result, different extents of the change yield different ratios between the numerical and the simulated discrepant areas. In an ideal case, the change does not affect the numerical results.

Table 5.6. The expected discrepant area of spatial features in 2D GIS.

Spatial feature	In the uniform case			In the normal case		
	Expected discrepant area		Ratio	Expected discrepant area		Ratio
	Numerical Result	Simulated Result		Numerical Result	Simulated Result	
Linear feature	7.4×10^4	8.0×10^4	0.92	6.0×10^4	5.8×10^4	1.03
Areal feature	1.9×10^5	1.8×10^5	1.05	1.3×10^5	1.3×10^5	1.00

Table 5.7. The expected discrepant area of spatial features in 3D GIS.

Spatial feature	In the uniform case			In the normal case		
	Expected discrepant area		Ratio	Expected discrepant area		Ratio
	Numerical Result	Simulated Result		Numerical Result	Simulated Result	
Line segment	9.3×10^{4}	7.5×10^{4}	1.23	6.2×10^{4}	5.9×10^{4}	1.05
Linear feature	1.2×10^{5}	1.1×10^{5}	1.02	8.6×10^{4}	8.9×10^{4}	0.96
Areal feature	2.3×10^{7}	2.4×10^{7}	0.99	1.5×10^{7}	1.9×10^{7}	0.82
Volumetric feature	6.4×10^{7}	5.7×10^{7}	1.12	3.8×10^{7}	4.4×10^{7}	0.85

In Tables 5.6 and 5.7, the expected discrepant areas of the spatial features in 2D and 3D GIS are recorded for both the numerical integration and the simulation techniques. The ratio of the results from the numerical model to that from the simulation model is in the range of 0.92 to 1.05 in 2D problem; the ratio is in the range of 0.82 to 1.23 in 3D problem. In an ideal situation, this ratio should be 1. A ratio varying from 1 is due to the approximation of the expected discrepant area for both techniques (numerical integration and simulation techniques).

CHAPTER 6
RELIABILITY OF BUFFER ANALYSIS

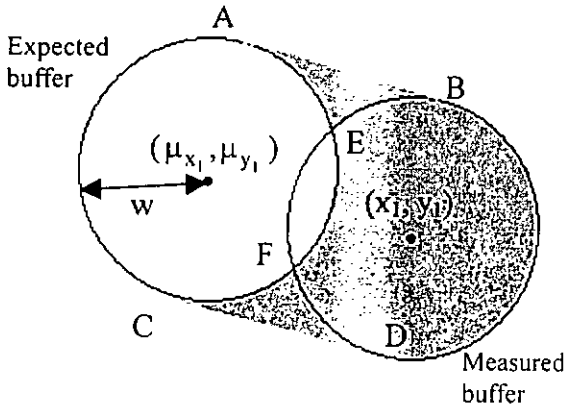
Spatial data in GIS is always expressed in terms of point feature, linear feature, or areal feature. Its reliability has been discussed in the previous chapters. Since the spatial data is not error-free, errors of the data will be transferred to the newly generated data via a GIS operation such as buffer operation. As a result, the derived spatial data may accumulate more errors and have different error characteristics from the original data. In this chapter, a reliability model of buffer analysis for the spatial data will be developed from two methods: simulation and numerical integration.

A discrepant area of the buffer around a spatial feature measures reliability in buffer spatial analysis for the spatial feature. The expected and the measured buffers bound the discrepancy of the buffer. The expected buffer and the measured buffer can be derived from the expected and the measured spatial features respectively. Then, it is assumed that nodal errors of the spatial feature are distributed "within" the nodal error ellipse. The discrepant area of the buffer can be estimated by simulation or numerical approach. In the following, buffers around point feature, linear feature and areal feature will be discussed.

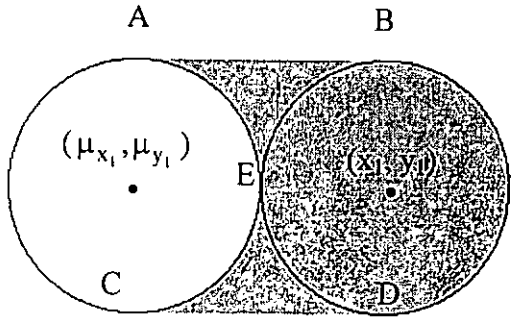
6.1 Point Feature

The discrepancy of the buffer around a point feature is composed of a region of which the boundary is the expected buffer, the measured buffer and the two tangents of both the expected and the measured buffers. This region is shaded in Figure 6.1.

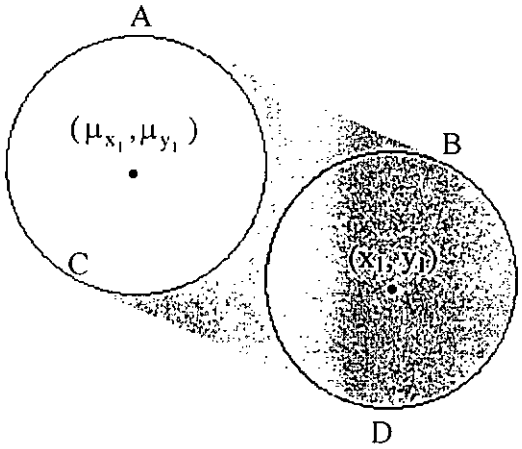
In Figure 6.1(a), (μ_{x_1}, μ_{y_1}) and (x_1, y_1) represent the expected and the measured point features respectively. Points A and C are on the expected buffer; points B and D are on the measured buffer; and circles on the left-hand side and on the right-hand side (the expected buffer and the measured buffer respectively) intersect at points E and F. The radii of these circles are the specified buffer size w. The line passing through points A and B is tangent to both the expected and the measured buffers; the line passing through points C and D is also tangent to the expected and the measured buffers. The two tangents which meet the left-hand circle at points A and C, and the right-hand circle at points B and D are parallel because these circles have the same radii. Given w, (μ_{x_1}, μ_{y_1}) and (x_1, y_1), the coordinates of points A, B, C, D, E and F can be calculated; the shaded area is subsequently calculated.



(a) Two intersecting points on the expected and the measured buffers around a point.



(b) One intersecting point on the expected and the measured buffers around a point.



(c) No intersecting point on the expected and the measured buffers around a point.

Figure 6.1. Discrepancy of the buffer around a point feature.

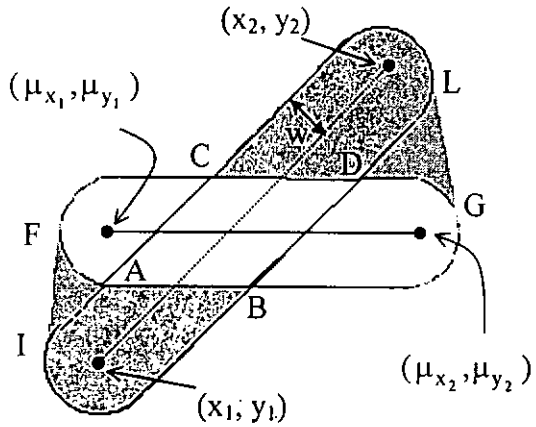
The shaded area becomes zero when the location of the expected buffer is the same as that of the measured buffer. This is one end of the spectrum of cases. Many different cases can arise. In Figure 6.1(a), the two buffers intersect at two points. In Figure 6.1(b), the expected and the measured buffers intersect only at one point, point E. In Figure 6.1(c), the two buffers do not intersect at all.

