Proceedings of the 8th International Symposium on
Spatial Accuracy Assessment in Natural Resources and Environmental Sciences
Shanghai, P. R. China, June 25-27, 2008, pp. 1-7

Uncertainty in Spatial Information, Analysis, and Applications: Chinese Perspectives

Deren Li

State Key Lab for Information Engineering in Surveying, Mapping and Remote Sensing
Wuhan University, 129 Luoyu Road, Wuhan 430079, China

Abstract. This paper reviews past research by Chinese researchers on error analysis in geomatic engineering. Currently, a trend towards uncertainty characterization in geoinformation, analysis, and applications is clearly visible in China. Future developments will be seen not only in theory and technology but also applications in terms of making contribution to national economy and social welfares, which, in turn, will reinforce the position of the research community of spatial uncertainty in the world.

Keywords: geomatic engineering, error analysis, uncertainty, geostatistics, validation, indeterminate objects

1. Error analysis in geomatic engineering

Along with science and technology progress and social sustainable developments, comes the need for information in a fast changing world. In 1980s, an information revolution made its impacts on various disciplines, causing disciplines to evolve from refinement to integration, promoting all the specialties within the discipline of surveying and mapping and related specialties to mix and integrate. For example, specialties, such as space science, computer science, earth sciences, geography, environmental sciences, urban science, and management science, were merging and integrated, forming a larger and interdisciplinary field: geographic information science. Surveying and mapping is widely known as geomatic engineering, which contains several disciplines and technology, including geodesy, land surveying, Global Position System (GPS), photogrammetry, remote sensing, cartography, and geographic information systems (GIS).

At present, national spatial databases at 1:1,000,000, 1:250,000, and 1:50,000 scales have been constructed. Digital products, including digital orthophoto maps (DOMs), digital elevation models (DEMs), digital line graphs (DLGs), digital raster graphs (DRGs), place names, and land use are made available. The 1:10,000 scale spatial databases for individual provinces have been or are being constructed, while spatial databases at scales ranging from 1:500 to 1:2,000 for large or medium cities have been constructed, becoming important basis for digital city and digital China.

In 2006, the State Bureau of Surveying and Mapping launched the project “western China mapping”, which utilized various new instruments, technology, and air-/space-borne remote sensing imagery. This project will rewrite the history that more than 2 million square kilometers territory in western China was never mapped to a scale of 1:50,000.

Error analysis has been one of the research foci in geomatic engineering for a long time. Solution of least squares and validation of results require good mastery of probability and error theory, as error modeling and analyses of residuals provide the mathematics for data handling. Concepts of precision and accuracy are usually exemplified by terms, such as standard deviation and root mean squared errors (RMSEs) (Li and Yuan, 2002). For examples, adjustment computation in traversals and leveling networks is usually based on measurements of distance, angles, and height differences, and provide most probable estimation of position and heights for geospatial objects. To quantify errors in the results, variance and covariance propagation is usually applied, whereby linearization through Taylor expansion is often necessary to facilitate computation.

ISBN: 1-84626-171-6, **ISBN13:** 978-1-84626-171-8

In photogrammetry, co-linearity equations are regarded as the foundation for reduction of objects position based on stereoscopic measurements of image coordinates. Photogrammetric mapping is implemented on the basis of proper inner and exterior orientation. In order to quantify the precision in photogrammetric point fixing, error propagation is performed. Aero-triangulation is commonly used for solution of large linear equation systems to derive unknown parameters, such as adjusted object-space coordinates. Image matching based on least squares that aims to replace human stereo-mapping in a photogrammetry workstation can also be analyzed with respect to precision and accuracy. It was found that precision in image matching is related to SNR and image textures (Wang, 1990). Li and Yuan (2002) provided a systematic review of error theory, hypothesis testing, and reliability issues in adjustment, especially those concerning analytical photogrammetry. Currently, feature extraction from high-resolution imagery is a research hot point. Error analysis will be a useful investment into information engineering in this regard (Wu *et al*., 2007).

