A REVIEW ON ACCURACY AND UNCERTAINTY OF SPATIAL DATA AND ANALYSES WITH SPECIAL REFERENCE TO URBAN AND HYDROLOGICAL MODELLING

Aarthi Aishwarya Devendran , Gnanappazham Lakshmanan*

Department of Earth and Space Sciences, Indian Institute of Space Science and Technology, Trivandrum, India –
aarthiaishwarya.14@iist.ac.in, gnanam@iist.ac.in

Commission VI, WG VI/4

KEY WORDS: Uncertainty, Spatial Data Quality, Modelling, Urban simulation models, CA, Hydrological models, SWAT, Error propagation

ABSTRACT:

Data quality for GIS processing and analysis is becoming an increased concern due to the accelerated application of GIS technology for problem solving and decision making roles. Uncertainty in the geographic representation of the real world arises as these representations are incomplete. Identification of the sources of these uncertainties and the ways in which they operate in GIS based representations become crucial in any spatial data representation and geospatial analysis applied to any field of application. This paper reviews the articles on the various components of spatial data quality and various uncertainties inherent in them and special focus is paid to two fields of application such as Urban Simulation and Hydrological Modelling. Urban growth is a complicated process involving the spatio-temporal changes of all socio-economic and physical components at different scales. Cellular Automata (CA) model is one of the simulation models, which randomly selects potential cells for urbanisation and the transition rules evaluate the properties of the cell and its neighbour. Uncertainty arising from CA modelling is assessed mainly using sensitivity analysis including Monte Carlo simulation method. Likewise, the importance of hydrological uncertainty analysis has been emphasized in recent years and there is an urgent need to incorporate uncertainty estimation into water resources assessment procedures. The Soil and Water Assessment Tool (SWAT) is a continuous time watershed model to evaluate various impacts of land use management and climate on hydrology and water quality. Hydrological model uncertainties using SWAT model are dealt primarily by Generalized Likelihood Uncertainty Estimation (GLUE) method.

1. INTRODUCTION

Data Quality is a very important factor in the process of effectively and accurately conveying the information about the spatial data. The quality of a data set can vary from person to person, organization to organization or from application to application. The quality measures and standards of a data might differ based on the application so it is the ultimate responsibility of the user to check and decide the quality of a data. Spatial dataset, which is suitable or appropriate for one project or application, need not be necessarily suitable for another. In spite of the growing concern for spatial database accuracy, most of the GIS software products do not provide tools for modelling the errors in individual data layers and for tracking the errors when numerous different data layers are combined and manipulated for GIS based spatial analysis (Forier & Canters, 1996). Errors in a spatial data may creep in at any stage of data acquisition and transformations. It is therefore crucial to control the errors at each and every transition from the observation to the final stage of data preparation and to check and ensure that the quality is maintained at all stages of data preparation and analysis. Maintaining data accuracy and eliminating the errors are the essential prerequisites for providing information about the spatial data quality. Error is inherent and inevitable in any GIS data. GIS dealing with different layers of data, collected from multiple sources, scales, dates and map projections, will have its complex error and the errors might get propagated even further. The key aspect of GIS functionality helps the users integrating and analysing data from different sources with

difference in scale, accuracy, resolution but does not provide information about the effects of combining these different levels of data uncertainty on the input maps and the output maps resulting from spatial querying and analysis (Forier & Canters, 1996). Though there are numerous researches on the topic of error in spatial databases, there is still a lack of proper tool for handling the errors in a spatial database. The uncertainties present at one or more stages of representing the real world get propagated throughout the geographical analysis affecting the final output thus making the decision uncertain. Thus, a correct conceptualization and modelling of the uncertainty is required and also modelling the uncertainties in individual layers and tracking them when data layers are combined and manipulated in GIS based spatial analysis become crucial.