Theoretically, the discrepant area of the buffer around the point feature should be calculated differently, depending on its case. However, the discrepant area of the buffer is equal to the area of the rectangle with vertices A, B, D and C.

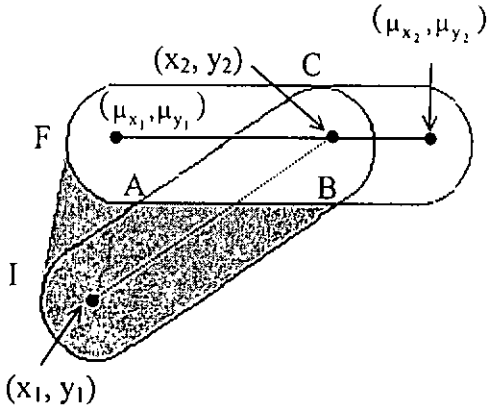
The two tangents in Figure 6.1 are parallel to the line passing through (μ_{x_l}, μ_{y_l}) and (x_l, y_l) because of the same radii of the two buffers. The normal to the expected buffer at point A and that at point C has equal slope. These two lines (normal) pass through (μ_{x_l}, μ_{y_l}) and so the line joined by points A and C passes through (μ_{x_l}, μ_{y_l}). Similarly, the line joined by points B and D passes through (x_l, y_l). A rectangle is formed and its vertices are points A, B, D and C. The discrepant area of the buffer is equal to adding the area of rectangle ABDC to the area of the semi-circle of the measured buffer and then subtracting the area of the semi-circle of the expected buffer. The area of the measured buffer is equal to that of the expected buffer. The discrepant area therefore should be the area of the rectangle ABDC.

6.2 Linear Feature

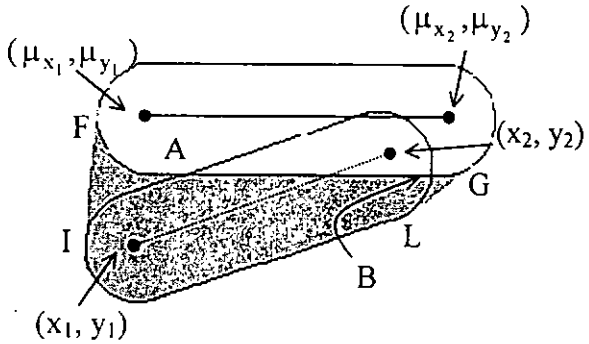
Let us consider the discrepancy of a buffer around a line segment. The discrepancy of the buffer around a line segment is defined as the region whose boundaries are composed of the expected buffer, the measured buffer and the tangents of the expected and the measured buffers. This region is shaded in Figure 6.2.



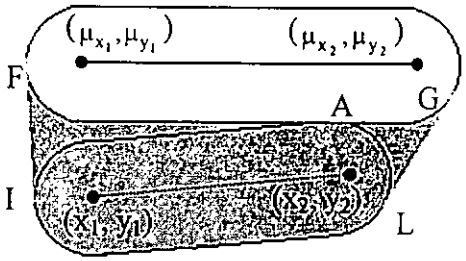
(a) Four intersecting points on the expected and the measured buffers around a line segment.



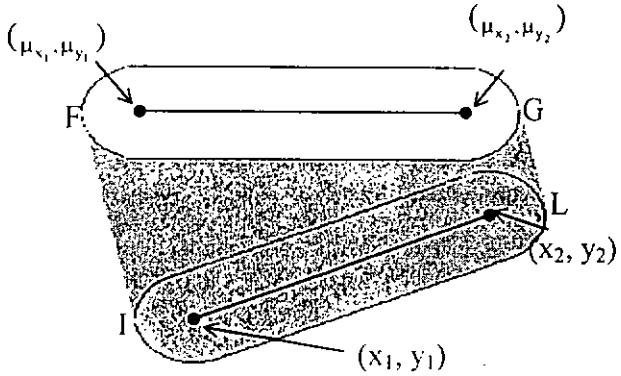
(b) Three intersecting points on the expected and the measured buffers around a line segment.



(c) Two intersecting points on the expected and the measured buffers around a line segment.



(d) One intersecting point on the expected and the measured buffers around a line segment.



(e) No intersecting point on the expected and the measured buffers around a line segment.

Figure 6.2. Discrepancy of the buffer around a line segment.

In Figure 6.2(a), the solid line segment represents the expected line segment of the two expected nodes (μ_{x_1}, μ_{y_1}) and (μ_{x_2}, μ_{y_2}); the dotted line segment represents the measured line segment of the two measured nodes (x_1, y_1) and (x_2, y_2). The band around the solid line segment is the expected buffer around the line segment; the band around the dotted line segment is the measured buffer; w is the buffer size. The expected and the measured buffers intersect at points A, B, C and D. The line passing through E and K is their tangent. Another tangent is the line of nodes L and G. These two tangents are tangents to the expected and the measured buffers around the node on the left-hand side, and the expected and the measured buffers around the node on the right hand side respectively. The expected and the measured buffers and their two tangents bound the discrepancy of the buffer. If the buffer size, the two expected nodes and the two measured nodes are given then the unknown points and the discrepant area can be determined. This is a case where the expected buffer intersects the measured buffer at four distinct points. Four other cases are shown in Figures 6.2(b)-6.2(e).

In Figure 6.2(b), the expected and the measured buffers intersect at three points. The tangent of the expected and the measured buffers around the right-hand sided node of the line segment is parallel to the expected line segment and then this tangent is not taken into accounts. The expected and the measured buffers intersect at two points and one point respectively in Figures 6.2(c) and 6.2(d) while they do not intersect in Figure 6.2(e). The amount of the intersecting points can be used to determine which discrepancy of the buffer is. Therefore, the appropriate case of discrepancy should be first determined in order to estimate the discrepant area exactly.

The discrepancy of a buffer around a linear feature is also defined as the region whose boundaries are composed of the expected buffer, the measured buffer and the tangents of the expected and the measured buffers. For example, an expected linear feature (the solid polyline in Figure 6.3) contains three expected nodes (μ_{x_1}, μ_{y_1}), (μ_{x_2}, μ_{y_2}) and (μ_{x_3}, μ_{y_3}). The corresponding measured linear feature is

the dotted polyline in which the measured nodes are (x_1, y_1), (x_2, y_2) and (x_3, y_3). The discrepancy of the buffer around the linear feature is shaded in Figure 6.3. The discrepant area is obtained after the intersecting points on the tangents; the expected and the measured linear features are derived.

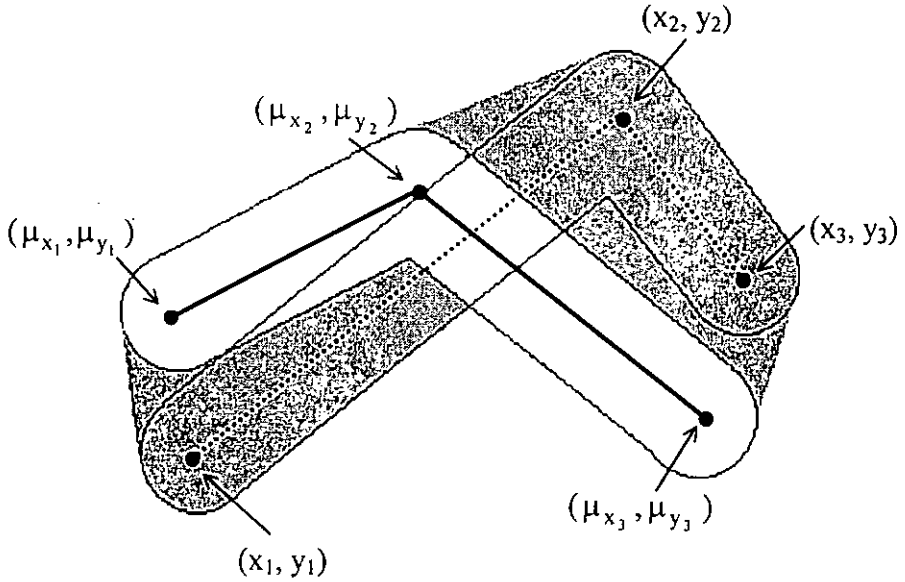


Figure 6.3. Discrepancy of the buffer around a linear feature with three nodes.