Accuracy assessment in remote sensing has become a popular topic for research, resulting in a large quantity of papers. Uncertainty in remote sensing was discussed by Li and Gao (1997). The concept of sensitivity matrices was proposed, which aims for quantitative analysis of error propagation in information retrievals. Since the 1990s, a series of important achievements have been gained by utilizing remote sensing for provision of global and local land cover and land use information (Yang and Zhu, 1999). Error matrices were commonly used to derive measures of classification accuracy, such as user's and producers' accuracy, overall accuracy, and Kappa index of coefficients. In remote sensing, image registration is facilitated by inversion of usually polynomials functions linking coordinates cross imagery. As reference control points (RCPs) used to establish models in registration may themselves contain errors, a consistent adjusted least squares (CALS) estimator and a relaxed consistent adjusted least squares (RCALS) method for registration were proposed by Ge *et al*. (2006), which can correct errors in the RCPs and propagate these errors to the corrected image with and without prior information. It is expected that research on validation in remotely sensed land cover and land use information will gain even greater progress (Yu, et al., 2004).

Computer cartography and geographic information systems (GIS) revolutionized the way map data were processed and visualized. Errors in map digitizing were quickly brought to research attention, as it was recognized that "slivers" occur due to digitizing errors, which hamper map overlaying and other geo-processing. Error ellipses and epsilon-error bands were introduced to model positional errors in discrete points and lines, respectively (Shi, 2005).

2. Error description *vs.* error modeling

GIS brought about major changes to spatial data handling, and provided major impetus to comprehensive research on error and accuracy in spatial data and analysis. A fundamental theory in GIS concerns how spatial data should be modeled, i.e., how the real world should be conceptualized. There are basically two models for spatial data: discrete entities/objects and fields, with former perceiving the world is populated with discrete points, lines, and areas, while in the latter the world is conceived of as consisting of a set of single-valued functions defined everywhere over space.

An entity with an ID can be represented as a tuple ID(x, *Attri*), where x and *Attri* stand for its position and attributes, respectively. As attribute data are usually taken care of by specialists in specific application domains, positional data remains the key responsibility of GIS specialists. Two kinds of variables are possible with fields: one is continuous variables exemplified by elevation, and the other is categorical variables of nominal or ordinal nature, such as land cover and residential areas of graded population density. For annotations, $\mathbf{Z}(x)$ stands for continuous fields, while $C(x)$ for categorical fields.

Depending on different data models adopted, error and accuracy can be discussed in alignment to those in objects and fields, respectively. As discussed previously, error in spatial entities positional errors in points are modeled by error ellipses, while those in lines are approached via epsilon-error bands. Errors in points and lines will be compounded in areas. Errors due to vector and raster data conversion were discussed by Chen *et al*. (2007).

Such error descriptors are useful. However, they may not be able to facilitate error propagation in measures, such as lengths and areas. This would require knowledge of spatial correlation in positional errors,

and how such information may be incorporated in error modeling. Error models are stochastic processes that can simulate the occurrences of errors, which are believed to be inherent to the phenomena being mapped. Often, realizations drawn from an error model are equal-probable, from which error statistics can be computed. These realizations emulate the differences or distributions that would be observed from a measurement process known to subject to error prescribed by the model. Error implies the existence of the true value for a quality or quantity, at least in principle. However, in many occasions, this is an unattainable assumption, because it is impossible to determine the truth, although we may reach consensus about the likely intervals wherein truth may fall into. Thus, it is more sensible to use the term uncertainty than error, as the former implies a distribution surrounding the truth and will be more suitable for conveying the meaning of vagueness (Wang *et al*., 2003), as will be seen later on.

Map digitizing may be taken as a good example for discussion of positional errors. Moving a digitizer along a smooth and continuous line could be regarded as a discrete time stochastic process consisting of trend motion and random motion. A stochastic stationary observation series of digitizing error may be generated by adopting a backward difference process. The stochastic motion may be simulated by using an autoregressive process in terms of time series analysis theory, resulting in an estimation model of digitizing error (Huang and Liu, 1997). Error propagation of buffer analysis in a vector-based geographical information system (GIS) is studied with the use of statistics and numerical analyses (Shi, *et al*., 2003). In this paper, such factors as the error of commission, the error of omission, the discrepant area and the normalized discrepant area are proposed as the error indicators of buffer analysis. Analytical expressions for the error indicators are developed as multiple integrals. A numerical integration method is recommended to find an approximation to the analytical expression.

