A General Framework f~~F~~or Error Analysis i~~I~~n Measurement-B~~b~~ased GIS, Part 4: Error Analysis i~~I~~n Length a~~A~~nd Area Measurements

Yee Leung
Department of Geography and Resource Management, Center for Environmental Policy
and Resource Management, and Joint Laboratory for Geoinformation Science,
The Chinese University of Hong Kong, Hong Kong
E-mail: yeeleung@cuhk.edu.hk

Jiang-Hong Ma
Faculty of Science, Xi'an Jiaotong University and Chang'an University, Xi'an, P.R. China
E-mail: jhmath@pub.xaonline.com

Michael F. Goodchild
Department of Geography, University of California, Santa Barbara, California, U.S.A.
E-mail: good@ncgia.ucsb.edu

Abstract. This paper is Part 4 of our four-part series of paper on the development of a general framework for error analysis in measurement-based geographic information systems (MBGIS). In this paper, we discuss the error analysis problems in length and area measurements under measurement error (ME) of the defining points. In line with the basic ME model constructed in Part 1 of this series, we formulate the ME models for length and area measurements. For length measurement and perimeter measurement, the approximate laws of error propagation are derived. For area measurement, the exact laws of error propagation are obtained under various conditions. An important result is that area measurement is distributed as a linear combination of independent noncentral chi-square variables when the joint ME vectors of vertices coordinates are normal. In addition, we also give a necessary and sufficient condition under which the area measurement estimator is unbiased. As a comparison, the approximate law of error propagation i~~o~~n area measurement is also considered and its approximation is substantiated by numerical experiments.

Keywords: Error propagation, geographic information systems, length and area measurement, measurement error, noncentral chi-square variable

1. Introduction

Error analysis in length and area measurement is an important problem in GIS and cartographic operations. Map accuracy has always been important whether map making is automated or manual.

However, its assessment has often been a confusing issue to the producer and user alike. Quite often map users do not understand, or ~~is~~ are completely unaware of, the accuracy level of the map being used. The accuracy of a map is normally expressed by reference to an accuracy standard, which may be stated on the map. Even when the accuracy standard is stated on the map, users usually cannot use that standard to determine the accuracy and precision of common map measurements such as polygon area or line length (Keefer et al., 1991).

Digitizing is one component of data input that introduces error into maps. Its magnitude and impact upon digital map accuracy has not been well studied (Chrisman, 1982). In fact, it is often overlooked or assumed to be negligible. A criterion (Lloyd, 1976) ~~is~~ has been presented for calculating several of the parameters used in the quantization error model in area measurement developed by Frolov and Maling (1969). This criterion yields larger parameter estimates but does not significantly alter the results obtained by the model. For quantifying the effect of digitizing errors, as pointed out in Keefer et al. (1991), a method is needed for evaluating digitizing standards in terms of how they affect the accuracy and precision of map measurements such as polygon area and line length. They utilized simulation techniques and the autoregressive model AR(1) to model stream-mode digitizing errors.

In GIS applications, one of the simplest functions on spatial objects is the computation of the area of any polygon. This is a classical yet important function of GIS. A model is developed in Frolov and Maling (1969) for assessing the accuracy of polygon area estimates derived from dot or grid cell planimetry. The model is based on the possible locations of polygon boundary segments passing through individual grid cells. The standard error of area estimates is shown to be a function of polygon size. Relative error declines as polygon size or the number of grid cells increases. This relationship may be used to select an appropriate cell size for estimating the area of a given polygon.

Since lengths and areas calculated from vector data can reflect positional errors in the points and lines defining them, error analysis in length and area measurements based on vector-based data has been studied over the years. Stolk and Ettershank (1987) used a standard algorithm for polygon area to calculate the area of an irregular shape. A simple statistical model for area measurement was proposed in Chrisman and Yandell (1988) under the assumption that error at each vertex is independently and identically distributed. Under less restrictive, fairly weak assumptions about point coordinate errors, Prisley et al. (1989) ~~has~~ developed an expression for the mean and variance of errors in polygon area. The derivation and expression may be verbose since too many coordinate components are involved. Griffith (1989) has also discussed the problems of digitization error and area calculation problems and outlined several topics for further research. Under three special cases, the expectations and variances of an area and physical centroid calculations are described in matrix form. However, the conditions are limiting and the expressions are still rather simple in mathematics.

Goodchild et al. (1992) have developed an error model for spatial categorical data and used it to provide estimates of the uncertainty associated with the products of GIS. The application of the proposed model in area measurement is also discussed. Kiiveri (1997) has considered a model for positional uncertainty in map with applications in GIS, and expressions for uncertainty in length, perimeters and area calculated from maps are given. The adopted statistical model for maps postulates random “distortions”, which ~~is~~ are modeled as a random linear combination of basis functions with independent random coefficient vectors in the *x* and *y* directions. The results depend on the choice of basis functions and boundary constraints of the map. A statistical procedure based on the familiar coordinate method for area computations is developed in Ghilani (2000) and the author proposes two less rigorous procedures that can be used to arrive at the same decision. Recently, Zhang and Kirby (2000) use a stochastic simulation approach to model positional errors and to estimate the polygonal areas using the formulae in Chrisman and Yandell (1988).

Although efforts such as the above have been made in length or area measurement, they seldom focus on error analysis in length and area measurements on vector-based data with measurement errors (ME). More importantly, some fundamental problems have not been completely or satisfactorily solved, for example, (1) ~~“Albeit~~ While some sufficient conditions have been given in the literature, what is the necessary and sufficient condition for unbiased area measurement?~~”,~~ (2) ~~“~~What is the statistical distribution of area measurement under certain conditions?~~”,~~ and (3) ~~“~~Is there a concise and general expression of variance for area measurement?~~”.~~ Based on results in the preceding parts of the present series (Leung et al., 2003a, b, c), we attempt to give an answer to each of these questions in this part of the ~~paper~~series.

In Section 2, we give the error analysis for length measurement of a line segment and perimeter measurement of a polygon. The corresponding approximate laws of error propagation are derived. In Section 3, we derive the error distribution of area measurement of a polygon when the ME distributions of the vertices coordinates are normally distributed. The exact laws of error propagation under different conditions are also derived. The statistical distribution of area measurement under normally distributed ME is further proposed. To make the theoretical analysis more complete, we also give the necessary and sufficient condition for an unbiased estimator of the area measurement. T~~t~~o substantiate the theoretical arguments, simulation experiments are performed and reported in Section 4. We then conclude our paper by giving a summary of the research results obtained in the four-part series of study, and outlining directions for further research.

2. Error analysis in length measurement

In addition to the study of absolute error for a single location, we will perform detailed analysis of relative error, in particular the covariance matrix in the determination of distance (or length of a line segment) between two locations with positional errors, on the basis of the discussions ~~made~~ in Kiiveri (1997) and Hunter and Goodchild (1996), and in light of the locational error models advanced in Leung and Yan (1998)~~ will perform detailed analysis of relative error, in particular the covariance matrix in the determination of distance (or length of a line segment) between two locations with positional errors~~.

We first consider the length measurement problem of a line segment. Let $V_1(\mathbf{x}_1)$ and $V_2(\mathbf{x}_2)$ be the endpoints of a line segment, $\mathbf{x}_i = (x_{i,1}, x_{i,2})^{\mathrm{T}}$, $i = 1, 2$, and l be the length of the line segment. Then

$$\begin{aligned} l^2 &= (x_{1,1} - x_{2,1})^2 + (x_{1,2} - X_{2,2})^2 = (\mathbf{x}_1 - \mathbf{x}_2)^{\mathrm{T}} (\mathbf{x}_1 - \mathbf{x}_2) \\ &= \mathbf{x}_{(2)}^{\mathrm{T}} [(\mathbf{e}_{2,1} - \mathbf{e}_{2,2})^{\mathrm{T}} \otimes \mathbf{I}_2]^{\mathrm{T}} [(\mathbf{e}_{2,1} - \mathbf{e}_{2,2})^{\mathrm{T}} \otimes \mathbf{I}_2] \mathbf{x}_{(2)} \\ &= \mathbf{x}_{(2)}^{\mathrm{T}} \{[(\mathbf{e}_{2,1} - \mathbf{e}_{2,2})(\mathbf{e}_{2,1} - \mathbf{e}_{2,2})^{\mathrm{T}}] \otimes \mathbf{I}_2\} \mathbf{x}_{(2)} = \mathbf{x}_{(2)}^{\mathrm{T}} \mathbf{G}_{(2)} \mathbf{x}_{(2)}, \end{aligned} \tag{2.1}$$

where $\mathbf{x}_{(2)}^{\mathrm{T}} \equiv (\mathbf{x}_1^{\mathrm{T}}, \mathbf{x}_2^{\mathrm{T}})$ is the 4×1 joint vector of $\mathbf{x}_1$ and $\mathbf{x}_2$, the subscript (2) indicates that there are two points, and $\mathbf{e}_{2,i}$ a two-dimensional unit column vector, and

$$\mathbf{G}_{(2)} \equiv [(\mathbf{e}_{2,1} - \mathbf{e}_{2,2})(\mathbf{e}_{2,1} - \mathbf{e}_{2,2})^{\mathrm{T}}] \otimes \mathbf{I}_2 = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes \mathbf{I}_2 .$$

The general definition of the term $\mathbf{e}_{n,i}$ is given by (2.7) in Leung et al. (2003c), that is

$$\underset{n\times 1}{\mathbf{e}_{n,i}} \equiv (\underset{1}{0} \ \underset{\cdots}{\cdots} \ \underset{i-1}{0} \ \underset{i}{1} \ \underset{i+1}{0} \ \underset{\cdots}{\cdots} \ \underset{n}{0})^{\mathrm{T}} .$$

Thus, (2.1) can be developed into a transformation function $l = f(\mathbf{x}_{(2)}) \equiv [\mathbf{x}_{(2)}^{\mathrm{T}} \mathbf{G}_{(2)} \mathbf{x}_{(2)}]^{1/2}$. When there are ME vectors $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$ in $\mathbf{x}_1$ and $\mathbf{x}_2$, what we measure are $\mathbf{X}_1$ and $\mathbf{X}_2$. Thus we have the length measurement model as follows:

$$\begin{cases} L = f(\mathbf{X}_{(2)}) & (2.2) \\ \mathbf{X}_{(2)} = \boldsymbol{\mu}_{(2)} + \boldsymbol{\varepsilon}_{(2)}, \ \boldsymbol{\varepsilon}_{(2)} \sim (\mathbf{0}, \boldsymbol{\Sigma}_{(2)}), & (2.3) \end{cases}$$

where $\mathbf{X}_{(2)}$ is the 4×1 joint positional vector given by $\mathbf{X}_{(2)}^{\mathrm{T}} \equiv (\mathbf{X}_1^{\mathrm{T}}, \mathbf{X}_2^{\mathrm{T}})$, $\boldsymbol{\mu}_{(2)}$ and $\boldsymbol{\varepsilon}_{(2)}$ are respectively the 4×1 corresponding joint true positional vector and ME vector defined by $\boldsymbol{\mu}_{(4)}^{\mathrm{T}} \equiv (\boldsymbol{\mu}_1^{\mathrm{T}} \ \boldsymbol{\mu}_2^{\mathrm{T}})$ and $\boldsymbol{\varepsilon}_{(2)}^{\mathrm{T}} = (\boldsymbol{\varepsilon}_1^{\mathrm{T}}, \boldsymbol{\varepsilon}_2^{\mathrm{T}})$, and $\boldsymbol{\varepsilon}_{(2)} \sim (\mathbf{0}, \boldsymbol{\Sigma}_{(2)})$ denotes the joint ME vector $\boldsymbol{\varepsilon}_{(2)}$ with zero mean and the covariance matrix $\boldsymbol{\Sigma}_{(2)} \equiv \mathrm{cov}(\boldsymbol{\varepsilon}_{(2)})$. In particular, if the ME vectors $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$ are independently distributed as bivariate normal distributions, $\boldsymbol{\varepsilon}_{(2)}$ as well as $\mathbf{X}_{(2)}$ ~~is~~ are then normal. Accordingly, $L^2 = \mathbf{X}_{(2)}^{\mathrm{T}} \mathbf{G}_{(2)} \mathbf{X}_{(2)}$ is a quadratic form in a normal vector $\mathbf{X}_{(2)}$. If the distribution of L^2 can be obtained, then so can the distribution of *L*.

For the distribution of a quadratic form in normal variables, relatively rich research can be found in the statistical literature (see for example Mathai and Provost, 1992). Various representations of the distribution function have been derived and several different procedures have been given for computing the distribution, including its series representations (power series expansions, Laguerre series expansions, expansions in central chi-squares densities, etc.), finite expressions and numerical methods. An important conclusion is that a quadratic form in a multivariate normal vector can be expressed ~~into~~ as a linear combination of independent noncentral chi-squares variables. For convenient reference, we give the result as a lemma and a simple proof in Appendix 1:

Lemma 2.1 Let $Q = \mathbf{X}^{\mathrm{T}}\mathbf{H}\,\mathbf{X}$ be a quadratic form in random column vector $\mathbf{X} \in R^{p}$, where $\mathbf{X} \sim N_p(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ and $\mathbf{H}$ is a $p \times p$ symmetric matrix (known). Then Q is distributed as a linear combination of independent noncentral chi-square variables $Z_i \sim \chi^2_{p_i,\delta_i^2}$, that is, Q can be expressed as

$$Q = \sum_{i=1}^{m} \lambda_i Z_i\,, \quad Z_i \sim \chi^2_{p_i,\delta_i^2}\,, \tag{2.4}$$

where λ_i are the distinct non-zero eigenvalues of $\mathbf{H}\boldsymbol{\Sigma}_x$ (for simplicity and without confusion, we henceforth use λ_i for both the singular and plural form of λ_i (i.e., rather than using λ_i's for plural) and the same applies to all other relevant symbols), p_i are their respective orders of multiplicity ($p_1 + \cdots + p_m = p$), the Z_i are independent chi-square variables with p_i degree of freedom and non-centrality parameter δ_i^2, and the notation $\chi^2_{p_i,\delta_i^2}$ denotes the noncentral chi-square distribution with p_i degree of freedom and non-centrality parameter δ_i^2.