6.3 Areal Feature

The boundary of an areal feature is a linear feature. As a result, the discrepancy of the buffer around the boundary of an areal feature has been studied in the previous section. The discrepancy of the buffer around the areal feature is distinct from the discrepancy of the buffer around the boundary of the areal feature because in the former case the interior of the buffer may contain errors. Here, the discrepancy of a buffer around an areal feature will be discussed.

The discrepancy of a buffer around an areal feature with four nodes is shaded in Figure 6.4. For $j = 1, 2, 3$ or 4, (μ_{x_j}, μ_{y_j}) and (x_j, y_j) represent the expected node and the measured node of the areal feature. The boundary of the expected areal feature is the solid polyline and that of the measured areal feature is the dotted

polyline. The region around the solid polyline is the buffer around the expected areal feature while the region around the dotted polyline is the buffer around the measured areal feature. The discrepancy of the buffer around the areal feature is formed by the expected buffer, the measured buffer and their tangents. The shaded region in Figure 6.4 represents the discrepancy of the buffer around the areal feature.

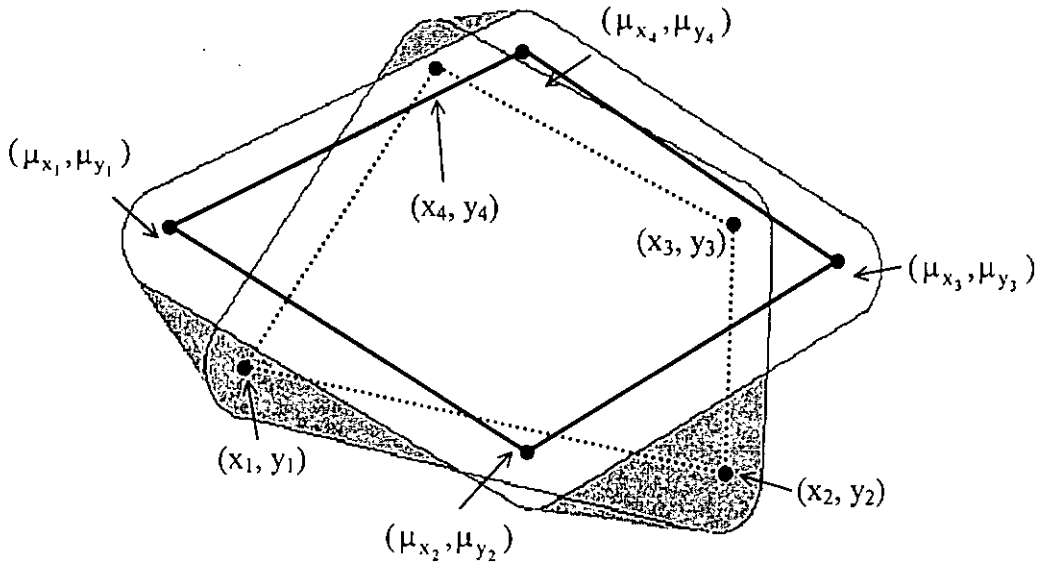


Figure 6.4. Discrepancy of the buffer around an areal feature with four nodes.

The discrepant area of the buffer around an areal feature is determined by the expected and the measured buffers. The buffer around the areal feature involves circles centered at the nodes of the areal feature (buffers around the nodes). In Figure 6.4, some of the buffers around the measured nodes such as (x_1, y_1), (x_2, y_2) and (x_4, y_4) are not completely inside the buffer around the expected areal feature but some such as (x_3, y_3) are. When the measured buffer around a node belongs to the former case, a tangent of both the measured and the expected buffers around the node is calculated and then the shaded area, which is caused by the discrepancy of the node, is determined. Otherwise, the discrepancy of the node is not considered. Hence, determining whether the buffer around a measured node of the areal feature is completely inside the expected buffer around the areal feature is the first step of estimating the shaded area. As a result, the shaded area of the areal feature is a sum of all shaded areas of the nodes of the areal feature.

6.4 Simulation Model

In order to study the effect of the positional error propagation in buffer spatial analysis, simulation is implemented based on the assumption. Given the error ellipse parameters, the measured buffer around a spatial feature is then generated and the discrepant area of the buffer around the spatial feature can be computed.

Based on the nodal error assumption, a node is generated for each expected node of the spatial feature; the measured buffer around the spatial feature and the discrepant area is then calculated. This is the first iteration. After the simulation is repeated N times, where N is a positive integer, an average of the discrepant area can be obtained.

6.5 Numerical Model

The concept of the numerical model has been demonstrated in the last chapter. Multiple integral has been used to estimate the expected discrepant area of a spatial feature. The Gaussian quadrature has been implemented to solve the multiple integral. In this session, the expected discrepant area of the buffer around a spatial feature will be derived in terms of multiple integral.

The expected discrepant area of the buffer around a point feature is as follows:

$$E = \iint_{D_I} f(x_I, y_I) \times A \, dx_I dy_I \tag{6.1}$$

where D_I is the error ellipse for the point feature

f is a joint probability density function of two random variables X_I and Y_I and

A is the area of the rectangle ABDC defined in Figure 6.1.

The expected discrepant area of the buffer around a line segment is as follows:

$$E = \int_{D_1 \cup D_2} \cdots \int f(x_1, y_1, x_2, y_2) \times A \, dx_1 dy_1 dx_2 dy_2 \tag{6.2}$$

where D_1 and D_2 are two error ellipses for the two expected nodes of the line segment

f is a joint probability density function of X_1, Y_1, X_2 and Y_2, and

A is the shaded area defined in Figure 6.2.

The interval of the multiple integral should be further divided because the shaded area in the line segment case cannot be expressed by only one equation. In Figures 6.2(a) and (b), the expected and the measured line segments intersect. The shaded area in Figure 6.2(a) contains two parts. The boundary of the upper part passes through points C, D, G and L; the boundary of the lower part passes through points A, B, I and F. The first region is called as region CDGL and the latter one is called as region ABIF. The region CDGL shown in Figure 6.2(a) is in the upper part but sometimes it is in the lower part. Let points A, B, C, D, F, G, I and L denote (Ax, Ay), (Bx, By), (Cx, Cy), (Dx, Dy), (Fx, Fy), (Gx, Gy), (Ix, Iy) and (Lx, Ly), a function A_{12}(Ax, Ay, Bx, By, Ix, Iy, Fx, Fy) denotes the area of the shaded region ABIF. Then, A_{12}(Cx, Cy, Dx, Dy, Gx, Gy, Lx, Ly) is the function used to denote the area of the shaded region CDGL. In Figures 6.2(c)-(e), the expected and the measured line segment do not intersect. Let A_{13} denotes the area of the region ABGLIF in which its boundary passes through points A, B, G, L, I and F. Afterward, the shaded band in Figure 6.5 is the buffer around the expected line segment. At the end of the expected line segment, there exist two error ellipses surrounding the two expected nodes. The measured line segment will intersect the expected line segment if its nodes are either inside D_1 and D_4 or inside D_2 and D_3. The measured line segment will not intersect the expected line segment if its nodes are either inside D_1 and D_3 or D_2 and D_4.

