Stochastic Simulation Models to Internally Validate Analytical Error Models of a Point and a Line Segment in GIS

Sungchul Hong[*a] and Joon Heo[b]

[a] Research Associate, School of Civil and Environmental Engineering, Yonsei University, 134 Shinchon-dong, Seodaemun-gu, Seoul 120-749, Korea; Tel: + 82-2-2123-2809; Fax: +82-2-2123-8209;
E-mail: sungchul@hotmail.com

[b] Professor, School of Civil and Environmental Engineering, Yonsei University, 134 Shinchon-dong, Seodaemun-gu, Seoul 120-749, Korea; Tel: + 82-2-2123-2809; Fax: +82-2-2123-8209;
E-mail: jheo@yonsei.ac.kr

KEY WORDS: GIS, Error Model, Spatial Data Quality, Point, Line

ABSTRACT: An error model in GIS is used to characterize positional errors in spatial data and to propagate the errors through spatial processes. Generally, there are two distinctive approaches for modeling positional errors in the spatial data: analytical and simulation. Analytical and simulation error models have the ability to describe (or realize) error-corrupted versions of spatial data. But the different approaches for modeling positional errors require internal validation that ascertains whether the analytical and simulation error models predict correct positional errors in a defined set of conditions. This paper presents stochastic simulation models of a point and a line segment to validate analytical error models, which are an error ellipse and an advanced error band model, respectively. The simulation error models populate positional errors by the Monte Carlo simulation, according to an assumed error distribution prescribed by given parameters of a variance-covariance matrix. In the validation process, a set of positional errors by the simulation models is compared to a theoretical description by the analytical error models. Results show that the proposed simulation models realize positional uncertainties of the same spatial data according to a defined level of positional quality.

1. INTRODUCTION

The quality of spatial data has become an important issue since Geographic Information System (GIS) began to be recognized as a practical tool for various areas such as infrastructure and resource management, and urban planning. Experienced GIS users have come to realize that different applications require spatial data at different accuracies and that dubious spatial data create problems rather than satisfying their needs. Thus, spatial data quality is one of the critical factors to be considered before utilizing GIS for spatial problem solving and decision making, leading to the need for error modeling that enables users to explore a variety of positional errors in spatial data. To deal with problems concerning the quality of spatial data, an error model in GIS is used to characterize positional errors in spatial data and to propagate the errors through spatial processes. A number of error models of a point and a line segment have been studied and proposed in the GIS community. Generally, there are two distinctive approaches for modeling positional errors in the spatial data: analytical and simulation. An analytical error model, generally derived from a variance-covariance matrix and the law of error propagation, describes a positional error distribution along spatial data while a simulation error model indicates a positional error distribution by generating error-corrupted versions of the same spatial data.

Therefore, it is essential to ascertain whether the analytical and simulation error models predict correct positional errors in a given set of conditions. This paper presents stochastic simulation models of a point and a line segment to internally validate analytical error models, which are an error ellipse and an advanced error band model, respectively. The simulation models based on Monte Carlo simulation populate positional errors according to an assumed error distribution prescribed by given parameters of a variance-covariance matrix. To test the models, a set of positional errors by the simulation models is compared to a theoretical description of positional errors by the analytical error models. The simulation and analytical error models will be internally validated when the differences between the simulated positional error and the predicted error fall within an acceptable tolerance.

2. PROPOSED ERROR MODELS FOR STOCHASTIC SIMULATION

The simulation model based on Monte Carlo simulation estimates positional errors by generating error corrupted versions of the same spatial data. Displacements imposed on spatial data indicate a positional uncertainty. Since samples generated from Monte Carlo simulation have a statistical property, the results can be analyzed with methods of statistical estimation and inference (Ang and Tang 1975).

2.1 Point

Positional errors can be simulated by generating random number pairs with the desired values of two means, two standard deviations, and a correlation coefficient (Campbell 1983). In simulation, Box and Muller method is used to generate a pair of independent random variables from a normal distribution with mean zero and unit variance (Box and Muller 1958). Since a distribution of random pairs with desired standard deviations and a correlation coefficient coincides with an error ellipse, the standard deviation with the x-y axes need to be transformed into the semi-major and minor axes of the error ellipse. The transformed standard deviations will be used as scale factors for the random variables from Box and Muller method. To generate error corrupted versions of point features, the adjusted pair of random numbers will be transformed back to the x-y axes again. An original point can be perturbed by adding the pairs to its coordinate components.