The non-centrality parameters can be determined by the following steps:

1). Compute the eigenvalues γ_i of $\boldsymbol{\Sigma}_x$. Determine the orthogonal matrix $\mathbf{S}$ and diagonal matrix $\boldsymbol{\Gamma}$ such that $\mathbf{S}\boldsymbol{\Sigma}_x\mathbf{S}^{\mathrm{T}} = \boldsymbol{\Gamma} = \mathrm{diag}(\gamma_1,\ldots,\gamma_p)$, and let $\mathbf{P} \equiv (\boldsymbol{\Gamma}^{\frac{1}{2}})^{-1}\mathbf{S}$.

2). For each distinct eigenvalues λ_i of $\mathbf{H}\boldsymbol{\Sigma}_x$, compute the corresponding normalized-orthogonal eigenvectors $\mathbf{t}_{i,1},\ldots,\mathbf{t}_{i,p_i}$ of $(\mathbf{P}^{-\mathrm{T}}\mathbf{H}\mathbf{P}^{-1})$ in order to form the matrix $\mathbf{T}_i \equiv (\mathbf{t}_{i,1},\ldots,\mathbf{t}_{i,p_i})^{\mathrm{T}}$. Then

$$\delta_i^2 = \boldsymbol{\mu}_{y,i}^{\mathrm{T}}\,\boldsymbol{\mu}_{y,i} = (\mathbf{T}_i\mathbf{P}\boldsymbol{\mu}_x)^{\mathrm{T}}\mathbf{T}_i\mathbf{P}\boldsymbol{\mu}_x = \sum_{j=1}^{p_i}\left(\mathbf{t}_{i,j}^{\mathrm{T}}\mathbf{P}\boldsymbol{\mu}_x\right)^2. \tag{2.5}$$

Although the distribution of Q is not so easy to compute as usual statistical distributions, it is still a very useful statistical distribution and has been studied intensively because of its wide applications and powerful descriptive capability. We find that it is useful in length and area measurements.

According to this lemma, $L^2 = \mathbf{X}_{(2)}^{\mathrm{T}}\mathbf{G}_{(2)}\mathbf{X}_{(2)}$ can be represented as

$$L^2 = \sum_{i=1}^{m} \lambda_i Z_i\,, \quad Z_i \sim \chi^2_{p_i,\delta_i^2}\,, \tag{2.6}$$

where λ_i are the distinct non-zero eigenvalues of $\mathbf{G}_{(2)}\mathbf{\Sigma}_{(2)}$, p_i their respective orders of multiplicity, and δ_i^2 are determined by (2.5).

In particular, when $\mathbf{\Sigma}_{(2)} = \sigma^2 \mathbf{I}_4$, a simple computation implies that the eigenvalues of $\mathbf{G}_{(2)}\mathbf{\Sigma}_{(2)} = \sigma^2 \mathbf{G}_{(2)}$ are $\lambda_{1,2} = 0$, $p_1 = 2$; $\lambda_{3,4} = \lambda_\sigma = 2\sigma^2$, $p_2 = 2$, and the corresponding normalized-orthogonal eigenvectors are respectively:

$$\mathbf{v}_1 = (0, -\tfrac{\sqrt{2}}{2}, 0, -\tfrac{\sqrt{2}}{2})^{\mathrm{T}},\ \mathbf{v}_2 = (-\tfrac{\sqrt{2}}{2}, 0, -\tfrac{\sqrt{2}}{2}, 0)^{\mathrm{T}},\ \mathbf{v}_3 = (-\tfrac{\sqrt{2}}{2}, 0, \tfrac{\sqrt{2}}{2}, 0)^{\mathrm{T}},\ \mathbf{v}_4 = (0, -\tfrac{\sqrt{2}}{2}, 0, \tfrac{\sqrt{2}}{2})^{\mathrm{T}}.$$

Let $\mathbf{V} \equiv (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4)$. Note that $\mathbf{G}_{(2)}\mathbf{V} = \mathbf{V}\mathrm{diag}(0,0,2,2)$. Then we have $\mathbf{P} = (\mathbf{\Gamma}^{\frac{1}{2}})^{-1}\mathbf{S} = \sigma^{-1}\mathbf{I}_4$, $(\mathbf{P}^{-\mathrm{T}}\mathbf{G}_{(2)}\mathbf{P}^{-1})\mathbf{V} = \sigma^2\mathbf{G}_{(2)}\mathbf{V} = \mathbf{V}\mathrm{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, and $\mathbf{T}_\sigma = (\mathbf{v}_3, \mathbf{v}_4)^{\mathrm{T}}$. Therefore, the non-centrality parameter δ_σ^2 corresponding to $\lambda_\sigma = 2\sigma^2$ is

$$\delta_\sigma^2 = \| \mathbf{T}_\sigma \mathbf{P} \boldsymbol{\mu}_x \|^2 = \sigma^{-2} \| (\mathbf{v}_3, \mathbf{v}_4)^{\mathrm{T}} \boldsymbol{\mu}_{(2)} \| = \sigma^{-2}[(\mathbf{v}_3^{\mathrm{T}}\boldsymbol{\mu}_{(2)})^2 + (\mathbf{v}_4^{\mathrm{T}}\boldsymbol{\mu}_{(2)})^2].$$

Accordingly,

$$L^2 = 2\sigma^2 Z,\quad Z \sim \chi^2_{2,\delta_\sigma^2}, \tag{2.7}$$

which is a non-central chi-square variable. Denote its density function by $f_{L^2}(x)$. For the computation of the distribution of a non-central chi-square variable, many methods have been proposed in the literature (see Ross, 1999). It is apparent that once $f_{L^2}(x)$ is obtained, the variance of L can be computed as follows:

$$\begin{aligned} E(L) &= \int_0^\infty x\, f_L(x)\,\mathrm{d}x = 2\int_0^\infty x^2 f_{L^2}(x^2)\,\mathrm{d}x, \\ E(L^2) &= \int_0^\infty x^2 f_L(x)\,\mathrm{d}x = 2\int_0^\infty x^3 f_{L^2}(x^2)\,\mathrm{d}x, \\ \sigma_l^2 &= \mathrm{var}(L) = E(L^2) - [E(L)]^2. \end{aligned} \tag{2.8}$$

That is, the error variance σ_l^2 of L in (2.2) is exactly given by (2.8), which can be viewed as the exact law of error propagation for the ME model (2.2).

In general, $\mathbf{\Sigma}_{(2)} = \sigma^2\mathbf{I}_4$ must not hold, ~~even~~ and it is possible that $\mathbf{\Sigma}_{(2)}$ is not diagonal, i.e., $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$ are dependent. It may be difficult to give an explicit expression to the non-central parameters. A general conclusion is that the number of nonzero eigenvalues of $\mathbf{G}_{(2)}\mathbf{\Sigma}_{(2)}$ is at most 2 (since

$\text{rank}(\mathbf{G}_{(2)}) = \text{rank}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \cdot \text{rank}(\mathbf{I}_2) = 2$, $\text{rank}(\mathbf{G}_{(2)}\mathbf{\Sigma}_{(2)}) \le \min(\text{rank}(\mathbf{G}_{(2)}), \text{rank}(\mathbf{\Sigma}_{(2)}) \le 2$, and the number of nonzero eigenvalues of a matrix is not larger than its rank). Thus, (2.6) may have two forms: (2.7) and the independent sum: $L^2 = \lambda_1 Z_1 + \lambda_2 Z_2$ ($Z_1 \sim \chi^2_{1,\delta_1^2}$ and $Z_2 \sim \chi^2_{1,\delta_2^2}$), and their distributions can be obtained from the noncentral chi-square distribution. So the exact law of error propagation can always be established.

For the approximate law for error propagation, from (2.1) we have $l\,(\mathrm{d}\,l) = \mathbf{x}^{\mathrm{T}}_{(2)}\mathbf{G}_{(2)}(\mathrm{d}\,\mathbf{x})$. So

$$\tilde{\sigma}_l^2 = \mathbf{B}_{\mu_{(2)}}\mathbf{\Sigma}_{(2)}\mathbf{B}^{\mathrm{T}}_{\mu_{(2)}}, \quad \text{where} \quad \mathbf{B}_{\mu_{(2)}} = (l^{-1}\mathbf{x}^{\mathrm{T}}_{(2)}\mathbf{G}_{(2)})_{\mu_{(2)}} = (\boldsymbol{\mu}^{\mathrm{T}}_{(2)}\mathbf{G}_{(2)}\boldsymbol{\mu}_{(2)})^{-1/2}\boldsymbol{\mu}^{\mathrm{T}}_{(2)}\mathbf{G}_{(2)}. \tag{2.9}$$

Now we investigate error propagation to the perimeter of a polygon under the effect of ME of the vertices coordinates. Assume that the vertices of a simple n-sided polygon are $V_i(\mathbf{x}_i)$, $i = 1,\ldots,n$, $V_{n+1} = V_1$. Let $\mathbf{x}^{\mathrm{T}}_{(n)} = (\mathbf{x}^{\mathrm{T}}_1, \cdots, \mathbf{x}^{\mathrm{T}}_n)$. Then its perimeter $l_{\mathrm{p}(n)}$ can be written as

$$\begin{aligned} l_{\mathrm{p}(n)} &= \sum_{i=1}^{n} l_{i,i+1} = \sum_{i=1}^{n}[(\mathbf{x}_i - \mathbf{x}_{i+1})^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_{i+1})]^{\frac{1}{2}} \\ &= \sum_{i=1}^{n}\{\mathbf{x}^{\mathrm{T}}_{(n)}[(\mathbf{e}_{n,i} - \mathbf{e}_{n,i+1})^{\mathrm{T}} \otimes \mathbf{I}_2]^{\mathrm{T}}[(\mathbf{e}_{n,i} - \mathbf{e}_{n,i+1})^{\mathrm{T}} \otimes \mathbf{I}_2]\mathbf{x}_{(n)}\}^{\frac{1}{2}} = \sum_{i=1}^{n}[\mathbf{x}^{\mathrm{T}}_{(n)}\mathbf{G}_{i(n)}\mathbf{x}_{(n)}]^{\frac{1}{2}}, \end{aligned} \tag{2.10}$$

where $\mathbf{G}_{i(n)} \equiv [(\mathbf{e}_{n,i} - \mathbf{e}_{n,i+1})(\mathbf{e}_{n,i} - \mathbf{e}_{n,i+1})^{\mathrm{T}}] \otimes \mathbf{I}_2$. Differentiating (2.10), we obtain

$$\mathrm{d}\,l_{\mathrm{p}(n)} = \mathrm{d}\sum_{i=1}^{n}[\mathbf{x}^{\mathrm{T}}_{(n)}\mathbf{G}_{i(n)}\mathbf{x}_{(n)}]^{\frac{1}{2}} = \sum_{i=1}^{n}[\mathbf{x}^{\mathrm{T}}_{(n)}\mathbf{G}_{i(n)}\mathbf{x}_{(n)}]^{-\frac{1}{2}}\mathbf{x}^{\mathrm{T}}_{(n)}\mathbf{G}_{i(n)}(\mathrm{d}\,\mathbf{x}_{(n)}).$$

It can be observed that (2.10) indeed determines a transformation function $l_{\mathrm{p}(n)} = f(\mathbf{x}_{(n)})$. When $\mathbf{x}_{(n)}$ has the ME vector $\boldsymbol{\varepsilon}_{(n)}$, the ME model for the perimeter of a simple n-sided polygon is:

$$\begin{cases} L_{\mathrm{p}(n)} = f(\mathbf{X}_{(n)}) & (2.11) \\ \mathbf{X}_{(n)} = \boldsymbol{\mu}_{(n)} + \boldsymbol{\varepsilon}_{(n)}, \quad \boldsymbol{\varepsilon}_{(n)} \sim (\mathbf{0}, \mathbf{\Sigma}_{(n)}), & (2.12) \end{cases}$$

where $\mathbf{X}_{(n)} \equiv (\mathbf{X}^{\mathrm{T}}_1, \cdots, \mathbf{X}^{\mathrm{T}}_n)^{\mathrm{T}}$, $\boldsymbol{\mu}_{(n)} \equiv (\boldsymbol{\mu}^{\mathrm{T}}_1, \cdots, \boldsymbol{\mu}^{\mathrm{T}}_n)^{\mathrm{T}}$, $\boldsymbol{\varepsilon}_{(n)} \equiv (\boldsymbol{\varepsilon}^{\mathrm{T}}_1, \cdots, \boldsymbol{\varepsilon}^{\mathrm{T}}_n)^{\mathrm{T}}$, and $\mathbf{\Sigma}_{(n)} \equiv \mathrm{cov}(\boldsymbol{\varepsilon}_{(n)})$. Thus, we have the following approximate law of error propagation for (2.11) and (2.12):

$$\tilde{\sigma}^2_{l_{\mathrm{p}(n)}} = \mathbf{B}_{\mu_{(n)}}\mathbf{\Sigma}_{(n)}\mathbf{B}^{\mathrm{T}}_{\mu_{(n)}}, \quad \text{where} \quad \mathbf{B}_{\mu_{(n)}} = \boldsymbol{\mu}^{\mathrm{T}}_{(n)}\sum_{i=1}^{n}[\boldsymbol{\mu}^{\mathrm{T}}_{(n)}\mathbf{G}_{i(n)}\boldsymbol{\mu}_{(n)}]^{-\frac{1}{2}}\mathbf{G}_{i(n)}. \tag{2.13}$$

3. Error analysis in area measurement

3.1 Error analysis for area measurement of a triangle

For area measurement, most research in the literature focuses on raster data. In addition, consideration has been given to classification error which may cause area estimates derived from pixel counts on satellite imagery to be inaccurate. For example, a method for correcting area estimates (Chrisman, 1982) has been proposed to account for classification error. This method involves multiplication of the normalized vector of area estimates by the inverse of the transposed classification error matrix. The variance of area estimates may also be computed and used to test hypotheses about the accuracy of area estimates.