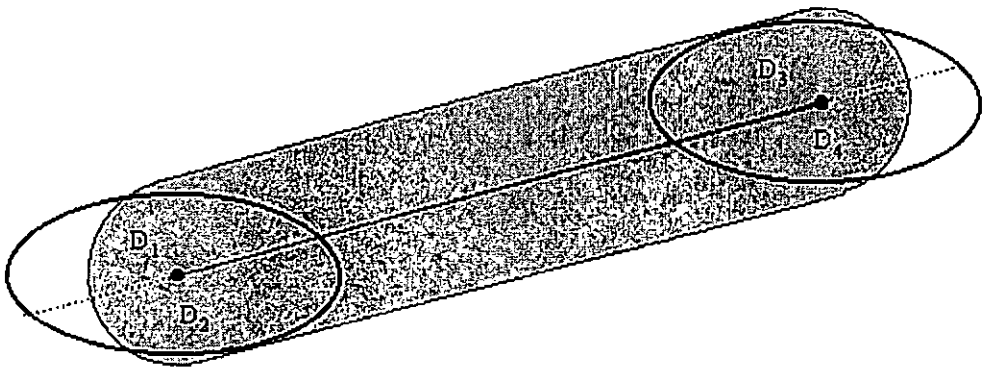


Figure 6.5. Interval of the expected discrepant area of the buffer around a line segment.

The discrepant area of the buffer around the line segment can be

$$
\begin{aligned}
E = & \int_{D_1 \cup D_4} \cdots \int f \times \left(A_{12}(Ax, Ay, Bx, By, Ix, Iy, Fx, Fy)\right) dx_1 dy_1 dx_2 dy_2 \\
& + \int_{D_1 \cup D_4} \cdots \int f \times \left(A_{12}(Cx, Cy, Dx, Dy, Gx, Gy, Lx, Ly)\right) dx_1 dy_1 dx_2 dy_2 \\
& + \int_{D_2 \cup D_3} \cdots \int f \times \left(A_{12}(Ax, Ay, Bx, By, Ix, Iy, Fx, Fy)\right) dx_1 dy_1 dx_2 dy_2 \\
& + \int_{D_2 \cup D_3} \cdots \int f \times \left(A_{12}(Cx, Cy, Dx, Dy, Gx, Gy, Lx, Ly)\right) dx_1 dy_1 dx_2 dy_2 \\
& + \int_{D_1 \cup D_3} \cdots \int f \times A_{13}\, dx_1 dy_1 dx_2 dy_2 + \int_{D_2 \cup D_4} \cdots \int f \times A_{13}\, dx_1 dy_1 dx_2 dy_2
\end{aligned}
\tag{6.3}
$$

The expected discrepant area of the buffer around a linear feature and an areal feature can be defined in the same manner.

6.6 Examples

Examples for a point, a line segment and an areal feature are given in the following session, where a meter is the unit of length. The simulated results will be discussed first. For the point, the expected node is (0, 0); $a_1 = b_1 = 100$; c_1 in the

uniform case = $\rho_{x_1y_1}$ in the normal case = 0.0; w = 500. If the correlation coefficients are assigned to be zero, the discrepant area of the buffer around the point is 65298.8m^2 in the uniform case and is 48665.7m^2 in the normal case. For the example of a line segment, two expected nodes are (0, 0) and (1000, 1000); $a_1 = b_1 = a_2 = b_2 =$ 100; c_1 in the uniform case = c_2 in the uniform case = $\rho_{x_1y_1}$ in the normal case = $\rho_{x_2y_2}$ in the normal case = 0.0; w = 500. The discrepant area of the buffer around the line segment is $3.5\text{x}10^5$ m^2 in the uniform case and $3.2\text{x}10^5$ m^2 in the normal case. For the areal feature, the expected nodes are (0,0), (1000,1000) and (2000, -1000); $a_1 = b_1 = a_2 = b_2 = a_3 = b_3 = 100$; $c_1 = c_2 = c_3 = \rho_{x_1y_1} = \rho_{x_2y_2} = \rho_{x_3y_3} = 0.0$; w = 500. The discrepant area of the buffer around the areal feature is $1.5\text{x}10^6$ m^2 in the uniform case and $5.3\text{x}10^5$ m^2 in the normal case

The discrepant area of the buffer around a spatial feature in the uniform case is significantly larger than that in the normal case. The discrepant area is affected by the measured location of the spatial feature. In this example, the measured spatial feature is generated by simulation under the assumption of uniformly or normally distributed nodal errors. The variance of the uniformly distributed nodal errors is greater than the variance of the normally distributed nodal errors.

The effect of the buffer size on the positional error propagation in buffer spatial analysis may be of interest. Tables 6.1, 6.2 and 6.3 show discrepant areas of the buffer around a point feature, of the buffer around a linear feature and of the buffer around an areal feature respectively with same data values as in the previous example except for the buffer size. The first column in these three tables shows the values for the buffer size. The next two columns record the mean and the 95% confidence interval for the average discrepant area of the buffer around a point feature in the uniform case and the last two columns record the mean and the confidence interval in the normal case.

Table 6.1. Discrepancy of the buffer around a point feature with different buffer sizes using simulation.

W	Discrepancy in a uniform case		Discrepancy in a normal case	
	mean	95% confidence interval	mean	95% confidence interval
500	$6.5x10^4$	$(4.9x10^4, 8.2x10^4)$	$4.9x10^4$	$(3.6x10^4, 6.1x10^4)$
1000	$1.3x10^5$	$(9.7x10^4, 1.6x10^5)$	$9.7x10^4$	$(7.2x10^4, 1.2x10^5)$
1500	$2.0x10^5$	$(1.5x10^5, 2.5x10^5)$	$1.5x10^5$	$(1.1x10^5, 1.8x10^5)$
2000	$2.6x10^5$	$(1.9x10^5, 3.3x10^5)$	$1.9x10^5$	$(1.4x10^5, 2.4x10^5)$
2500	$3.3x10^5$	$(2.4x10^5, 4.1x10^5)$	$2.4x10^5$	$(1.8x10^5, 3.1x10^5)$

Table 6.2. Discrepancy of the buffer around a linear feature with different buffer sizes using simulation.

W	Discrepancy in a uniform case		Discrepancy in a normal case	
	mean	95% confidence interval	mean	95% confidence interval
500	$3.5x10^5$	$(2.6x10^5, 4.3x10^5)$	$3.2x10^5$	$(2.4x10^5, 4.0x10^5)$
1000	$7.8x10^5$	$(5.8x10^5, 9.8x10^5)$	$7.4x10^5$	$(5.6x10^5, 9.3x10^5)$
1500	$1.2x10^6$	$(9.1x10^5, 1.5x10^6)$	$1.2x10^6$	$(8.8x10^5, 1.5x10^6)$
2000	$1.7x10^6$	$(1.3x10^6, 2.2x10^6)$	$1.7x10^6$	$(1.3x10^6, 2.1x10^6)$
2500	$2.5x10^6$	$(1.8x10^6, 3.1x10^6)$	$2.4x10^6$	$(1.8x10^6, 3.0x10^6)$

Table 6.3. Discrepancy of the buffer around an areal feature with different buffer sizes using simulation.