2. SPATIAL DATA QUALITY

The U.S. Spatial Data Transfer Standard lists 5 components of a data quality which provide necessary information to assess its suitability. They are *data lineage, positional accuracy, attribute accuracy, logical consistency and completeness. Data lineage* describes the source of the derived data, methods involved in the derivation and all the transformations that are employed in producing the final data set. It is important to know the lineage of the data as most of the errors are the results of errors in the field. *Positional accuracy* describes the comparison of the given spatial data with an independent and more accurate source of data. It must be determined by comparing the spatial data to an

---

*Corresponding author.

independent source of higher accuracy. *Attribute or thematic accuracy* is the accuracy of the attributes of its data. *Logical consistency* determines the faithfulness of the data set. Consistency refers to the uniformity or the agreement among all the parts of a given dataset. Consistency problems arise in spatial data when maps are overlaid or data from two or more maps are combined into one. When two different maps of different map scale are combined there might be an increased chance of losing a certain degree of consistency. *Completeness* refers to the aspects of a data set that characterizes it as a whole and not as a specific or individual element. Completeness can be divided into two types: *feature or entity completeness, value completeness* and *attribute completeness. Feature completeness* is further divided into *data and model completeness. Data completeness* is a measure of check on data quality. It is defined as the measurable error observed between the database and the specification. If the database contains all the objects with their specifications, then the database is considered complete. *Model completeness* is the agreement between the database and the abstract universe required for a particular database application. *Attribute completeness* is the measure of degree to which all the relevant attributes of a feature have been included. *Value completeness* is the degree to which all the values of attributes are present.

The generic measures of a spatial data quality are *accuracy, precision, error and uncertainty* having varying degrees of meaning. *Accuracy* will be the first and foremost consideration when a data is analysed and evaluated. It is the extent to which an estimated value approaches the true value. Data accuracy can be divided into lineage, positional, thematic and temporal accuracies. *Precision* is the deviation or dispersion of a data from the actual data. It is estimated in terms of standard deviation of the observations over the mean. Positional accuracy determines the way in which the geographic database can be ultimately used (Ries, 1993). Positional accuracy can be described into two components: absolute and relative positional accuracy. Absolute positional accuracy describes how closely all positions on a map or data layer match, corresponding to the positions of features represented on the ground in a desired map projection system. Relative positional accuracy of a map considers how closely all the positions on a map or data layer represent their corresponding geometrical relationships on the ground (Chong et al. 1997). *Error* is the difference between the true and estimated values. *Uncertainty* can be due to lack of data, measurement errors or sampling errors. GIS users and decision makers using maps containing uncertainties face serious risks. Thus, quality control of the products is very crucial for decision-making (Wang et al., 2005).

3. ERRORS IN SPATIAL DATA

Quality is more difficult to define for data. Unlike manufactured products, data do not have physical characteristics that allow quality to be easily assessed. The origin of errors in a spatial dataset might be either measurement errors or data acquisition errors or data processing errors (Rasdorf, 2000). Some methodologies can be implemented to reduce the impact of certain types of errors but they can never be completely eliminated. The greater the degree of error in a data, the lesser is the reliability of results of analyses based on them. This is referred as *GIGO* [Garbage In Garbage Out]. Error in a spatial dataset is inherent and is of two types: *inherent or source error and operational or introduced error* which contribute to the reduction of quality of data products. *Inherent error* is the error

present in the source documents and data. *Operational error* is the amount of error produced through the data capture and manipulation functions of GIS. *Inherent error* is almost impossible to avoid whereas operational error resulting from geo-spatial analysis could be avoided by the appropriate choice of instruments for measurements, algorithms for processing and models for analysis (Buckley, 1997). In order to minimize the error, information about the origin and the properties of errors should be analysed if not by the end of the analysis process, the original error, which is easy to identify and eliminate at the first step, would become more complex and become almost impossible to remove. Bad data should, thus, be identified early (Rasdorf, 2000). As no model is perfect and since the choice of the model is based on the user, the error rate rises if the appropriate model is not selected.

While error will always exist in any scientific process, the aim within GIS processing should be to identify existing error in data sources and minimize the amount of error added during processing. Because of cost constraints it is often more appropriate to manage error than attempt to eliminate it. There is a trade-off between reducing the level of error in a data base and the cost to create and maintain the database (Buckley, 1997). There is a need for developing error statements for data contained within geographic information systems (Vitek et al, 1984). The integration of data from different sources and in different original formats (e.g. points, lines, and areas), at different original scales, and possessing inherent errors can yield a product of questionable accuracy (Vitek et al, 1984). The accuracy of a GIS derived product is dependent on characteristics inherent errors in the source products, and on user requirements, such as scale of the desired output products and the method and resolution of data encoding (Marble and Peuquet, 1983). As the number of layers in an analysis increases, the number of possible opportunities for error also increases (Newcomer, Szajgin, 1984).