The accuracy of the direct overlay method of areal interpolation ~~is~~ has been examined (Lam, 1982). Accuracy is a function of the differences in the values of neighboring source zones and the number and size of split source zones (i.e., source zones that overlap more than one target zone). An error model incorporating these factors was tested for four fractal surfaces. Strong correlations were observed between actual and predicted levels of error for the surfaces. The highest correlations occurred for surfaces of low dimensionality when the number of source zones was large relative to the number of target zones.

In this subsection, we derive the error propagation equations mostly from a statistical point of view. We intend to obtain the propagation relationship and establish a description for the statistical distribution of relative errors.

Let $V_i^0(\boldsymbol{\mu}_i)$ be the true vertex positions of the underlying triangle $\Delta V_1^0 V_2^0 V_3^0$, $V_i(\mathbf{X}_i)$ be the corresponding random positions, and $\boldsymbol{\varepsilon}_i$ be the ME vectors, $\boldsymbol{\varepsilon}_i \sim (\mathbf{0}, \boldsymbol{\Sigma}_i)$, $\boldsymbol{\Sigma}_i \equiv \mathrm{cov}(\boldsymbol{\varepsilon}_i)$, $i = 1, 2, 3$. By the results derived in Leung et al. (2003b), the *signed* area $A_{(3)}$ of the random triangle $\Delta V_1 V_2 V_3$ can be represented as

$$A_{(3)} = \tfrac{1}{4}\mathbf{X}_{(3)}^{\mathrm{T}} \overline{\mathbf{H}}_0 \, \mathbf{X}_{(3)}, \tag{3.1}$$

where $\overline{\mathbf{H}}_0 \equiv \boldsymbol{\Delta}_0 \otimes \mathbf{H}_0$ is symmetric, $\boldsymbol{\Delta}_0$, $\mathbf{H}_0$ and the joint coordinate vector $\mathbf{X}_{(3)}$ of vertices of the triangle are respectively given as follows:

$$\boldsymbol{\Delta}_0 = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}, \quad \mathbf{H}_0 \equiv \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{X}_{(3)}^{\mathrm{T}} \equiv (\mathbf{X}_1^{\mathrm{T}} \;\; \mathbf{X}_2^{\mathrm{T}} \;\; \mathbf{X}_3^{\mathrm{T}}).$$

When the coordinates vectors $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ of the three vertices of the triangle are ~~circulated~~ ordered in counter-clockwise fashion, $A_{(3)}$ is positive. Furthermore, we can easily check that the eigenvalues and the corresponding orthogonal normalized eigenvectors of $\overline{\mathbf{H}}_0$ are as follows:

$$(\bar{\lambda}_1 \;\; \bar{\lambda}_2 \;\; \bar{\lambda}_3 \;\; \bar{\lambda}_4 \;\; \bar{\lambda}_5 \;\; \bar{\lambda}_6) = (\sqrt{3} \;\; \sqrt{3} \;\; -\sqrt{3} \;\; -\sqrt{3} \;\; 0 \;\; 0), \tag{3.2}$$

$$9\begin{pmatrix} -\frac{\sqrt{3}}{2} & -\frac{1}{2} & \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 & 1 \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} & -\frac{1}{2} & -\frac{\sqrt{3}}{2} & 1 & 0 \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} & -\frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 & 1 \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} & -\frac{1}{2} & \frac{\sqrt{3}}{2} & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

$$\underset{6\times 6}{\mathbf{V}} \equiv \begin{pmatrix} \underset{6\times 1}{\mathbf{v}_1} & \underset{6\times 1}{\mathbf{v}_2} & \underset{6\times 1}{\mathbf{v}_3} & \underset{6\times 1}{\mathbf{v}_4} & \underset{6\times 1}{\mathbf{v}_5} & \underset{6\times 1}{\mathbf{v}_6} \end{pmatrix} = \frac{1}{\sqrt{3}} \quad , \tag{3.3}$$

that is, $\overline{\mathbf{H}}_0 \mathbf{v}_i = \overline{\lambda}_i \mathbf{v}_i$, $i = 1,\ldots,6$.

Under the effect of ME, the area measurement model for a triangle is

$$\begin{cases} A_{(3)} = f(\mathbf{X}_{(3)}) = \frac{1}{4}\mathbf{X}_{(3)}^{\mathrm{T}}\overline{\mathbf{H}}_0\, \mathbf{X}_{(3)}, & (3.4) \\ \mathbf{X}_{(3)} = \boldsymbol{\mu}_{(3)} + \boldsymbol{\varepsilon}_{(3)}, \quad \boldsymbol{\varepsilon}_{(3)} \sim (\mathbf{0}, \boldsymbol{\Sigma}_{(3)}), & (3.5) \end{cases}$$

where $\boldsymbol{\mu}_{(3)}^{\mathrm{T}} \equiv (\boldsymbol{\mu}_1^{\mathrm{T}}\ \boldsymbol{\mu}_2^{\mathrm{T}}\ \boldsymbol{\mu}_3^{\mathrm{T}})$ and $\boldsymbol{\varepsilon}_{(3)}^{\mathrm{T}} \equiv (\boldsymbol{\varepsilon}_1^{\mathrm{T}}\ \boldsymbol{\varepsilon}_2^{\mathrm{T}}\ \boldsymbol{\varepsilon}_3^{\mathrm{T}})$ are respectively the joint true coordinates vector and the joint ME vector for three vertices of the triangle. The covariance matrix of $\boldsymbol{\varepsilon}_{(3)}$ is $\boldsymbol{\Sigma}_{(3)}$. In particular, when the ME vectors $\boldsymbol{\varepsilon}_i$ for the vertices are independent, $\boldsymbol{\Sigma}_{(3)} = \mathrm{diag}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3)$.

According to Lemma 2.1, the area of the triangle $\Delta V_1V_2V_3$, $A_{(3)}$, is distributed as a linear combination of independent noncentral chi-square variables under the assumption that the joint ME $\boldsymbol{\varepsilon}_{(3)}$ are normal. That is, we have:

Proposition 3.1 Assume that $\boldsymbol{\varepsilon}_{(3)} \sim N_6(0, \boldsymbol{\Sigma}_{(3)})$. Then $A_{(3)}$ can be represented as

$$A_{(3)} = \frac{1}{4}\sum_{i=1}^{m} \lambda_i Z_i, \quad Z_i \sim \chi^2_{p_i,\delta_i^2}, \tag{3.6}$$

where Z_i are independent, $\lambda_i \equiv \lambda_i(\overline{\mathbf{H}}_0\boldsymbol{\Sigma}_{(3)})$ are the distinct non-zero eigenvalues of $\overline{\mathbf{H}}_0\boldsymbol{\Sigma}_{(3)}$, p_i their respective orders of multiplicity, and $\delta_i^2 \equiv \delta_i^2(\boldsymbol{\Sigma}_{(3)}, \boldsymbol{\mu}_{(3)})$ are determined by $\boldsymbol{\Sigma}_{(3)}$ and $\boldsymbol{\mu}_{(3)}$ (see (2.5)).

From Proposition 3.1 and the properties of the noncentral chi-square distribution, we further have

$$E(A_{(3)}) = \frac{1}{4}\sum_{i=1}^{m} \lambda_i (p_i + \delta_i^2), \tag{3.7}$$

$$\mathrm{var}(A_{(3)}) = \frac{1}{8}\sum_{i=1}^{m} \lambda_i^2 (p_i + 2\delta_i^2). \tag{3.8}$$

Equation (3.8) is indeed the exact law of error propagation for area measurement of a triangle. To express the propagation relationship more clearly, the variance of the area $\sigma^2_{\mathrm{area}(3)} \equiv \mathrm{Var}(A_{(3)})$ in (3.8) can be written as

$$\sigma^2_{\mathrm{area}(3)} = F(\boldsymbol{\Sigma}_{(3)}; \boldsymbol{\mu}_{(3)}) \equiv \frac{1}{8}\sum_{i=1}^{m} \lambda_i^2(\overline{\mathbf{H}}_0\boldsymbol{\Sigma}_{(3)})[p_i + 2\delta_i^2(\boldsymbol{\Sigma}_{(3)}, \boldsymbol{\mu}_{(3)})], \tag{3.9}$$

where $F(\cdot;\cdot)$ is a generic notation for representing the dependence of the area error variance $\sigma^2_{\mathrm{area}(3)}$ on the joint ME covariance matrix $\boldsymbol{\Sigma}_{(3)}$ and the joint true vector $\boldsymbol{\mu}_{(3)}$ of the triangle.

From Proposition 3.1, the density function of $A_{(3)}$ can also be obtained by the methods in Mathai and Provost (1992).

In particular, when three vertices of a triangle have homogeneous ME covariance structure $\sigma^2\mathbf{I}_2$ and are independent, then $\mathbf{\Sigma}_{(3)} = \sigma^2\mathbf{I}_6$, $\lambda_i(\overline{\mathbf{H}}_0\mathbf{\Sigma}_{(3)}) = \sigma^2\lambda_i(\overline{\mathbf{H}}_0) = \pm\sqrt{3}\sigma^2$, $p_1 = p_2 = 2$, and from (2.5), we have

$$\delta_1^2 = \frac{1}{\sigma^2}\left[\begin{pmatrix}\mathbf{v}_1^{\mathrm{T}}\\ \mathbf{v}_2^{\mathrm{T}}\end{pmatrix}\boldsymbol{\mu}_{(3)}\right]^{\mathrm{T}}\begin{pmatrix}\mathbf{v}_1^{\mathrm{T}}\\ \mathbf{v}_2^{\mathrm{T}}\end{pmatrix}\boldsymbol{\mu}_{(3)} = \frac{1}{\sigma^2}\left[(\boldsymbol{\mu}_{(3)}^{\mathrm{T}}\mathbf{v}_1)^2 + (\boldsymbol{\mu}_{(3)}^{\mathrm{T}}\mathbf{v}_2)^2\right], \tag{3.10}$$

$$\delta_2^2 = \frac{1}{\sigma^2}\left[\begin{pmatrix}\mathbf{v}_3^{\mathrm{T}}\\ \mathbf{v}_4^{\mathrm{T}}\end{pmatrix}\boldsymbol{\mu}_{(3)}\right]^{\mathrm{T}}\begin{pmatrix}\mathbf{v}_3^{\mathrm{T}}\\ \mathbf{v}_4^{\mathrm{T}}\end{pmatrix}\boldsymbol{\mu}_{(3)} = \frac{1}{\sigma^2}\left[(\boldsymbol{\mu}_{(3)}^{\mathrm{T}}\mathbf{v}_3)^2 + (\boldsymbol{\mu}_{(3)}^{\mathrm{T}}\mathbf{v}_4)^2\right]. \tag{3.11}$$

In this case, (3.6), (3.7) and the error propagation equation (3.8) become respectively,

$$A_{(3)} = \tfrac{1}{4}\sum_{i=1}^{2}\lambda_i Z_i = \tfrac{\sqrt{3}}{4}(Z_1 - Z_2)\sigma^2,\quad Z_1 \sim \chi^2_{2,\delta_1^2},\quad Z_2 \sim \chi^2_{2,\delta_2^2}, \tag{3.12}$$

$$E[A_{(3)}] = \tfrac{\sqrt{3}}{4}\left[(\boldsymbol{\mu}_{(3)}^{\mathrm{T}}\mathbf{v}_1)^2 + (\boldsymbol{\mu}_{(3)}^{\mathrm{T}}\mathbf{v}_2)^2 - (\boldsymbol{\mu}_{(3)}^{\mathrm{T}}\mathbf{v}_3)^2 - (\boldsymbol{\mu}_{(3)}^{\mathrm{T}}\mathbf{v}_4)^2\right], \tag{3.13}$$

and

$$\sigma^2_{\mathrm{area}(3)} = \overline{F}(\sigma^2;\boldsymbol{\mu}_{(3)}) = \tfrac{3}{2}\sigma^4 + \tfrac{3}{4}\sigma^2\sum_{i=1}^{4}(\boldsymbol{\mu}_{(3)}^{\mathrm{T}}\mathbf{v}_i)^2\,. \tag{3.14}$$

This equation indicates that the variance of area measurement error of a triangle is not only related to the variance of the locational error of the triangle vertices, but also related to the true locations of the vertices. This conclusion may give a valuable insight ~~on~~ into area measurement of a triangle. In addition, it can be observed that (3.13) is independent of the variance σ^2 of the coordinate ME and may change its sign when the order ~~in which~~ of the three true vertices forming the joint true vector $\boldsymbol{\mu}_{(3)}$ is changed. In the following subsection, we will give more general results on these observations.