Uncertainty characterization has become increasingly recognized as an integral component in thematic mapping based on remotely sensed imagery, and descriptors such as percent correctly classified pixels (PCC) and Kappa coefficients of agreement have been devised as thematic accuracy metrics. However, such spatially averaged measures about accuracy neither offer hints about spatial variation in misclassification, nor are they useful for quantifying error margins in derivatives, such as areal extents of different land cover types and land cover change statistics. Such limitations originate from the deficiency that spatial dependency is not accommodated in the conventional methods for error analysis. Geostatistics provides a good framework for uncertainty characterization in land cover information. Methods for predicting and propagating misclassification were developed on the basis of indicator samples and covariates, such as spectrally derived posteriori probabilities (Zhang and Sun, 2006). It was found that significant biases result from applying joint probability rules assuming temporal independence between misclassifications across time, thus consolidating the need for stochastic simulation in error modeling. Further investigations are anticipated incorporating indicators and probabilistic data for mapping and propagating misclassification.

Digital elevation models (DEMs) have found wide applications in disciplines such as civil engineering, geology, geomorphology, planning, and communications (Li and Zhu, 2005). Topics, such as sampling, data acquisition, interpolation, multi-scale representation, derivation of various types of terrain factors, visualization, and applications, all have attracted research efforts. A comparative study of the accuracies of DEMs derived from four different data models, namely, contour data only, contour data with additional feature-specific data (peaks, pits, points along ridges, points along ravines, and points along break lines, etc.), square-grid data only, and square-grid data with additional feature-specific data (Li, 1994). It has been found that the accuracy of DTMs derived from photogrammetrically measured contour data is related to fractions of contour intervals, depending on the characteristics of the terrain topography, and that additional feature-specific data can increase accuracy. A recent review of terrain analysis and error issues was provided in Hu *et al*. (2007), where a more rigorous treatment of accuracy and fidelity in terrain modeling is provided.

Topography is important to the description, quantification, and interpretation of many hydrological processes, such as surface runoff and water storage, energy fluxes, evapo-transpiration, soil erosion, and snow metamorphosis. Extracting topographic information for a watershed by traditional, manual techniques can be a tedious, time consuming, subjective, and error-prone task, particularly for large watersheds. Research over the past decades has demonstrated the feasibility of extracting topographic information directly from raster DEMs through digital terrain analysis. In the field of distributed hydrological modeling,

automated evaluation of DEMs has focused on watershed segmentation, definition of drainage divides and identification of the drainage networks. This automated extraction of network and sub-watershed properties from DEMs represents a convenient and rapid way to parameterize a watershed. This technique also has the advantage of generating digital data that can be readily managed and analyzed by GIS, and the extracted topographic features also can be directly input for hydrological models. Li *et al*. (2002) undertook experiments with DEMs-based hydrological modeling, and found that the drainage network and sub-basins extracted from DEM are acceptable as compared with that of manual digitization from 1:100,000 scale topographic maps. In deed, hydrology provides an excellent example for geomatics specialists who are interested in how to add values to data they derive from various sources, and who care about the accuracy of spatial information products and the analyses performed on the basis of these products.

3. Current developments

It becomes obvious that geostatistics has much to offer to the research community of spatial uncertainty, as it provides the techniques, i.e., stochastic simulation, which can generate equal-probable realizations, each of which reproduces the histogram and moments observed in the empirical data. Such realizations will be fed to specific geo-processing routines to facilitate error propagation in the results. Variogram modeling is itself important for determining suitable scale in a ground or remote surveys (Bo and Wang, 2003). However, geostatistics is a specialized field, which may be highly advanced for some geospatial professionals. Efforts must be made to popularize its applications by providing user-friendly interface to geostatistical software systems that are built-in with error modeling functions.

For spatial categorical information, it is not adequate to just look at classification accuracy. Often, the misclassification is actually due to different strategies followed or class naming systems applied. As a matter of fact, various inconsistencies exist in classification schemes adopted in different agencies, projects, and/or professionals for different problem domains. Therefore, research should de directed toward semantics and increased interoperability cross categorical information.