W	Discrepancy in a uniform case		Discrepancy in a normal case	
	mean	95% confidence interval	mean	95% confidence interval
500	$1.5x10^6$	$(2.5x10^5, 2.7x10^6)$	$5.3 x10^5$	$(2.7 x10^5, 7.9 x10^5)$
1000	$3.2 x10^6$	$(9.6 x10^5, 5.4 x10^6)$	$1.5 x10^6$	$(5.5 x10^5, 2.4 x10^6)$
1500	$6.0 x10^6$	$(1.6 x10^6, 1.0 x10^7)$	$3.1 x10^6$	$(1.1 x10^6, 5.0 x10^6)$
2000	$9.2 x10^6$	$(1.7 x10^6, 1.7 x10^7)$	$5.2 x10^6$	$(1.7 x10^6, 8.7 x10^6)$
2500	$1.7 x10^7$	$(1.0 x10^7, 2.3 x10^8)$	$1.1 x10^7$	$(7.9 x10^6, 1.5 x10^7)$

In Tables 6.1, 6.2 and 6.3, it is observed that the discrepant area of the buffer around a spatial feature will increase if the buffer size increases. In Figures 6.1, 6.2 6.3 and 6.4, the discrepancy of the buffer around the spatial feature is considered when the intersection of the measured buffer and the expected buffer is not empty. In other words, the area of the measured buffer also affects the discrepant area to some extent.

Let us consider the relationship between the discrepant area of the buffer and its buffer size. A linear regression model is implemented to study their relationships in different cases (uniform and normal cases). Table 6.4 shows linear regression models for the discrepant area of the buffer around spatial features in both uniform and normal cases. R squared is a goodness-of-fit measure of a linear model (sometimes called the coefficient of determination). It represents the dependent variables (the discrepant area) and ranges in value from 0 to 1. Small values indicate that the model does not fit the data well. The last two columns show the relationship between the dependent variables (the discrepant area) and the independent variable (the buffer size). Each observation of the discrepant area y can be described by the model

$$y = b_0 + b_1 w + \epsilon . \tag{6.4}$$

where ϵ is a random error with mean zero and variance σ^2.

Table 6.4. Linear regression model for the simulated discrepant area of the buffer.

		R squared	b_0	B_1
Point feature in	uniform case	1.00	18.7	130.5
	normal case	1.00	-0.0	97.3
Linear feature in	uniform case	0.99	-2.5×10^5	1035.9
	normal case	0.99	-2.6×10^5	1019.4
Areal feature in	uniform case	0.92	-3.6×10^6	7314.4
	normal case	0.87	-3.3×10^6	5106.3

Table 6.5. Comparison between the numerical and the simulated expected discrepant area of a buffer.

		w	Numerical result	Simulated result	Ratio
Point feature	In uniform case	500	6.7×10^4	6.5×10^4	1.02
		1000	1.3×10^5	1.3×10^4	1.02
		1500	2.0×10^5	1.9×10^5	1.02
		2000	2.7×10^5	2.6×10^5	1.02
		2500	3.3×10^5	3.3×10^5	1.02
	In normal case	500	4.6×10^4	4.9×10^4	0.94
		1000	9.1×10^4	9.7×10^4	0.94
		1500	1.4×10^5	1.5×10^5	0.94
		2000	1.8×10^5	1.9×10^5	0.94
		2500	2.3×10^5	2.4×10^5	0.94
Linear feature	In uniform case	500	3.9×10^5	3.4×10^5	1.12
		1000	6.6×10^5	7.8×10^5	0.85
		1500	1.2×10^6	1.2×10^6	0.98
		2000	1.6×10^6	1.7×10^6	0.93
		2500	2.5×10^6	2.5×10^6	1.00
	In normal case	500	3.4×10^5	3.2×10^5	1.07
		1000	6.4×10^5	7.4×10^5	0.86
		1500	1.2×10^6	1.2×10^6	0.99
		2000	1.5×10^6	1.7×10^6	0.89
		2500	2.3×10^6	2.4×10^6	0.95
Areal feature	In uniform case	500	1.3×10^6	1.5×10^6	0.89
		1000	3.1×10^6	3.2×10^6	0.99
		1500	5.5×10^6	6.0×10^6	0.92
		2000	9.2×10^6	9.2×10^6	0.99
		2500	1.7×10^7	1.7×10^6	0.99
	In normal case	500	5.3×10^5	5.3×10^5	0.99
		1000	1.5×10^6	1.5×10^6	1.03
		1500	2.8×10^6	3.1×10^6	0.92
		2000	5.0×10^6	5.2×10^6	0.96
		2500	1.1×10^7	1.1×10^7	0.96

In Table 6.4, R squared is close to 1 in both the uniform and the normal cases and so it is concluded that the discrepant area of the buffer around a spatial feature is directly proportional to the buffer size.

Now numerical results are going to be compared with simulated results. The first column in Table 6.5 shows the spatial feature discussed in this example, in which the data values are equal to the previous simulation model. The second column shows the assumption of the nodal errors. The next one records the change of the buffer size. The following two columns record the numerical and the simulated results respectively. The remaining column tabulates the ratio of a numerical result to the corresponding simulated result.

It is noticed that the ratio (in the range of 0.85 and 1.12) is close to 1. As mentioned in Chapter 5, the numerical and the simulation models provide an approximation of the expected discrepant area (of the buffer around a spatial feature). Although accuracy of both methods is unknown, it is observed that they have similar results in the reliability model. Therefore, both methods are suitable to study the reliability model.

From my point of view, the simulation approach will be potentially implemented in GIS. The numerical method in fact has some limitations on combination of the correlation coefficients of the nodal errors. It is because the probability density function of the multivariate normal distribution f must be exited and hence the determinant of the corresponding covariance matrix must be non-zero.

6.7 Summary

The error propagation model in the buffer spatial analysis has been derived by two different approaches: simulation and numerical methods. Similar to the result in the previous comparison between these two methods, the numerical result

approximates to the simulated result, and vice versa. Moreover, it is noticed that the buffer size affects the discrepancy of the buffer around a spatial feature. When the buffer size increases, the measure of the discrepancy increases. In the regression model on the discrepant measure with the buffer size being an independent variable, it is proved that they are directly proportional to each other.

CHAPTER 7
CONCLUSIONS AND RECOMMENDATIONS

The quality of spatial data in a GIS database is significant for their applications. Accuracy information related to spatial data should attach to each spatial feature depending on its type and source of error. In this study, positional error models for spatial features and error propagation models in buffer spatial analysis were derived based on two means: simulation and numerical analysis.

7.1 Summary of the Study

This study modeled positional errors in vector-based Geographical Information Systems, a research topic in many organizations such as the International Standards Organization. It is an important element for potential users to decide whether the database in hand fit the intended use. This investigation included modeling positional error in either 2D or 3D spatial features and positional error propagation in buffer spatial analysis. The reliability of 2D spatial features including linear features and areal features and that of 3D spatial features including linear features, areal features and volumetric features was studied first using the simulation technique. The positional error was measured by a discrepancy. The larger the discrepancy, the more the positional error. Since the discrepancy depended on the positional error of a 2D (or 3D) spatial feature, it was assumed that the nodal error of the spatial feature was either uniformly or normally distributed "inside" an error ellipse (or ellipsoid). Then, the positional error was simulated. Moreover, because of the limitation of the simulation method, the numerical integration was implemented in order to check the accuracy of the simulation model. Furthermore, errors in source data will propagate over GIS operations and so the error propagation model was proposed by the simulation technique or numerical integration.