3.2 Error analysis for area measurement of a polygon

Area measurement of a polygon (simple) is a basic problem in GIS applications. For a complex polygon, its area is always divided into the sum of areas of the constituent simple polygons. In general, the area of any polygon, whether it is convex or not, can be computed by summing the areas of a series of triangles generated by connecting all the edges of the polygon to a single point. It should be noted that the area of the triangle/parallelogram returned by this algorithm is signed, that is, we will have to sum a series in which both positive and negative values may exist. However, it is also feasible to compute directly the area of a polygon.

First, the signed area of a polygon can be similarly defined. The area of a polygon is defined to be positive if the points are arranged in a counterclockwise order, and negative if they are in a clockwise order (Beyer, 1987).

If the coordinates of vertex V_i of a simple n-sided polygon are (x_{i1}, x_{i2}), $i=1,\ldots,n$, the signed area of this polygon is given by

$$A_{(n)} = \frac{1}{2}\sum_{i=1}^{n}(x_{i1}x_{i+1,2} - x_{i2}x_{i+1,1}), \tag{3.15}$$

where n is the number of vertices of the polygon. Nevertheless, we have to be careful to "close the path" by replacing $V_{n+1} = V_1$ (or replacing $x_{n+1,1}$ by x_{11}, and $x_{n+1,2}$ by x_{12}). It should be noted that we need a simple closed curve which is a curve that does not intersect itself. Otherwise parts of the curve might not traverse in a counter-clockwise direction, which would give the wrong signed area. The proof of this can be obtained by induction.

The same formula can also be used to calculate the *signed area* of complex polygons: follow the sequence of points and count area to the left of your path positive, to the right negative. An advantage of the formula (3.15) for computing the signed area is that it can be performed in linear time by a simple sum.

To derive the error propagation equation for area measurement of a polygon, we need to generalize the result obtained in a triangle. Consider an n-sided polygon with the true vertices $V_i^0(\boldsymbol{\mu}_i)$ and the random vertices $V_i(\mathbf{X}_i)$ under the ME vectors $\boldsymbol{\varepsilon}_i \equiv (\varepsilon_{i1}, \varepsilon_{i2})^{\mathrm{T}}$, $\boldsymbol{\mu}_i \equiv (\mu_{i1}, \mu_{i2})^{\mathrm{T}}$, $\mathbf{X}_i \equiv (X_{i1}, X_{i2})^{\mathrm{T}}$, $i=1,\ldots,n$, $V_{n+1} = V_1$. The corresponding joint vectors are denoted by $\mathbf{X}_{(n)} \equiv (\mathbf{X}_1^{\mathrm{T}}, \cdots, \mathbf{X}_n^{\mathrm{T}})^{\mathrm{T}}$, $\boldsymbol{\mu}_{(n)} \equiv (\boldsymbol{\mu}_1^{\mathrm{T}}, \cdots, \boldsymbol{\mu}_n^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{\varepsilon}_{(n)} \equiv (\boldsymbol{\varepsilon}_1^{\mathrm{T}}, \cdots, \boldsymbol{\varepsilon}_n^{\mathrm{T}})^{\mathrm{T}}$. First we express each $\mathbf{X}_i$ by the joint coordinate vector $\mathbf{X}_{(n)}$:

$$\mathbf{X}_i = (\mathbf{0}_2 \ \cdots \ \mathbf{0}_2 \ \mathbf{I}_2 \ \mathbf{0}_2 \ \cdots \ \mathbf{0}_2)\,\mathbf{X}_{(n)} \equiv \underset{2\times(2n)}{(\mathbf{e}_{n,i}^{\mathrm{T}} \otimes \mathbf{I}_2)}\,\mathbf{X}_{(n)}. \tag{3.16}$$

By (2.6) in Leung et al. (2003c), we can obtain

$$\sum_{i=1}^{n}(X_{i1}X_{i+1,2} - X_{i2}X_{i+1,1}) = \sum_{i=1}^{n}\begin{vmatrix} X_{i1} & X_{i+1,1} \\ X_{i2} & X_{i+1,2} \end{vmatrix} = \sum_{i=1}^{n}\left| \mathbf{X}_i \quad \mathbf{X}_{i+1} \right| = \sum_{i=1}^{n} f_{\det}(\mathbf{X}_i, \mathbf{X}_{i+1})$$

$$= \sum_{i=1}^{n}\mathbf{X}_i^{\mathrm{T}}\mathbf{H}_0\mathbf{X}_{i+1} = \sum_{i=1}^{n}\mathbf{X}_{(n)}^{\mathrm{T}}(\mathbf{e}_{n,i}^{\mathrm{T}} \otimes \mathbf{I}_2)^{\mathrm{T}}\mathbf{H}_0(\mathbf{e}_{n,i+1}^{\mathrm{T}} \otimes \mathbf{I}_2)\mathbf{X}_{(n)}$$

$$= \mathbf{X}_{(n)}^{\mathrm{T}}\left[\sum_{i=1}^{n}(\mathbf{e}_{n,i} \otimes \mathbf{I}_2)\,\mathbf{H}_0\,(\mathbf{e}_{n,i+1}^{\mathrm{T}} \otimes \mathbf{I}_2)\right]\mathbf{X}_{(n)} = \mathbf{X}_{(n)}^{\mathrm{T}}\left[\sum_{i=1}^{n}(\mathbf{e}_{n,i} \otimes \mathbf{H}_0)\,(\mathbf{e}_{n,i+1}^{\mathrm{T}} \otimes \mathbf{I}_2)\right]\mathbf{X}_{(n)}$$

$$= \mathbf{X}_{(n)}^{\mathrm{T}} \left[\sum_{i=1}^{n} (\mathbf{e}_{n,i}\ \mathbf{e}_{n,i+1}^{\mathrm{T}}) \otimes (\mathbf{H}_0 \mathbf{I}_2) \right] \mathbf{X}_{(n)} = \mathbf{X}_{(n)}^{\mathrm{T}} \left[\left(\sum_{i=1}^{n} \mathbf{e}_{n,i}\ \mathbf{e}_{n,i+1}^{\mathrm{T}} \right) \otimes \mathbf{H}_0 \right] \mathbf{X}_{(n)} .$$

Since

$$\sum_{i=1}^{n} \mathbf{e}_{n,i} \mathbf{e}_{n,i+1}^{\mathrm{T}} = \sum_{i=1}^{n} \begin{pmatrix} 0 & & & \cdots & & & 0 \\ & \ddots & & & & \cdot^{\cdot^{\cdot}} & \\ & & 0 & 0 & 0 & & \\ \vdots & & 0 & 1 & 0 & & \vdots \\ & & 0 & 0 & 0 & & \\ & \cdot^{\cdot^{\cdot}} & & & & \ddots & \\ 0 & & & \cdots & & & 0 \end{pmatrix} \begin{matrix} 1 \\ \vdots \\ i-1 \\ i \\ i+1 \\ \vdots \\ n \end{matrix} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ & 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & & & \ddots & 1 \\ 1 & 0 & \cdots & & 0 \end{pmatrix} \equiv \mathbf{G}_0 \tag{3.17}$$

(columns: $1 \quad \cdots \quad i \quad i+1 \quad i+2 \quad \cdots \quad n$)

($\mathbf{G}_0$ is just the forward-shift matrix $\mathbf{\Pi}_1$ (Turkington, 2002) and is also a permutation matrix.), we have

$$A_{(n)} = \tfrac{1}{2} \mathbf{X}_{(n)}^{\mathrm{T}} (\mathbf{G}_0 \otimes \mathbf{H}_0)\, \mathbf{X}_{(n)} .$$

It can be observed that $A_{(n)}$ may also be denoted as $A_{(n)} = \tfrac{1}{2} \mathbf{X}_{(n)}^{\mathrm{T}} (\mathbf{G}_0 \otimes \mathbf{H}_0)^{\mathrm{T}}\ \mathbf{X}_{(n)}$. Then $A_{(n)}$ can be expressed as a quadratic form:

$$A_{(n)} = \tfrac{1}{4} \mathbf{X}_{(n)}^{\mathrm{T}} [(\mathbf{G}_0 \otimes \mathbf{H}_0) + (\mathbf{G}_0 \otimes \mathbf{H}_0)^{\mathrm{T}}]\, \mathbf{X}_{(n)} = \tfrac{1}{4} \mathbf{X}_{(n)}^{\mathrm{T}} [(\mathbf{G}_0 \otimes \mathbf{H}_0) + (\mathbf{G}_0^{\mathrm{T}} \otimes \mathbf{H}_0^{\mathrm{T}})]\, \mathbf{X}_{(n)}$$

$$= \tfrac{1}{4} \mathbf{X}_{(n)}^{\mathrm{T}} [(\mathbf{G}_0 - \mathbf{G}_0^{\mathrm{T}}) \otimes \mathbf{H}_0]\, \mathbf{X}_{(n)} = \tfrac{1}{4} \mathbf{X}_{(n)}^{\mathrm{T}} \mathbf{H}_{(n)} \mathbf{X}_{(n)} , \tag{3.18}$$

where $\mathbf{H}_{(n)} \equiv \mathbf{\Delta}_{(n)} \otimes \mathbf{H}_0$,

$$\mathbf{\Delta}_{(n)} \equiv \mathbf{G}_0 - \mathbf{G}_0^{\mathrm{T}} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & -1 \\ -1 & 0 & 1 & \ddots & \vdots & 0 \\ 0 & -1 & 0 & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \vdots & \ddots & \ddots & 0 & 1 \\ 1 & 0 & \cdots & 0 & -1 & 0 \end{pmatrix}_{n \times n} , \quad n \geq 3, \qquad \mathbf{H}_0 \equiv \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{3.19}$$

which is an anti-symmetric matrix. It can easily be checked that $\mathbf{H}_{(n)}$ is symmetric. In particular, when $n = 3$, $\mathbf{\Delta}_{(n)}$ becomes $\mathbf{\Delta}_0$ in (3.1). Therefore, (3.18) is the generalization of (3.1). And the area measurement model for a polygon can be formulated as:

$$\begin{cases} A_{(n)} = f(\mathbf{X}_{(n)}) = \tfrac{1}{4} \mathbf{X}_{(n)}^{\mathrm{T}} \mathbf{H}_{(n)}\ \mathbf{X}_{(n)} , & (3.20) \\ \mathbf{X}_{(n)} = \boldsymbol{\mu}_{(n)} + \boldsymbol{\varepsilon}_{(n)} ,\ \ \boldsymbol{\varepsilon}_{(n)} \sim (\mathbf{0},\ \mathbf{\Sigma}_{(n)}) . & (3.21) \end{cases}$$

The first problem to be answered is that whether area measurement $A_{(n)}$ is an unbiased estimator of its true value $A_{(n)}^0 = \tfrac{1}{4} \boldsymbol{\mu}_{(n)}^{\mathrm{T}} \mathbf{H}_{(n)} \boldsymbol{\mu}_{(n)}$, i.e., whether $E(A_{(n)}) = A_{(n)}^0$ holds. For the unbiasedness of area measurement, there are many discussions in the literature (see for example, Chrisman and Yandell, 1988, Griffith, 1989, Prisley et al., 1989). However, the conditions given in these works are only sufficient conditions for unbiased area measurement and they need the independence assumptions or some special correlation structure (e.g. autocorrleation) among vertices. As a generality, we now give a necessary and sufficient condition for unbiased area measurement, without having to impose additional assumptions on the error covariance structure. Since area measurement $A_{(n)}$ can be expressed by quadratic forms, the derivation of this condition is very simple. In fact, Seber (1977) has shown that for any symmetric matrix $\mathbf{A}$ and random vector $\mathbf{X}$ with $E(\mathbf{X}) = \boldsymbol{\mu}$ and $\operatorname{cov}(\mathbf{X}) = \boldsymbol{\Sigma}$, the following equality always holds:

$$E(\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X}) = \boldsymbol{\mu}^{\mathrm{T}}\mathbf{A}\boldsymbol{\mu} + \operatorname{tr}(\mathbf{A}\boldsymbol{\Sigma}) .$$

Thus, from (3.20), we have

$$E(A_{(n)}) = \tfrac{1}{4}E(\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{H}_{(n)}\mathbf{X}_{(n)}) = \tfrac{1}{4}[\boldsymbol{\mu}_{(n)}^{\mathrm{T}}\mathbf{H}_{(n)}\boldsymbol{\mu}_{(n)} + \operatorname{tr}(\mathbf{H}_{(n)}\boldsymbol{\Sigma}_{(\mathrm{n})})] = A_{(n)}^{0} + \tfrac{1}{4}\operatorname{tr}(\mathbf{H}_{(n)}\boldsymbol{\Sigma}_{(\mathrm{n})}) .$$

Proposition 3.2 Under the effect of ME, the area estimate $A_{(n)}$ given by (3.20) is unbiased if and only if $\operatorname{tr}(\mathbf{H}_{(n)}\boldsymbol{\Sigma}_{(\mathrm{n})}) = 0$. When $\operatorname{tr}(\mathbf{H}_{(n)}\boldsymbol{\Sigma}_{(\mathrm{n})}) \neq 0$, $A_{(n)}$ is biased and the bias is

$$\operatorname{bias}(A_{(n)}) = \tfrac{1}{4}\operatorname{tr}(\mathbf{H}_{(n)}\boldsymbol{\Sigma}_{(\mathrm{n})}) . \tag{3.22}$$

Remark 1. (3.22) shows that the bias in area measurement of a polygon depends only on the covariance matrix of the joint ME vector of the vertices and is independent of the true joint locational vector of the polygon.