The relationships between the concept of uncertainty in GIS and remote sensing and that in the information theory were discussed in the context of GIS and remote sensing by Lin and Zhang (2006). This was meant to enable evaluation of the degree of information and to quantify the effects of uncertainty upon maps and imagery. Information content of GIS data can be inferred based on the definition of uncertainty in information theory. With this, the degrees of information in positional and attribute data can be measured in the unit of bits, and the information of any GIS objects is measurable by considering the doubtfulness in position and attributes. Information in remote sensing imagery can also be discussed, whereby spatial correlation and cross-band correlation should be accommodated for proper quantification of degrees of information. With multi-temporal imagery becoming available, change detection is widely applied for monitoring landscape dynamics. Research on mis-registration and its impacts on information contents and accuracy of change detection will be important.

Remote sensing products are accumulating at an unprecedented speed. For scientific applications of such products, it is important to asses their fitness for use. The Committee on Earth Observation Satellites (CEOS) (http://www.ceos.org) and the Group on Earth Observations (GEO) (http://earthobservations.org) are sponsoring workshops on “Quality Assurance of Calibration & Validation Processes”. The workshops were established to address the urgent need to harmonise and standardise satellite data handling and information exchange across the international community, with GEOSS’s ultimate goal of “the seamless delivery of comprehensive, global knowledge /information products in a timely manner” to users. China will play an active role in this drive as it has established its Earth observing and applications systems. Nowadays, National Spatial Data Infrastructure is a great venture for China. Various applications, such as cadastral surveys, environmental modeling, and insurance policy-making, are making use of spatial information. A science of geospatial information needs better understanding of spatial uncertainty and should provides guidance for quality assurance in information production and use. Quality assurance for information services will enable best quality of service for any client, any connection, and any server or server composition (Wu and Zhang, 2007).

Virtual globe systems, such as Google Earth, puts a planet's worth of imagery and other geographic information right on your desktop. It is easy to view exotic places like Paris, and to check points of interest such as local restaurants, hospitals, and schools. With virtual globe, you can fly from space to your neighborhood, get driving directions, and tilt and rotate the view to see 3D terrain and buildings. A Chinese system for geospatial information services (http://geoglobe.whu.edu.cn) provides information management and visualization functions, and has drawn international attention. The research community of spatial uncertainty will face challenges when virtual globe becomes a living phenomenon and poses a realistic concern with respect to the accuracy in the information it conveys.

4. Research issues

Despite China's remarkable developments in research on spatial accuracy and uncertainty, there exist some fundamental issues that have been largely unsolved. I will discuss them below, hoping to clarify some crucial conceptual and theoretical issues to stimulate further research.

4.1. Differentiation between determinate and indeterminate objects

There are two types of objects, natural and man-made ones. Many of them are well-defined, such as sport fields, buildings, which have determinate areal extents so that repeated measurements will give rises to different values. In statistics, it can be seen that means of infinite number of measurements will approach the true values with arbitrarily infinitesimal differences, but means of limited number of measurements will be estimation of the true values. Error analysis for such geographic objects may be conducted in the framework of least squares and robust estimation methods so that precision and reliability in measurements, adjusted results and functions of adjusted results can be computed with ease. Nothing indeterminate poses a concern there. In the contrary, there are many other objects in GIS, mostly natural objects, such as coastlines, river shorelines, lines describing land forms, laves, and eco-tones between forests and grasslands, which are fractal, i.e., non-differentiable due to discontinuity and non-smoothness along their extents. They are termed indeterminate objects (Burrough and Frank, 1996), and may be handled with fractals, fuzzy sets, rough sets, and cloud models that integrate statistics and fuzzy logic, as well as statistics.

4.2. Scale-dependent definitions of geospatial objects

It is well known that points, lines, and areas are used for representing real-world entities in GISs, whose geometries can be well approximated by the corresponding object metaphors. Such objects are, however, scale-dependant in the sense that apparent point objects on a smaller scale map may become areas of finite extents at larger scale while lines or areas may be reduced to points when up-scaled. Mis-specification of objects in terms of geometry may lead to error and can cause ambiguity and confusion in spatial information processing. Thus, scale is an issue closely related to uncertainty.

A promising concept is the so-called modality. Instead of pure geometric points and lines, modal points and modal lines are accommodated so that their geometry, attributes, time, semantics, and uncertainty are integrated for abstraction and description of geospatial entities. Unlike conventional points or lines, modal points and lines possess spatial extents so that hierarchies, divisibility, and temporality are well supported. Further research along this line will shed new light unto developments related to uncertainty.