2.2 Line Segment

A stochastic simulation model based on Cholesky decomposition is proposed to realize an error-corrupted line segment by perturbing its vertices according to prescribed conditions. Cholesky decomposition commonly used in Monte Carlo simulation generates multiple correlated random numbers according to the correlation matrix. The correlated random numbers are treated as representing spatially correlated errors in vertices of a line segment. An error-corrupted line segment can be simulated by adding each pair of the correlated numbers to its corresponding vertices. In simulation, Cholesky decomposition is used to factorize the correlation matrix into the product of the lower triangular matrix and its conjugate transpose. Correlated random numbers can then be generated by multiplying the random number matrix with the lower triangular matrix. However, in Cholesky decomposition, the correlation matrix should be a symmetric positive-definite. To overcome the limitation, an eigen-decomposition (or spectral decomposition) is utilized. Since elements in the eigen value diagonal matrix should be non-negative, the correlation matrix needs to be a positive semi-definite matrix.

3. INTERNAL VALIDATION OF ANALYTICAL AND SIMULATION ERROR MODELS

Internal validation aims to ascertain whether analytical and simulation error models predict correct errors in a given set of conditions. While an analytical model describes a positional error distribution along spatial data, a simulation model perturbs spatial data according to an assumed error distribution. To accomplish the purpose, a theoretical description of positional errors by an analytical error model will be compared to a set of positional errors by a simulation model.

3.1 Point

A set of errors populated by the simulation model for point data is compared to an error description by an error ellipse. The simulation model generates positional errors according to the bivariate normal distribution along with the x and y axes. Therefore, a probability of perturbed points that would fall within an error ellipse should be identical to a confidence region of the error ellipse (Figure 1).

In the validation process, the simulation model is tested with varying correlation coefficients from -1 to 1, maintaining the given variances constant. The percentage of points that would fall within an error ellipse is computed after simulating 10,000 points according to identical properties of the error ellipse. The simulation results indicate that all average percentages are close to their corresponding confidence regions (within 0.5 %). Although curves in charts fluctuate in random pattern, the differences between the average percentage and the confidence region range from -0.28% to 0.49%. When the simulation results are analyzed with regards to confidence regions of an error ellipse, the small differences indicate that a set of simulated positional errors reflect the assumed error distributions. Also, when they are analyzed in terms of variances and correlation coefficients, the small differences show that the simulation model realizes the intended nature of positional errors in and between x and y components of points.

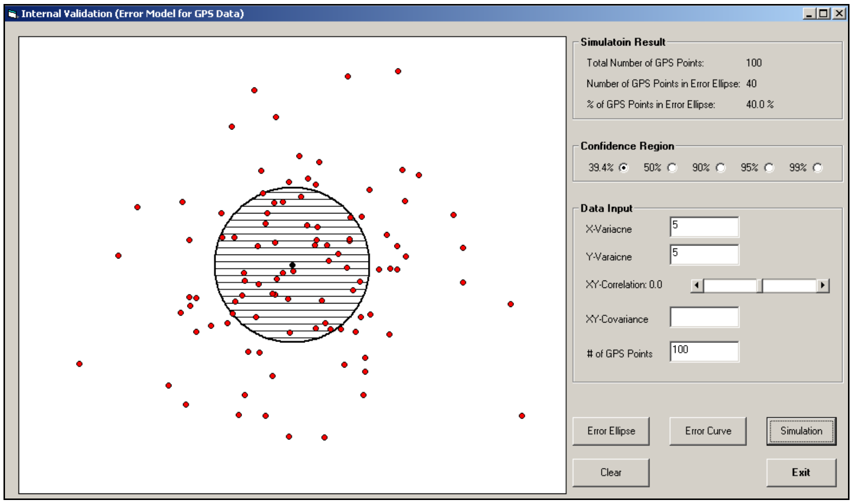
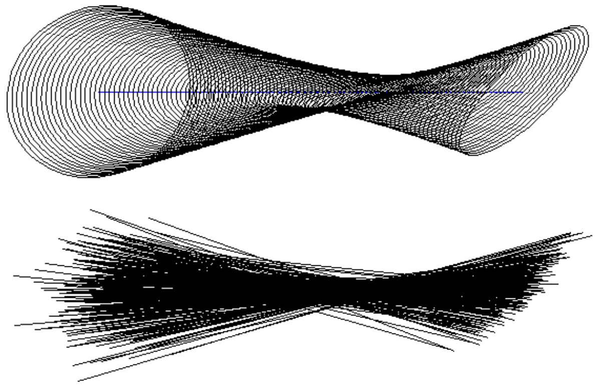


Figure 1 Analytical and Simulation Error Models for a Point

3.2 Line Segment

In the validation, an advanced error band model is adopted to theoretically describe positional errors along a line segment (Leung et al. 2004). A set of simulated results by the stochastic simulation model is compared with an error description by the advanced error band model that has various shapes depending on parameters in a variance-covariance matrix (Figure 2). Since the simulation model distorts a line segment conforming to the distribution prescribed by given parameters in the covariance matrix, displacements imposed on the line segment should be identical to the error band model. To test whether the simulation model realizes correct positional uncertainties, a group of 300 simulation results are compared with the error band model with a 99% confidence level. Simulation results indicate that the stochastic simulation model realizes positional uncertainties in a line segment conforming to prescribed conditions in a covariance matrix. The sets of displacements imposed on the original line segment follow the predicted error distributions by the advanced error band model. Due to the simplicity of implementation, this simulation model can be easily applied to a line with multiple points.