In particular, when the ME vectors $\boldsymbol{\varepsilon}_i$ at each vertex are independent and identically distributed, the joint ME covariance matrix $\boldsymbol{\Sigma}_{(n)}$ can be written as

$$\boldsymbol{\Sigma}_{(n)} = \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\Sigma} \end{pmatrix} = \mathbf{I}_n \otimes \boldsymbol{\Sigma} .$$

Accordingly, $\operatorname{tr}(\mathbf{H}_{(n)}\boldsymbol{\Sigma}_{(\mathrm{n})}) = \operatorname{tr}[(\boldsymbol{\Delta}_{(n)} \otimes \mathbf{H}_0)(\mathbf{I}_n \otimes \boldsymbol{\Sigma})] = \operatorname{tr}[\boldsymbol{\Delta}_{(n)} \otimes (\mathbf{H}_0\boldsymbol{\Sigma})] = \operatorname{tr}(\boldsymbol{\Delta}_{(n)}) \cdot \operatorname{tr}(\mathbf{H}_0\boldsymbol{\Sigma}) = 0$ (see (3.19)). By Proposition 3.2, the area measurement $A_{(n)}$ is unbiased. This is actually a result in Chrisman and Yandell (1988).

When

$$\boldsymbol{\Sigma}_{(n)} = \begin{pmatrix} \sigma_1^2\mathbf{I}_2 & \sigma_{12}\mathbf{I}_2 & \mathbf{0} & \cdots & \sigma_{n,1}\mathbf{I}_2 \\ \sigma_{12}\mathbf{I}_2 & \sigma_2^2\mathbf{I}_2 & \ddots & \ddots & \vdots \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \sigma_{n-1}^2\mathbf{I}_2 & \sigma_{n-1,n}\mathbf{I}_2 \\ \sigma_{n,1}\mathbf{I}_2 & \cdots & \mathbf{0} & \sigma_{n-1,n}\mathbf{I}_2 & \sigma_n^2\mathbf{I}_2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & 0 & \cdots & \sigma_{n,1} \\ \sigma_{12} & \sigma_2^2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \sigma_{n-1}^2 & \sigma_{n-1,n} \\ \sigma_{n,1} & \cdots & 0 & \sigma_{n-1,n} & \sigma_n^2 \end{pmatrix} \otimes \mathbf{I}_2 \equiv \mathbf{S} \otimes \mathbf{I}_2 ,$$

$\operatorname{tr}(\mathbf{H}_{(n)}\boldsymbol{\Sigma}_{(\mathrm{n})}) = \operatorname{tr}[(\boldsymbol{\Delta}_{(n)} \otimes \mathbf{H}_0)(\mathbf{S} \otimes \mathbf{I}_2)] = \operatorname{tr}[(\boldsymbol{\Delta}_{(n)}\mathbf{S}) \otimes \mathbf{H}_0] = \operatorname{tr}(\boldsymbol{\Delta}_{(n)}\mathbf{S}) \cdot \operatorname{tr}(\mathbf{H}_0) = 0$ (see the definition of $\mathbf{H}_0$ in (3.1)). Therefore, by Proposition 3.2 the area measurement $A_{(n)}$ is unbiased. This actually is a result in Prisley et al. (1989). It is not difficult to see that the sufficient conditions discussed in

Griffith (1989) may be derived from Proposition 3.2. It can be observed from the above discussion that a wider and weaker sufficient condition than that of Prisley et al. (1989) can be obtained. That is, as long as $\mathbf{\Sigma}_{(n)}$ can be written as $\mathbf{\Sigma}_{(n)} = \mathbf{M} \otimes \mathbf{I}_2$, where $\mathbf{M}$ is a certain matrix (all of whose elements may be nonzero, unlike **S**), the area measurement $A_{(n)}$ is then unbiased.

Although some matrix expressions for area measurements have been obtained in Griffith (1989), they are not quadratic forms. Accordingly, the related derivations and expressions are not concise, especially for variance of area measurement. A more complex derivation for variance of area measurement is given by Prisley et al. (1989). Their approach is as follows: first, the variance of area of a triangle is determined. Second, the covariance of area between adjacent triangles is derived. By summing the triangle variances and covariances, the polygon variance is obtained. Now we give a concise expression for variance of area measurement, and the condition imposed does not involve any assumption ~~on~~ about the error distribution but only moments of error are of relevance.

Proposition 3.3 If the ME ε_{ij} are independent and have ~~the~~ common second, third and fourth moments about zero~~s~~, i.e., $\mathbf{\Sigma}_{(n)} = \sigma^2 \mathbf{I}_{2n}$, $E(\varepsilon_{ij})^3$ and $E(\varepsilon_{ij})^4$ are constants independent of the subscripts i and j, then

$$\sigma^2_{\text{area}(n)} \equiv \text{var}(A_{(n)}) = \tfrac{n}{2}\sigma^4 + \tfrac{1}{4}\sigma^2 \boldsymbol{\mu}^{\text{T}}_{(n)} \mathbf{H}^2_{(n)} \boldsymbol{\mu}_{(n)} . \tag{3.23}$$

(the proof is given in Appendix 2).

Equality (3.23) is indeed the exact law of error propagation for area measurement under the above conditions, where σ^2 is the variance of the ME ε_{ij} of the vertex coordinates. Obviously, (3.23) can be expressed as a function of σ^2, i.e., $\sigma^2_{\text{area}(n)} = f_n(\sigma^2; \boldsymbol{\mu}_{(n)})$, which indicates an apparent error propagation relation. When $n = 3$, (3.23) becomes (3.14).

In general, under the condition that no specific distribution is assumed for the joint ME vector $\boldsymbol{\varepsilon}_{(n)}$, any moments of area measurement with nonsingular $\mathbf{\Sigma}_{(n)}$ may be obtained from the results in Mathai and Provost (1992), that is,

Comment [v1]: Page 49

$$E[A_{(n)}]^r = \frac{1}{4^r}\sum_{r_1+\cdots+r_{2n}=r}\cdots\sum \frac{r!\lambda_1^{r_1}\cdots\lambda_{2n}^{r_{2n}}}{r_1!\cdots r_{2n}!}E[Z_1^{r_1}\cdots Z_{2n}^{r_{2n}}],$$

where $\lambda_1,\cdots,\lambda_{2n}$ are the eigenvalues of $\mathbf{H}_{(n)}\mathbf{\Sigma}_{(n)}$, $Z_i=(Y_i+b_i)^2$, Y_i 's are non-correlated with $EY_i=0$ and $\mathrm{var}(Y_i)=1$, $\mathbf{b}^{\mathrm{T}}=(b_1,\cdots,b_{2n})=\boldsymbol{\mu}_{(n)}^{\mathrm{T}}\mathbf{\Sigma}_{(n)}^{-\frac{1}{2}}\mathbf{C}$, $\mathbf{CC}^{\mathrm{T}}=\mathbf{I}$.

Furthermore, if we assume that the distribution of $\boldsymbol{\varepsilon}_{(n)}$ is normal, then not only can we obtain the distribution of area measurement but the exact law of error propagation can also be ascertained (see the following proposition which is similar to Proposition 3.1).

Proposition 3.4 Assume that $\boldsymbol{\varepsilon}_{(n)} \sim N_{2n}(\mathbf{0},\mathbf{\Sigma}_{(n)})$. Then the area of polygon, $A_{(n)}$, in (3.20) can be expressed as a linear combination of independent noncentral chi-square random variables Z_i, i.e.,

$$A_{(n)}=\frac{1}{4}\sum_{i=1}^{m}\lambda_i Z_i\,,\quad Z_i\sim\chi^2_{p_i,\delta_i^2}\,, \tag{3.24}$$

where $\lambda_i\equiv\lambda_i(\mathbf{H}_{(n)}\mathbf{\Sigma}_{(n)})$ are the distinct non-zero eigenvalues of $\mathbf{H}_{(n)}\mathbf{\Sigma}_{(n)}$, p_i and $\delta_i^2\equiv\delta_i^2(\mathbf{\Sigma}_{(n)},\boldsymbol{\mu}_{(n)})$ are similarly determined by $\mathbf{\Sigma}_{(n)}$ and $\boldsymbol{\mu}_{(n)}$, just like Proposition 3.1.

Remark 2. Although the normality condition is still assumed in the proposition, this condition is different from the normality assumption usually used in the literature (e.g., Prisley et al. (1989)), which often means that the ME vectors $\boldsymbol{\varepsilon}_i$ at each vertex are independent and each $\boldsymbol{\varepsilon}_i$ is normal. Our normality condition is that the joint ME vector $\boldsymbol{\varepsilon}_{(n)}$ is normal. It implies that $\boldsymbol{\varepsilon}_i$ may not be independent. In other words, under the condition of Proposition 3.4, the ME vectors among vertices may or may not be independent. We think that the normality assumption for error distributions may be more reasonable than the independence assumption in practice.

Under the condition that the joint ME vector is normal, the expectation and variance formulae of $A_{(n)}$ are unchanged just as (3.7) and (3.8), only the involved parameters are replaced by the present corresponding values. In this case, the resulting exact law of error propagation can be written as

$$\sigma^2_{\mathrm{area}(n)}=F_n(\mathbf{\Sigma}_{(n)};\boldsymbol{\mu}_{(n)})\equiv\frac{1}{8}\sum_{i=1}^{m}\lambda_i^2(\mathbf{H}_{(n)}\mathbf{\Sigma}_{(n)})[p_i+2\delta_i^2(\mathbf{\Sigma}_{(n)},\boldsymbol{\mu}_{(n)})], \tag{3.25}$$

where $F_n(\cdot;\cdot)$ is similar to $F(\cdot;\cdot)$ in (3.9). Since (3.25) involves the computation of eigenvalues and eigenvectors, it is not so convenient to apply and the relation of error propagation is not so clear or apparent to understand. When our concern is the variance of area measurement rather than the error distribution of area measurement, a simpler expression without the computation of eigenvalues and eigenvectors is needed. In fact, according to Mathai and Provost (1992), (3.25) can be expressed as

Comment [v2]: Page 53 Theorem 3.2b.2

$$\sigma^2_{\text{area}(n)} = F_n(\mathbf{\Sigma}_{(n)}; \boldsymbol{\mu}_{(n)}) \equiv \tfrac{1}{8}\text{tr}[(\mathbf{H}_{(n)}\mathbf{\Sigma}_{(n)})^2] + \tfrac{1}{4}\boldsymbol{\mu}^{\text{T}}_{(n)}\mathbf{H}_{(n)}\mathbf{\Sigma}_{(n)}\mathbf{H}_{(n)}\boldsymbol{\mu}_{(n)}, \tag{3.26}$$

which satisfies the above requirement. And it can be seen that (3.23) is a special case of (3.26). This equation shows again that under the assumption that the joint ME vector is normal, the error variance $\sigma^2_{\text{area}(n)}$ for area measurement of an *n*-sided polygon is not only related to the joint locational error covariance matrix $\mathbf{\Sigma}_{(n)}$ of its vertices, but also to the true locations $\boldsymbol{\mu}_{(n)}$ of its vertices. Therefore, the important revelation is that in the study of error analysis for area measurement, the methods which only consider the ME vectors at vertices without involving the true locations are problematic.

We summarize this discussion in~~to~~ the following statement:

Proposition 3.5 Assume that $\boldsymbol{\varepsilon}_{(n)} \sim N_{2n}(\mathbf{0}, \mathbf{\Sigma}_{(n)})$. Then the exact law of error propagation for area measurement $A_{(n)}$ is given by (3.26).

In particular, when all ME vectors are normal, independent and have homogeneous ME covariance structure $\mathbf{\Sigma}_\sigma$, i.e., $\mathbf{\Sigma}_i = \mathbf{\Sigma}_\sigma$, then $\mathbf{\Sigma}_{(n)} = \sigma^2\mathbf{I}_{2n}$. If the error distribution (3.24) is of concern, the eigenvalues and eigenvectors should be computed by $\lambda_i(\mathbf{H}_{(n)}\mathbf{\Sigma}_{(n)}) = \sigma^2\lambda_i(\mathbf{H}_n)$ and

$$\delta_i^2 = \frac{1}{\sigma^2}\sum_{k=1}^{p_i}[\boldsymbol{\mu}^{\text{T}}_{(n)}\mathbf{v}_k(\mathbf{H}_{(n)})]^2, \quad i = 1, \ldots, m,$$

where $\lambda_i(\mathbf{H}_{(n)})$ and $\mathbf{v}_k(\mathbf{H}_{(n)})$, independent of the observation locations of a *n*-sided polygon, are respectively the non-zero eigenvalues and the corresponding orthogonal normalized eigenvectors of $\mathbf{H}_{(n)}$. Meanwhile, in this case the exact law of error propagation (3.26) becomes

$$\sigma^2_{\text{area}(n)} = f_n(\sigma^2; \boldsymbol{\mu}_{(n)}) \equiv \tfrac{1}{8}\sigma^4\text{tr}(\mathbf{H}^2_{(n)}) + \tfrac{1}{4}\sigma^2\boldsymbol{\mu}^{\text{T}}_{(n)}\mathbf{H}^2_{(n)}\boldsymbol{\mu}_{(n)} = \tfrac{n}{2}\sigma^4 + \tfrac{1}{4}\sigma^2\boldsymbol{\mu}^{\text{T}}_{(n)}\mathbf{H}^2_{(n)}\boldsymbol{\mu}_{(n)}, \tag{3.27}$$

which is consistent with (3.23).