7.2 Discussion and Analysis

7.2.1 Reliability of a Spatial Feature on Scale Map

In this study, the discrepancy of a spatial feature is used to measure its reliability based on the assumption of nodal error distribution. The reliability of a spatial feature on different scale maps may be worried. For example, the reliability of a land parcel on a 1:5,000-scale map should vary from that on a 1:25,000-scale map. The discrepant area (or volume) of a spatial feature is different on different scale maps. It is due to the fact that the discrepant area is computed based on the assumption of nodal error distribution and the distribution's properties such as variance are affected by scale. Second, the discrepant area is an indicator for positional error description. Whether a spatial data is reliable or not is determined by an acceptant level and potential users may have different acceptant level for different scale maps.

7.2.2 Error of Commission

The discrepancy of a point can be measured by distance between the measured location and the expected location of the point. This concept was generalized to define the discrepancy of a line segment, a region bounded by the measured location and the expected location of the line segment. Similarly, the discrepancy of a polygon was bounded by the measured location and the expected location of the polygon and their tangents. Both error of commission and error of omission should be considered to be the sum of the discrepancy of the polygon. However, the discrepancy I mentioned in this study was the error of commission. This simplified the reliability problem and so the error of omission should be considered too.

7.2.3 Effects of Nodal Error Distribution

In the simulated examples for the 2D line segment, it was concluded that the existing simulation-based models might require some adjustments. It was ascertained that when the nodal errors are independent, decreasing either a_1 or a_2 would increase the area of discrepancy. Thus, it was necessary to consider the error ellipse model rather than the simplified error circle model. Furthermore, the discrepant area of the line segment in the uniform case was significantly greater than that in the normal case. This fitted the fact that most of the normal random nodal errors distributed centrally near the measured nodes. The uniform case seemed to be the normal case with a very large variance. Finally, it was noticed that different correlation matrices of the nodal errors yielded different reliability of a line segment. Therefore, users should be concerned both with positional errors within the nodes and between the nodes. In the simulated examples for the 3D line segment, similar results in the 2D problem could be achieved. The discrepancy of a spatial feature in the uniform case was greater than that in the normal case. It was also observed that the parameters of the nodal error ellipsoid might affect the reliability of a spatial feature.

7.2.4 Effect of Buffer Size in Buffer Spatial Analysis

In the buffer spatial analysis, it was noticed that the positional error propagation was affected by the buffer size. The reliability model for spatial features was further developed in order to investigate the error propagation in the buffer spatial analysis. Hence, the propagated error was simulated. Buffers around a point feature, a linear feature and an areal feature were considered. Among these three features, it was ascertained that the buffer size and the discrepant area had a linear relationship. Therefore, the buffer size affected the reliability to a certain degree.

7.2.5 Choice of Methods: Simulation verse Numerical Integration

The weakness of the simulation model was that it is very time-consuming and might not be very accurate. A newly developed analytical model provided alternative numerical solutions to validate the simulated results. The analytical solution (in the form of a multiple integral) for the expected discrepant area (or volume) was given. A numerical integration (Gaussian quadrature) was implemented to provide numerical solutions. Then, the proposed numerical models were compared with the simulation models. From our results, both models were able to approximate a similar value of the discrepancy.

The simulation model was considered the preferred approach in GIS. The expected discrepancy in the numerical model was expressed in term of a multiple integral in which the interval should be defined regarding the discrepancy case. For a point feature and a line segment, the interval could be defined in general terms. However, for a linear feature, an areal feature and a volumetric feature in 2D and 3D, the interval was determined by the amount of the nodes of the spatial feature. If the number of the nodes of the spatial feature increases, the possible discrepancy cases will increase. It was difficult to express the discrepant area (or volume) mathematically with only one equation. Therefore, the numerical model was not suitable to provide the error description of GIS spatial data. Moreover, a limitation of the numerical solution was its inability to handle the case of a correlation coefficient being 1 or -1, mainly due to the undefined joint probability density function.

7.2.6 Position Random Errors and Modeling Errors

In this study, it was noticed that the positional errors were smallest midway between the two nodes of the line segment, but one might expect that the positional errors would be the largest at the furthest distance from the given nodes. It was

mainly due to modeling the reliability of the line segment regardless of modeling errors. If the variance of the modeling errors is added to the measurement errors in perpendicular direction of the line segment, the positional errors at the midpoint position will be larger than that at the given nodes of the line segment. Therefore, the proposed model assumed that the modeling errors were zero.

7.3 Conclusions

This study attempted to clarify how to model positional errors of spatial features and its analysis in vector-based GIS. The proposal models are based on two approaches: simulation and numerical integration method. After weaknesses of these two methods were compared, the simulation model is preferable in GIS. It is potentially implemented to simulate positional errors of GIS-based data and even error propagation over GIS operations. The results are thus applicable to all features of GIS for error description. This model provides a measure tool to identity reliability of spatial features in GIS.

7.4 Contributions of this Study

The proposed simulation model modified existing simulation-based models for a line segment. In the simulation model, most possible measured spatial features were generated based on our assumption. The area (or volume) of the discrepancy for each measured spatial feature was computed. The average discrepant area (or volume) was an indicator of the reliability. Differing from existing simulation-based models on the reliability of a 2D line segment, which assumed the nodal errors uniformly distributed within error circles, the proposed simulation model further investigated the reliability of the line segment. Both uniform and normal distribution cases of the nodal errors were considered. Furthermore, the error circle model was extended to the error ellipse model, which is a more realistic model to describe the

positional errors of the real world features. In addition to the reliability model for a 2D line segment, the reliability model was generalized to the model on the reliability of GIS spatial features.

7.5 Recommendations for Further Studies

In the numerical model, the Gaussian quadrature was implemented to approximate the expected discrepant area (or volume). An error term, which is the difference between the true value and the approximation, can be derived in the numerical analysis and provide much information about the accuracy of this approximation. Existing estimations of the error term consider that the interval of the integral should be divided into equal subintervals. However, the subintervals in the Gaussian quadrature are not in equal length in order to minimize the error term. If this numerical model is implemented in the future, one should pay attention to the error term.

Second, the discrepant area of spatial features defined in this study depended on a characteristic of the spatial features such as length, area, and so forth. Dividing the discrepant area by the length of a line segment in 2D, for example, can normalize this indicator. In the analysis of the reliability of a line segment using the simulation approach, the comparison of the discrepant area in different situations was unaffected, because the length of the spatial features was the same in each comparison. For a more general study in which the length of the features is not equal to each other, the normalization approach may be applied.

This research study focused on modeling positional random errors, however modeling errors (or generalization errors) is also significant in GIS. Although a map is a traditional approach to identify a destination, it is impossible to represent the reality completely on a map and so modeling errors exists in the map. If the map is digitized, modeling errors will be generated. Therefore, modeling errors may be due

to the difference between (a) the reality and its digital representation in GIS; (b) a map and its digital representation in GIS; and (c) a digital and its geographical representations in GIS. It is quite difficult to model them because modeling these errors is partly in the field of psychology. Hence, a further research in this circle is necessary.