Covariance Matrix

$$\begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1y_1} & \sigma_{x_1x_2} & \sigma_{x_1y_2} \\ \sigma_{x_1y_1} & \sigma_{y_1}^2 & \sigma_{y_1x_2} & \sigma_{y_1y_2} \\ \sigma_{x_1x_2} & \sigma_{y_1x_2} & \sigma_{x_2}^2 & \sigma_{x_2y_2} \\ \sigma_{x_1y_2} & \sigma_{y_1y_2} & \sigma_{x_2y_2} & \sigma_{y_2}^2 \end{bmatrix} = \begin{bmatrix} 2 & 0.2 & 0.57 & 0.28 \\ 0.2 & 2 & -0.85 & -1.27 \\ 0.57 & -0.85 & 1 & 0.8 \\ 0.28 & -1.27 & 0.8 & 1 \end{bmatrix}$$

Correlation Matrix

$$\begin{bmatrix} 1 & \rho_{x_1y_1} & \rho_{x_1x_2} & \rho_{x_1y_2} \\ \rho_{x_1y_1} & 1 & \rho_{y_1x_2} & \rho_{y_1y_2} \\ \rho_{x_1x_2} & \rho_{y_1x_2} & 1 & \rho_{x_2y_2} \\ \rho_{x_1y_2} & \rho_{y_1y_2} & \rho_{x_2y_2} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.1 & 0.4 & 0.2 \\ 0.1 & 1 & -0.6 & -0.9 \\ 0.4 & -0.6 & 1 & 0.8 \\ 0.2 & -0.9 & 0.8 & 1 \end{bmatrix}$$

Figure 2 Analytical Error Model and Stochastic Simulation Model for a Line Segment

4. CONCLUSION AND FUTURE STUDY

This paper presents stochastic simulation models of a point and a line segment that generate positional errors by Monte Carlo simulation, according to an assumed error distribution. Analytical error models for a point and a line segment, which are an error ellipse and an advanced error band model, respectively, were employed and compared to test whether the simulation models realize a correct positional error distribution. Results from an internal validation indicated that each simulation model realizes positional uncertainties of the same spatial data conforming to prescribed conditions in a variance-covariance matrix. The simulation and analytical error models for a point and a line segment are suitable for simulating as well as describing positional errors. Particularly, due to the simplicity of implementation, the simulation model for a line segment can be easily applied to a line with multiple points.

The simulation and analytical error models for a point and a line segment are suitable for simulating as well as describing positional errors. Particularly, due to the simplicity of implementation, the simulation model for a line segment can be easily applied to a line with multiple points. But, this paper was concerned only about the simulation error models that validate the analytical error models based on a stochastic process theory. When a line is constructed from a digitization of photogrammetric or cartographic data, carefully and intentionally selected points to characterize the line tend to be non-random and highly correlated. Furthermore, during a generalization process, a line is simplified and smoothed, especially when the line is curved in reality. Thus, it is necessary to develop a new approach for modeling positional uncertainties that accounts for spatial dependencies in positional errors as well as the nature of map generalization.

ACKNOWLEDGEMENT

This research was supported by a grant (07KLSGC04) from Cutting-edge Urban Development - Korean Land Spatialization Research Project funded by Ministry of Land, Transport and Maritime Affairs.

REFERENCE

ANG, A. H. and TANG, W. H., eds., 1975. *Probability Concepts in Engineering Planning and Design.* New York: John Wiley & Sons.

BOX, G. E. P. and MULLER, M. E. 1958. A Note on the Generation of Random Normal Deviates. *The Annals of Mathematical Statistics,* 29(2), 610-611.

CAMPBELL, C. W. 1983. Efficient algorithm for generating random number pairs drawn from a bivariate normal distribution. *NTIS, SPRINGFIELD, VA(USA), 1983, 17*.

LEUNG, Y., MA, J. H. and GOODCHILD, M. F. 2004. A general framework for error analysis in measurement-based GIS Part 1: The basic measurement-error model and related concepts. *Journal of Geographical Systems,* 6(4), 325-354.

WATKINS, D. S., 2002. *Fundamentals of matrix computations.* John Wiley and Sons.