It should be noted that the exact law of error propagation (3.26) holds for a general ME covariance matrix $\mathbf{\Sigma}_{(n)}$ as long as $\boldsymbol{\varepsilon}_{(n)} \sim N_{2n}(\mathbf{0}, \mathbf{\Sigma}_{(n)})$. In other words, as indicated in Remark 1, this assumption allows the ME vectors at vertices of the polygon to be dependent. Thus it is suitable for the analysis of error in area measurement for the overlaid polygons.

In practice, when there is little knowledge about the error characteristics, we have to appeal to the approximate law of error propagation in Leung et al. (2003a). Since area measurement $A_{(n)} = f(\mathbf{X}_{(n)})$ in (3.20) is a quadratic form in $\mathbf{X}_{(n)}$, the transformation function *f* is nonlinear. The approximate law of error propagation can be derived.

From (3.20), it is clear that $\mathrm{d}(A_{(n)}) = \mathrm{d}[f(\mathbf{X}_{(n)})] = \frac{1}{4}\mathrm{d}[\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{H}_{(n)}\ \mathbf{X}_{(n)}] = \frac{1}{2}\mathbf{X}_{(n)}^{\mathrm{T}}\mathbf{H}_{(n)}\ \mathrm{d}(\mathbf{X}_{(n)})$. So the corresponding Jacobian matrix of *f* at the true vector $\boldsymbol{\mu}_{(n)}$ is

$$\mathbf{B}_{\mu_{(n)}} \equiv \tfrac{1}{2}\boldsymbol{\mu}_{(n)}^{\mathrm{T}}\mathbf{H}_{(n)}. \tag{3.28}$$

According to the approximate law of error propagation in Leung et al. (2003a), the approximate law of error propagation for area measurement of a *n*-sided polygon is

$$\tilde{\sigma}_{\mathrm{area}(n)}^2 = \mathbf{B}_{\mu_{(n)}}\boldsymbol{\Sigma}_{(n)}\mathbf{B}_{\mu_{(n)}}^{\mathrm{T}} = \tfrac{1}{4}\boldsymbol{\mu}_{(n)}^{\mathrm{T}}\mathbf{H}_{(n)}\boldsymbol{\Sigma}_{(n)}\mathbf{H}_{(n)}\boldsymbol{\mu}_{(n)}. \tag{3.29}$$

Under the condition that the joint ME vector is normal, it can be observed that the difference between the exact law (3.26) and the approximate law (3.29) is $\frac{1}{8}\mathrm{tr}[(\mathbf{H}_{(n)}\boldsymbol{\Sigma}_{(n)})^2]$. Furthermore, when $\boldsymbol{\Sigma}_{(n)} = \sigma^2\mathbf{I}_{2n}$, the difference becomes $\frac{n}{2}\sigma^4$. Thus the bigger ~~is~~ σ^2, the bigger the difference becomes.

For the approximate law of error propagation for area measurement formulated in a component-wise way, one can find some studies in the literature (see for example, Zhang and Goodchild, 2002, and Ghilani, 2000).

4. Simulation experiments

Example 4.1 (continued from Example 3.1 in Leung et al. (2003b)) Let $V_1^0\,(0,0)$, $V_2^0\,(3,-1)$, and $V_3^0\,(1,2)$ be vertices of a true triangle. Then the true location vector is $\boldsymbol{\mu}_{(3)} = (0,\ 0,\ 3,\ -1,\ 1,\ 2)^{\mathrm{T}}$. Under the effect of ME, what we observe are $V_i(\mathbf{X}_i)$, $i = 1, 2, 3$. Assume further the corresponding ME vectors $\boldsymbol{\varepsilon}_i$ are independent and $\boldsymbol{\varepsilon}_i \sim N_2(\mathbf{0}, \sigma^2\mathbf{I}_2)$, $i = 1, 2, 3$, that is, $\boldsymbol{\Sigma}_{(3)} = \sigma^2\mathbf{I}_6$. First, we simulate the error propagation for perimeter measurement of this triangle. According to (2.13), we obtain

$$\mathbf{B}_{\mu_{(3)}} = (-1.395897,\ -0.5781994,\ 1.503383, -1.148278,\ -0.1074866,\ 1.726477)^{\mathrm{T}}.$$

So $\tilde{\sigma}^2_{l_{p(3)}}$ can be obtained for different σ^2. Their values and estimation results of 1000 simulations are reported in Table 4.1. It can be seen that as the ME variance σ^2 increases, the difference between $\tilde{\sigma}^2_{l_{p(3)}}$ (corresponding to "Comp." column) and sample variance estimates (corresponding to "Est." column) increases. It reflects to a certain extent the local property of the approximate law. On the other hand, the means of sample perimeters of the random triangles become bigger when σ^2 becomes bigger. It shows that the estimate of the perimeter appears to be biased. This observation is consistent with that of Keefer et al. (1991).

Then we consider the error propagation of area measurement. We still perform the simulation experiment with the sample size of 1000 for each case. The sample variance estimates (corresponding to "Est." column) and the exact variances (corresponding to "Comp." column) given by (3.14) or (3.26) are listed in Table 4.1 for different values of σ^2. Obviously, there is no apparent difference between these two classes. It is evident that the exact law of error propagation for area measurement is effective.

Furthermore, we can obtain the distribution of area measurement $A_{(3)}$. For example, when $\sigma^2 = 0.1$, the non-zero eigenvalues of $\overline{\mathbf{H}}_0 \mathbf{\Sigma}_{(3)}$ and the corresponding non-central parameters are obtained respectively from (3.10) and (3.11) as follows:

$$\lambda_1 = -\lambda_2 = 0.1 \times \sqrt{3}, \qquad p_1 = p_2 = 2,$$
$$\delta_1^2 = 10(14 + 7\sqrt{3})/3 \approx 87.08, \quad \delta_2^2 = 10(14 - 7\sqrt{3})/3 \approx 6.25.$$

If we use (3.7) and (3.8), it can be obtained again that $E(A_{(3)}) = 7/2 = 3.5$, which is just the real area of the true triangle, and $\mathrm{Var}(A_{(3)}) = 143/200 = 0.715$ (or using (3.14)). In this case, the area measurement can be expressed ~~into~~ as a linear combination of two independent noncentral chi-square variables Z_1 and Z_2, i.e.,

$$A_{(3)} = \tfrac{\sqrt{3}}{40}(Z_1 - Z_2), \text{ where } Z_1 \sim \chi^2_{2,87.08} \text{ and } Z_2 \sim \chi^2_{2,6.25}.$$

If the normality assumption is replaced by $\boldsymbol{\varepsilon}_i \sim N_2(\mathbf{0}, \mathbf{\Sigma})$, $\mathbf{\Sigma} = \begin{pmatrix} 0.01 & 0.018 \\ 0.018 & 0.09 \end{pmatrix}$, then we have the non-zero eigenvalues:

$$\lambda_1 = -\lambda_2 \approx 0.0416, \quad p_1 = p_2 = 2, \quad \delta_1^2 = 646.4031, \quad \delta_2^2 = 309.6154.$$

In this case, $E(A_{(3)})$ is unchanged, $\sigma^2_{\text{area(3)}} = 0.4139$ and

$$A_{(3)} = 0.0416(Z_1 - Z_2), \text{ where } Z_1 \sim \chi^2_{2,646.4031} \text{ and } Z_2 \sim \chi^2_{2,309.6154}.$$

Table 4.1 Comparison of simulation and propagation results for different variances of ME*

	$\sigma^2 = 0.01$		$\sigma^2 = 0.05$		$\sigma^2 = 0.10$		$\sigma^2 = 0.15$		$\sigma^2 = 0.20$		$\sigma^2 = 0.25$	
perimater	Est.	Comp.	Est.	Comp.	Est.	Comp.	Est.	Comp.	Est.	Comp.	Est.	Comp.
mean	9.0193	9.0039	9.0938	9.0039	9.1336	9.0039	9.1876	9.0039	9.2560	9.0039	9.3239	9.0039
var	0.0932	0.0885	0.4513	0.4427	0.8900	0.8854	1.2346	1.3281	1.6752	1.7708	1.9799	2.2135
area												
mean	3.4964	3.5000	3.5064	3.5000	3.4968	3.5000	3.5004	3.5000	3.5173	3.5000	3.4249	3.5000
var	0.0667	0.0702	0.3436	0.3538	0.6607	0.7150	1.0306	1.0838	1.5134	1.4600	1.8429	1.8438

* The “Est.” column represents the estimation results and the “Comp. ” column represents the proposed propagation results using (2.13) or (3.14).

Example 4.2 (continued from Example 3.3 in Leung et al (2003b)) Consider a true convex pentagon consisting of five points: V_1^0 (0, 0), V_2^0 (1, 2), V_3^0 (3, –1), V_4^0 (2, –2), and V_5^0 (–0.2, –1.5), and a true concave polygon which has the same vertices as the convex polygon except that V_5^0 is replaced by V_5^0 (1.6, –0.5). Assume that the corresponding ME vectors are independently distributed as $\boldsymbol{\varepsilon}_i \sim N_2(\mathbf{0}, \sigma^2 \mathbf{I}_2)$, $\sigma^2 = 0.1$, $i = 1,...,5$. Thus $\boldsymbol{\Sigma}_{(5)} = \sigma^2 \mathbf{I}_{10}$.

For any 5-sided polygon, the non-zero eigenvalues of $\mathbf{H}_{(5)}$ in (3.20) are:

$$\lambda_{1,2}(\mathbf{H}_{(5)}) = \frac{\sqrt{10+2\sqrt{5}}}{2} = 1.90211, \quad p_{1,2} = 2; \quad \lambda_{3,4}(\mathbf{H}_{(5)}) = \frac{\sqrt{10-2\sqrt{5}}}{2} = 1.17557, \quad p_{3,4} = 2;$$

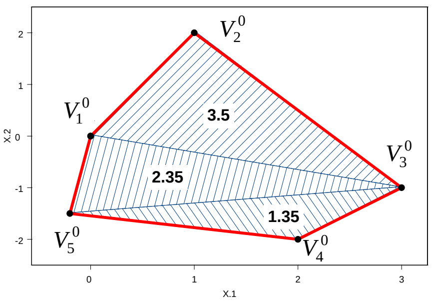
$$\lambda_{5,6}(\mathbf{H}_{(5)}) = -\frac{\sqrt{10-2\sqrt{5}}}{2} = -1.17557, \quad p_{5,6} = 2; \quad \lambda_{7,8}(\mathbf{H}_{(5)}) = -\frac{\sqrt{10+2\sqrt{5}}}{2} = -1.90211, \quad p_{7,8} = 2.$$

However, it is relatively complex to express exactly the corresponding orthogonal normalized vectors. They, nevertheless, can be obtained from some general purpose mathematics software packages if the distribution of area measurement is needed. As an illustration, in Fig. 4.2 we plot respectively the empirical histograms of 10000 simulation data ($\sigma^2 = 0.25$) and the fitting densities given by (3.24) for convex and concave pentagons, where the density (3.24) is obtained by the classical Pearson’s three-moment chi-square approximation (Pearson, 1959). For the purpose of error analysis, we consider only the exact law of error propagation for area measurement. Thus the eigenvalues and eigenvectors are not needed, as shown in (3.26) or (3.27) and (3.29). In Table 4.2, we list only the simulation results with a sample of size 1000 for each case, and the calculated results from (3.27) or (3.26). It should be noted that the estimated and calculated areas of the convex polygon are all negative is because its vertices are arranged in a clockwise order (see Fig. 4.1(a)), that is, $\boldsymbol{\mu}_{(5)} \equiv (\boldsymbol{\mu}_1^{\mathrm{T}}, \cdots, \boldsymbol{\mu}_5^{\mathrm{T}})^{\mathrm{T}} = (0, 0, 1, 2, 3, -1, 2, -2, -0.2, -1.5)^{\mathrm{T}}$. We can also see that when the vertices of the concave polygon are arranged in a counterclockwise order, i.e., $\boldsymbol{\mu}_{(5)} \equiv (\boldsymbol{\mu}_1^{\mathrm{T}}, \boldsymbol{\mu}_5^{\mathrm{T}}, \boldsymbol{\mu}_4^{\mathrm{T}}, \boldsymbol{\mu}_3^{\mathrm{T}}, \boldsymbol{\mu}_2^{\mathrm{T}})^{\mathrm{T}} = (0, 0, 1.6, -0.5, 2, -2, 3, -1, 1, 2)^{\mathrm{T}}$, the estimated and calculated areas are positive.

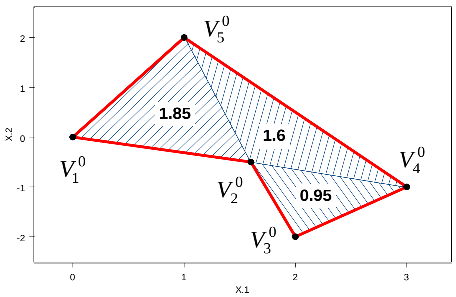
Table 4.2 Comparison of simulations and propagation results for different variances of ME in area measurement *

	$\sigma^2=0.01$		$\sigma^2=0.05$		$\sigma^2=0.10$		$\sigma^2=0.15$		$\sigma^2=0.20$		$\sigma^2=0.25$	
convex	Est.	Comp.	Est.	Comp.	Est.	Comp.	Est.	Comp.	Est.	Comp.	Est.	Comp.
mean	-7.1982	-7.2000	-7.2439	-7.2000	-7.1910	-7.2000	-7.2019	-7.2000	-7.1815	-7.2000	-7.1583	-7.2000
var	0.1460	0.1482	0.7454	0.7460	1.3924	1.5045	2.3099	2.2755	3.0688	3.0590	3.7880	3.8550
concave												
mean	4.3967	4.4000	4.3581	4.4000	4.3794	4.4000	4.3945	4.4000	4.4091	4.4000	4.3402	4.4000
var	0.1067	0.1098	0.5522	0.5540	1.0401	1.1205	1.6202	1.6995	2.4086	2.2910	2.9381	2.8950

* The “Est.” column represents the estimation results and the “Comp. ” Column represents the proposed propagation results.