REFERENCES

Aalders, H.J.G.L. "The Registration of Quality in a GIS", *Proceedings of the International Symposium on Spatial Data Quality '99*, Hong Kong, China, pp. 23-32 (1999).

Alesheikh, A.A. *Modeling and Managing Uncertainty in Object-Based Geospatial Information System*, Ph.D. Thesis, Department of Geomatics Engineering, The University of Calgary, Alberta, Canada (1998).

Allan, A.L. *Practical Surveying and Computations*. 2nd ed., Butterworth-Heinemann Ltd., Oxford, 551pp (1993)

Arbia, G., Griffith, D. and Haining, R. "Error Propagation Modeling in Raster GIS: Overlay Operations". *International Journal of Geographical Information Science*, Vol. 12, No. 2, pp. 145-167 (1998).

Blakemore, M. "Generalization and Error in Spatial Data Bases". *Cartographica*, Vol. 21, pp. 131-139 (1984).

Bolstad, P.V., Gessler, P. and Lillesand, T.M. "Positional uncertainty in manually digitized map data". *International Journal of GIS*, Vol. 4, pp. 399-412 (1990).

Burden, R.L. and Faires, J.D. *Numerical Analysis*, International Thomson Publishing, pp. 211-222 (1993)

Burrough, P.A. *Principles of Geographical Information Systems for Land Resources Assessment*, Oxford: Clarendon Press, pp. 103-135 (1986).

Caspary, W. and Scheuring, R. "Positional Accuracy in Spatial Databases". *Computers Environment and Urban Systems*, Vol. 17, pp.103-110 (1993)

Cassettaari, S. *Introduction to Integrated Geo-information Management*, Chapman & Hall, pp. 2-3 (1993).

Chapman, M.A., Alesheikh A. and Karimi, H. "Error Modeling and Management for Data in Geospatial Information Systems". *Proceedings of Coast GIS'97*, pp. 1-15 (1997)

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

Cromley, R.G. and Campbell, G.M. "A Geometrically Efficient Bandwidth Line Simplification Algorithm". *Proceedings of the 4th International Symposium on Spatial Data Handling*, Zurich, August 1990, Vol. 1, pp. 77-84 (1990).

Cromley, R.G. and Campbell, G.M. "Integrating Quantitative and Qualitative Aspects of Digital Line Simplification". *The Cartographic Journal*, Vol. 29, No. 1, pp. 25-30 (1992).

Dunn, R., Harrison, A.R. and White, J.C. "Positional Accuracy and Measurement Error in Digital Databases of Land Use: An Empirical Study". *International Journal of GIS*, Vol. 4, pp. 385-98 (1990).

Dutton, G. "Handling Positional Uncertainty in Spatial Databases". *Proceedings of the 5th International Symposium on Spatial Data Handling*, South Caroline, U.S.A., 3-7 August 1992, pp. 460-469 (1992)

Dutton, G. "Scale, Sinuosity, and Point Selection in Digital Line Generalization". *Cartography and Geographic Information Science*, Vol. 26, No. 1, pp. 33-53 (1999).

Easa, S.M. "Discussion of: a Generalization on Line Segment Reliability Measures". *Surveying and Land Information Systems*, Vol. 54, No.2, pp.123 (1994)

Easa, S.M. "Estimating Line Segment Reliability using Monte Carlo Simulation". *Surveying and Land Information Systems*, Vol. 55, No. 3, pp. 136-141 (1995)

Fisher, P.F. "First Experiments in Viewshed Uncertainty: Simulating Fuzzy Viewsheds". *Photogrammetric Engineering and Remote Sensing*, Vol. 58, No. 3, pp. 345-352 (1992)

Goodchild, M.F. "Keynote Address: Symposium on Spatial Database Accuracy". *Proceedings: Symposium on Spatial Database Accuracy*, pp. 1-16 (1991).

Goodchild, M.F., Sun, G. and Yang, S. "Development and Test of an Error Model for Categorical Data". *International Journal of Geographical Information Systems*, Vol. 6, No. 2, pp. 87-104 (1992)

Guptill, S. and Morrison, J. *The Elements of Spatial Data Quality*. Elsevier (1995).

Haining, R. and Arbia, G. "Error Propagation through Map Operations". *Technometrics*, Vol. 35, No. 3, pp. 293-305 (1993).

Hansen, M.H. *Sample Survey Methods and Theory*, Vol. 1.Wiley, New York, N.Y. (1993)

Heuvelink, G.B.M. *Error Propagation in Environmental Modeling with GIS*, Taylor & Francis Ltd., 4pp (1998).

Heuvelink, G.B.M. and Burrough, P.A. "Error Propagation in Cartographic Modeling using Boolean Logic and Continuous Classification". *International Journal of Geographical Information Systems*, Vol. 7, No. 3, pp. 231-246 (1993).

Heuvelink, G.B.M., Burrough, P.A. and Stein, A. "Propagation of Errors in Spatial Modeling with GIS". *International Journal of Geographical Information Systems*, Vol. 3, No. 4, pp. 303-322 (1989).

Keefer, B.J., Smith, J. L. and Gregoire, T. G. "Simulating manual digitizing error with statistical models". *GIS/LIS'88*, pp. 475-483 (1988).

Keefer, B.J., Smith, J.L. and Gregoire, T.G. "Modeling and Evaluating the Effects of Stream Mode Digitizing Errors on Map Variables". *Photogrammetric Engineering & Remote Sensing*, Vol. 57, pp. 957-963 (1991).

Maffini, G., Arno, M. and Bityterlich, W. "Observations and comments on the generation of error in digital GIS data". *Accuracy of Spatial Databases*, edited by Goodchild and Gopal, pp. 55-67 (1989).

McGrew, J.C. *An Introduction to Statistical Problem Solving in Geography*. Wm. C. Brown Publishers, Dubuque, Iowa, pp. 21-22 (1993)

Moellering, H. "Digital Cartographic Data Standards: An Interim Proposed Standard", *Report #6, ACSM* (1985).

Moellering, H. "A Draft Proposed Standards for Digital Cartographic Data", *Report #8, ACSM* (1987).

Muller, J.C. "Minimum Point Density and Compaction Rates for the Representation of Geographic Lines". *Proceedings of Auto Carto 8*, ACSM/ASPRS, Bethesda, Md., pp. 221-230 (1987).

Næsset, E. "Use of the Weighted Kappa Coefficient in Classification Error Assessment of Thematic Maps". *International Journal of Geographical Information Systems*, Vol. 10, No. 5, pp. 591-604 (1996)

Newcomer, J.A. and Szajgin, J. "Accumulation of Thematic Map Errors in Digital Overlay Analysis". *The American Cartographer*, Vol. 11, No. 1,I pp. 58-62 (1984).

Perkal, J. *On the Length of Empirical Curves: Discussion Paper 10*, Ann Arbor, Michigan Inter-University Community of Mathematical Geographers (1966)

Shannon, R.E. *System Simulation: the Art and Science*. Prentice-Hall, Englewood Cliffs, N.J., 2pp (1975)

Shi, W.Z. *Modeling Positional and Thematic Uncertainty in Integration of GIS and Remote Sensing*, Ph.D. Dissertation, University of Osnabrück-Vechta, Enschede, NL: ITC publication 22 (1994)

Shi, W.Z. "A Generic Statistical Approach for Modeling Error of Geometric Features in GIS". *International Journal of Geographical Information Science*, Vol.12, No 2, pp.131-143 (1998)

Shi, W.Z. and Ehlers, M. "Determining Uncertainties and Their Propagation in Dynamic Change Detection based on Classified Remotely-sensed Images". *International Journal of Remote Sensing*, Vol. 17, No. 14, pp. 2729-2741 (1996).