4.3. Error in approximating complex curves and surfaces by discrete points *vs.* error in measurements at these points and its propagation.

Descriptions of points, lines, areas, and volumes by discrete points are necessary for computer storage and management of complex objects, with the adverts of computer and electronic atlas. The approximation of complex curves and surfaces by discrete points thus becomes an outstanding issue. In 1970s through 1990s, there was research of breadth and at depth on this issue, resulting in many papers. Take DEMs as an example. There were the progressive sampling method of Markovic (Makarovic, 1973) and the pyramid method for terrain analysis based on dense image match points up to few terrain feature points of Ackermann et al. (1992). Much can be learnt a from such research achievements.

As seen in Figure 1, points 1, 2, and 3 stand for discrete points approximating the true surface profile, where ΔM is clearly the maximum error in approximation. There are measurement errors at points 1 and 2, whose impacts to point M can be quantified through variance and covariance propagation. Measurement

errors are different from approximation errors, which should not be confused, nor should measurement error propagation be taken as substitute for error in representations of curves and surfaces.

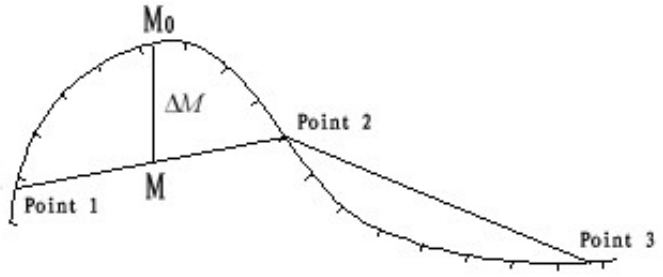

Fig. 1: Error in curve approximation by discrete points

4.4. Positional uncertainty, attribute (semantic) uncertainty, and temporal uncertainty

GIS is more informative than traditional surveying and mapping techniques, and can provide answers to questions that are more wide-reaching and complex. For example, answers to such queries as the total lengths of China's railways and roads can obtained by identifying and summarizing rails and roads based on aerial photographs and satellite imagery. The accuracy in the lengths will depend on not only error in rails and roads measurement but also error in identification of these features from image sources. An example for the latter may be mis-identification of aqueducts to roads, or vice versa. Also, for census of China's arable lands, the accuracy in the sum will be related to not only measurement of each individual piece of arable land, but also misclassification of land use or land use change due to temporality. For the latter, examples can be change of arable land to buildt-up areas, to forestry, or to lakes. The affects of such errors will be more severe than those in measurements of arable land use boundaries, provided the measurements are up to specifications. Thus, it is important to direct research into semantic and temporal inconsistency in GIS, and to homogenization in semantics and temporality (Li, 2005).

4.5. Spatial data quality *vs*. quality of spatial information services.

Past research has largely focused upon spatial data quality. With the developments of Web-GIS, mobile GIS, and grid-GIS, it is high time that we extended our research on spatial data quality to research on quality in spatial information services. In web-GISs, the quality of spatial information services lies in delivery of most useful information to those in needs within the shortest time and at minimized costs. Thus, the quality of spatial information services is influenced by the factors, such as the level of understanding of users' requirements for spatial information, sophistication of semantics models in spatial information services, quality of data and algorithms involved in a task, quality, speed, and integrity of Internet communication, and quality of user interfaces, in particular, visualization, of spatial information services.

4.6. Applications

In photogrammetry, research on precision in 1960s and reliability and separability in 1980s followed the principle of theory and practice combination. The transformation of research achievements to operational use provided impetus to developments in compensation for systematic error, detection of gross error, and photogrammetric adjustment system designs.

However, research on spatial uncertainty has been largely confined to pure theoretical and mathematical exploration. Although the research issues concerned are become increasingly complicated, there is little relevance or reference to applications. During an ISPRS Commission workshop held in Beijing in 2005, the topic of "uncertainty in research on uncertainty in spatial data" was even proposed by some delegate. Thus, it is crucial to make the transition from mathematical expressions concerning uncertainty in spatial data to measures making obvious sense to and operable by users, such as maximum and mean risks, who can make informed decisions about risks in land purchase based on these quality measures in spatial data and information services.