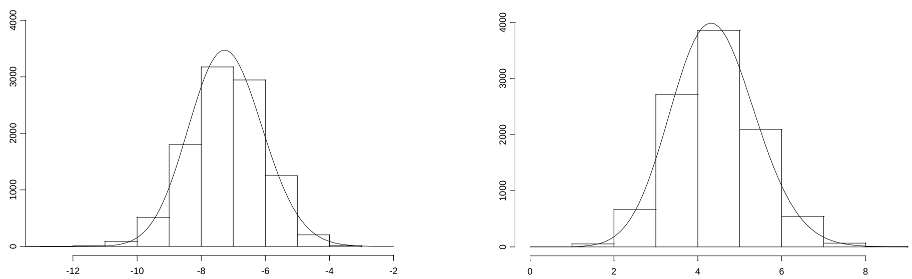


(a) Convex pentagon



(b) Concave pentagon

Fig. 4.1 Area measurements for concave and convex pentagon



(a) For convex pentagon (b) For concave pentagon

Fig. 4.2 Empirical histograms and the fitting densities

To investigate the effect of the joint ME vector with different and dependent structures on area measurement, we assume that the joint ME vector is $\boldsymbol{\varepsilon}_{(5)} \sim N_{10}(\mathbf{0},\ \boldsymbol{\Sigma}_{(5)})$ and choose $\boldsymbol{\Sigma}_{(5)}$ as follows:

$$\boldsymbol{\Sigma}_{(5)} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & & & & \\ & \boldsymbol{\Sigma}_2 & & & \\ & & \boldsymbol{\Sigma}_3 & & \\ & & & \boldsymbol{\Sigma}_4 & 0.01\mathbf{I}_2 \\ & & & 0.01\mathbf{I}_2 & \boldsymbol{\Sigma}_5 \end{pmatrix}, \text{ where } \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}_3 = \begin{pmatrix} 0.01 & 0.018 \\ 0.018 & 0.09 \end{pmatrix},\ \boldsymbol{\Sigma}_4 = \boldsymbol{\Sigma}_5 = \begin{pmatrix} 0.05 & 0 \\ 0 & 0.05 \end{pmatrix}.$$

That is, the ME vectors corresponding to V_1^0, V_2^0, and V_3^0 have an elliptical covariance matrix and are independent; the ME vectors corresponding to V_4^0 and V_5^0 have a circular covariance matrix and are dependent; and any other pair of ME vectors are independent.

According to (3.22), $\text{bias}(A_{(5)}) = \frac{1}{4}\text{tr}(\mathbf{H}_{(5)}\boldsymbol{\Sigma}_{(5)}) = 0.008552$. So the area measurement $A_{(5)}$ is biased. For each pentagon, we run simulations with a sample of size 10000 and the results are tabulated in Table 4.3, where $\overline{A}_{(5)}$ and $\hat{\sigma}^2_{\text{area}(5)}$ denote respectively the sample mean and sample variance of $A_{(5)}$. It can be observed that although the area measurement is biased but it is very small, the mean estimators are still very close to the true area in the two cases. The sample variances of area measurement are consistent with the corresponding exact values given by (3.26) even if the ME vectors at vertices have different structures of the covariance matrix and there are dependences among them. Therefore, the proposed exact law of error propagation for area measurement is effective and powerful for a variety of ME structures.

Table 4.3 Comparison of simulations results and the exact results in area measurement

True	Convex pentagon			True	Concave pentagon		
$A^0_{(5)}$	$\overline{A}_{(5)}$	$\hat{\sigma}^2_{\text{area}(5)}$	$\sigma^2_{\text{area}(5)}$	$A^0_{(5)}$	$\overline{A}_{(5)}$	$\hat{\sigma}^2_{\text{area}(5)}$	$\sigma^2_{\text{area}(5)}$

–7.2	–7.189183	0.6157093	0.618126	4.4	4.393621	0.5999042	0.601026

5. Conclusion

We have discussed error analysis in length and area measurements under ME in locational coordinates. The approximate laws of error propagation in length measurement of a line segment and perimeter measurement of a polygon have been derived in a concise matrix form. Based on the proposed quadratic form expression of area measurement, the necessary and sufficient condition (see Proposition 3.2) for the unbiased area measurement has been obtained and problem (1) raised at the end of Section 1 has been answered. Under the normality assumption of the joint ME vector, we have also derived the statistical distribution of area measurement and the exact law of error propagation for area measurement of a polygon. An important result is that area measurement under the normality assumption is distributed as a linear combination of independent noncentral chi-square variables (see Proposition 3.4) and it forms our answer to problem (2) raised at the end of Section 1. Furthermore, the exact law of error propagation has a simple and clear analytic expression indicating the relation between the variance of area measurement and the covariance matrix of the joint normal ME vector (see (3.26) or Proposition 3.5). Our answer to problem (3) is thus given by Propositions 3.3 and 3.5. These results have largely been substantiated by simulation experiments. While more research is necessary for their further development and applications, the study of the effect of spatial correlation of positional ME on length and area measurements may be of immediate interest.

As a summary, the four-part series of papers focuses on the development of a general framework for error analysis in measurement-based geographic information systems (MBGIS). Major findings are condensed into Fig. 5.1. Overall, the study provides a rigorous statistical approach to measurement error analysis and error propagations through basic GIS and spatial operations. We first construct in Part 1 a basic measurement error model from which relevant concepts such as the approximate law of error propagation, covariance-based error band and maximal allowable limits for positional error are developed. Extend~~ed on~~ing the basic measurement error model, a geodetic model for MBGIS is also constructed for the analysis of MEs under different levels of measurements. The research then proceeds to Part 2 to solve the point-in-polygon analysis under measurement errors. We give a condition under which the point-in-polygon problem is well-defined and a conditional probability can be given. The algebra-based probability model is based on the quadratic forms in the joint coordinate vectors by which whether a point is inside a polygon can be determined. Solution to the general point-in-polygon problem under ME is obtained by triangulating a polygon through which the point-in-triangle problem can first be solved, and the point-in-polygon problem is then solved by summing up the results of the relevant point-in-triangle problems. The algebra-based probability model is simple but rigorous and can circumvent the complexities surrounding the geometric relations between points and convex or non-convex polygons. The quadratic forms are not only used in point-in-polygon analysis under ME, but are also utilized as an identification of intersection points by which polygon-on-polygon problems are solved in Part 3. As a consequence, simple analytic expressions and an approximate law of error propagation for intersection points and polygon-on-polygon overlay are established. The idea of the quadratic forms is further extended to analyze errors in length and area measurements in Part 4. Approximate laws of error propagation in length measurement can again be obtained in concise matrix form. The sStatistical distribution of area measurement as well as the exact and approximate laws of error propagation are derived. Necessary and sufficient conditions for an unbiased area measurement is are also established.

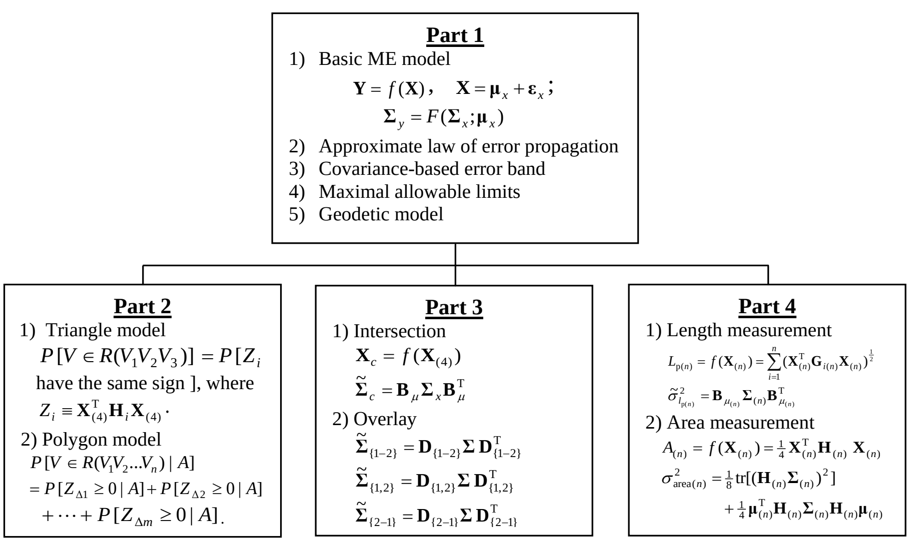


Fig. 5.1 A summary of our main results in the four-part series of papers

It is apparent that the proposed general framework facilitates a formal and practical error analysis in MBGIS. Starting from the basic ME model, we can build ME models and formulate exact or approximate laws of error propagation in a variety of interrelated GIS and spatial operations in a consistent and effective manner. The results are not only applicable to error analysis in MBGIS, ~~they~~ but are also suitable for general error analysis in GIS.

In addition to the study of the effect of spatial correlation of positional ME on length and area measurements mentioned above, directions for further research may include the derivation of the analytical error distribution of the coordinate vectors of the intersection point so that the exact law of error propagation for the intersection point of two random line segments can be obtained; the study ~~for~~ of the analytical distribution of the proposed quadratic form variables under ~~a~~ certain conditions in order to obtain ~~the~~ an analytical expression of the probability model for point-in-polygon problem with ME; and extensive applications of the results derived in this series of studies in a variety of situations.

Appendix 1

The proof of Lemma 2.1

Suppose that $\mathbf{\Sigma}_x$ is positive definite. Since it is symmetric, there exists an orthogonal matrix $\mathbf{S}$ such that $\mathbf{S\Sigma}_x\mathbf{S}^{\mathrm{T}} = \mathbf{\Gamma}$, where $\mathbf{\Gamma} \equiv \mathrm{diag}(\gamma_1,\ldots,\gamma_p)$ is a diagonal matrix with diagonal elements γ_i, and γ_i (≥ 0) are the eigenvalues of $\mathbf{\Sigma}_x$. Let $\mathbf{\Gamma}^{\frac{1}{2}} \equiv \mathrm{diag}(\gamma_1^{\frac{1}{2}},\ldots,\gamma_p^{\frac{1}{2}})$. Then $(\mathbf{\Gamma}^{\frac{1}{2}})^{-1}\mathbf{S\Sigma}_x\mathbf{S}^{\mathrm{T}}(\mathbf{\Gamma}^{\frac{1}{2}})^{-1} = \mathbf{I}$. Accordingly,

$$\mathbf{P\Sigma}_x\mathbf{P}^{\mathrm{T}} = \mathbf{I}, \text{ or } \mathbf{\Sigma}_x = \mathbf{P}^{-1}\mathbf{P}^{-\mathrm{T}}, \tag{A.1}$$

where $\mathbf{P} \equiv (\mathbf{\Gamma}^{\frac{1}{2}})^{-1}\mathbf{S}$. Similarly, since $\mathbf{P}^{-\mathrm{T}}\mathbf{HP}^{-1}$ is also symmetric, there exists an orthogonal matrix $\mathbf{T}$ such that $\mathbf{T}(\mathbf{P}^{-\mathrm{T}}\mathbf{HP}^{-1})\mathbf{T}^{\mathrm{T}} = \mathbf{\Lambda}$, where the diagonal matrix $\mathbf{\Lambda} = \mathrm{diag}(\tilde{\lambda}_1,\ldots,\tilde{\lambda}_p)$ consists of eigenvalues $\tilde{\lambda}_i$ of $\mathbf{P}^{-\mathrm{T}}\mathbf{HP}^{-1}$. It can easily be observed from (A.1) that for any λ,

$$|\mathbf{P}^{-\mathrm{T}}\mathbf{HP}^{-1} - \lambda\mathbf{I}| = |\mathbf{P}^{-\mathrm{T}}|\cdot|\mathbf{HP}^{-1}\mathbf{P}^{-\mathrm{T}} - \lambda\mathbf{I}|\cdot|\mathbf{P}^{\mathrm{T}}| = |\mathbf{H\Sigma}_x - \lambda\mathbf{I}|.$$

Thus we have shown that the diagonal elements $\tilde{\lambda}_i$ of $\mathbf{\Lambda}$ are also eigenvalues of $\mathbf{H\Sigma}_x$.

Let $\mathbf{Y} \equiv \mathbf{TPX}$. Then we can obtain $\mathbf{Y} \sim N_p(\mathbf{TP\mu}_x, \mathbf{TP\Sigma}_x(\mathbf{TP})^{\mathrm{T}})$ from the normality of **X**. According to the orthogonality of **T** and (A.1), $\mathbf{Y} \sim N_p(\boldsymbol{\mu}_y, \mathbf{I})$, where $\boldsymbol{\mu}_y \equiv \mathbf{TP\mu}_x$. In addition, using $\mathbf{T}^{-1} = \mathbf{T}^{\mathrm{T}}$, we have

$$Q = \mathbf{X}^{\mathrm{T}}\mathbf{H}\,\mathbf{X} = \mathbf{Y}^{\mathrm{T}}(\mathbf{TP})^{-\mathrm{T}}\mathbf{H}(\mathbf{TP})^{-1}\mathbf{Y} = \mathbf{Y}^{\mathrm{T}}\mathbf{\Lambda Y} = \sum_{i=1}^{p}\tilde{\lambda}_i y_i^2, \tag{A.2}$$

where $\tilde{\lambda}_i$ are eigenvalues of $\mathbf{H\Sigma}_x$.