Shi, W.Z. and Liu, W.B. "A Stochastic Process-based Model for Positional Error of Line Segments in GIS". *International Journal of Geographical Information Science*, Vol. 14, No. 1, pp. 51-66 (2000)

Stanfel, L.E. "Reply for: Discussion of: a Generalization on Line Segment Reliability Measures". *Surveying and Land Information Systems*, Vol. 56, No. 1, pp. 56-57 (1996)

Stanfel, L.E., Conerly M. and Stanfel, C.M. "Reliability of Polygonal Boundary of Land Parcel". *Journal of Surveying Engineering*, Vol. 121, No. 4, pp. 163-176 (1995)

Stanfel, L.E. and Stanfel, C.M. "A Model of the Reliability of a Line Connecting Uncertain Points". *Surveying and Land Information Systems*, Vol. 53, No. 1, pp. 49-52 (1993)

Stanfel, L.E. and Stanfel, C.M. "A Generalization on Line Segment Reliability Measures". *Surveying and Land Information Systems*, Vol. 54, No. 1, pp. 41-44 (1994)

Stanislawski, L.V., Dewitt, B.A. and Shrestha, R.L. "Estimating Positional Accuracy of Data Layers within a GIS through Error Propagation". *Photogrammetric Engineering & Remote Sensing*, Vol. 62, No. 4, pp. 429-433 (1996).

Stroud, A.H. and Secrest, D. *Gaussian Quadrature Formulas*. Prentice Hall, Englewood Cliffs, N.J., 374pp (1966)

Topfer, F. and Pillewizer, W. "The Principle of Selection". *The Cartographic Journal*, Vol. 3, pp. 10-16 (1966).

Uren, J. and Price, W.F. *Surveying for Engineers*. 3rd ed., The Macmillan Press Ltd., Basingstoke, 210pp (1994)

Veregin, H. "Developing and Testing of an Error Propagation Model for GIS Overlay Operations". *International Journal for Geographical Information Systems*, Vol. 9, No. 6, pp. 595-619 (1995)

Wolf, P.R. and Brinker, R.C. *Elementary Surveying*. 9th ed., HarperCollins College Publishers, 24pp (1994).

Zhan, F.B. and Buttenfield, B.P. "Multi-scale Representation of a Digital Line". *Cartography and Geographic Information Systems*, Vol. 23, No. 4, pp. 206-228 (1996).

Zhang, B., Zhu, L. and Zhu, G. "The uncertainty propagation model of vector data on buffer operation in GIS". *ACTA Geodaetica et Cartographic Sinica*, Vol. 27, No. 3, pp. 259-266 (1998) (in Chinese).

APPENDIX
MATHEMATICAL PRELIMINARIES AND STATISTICAL THEOREMS

Error distribution of a point may involve a matrix and some matrix operations (such as matrix arithmetic, product of matrices, and so on), which are common, will not be shown here but definitions of some matrices will be elaborated in the first session because these definitions are used when the error distribution is defined. The next session will state some statistical techniques used in this study.

Mathematical Preliminaries

Eigenvalue of Matrix

Systems of n linear equations in n unknowns are usually expressed in the form

$$Ax = \lambda x , \tag{1}$$

where A is a n × n square matrix, x is a n × 1 matrix and λ is a scalar.

Equation (1) can be rewritten as

$$(\lambda I - A)x = 0 , \tag{2}$$

where I is an identity matrix and 0 is a zero vector.

Those values of λ will be determined for which the system has a nontrivial solution (non-zero solution). Such a value of λ is called a characteristic value or an eigenvalue of A.

Positive Definite of Matrix

A symmetric matrix A is positive definite if, and only if, all the eigenvalues of A are positive.

Statistical Theorems

The Central Limit Theorem

If samples $X_1, X_2, \ldots, X_n$ are selected independently from identical populations which are normally distributed with mean μ and standard derivation σ, the distribution of sample means ($\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$) is normally distributed with mean $\mu_{\bar{x}} = \frac{\mu_{x_1} + \mu_{x_2} + \ldots + \mu_{x_n}}{n} = \mu$ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. In probability theory, there is an important theorem called the central limit theorem. In sampling from a large population of any distribution, the sample means have a normal distribution whenever the sample size is large. The distribution of the sample means has mean $\mu_{\bar{x}} = \mu$ and variance $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. A special case of this theorem asserts that if $X_1, X_2, \ldots, X_n$ denote the random sample from any distribution having mean μ and positive standard derivation σ, the random variable $\sqrt{n}\left(\overline{X_n} - \mu\right)/\sigma$ has an approximate normal distribution with mean zero and variance 1 when the sample size n is large. That is, the sum of n independent and identically distributed random variables is approximately normally distributed.

Confidence Interval for the Population Mean

Suppose that the numerical outcome X of a random experiment is a random variable having a univariate normal distribution with known variance σ^2 but unknown mean μ. Here μ is constant but its value is unknown. To estimate μ, the random experiment should be repeated n times independently under identical conditions where n is a fixed positive integer. Let the random variables $X_1, X_2, \ldots, X_n$ denote the outcomes on these n repetitions of the experiment respectively. Then, the distribution of the random variable X_i (i = 1, .., n) is normal with unknown mean μ and known variance σ^2. The distribution of the sample mean $\bar{X}$ is normal with mean μ and variance σ^2/n and the distribution of the random variable $\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)$ is normal with mean zero and variance 1. Since 95% of the area under the standard normal curve is between z = -1.96 and z = 1.96 (see Figure 1) and $\bar{X}$ has a standard normal distribution, we have

$$P\left(-1.96 < \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} < 1.96\right) = 0.95 \ . \qquad (3)$$

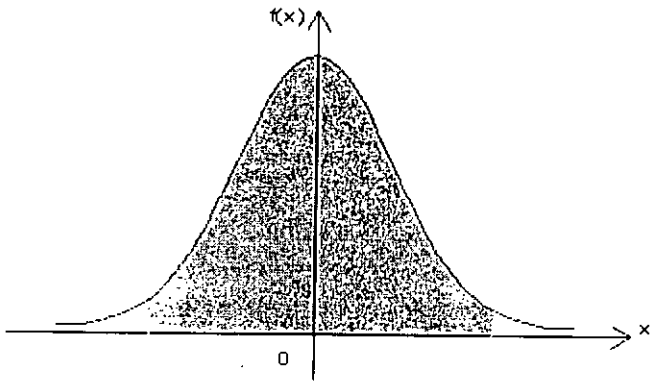


Figure A1. 95% of the area under the standard normal curve.

The inequality in Equation (3) is simplified in Equation (4).

$$\bar{x} - 1.96\sigma_{\bar{x}} < \mu < \bar{x} + 1.96\sigma_{\bar{x}} \ . \qquad (4)$$

This interval is called a 95% confidence interval for the population mean, μ. The general form for the interval is shown in Equation (5) where $Z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution. The confidence coefficient (confidence level) is $1-\alpha$ where $0 \leq \alpha \leq 1$. The $100(1-\alpha)$ % confidence interval for the population mean is

$$\bar{x} - Z_{\alpha/2}\sigma_{\bar{x}} < \mu < \bar{x} + Z_{\alpha/2}\sigma_{\bar{x}} \ . \qquad (5)$$