Only then, research on spatial uncertainty will become operation-oriented and hence fruitful. As said by Prof Ackermann 20 years ago, theoretical research lies in solving problems in practice. In other words, only

with an aim to solve practical problems, can theoretical research find its place in the world and fulfill its values.

5. Acknowledgements

This research is partially supported by a grant from China's "973 Program" (project No. 2006CB701302).

6. References

[1] Ackermann, F., and Schneider, W., Experience with automatic DEM Generation, *International Archives of Photogrammetry and Remote Sensing*, Vol. XXIX Part B4, Comm.IV, Washington, D.C. 1992: 986-989.

[2] Bo, Y., and Wang, J., Uncertainty in Remote Sensing Information: *Classification and Effects of Scale*. Geology Press, 2003.

[3] Burrough, P.A., and Frank, A.U. (Eds.), Geographic Objects with Indeterminate Boundaries London: *Taylor & Francis*, 1996.

[4] Chen J., Zhou C., Cheng W., Area error analysis of vector to raster conversion of areal features in GIS, *Acta Geodaetica et Cartographica Sinica*, 2007, **36**(3): 344-350.

[5] Ge Y., Liang Y., Ma J., Wang J., Error propagation model for registration of remote sensing image and simulation analysis, *Journal of Remote Sensing*, 2006, **10**(3): 299-305.

[6] Hu, P., Yang, C., Wu, Y., and Hu, H., New Digital Elevation Models: *Theory, Methods, Standards, and Applications*. Surveying and Mapping Press, 2007.

[7] Huang Y., and Liu W., Building the estimation model of digitizing Error. *Photogrammetric Engineering and Remote Seneing*, 1997, **63**(10): 1203-1209.

[8] Li, D., On generalized spatial information grid and specialized spatial information grid. *Journal of Remote Sensing*, 2005, **9**(5): 513-519.

[9] Li, D., and Yuan X., *Error processing and reliability theory*. Wuhan University Press, 2002.

[10] Li S., Zeng Z., Zhang Y., Application of digital terrain analysis technology for distributed hydrological modeling, *Advance in Earth Science*, 2002, **17**(5): 769-775

[11] Li, X., and Gao, F., Uncertainty and sensitivity matrices in parameter inversion by remote sensing. *Journal of Remote Sensing*, 1997, **1**(1): 5-14.

[12] Li Z., A comparative study of the accuracy of digital terrain models (DTMs) based on various data models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 1994, **49**(1): 2-11.

[13] Li Z., and Zhu, Q., *Digital Terrain Modeling: Principles And Methodology*, CRC Press, 2005.

[14] Lin, Z., and Zhang, Y., Measurement of information and uncertainty of remote sensing and GIS data, *Geomatics and information science of Wuhan University*, 2006, **31**(7): 569-572.

[15] Makarovic, B., Progressive Sampling for Digital Terrain Models. *ITC Journal*, 1973, **3**: 145-153.

[16] Shi, W., *Principle of modeling uncertainties in spatial data and analysis*. Beijing: Science Press, 2005.

[17] Shi, W., Cheung C., and Zhu C., Modelling error propagation in vector-based buffer analysis. *International Journal of Geographical Information Science*, 2003, **17** (3): 251-271.

[18] Wang, X., Shi, W., Wang, S., *Fuzziness in Spatial Information*. Wuhan University Press, 2003.

[19] Wu, C., Lu, G., Shu, F., Research on Quality Checking Method Based on Knowledge and Rule to Cadastral Data. *Geography and Geo-Information Science*, 2007, **23** (5): 22-30.

[20] Wu, H., and Zhang, H., QoGIS: Concept and research framework, *Geomatics and Information Science of Wuhan University*, 2007, **32**(5): 385-388.

[21] Yang, L., and Zhu, Z., The status quo and expectation of global and local land cover and land use remote sensing research. *Journal of Natural Resources*, 1999, **14**(4): 340-344.

[22] Yu, T., Gu X., Tian G., Legrand M., Baret F. Hanocq J.F., Bosseno, R., Zhang, Y., Modeling directional brightness temperature over a maize canopy in row structure. *IEEE transactions on geoscience and remote sensing*, 2004, **42**(10): 2290-2304.

[23] Zhang, J., and Sun J., Uncertainty characterization in remotely sensed land cover information. *Proceedings Accuracy*, 2006: 663-672.