Let λ_i be the distinct non-zero eigenvalues of $\mathbf{H\Sigma}_x$, p_i be their respective orders of multiplicity, and $y_{i,1},\ldots,y_{i,p_i}$ be the p_i values of *y* corresponding to λ_i. From $\mathbf{Y} \sim N_p(\boldsymbol{\mu}_y, \mathbf{I})$, it can be derived that $y_i,\ldots,y_p$ are independent and normally distributed: $y_i \sim N((\boldsymbol{\mu}_y)_i, 1)$. Thus by the definition of noncentral χ^2 - distribution, it implies that

$$Q = \sum_{i=1}^{p}\tilde{\lambda}_i y_i^2 = \sum_{i=1}^{m}\lambda_i\left(\sum_{j=1}^{p_i} y_{i,j}^2\right) = \sum_{i=1}^{m}\lambda_i\chi^2_{p_i,\delta_i^2},$$

where $\delta_i^2 \equiv (\boldsymbol{\mu}_y)_{i,1}^2 + \cdots + (\boldsymbol{\mu}_y)_{i,p_i}^2$ is the sum of squares of the *y*-means corresponding to λ_i. ف

The computation of noncentral parameters

We can describe in details the computation of the noncentral parameters δ_i^2 in (2.5).

For each eigenvalues λ_i of $\mathbf{H\Sigma}_x$ with p_i orders of multiplicity, the corresponding orthogonal eigenvectors $\mathbf{t}_{i,1},\ldots,\mathbf{t}_{i,p_i}$ are first computed in order to form the $p_i \times p$ matrix $\mathbf{T}_i \equiv (\mathbf{t}_{i,1},\ldots,\mathbf{t}_{i,p_i})^{\mathrm{T}}$. Then we have $(\mathbf{P}^{-\mathrm{T}}\mathbf{HP}^{-1})\mathbf{T}_i^{\mathrm{T}} = \lambda_i\mathbf{T}_i^{\mathrm{T}}$, $i = 1,\ldots,m$. Consequently, $(\mathbf{P}^{-\mathrm{T}}\mathbf{HP}^{-1})\cdot(\mathbf{T}_1^{\mathrm{T}},\ldots,\mathbf{T}_m^{\mathrm{T}}) =$ $(\mathbf{T}_1^{\mathrm{T}},\ldots,\mathbf{T}_m^{\mathrm{T}})\ \mathrm{diag}(\lambda_1\mathbf{I}_{p_1},\ldots,\lambda_m\mathbf{I}_{p_m})$, that is,

$$\mathbf{T}(\mathbf{P}^{-\mathrm{T}}\mathbf{HP}^{-1})\mathbf{T}^{\mathrm{T}} = \begin{pmatrix}\lambda_1\mathbf{I}_{p_1} & & \\ & \ddots & \\ & & \lambda_m\mathbf{I}_{p_m}\end{pmatrix}, \text{ where } \mathbf{T} = \begin{pmatrix}\mathbf{T}_1\\ \vdots \\ \mathbf{T}_{\mathrm{m}}\end{pmatrix}.$$

It can easily be checked that the matrix **T** in such a way is orthogonal. If we let $\boldsymbol{\mu}_{y,i} \equiv ((\boldsymbol{\mu}_y)_{i,1},\ldots,(\boldsymbol{\mu}_y)_{i,1})^{\mathrm{T}}$, we have further $\boldsymbol{\mu}_{y,i} = \mathbf{T}_{\mathrm{i}}\mathbf{P\mu}_x$. Thus

$$\delta_i^2 = \boldsymbol{\mu}_{y,i}^{\mathrm{T}}\,\boldsymbol{\mu}_{y,i} = (\mathbf{T}_i\mathbf{P\mu}_x)^{\mathrm{T}}\mathbf{T}_i\mathbf{P\mu}_x = \sum_{j=1}^{p_i}\left(\mathbf{t}_{i,j}^{\mathrm{T}}\mathbf{P\mu}_x\right)^2, \tag{A.3}$$

which results in the computational steps for δ_i^2 described in Section 2. ف

Appendix 2

Proof of Proposition 3.3

Since all of the diagonal elements of $\mathbf{\Delta}_{(n)}$ and $\mathbf{H}_0$ given in (3.19) are zero, all of the diagonal elements of the symmetric matrix $\mathbf{H}_{(n)} \equiv \mathbf{\Delta}_{(n)} \otimes \mathbf{H}_0$ are zero. According to Theorem 1.8 in Seber (1977) and (3.20), we have immediately

$$\operatorname{var}(\mathbf{X}_{(n)}^{\mathrm{T}} \mathbf{H}_{(n)} \mathbf{X}_{(n)}) = 2\sigma^4 \operatorname{tr}(\mathbf{H}_{(n)}^2) + 4\sigma^2 \boldsymbol{\mu}_{(n)}^{\mathrm{T}} \mathbf{H}_{(n)}^2 \boldsymbol{\mu}_{(n)} .$$

Now we compute the trace of the matrix $\mathbf{H}_{(n)}^2$, $\operatorname{tr}(\mathbf{H}_{(n)}^2)$. Due to

$$\operatorname{tr}(\mathbf{H}_{(n)}^2) = \operatorname{tr}[(\mathbf{\Delta}_{(n)} \otimes \mathbf{H}_0)(\mathbf{\Delta}_{(n)} \otimes \mathbf{H}_0)] = \operatorname{tr}(\mathbf{\Delta}_{(n)}^2 \otimes \mathbf{H}_0^2) = \operatorname{tr}(\mathbf{\Delta}_{(n)}^2) \cdot \operatorname{tr}(\mathbf{H}_0^2) = -2 \cdot \operatorname{tr}(\mathbf{\Delta}_{(n)}^2),$$

only $\operatorname{tr}(\mathbf{\Delta}_{(n)}^2)$ needs to be calculated. From (3.17) and (3.19), we obtain

$$\mathbf{\Delta}_{(n)}^2 = \left(\sum_{i=1}^{n} (\mathbf{e}_i \mathbf{e}_{i+1}^{\mathrm{T}} - \mathbf{e}_{i+1} \mathbf{e}_i^{\mathrm{T}}) \right)^2 = \sum_{i,j}^{n} (\mathbf{e}_i \mathbf{e}_{i+1}^{\mathrm{T}} - \mathbf{e}_{i+1} \mathbf{e}_i^{\mathrm{T}})(\mathbf{e}_j \mathbf{e}_{j+1}^{\mathrm{T}} - \mathbf{e}_{j+1} \mathbf{e}_j^{\mathrm{T}}) .$$

(To simplify the notations, the subscript "*n*" in $\mathbf{e}_{n,i}$ is omitted). Note that

$$\mathbf{e}_i^{\mathrm{T}} \mathbf{e}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

By simple calculations, $\mathbf{\Delta}_{(n)}^2 = \sum_{i=1}^{n} (\mathbf{e}_i \mathbf{e}_{i+2}^{\mathrm{T}} - \mathbf{e}_{i+1} \mathbf{e}_{i+1}^{\mathrm{T}} - \mathbf{e}_i \mathbf{e}_i^{\mathrm{T}} + \mathbf{e}_{i+1} \mathbf{e}_{i-1}^{\mathrm{T}})$. Therefore,

$$\begin{aligned} \operatorname{tr}(\mathbf{\Delta}_{(n)}^2) &= \sum_{i=1}^{n} [\operatorname{tr}(\mathbf{e}_i \mathbf{e}_{i+2}^{\mathrm{T}}) - \operatorname{tr}(\mathbf{e}_{i+1} \mathbf{e}_{i+1}^{\mathrm{T}}) - \operatorname{tr}(\mathbf{e}_i \mathbf{e}_i^{\mathrm{T}}) + \operatorname{tr}(\mathbf{e}_{i+1} \mathbf{e}_{i-1}^{\mathrm{T}})] \\ &= \sum_{i=1}^{n} [\operatorname{tr}(\mathbf{e}_{i+2}^{\mathrm{T}} \mathbf{e}_i) - \operatorname{tr}(\mathbf{e}_{i+1}^{\mathrm{T}} \mathbf{e}_{i+1}) - \operatorname{tr}(\mathbf{e}_i^{\mathrm{T}} \mathbf{e}_i) + \operatorname{tr}(\mathbf{e}_{i-1}^{\mathrm{T}} \mathbf{e}_{i+1})] = -2n, \end{aligned}$$

and

$$\sigma_{\mathrm{area}(n)}^2 = \tfrac{1}{16} \operatorname{var}(\mathbf{X}_{(n)}^{\mathrm{T}} \mathbf{H}_{(n)} \mathbf{X}_{(n)}) = \tfrac{1}{8} \sigma^4 \operatorname{tr}(\mathbf{H}_{(n)}^2) + \tfrac{1}{4} \sigma^2 \boldsymbol{\mu}_{(n)}^{\mathrm{T}} \mathbf{H}_{(n)}^2 \boldsymbol{\mu}_{(n)} = \tfrac{n}{2} \sigma^4 + \tfrac{1}{4} \sigma^2 \boldsymbol{\mu}_{(n)}^{\mathrm{T}} \mathbf{H}_{(n)}^2 \boldsymbol{\mu}_{(n)} .$$ ڤ

References

Beyer, W. H. 1987. *CRC Standard Mathematical Tables*, 28th ed. Boca Raton, FL: CRC Press, pp. 124-125 and 196.

Chrisman, N.R. 1982. A theory of cartographic error and its measurement in digital data bases. *Proc. Aut-Carto* 5. Crystal City, Virginia, pp. 159-168.

Chrisman, N.R. and B.S. Yandell. 1988. Effects of point error on area calculations: a statistical model. *Surveying and Mapping*, 48: 241-246.

Frolov, Y.S. and D.H. Maling. 1969. The accuracy of area measurement by point counting techniques. *The Cartographic Journal*, 6(1), 21-35

Ghilani, C.D. 2000. Demystifying area uncertainty: more or less. *Surveying and Land Information Systems*, 60(3): 177-182.

Goodchild, M.F., G. Sun, and S. Yang. 1992. Development and test of an error model for categorical data. *Int. J. Geograpical Information Systems*, 6(2), 87-104

Griffith, D.A. 1989. Distance calculations and errors in geographic databases. In Goodchild, M.F. and Gopal, S. (Eds). *Accuracy of Spatial Databases*, London: Taylor & Francis. pp. 81-90.

Hunter, G.J., and M.F. Goodchild. 1996. A new model for handling vector data uncertainty in geographical information systems. *Journal of the Urban and Regional Information Systems Association*, 8(1), 51-57.

Keefer, B.J., J.L. Smith and T.G. Gregoire. 1991. Modeling and evaluating the effects of stream mode digitizing errors on map variables. *Photogrammetric Engineering and Remote Sensing*, 57(7), 957-963.

Kiiveri, H.T. 1997. Assessing, representing and transmitting positional uncertainty in maps. *Int. J. Geographical Information Science*, 11(1), 33-52

Lam, N. S.-N. 1982. Areal interpolation using map overlay. *Modeling and Simulations*, 13, 953-959.

Leung, Y., and J.P. Yan. 1998. A locational error model for spatial features. *Int. J. Geographical Information Science*, 12, 607-620.

Leung, Y., J. H. Ma, and M.F. Goodchild. 2003a. A general framework for error analysis in measurement-based GIS, Part 1: the basic measurement-error model and related concepts. (unpublished paper)

Leung, Y., J. H. Ma, and M.F. Goodchild. 2003b. A general framework for error analysis in measurement-based GIS, Part 2: the algebra-based probability model for point-in-polygon analysis. (unpublished paper)

Leung, Y., J. H. Ma, and M.F. Goodchild. 2003c. A general framework for error analysis in measurement-based GIS, Part 3: error analysis in intersection and overlay. (unpublished paper)

Lloyd, P.R. 1976. Quantisation error in area measurement. *The Cartographic Journal*, 13(1), 22-25.

Mathai, A.M. and S.B. Provost. 1992. *Quadratic Forms In Random Variables: Theory and Applications*. New York: Marcel Dekker.

Pearson, E.S. Note on an approximation to the distribution of non-central χ^2. *Biometrika*, 46, 364.

Prisley, S.P., T.G. Gregoire, and J.L. Smith. 1989. The mean and variance of area estimates computed in an arc-node Geographical Information System. *Photogrammetric Engineering and Remote Sensing*, 55: 1601-1612.

Ross, A.H.M. 1999. Algorithm for calculating the noncentral chi-squares distribution. *IEEE Trans. Information Theory*, 45(4), 1327-1333.

Seber, Q.A.F. 1977. *Linear regression analysis.* New York: Wiley

Stolk, R. and G. Ettershank. 1987. Calculating the area of an irregular shape. *BYTE*, February, 135-136.

Turkington, D.A. 2002. *Matrix Calculus and Zero-One Matrices*. Cambridge, UK: Cambridge University Press.

Zhang, J. and R.P. Kirby. 2000. A geostatistical approach to modeling positional error in vector data. *Transactions in GIS*, 4(2), 145-159.

Zhang, J. and M.F. Goodchild. 2002. *Uncertainty In Geographical Information*. New York: Taylor & Francis.