3.1 Error Propagation in GIS Analysis

Error propagation in spatial analysis is defined as the process whereby error is propagated from the original dataset to the resulting dataset generated by a spatial operation. The two approaches for modelling error propagation in spatial analyses are the analytical approach such as error propagation law in statistics and simulation approach such as Monte Carlo simulation for either raster based or vector based spatial analysis environment (Abbaspour at al. 2003). When a spatial dataset is generated, errors and uncertainties starting from data collection, processing, analysis of data and model development, get accumulated and propagated to the final output maps. Thus the data products possess many sources of uncertainties varying spatially and temporally. To control and maintain the quality of a spatial data, it is crucial to spatially identify the sources of uncertainties, model their accumulations and propagation and finally to quantify them. Lunetta et al (1991) systematically discussed the sources and classification of the errors and uncertainties and suggested priorities for error quantification. Several methods were used for assessing the sources of uncertainty, including Monte Carlo methods (Heuvelink, 1998; Openshaw,1992), Fourier Amplitude Sensitivity Test [FAST] (Cukier et al., 1973), Taylor Series (Gertner et al., 1995), Polynomial Regression (Gertner et al. 1996) and Sobol's method (Sobol, 1993).

If uncertainty analysis is ignored, there might be incorrect predictions and analyses that are fatally flaw. A sensitivity

analysis can be applied for studying propagation of attribute uncertainty in a spatial analysis. If errors of input data are unknown, a sensitivity analysis can be applied to study the relationships between input and output errors in a function transformation. One limitation of a sensitivity analysis is that it cannot give the mathematical expression of attribute uncertainty propagation analytically. Lodwick et al. (1988) reviewed methods and applications of sensitivity analysis for attribute uncertainties analysis. Basically for each source layer in a GIS operation the uncertainty which is present in the data will be propagated through the operation becomes the input of another.

4. UNCERTAINTIES IN SPATIAL DATA

'Uncertainty' can refer to *imprecision, vagueness, ambiguity* or anything that is undetermined. *Imprecision* refers to the level of variation associated with a set of measurements or the lack of quality precision. *Vagueness* refers to a lack of clarity in meaning and is normally associated with the difficulty of making a sharp or precise distinction in relation to an object in the real world. *Ambiguity* is associated with either one or many relationships, or with a form of lack of clarity, which implies one or more meanings. Unlike error, uncertainty may be caused by mistakes and it may also be caused by incomplete information. The uncertainties in spatial data and analysis result from four major factors: (i) inherent uncertainty in the real world, (ii) the limitation of human knowledge in cognition of the real world, (iii) the limitation of the measurement technologies for obtaining spatial data, and (iv) the potential of generating and propagating uncertainty in the spatial data processing and analysis (Shi, 2010). During the past two decades, uncertainty modelling and the quality of the spatial data had been studied by many researchers and organizations; for example, Quality elements of spatial data were addressed by Shi et al. (2002). Different aspects of uncertainty are addressed by correspondingly different mathematical theories. For instance, probability and statistical theory are employed to describe imprecision. It has been demonstrated that fuzzy measurements can be used to describe vagueness, whereas ambiguity can be quantified by discordance measures, confusion measures, and non-specificity measures (Shi, 2010). A measurement of information was introduced by Hartley in 1928 (known as Hartley's measure) to describe ambiguity in a crisp set. Shannon's entropy measures uncertainty in information theory, based on the probability theory. Fuzzy topology theory was used to model uncertain topologic relationships between spatial objects (Liu & Shi, 2006).

4.1 Modelling Uncertainties

The process of handling uncertainty in spatial data involves three steps. The first step involves the identifying the kind of error information available in each of the GIS data layer. The type of error information will determine which assumptions have to be made when error is modelled (Lanter & Veregin, 1992). The second step is the definition of a conceptual model of error. The choice of the error model depends mainly on the kind of error information that is available and also on the error propagation technique that will be applied. Finally an appropriate error propagation function has to be selected to verify how error is propagated through the GIS operations. One of the most important hurdles in handling uncertainty in spatial database is the lack of knowledge about the error which is present in the source data. The type of error information which is available will define what kind of error modelling can be

applied, whether error can be differentiated thematically and/or spatially, and as such will largely determine the quality of the whole error modelling process. Due to a potentially large number of parameters and the heterogeneity of data sources, the uncertainty of the results are difficult to quantify (Feizizadeh et al., 2014). With increasing interest in analysis of uncertainty, many methods have been developed to estimate the model uncertainties. Choice between the methods is according to the level of model's complexity and preferences of a modeller. Shrestha & Solomantine, (2008) proposed various methods including analytical methods, approximation methods, sampling based methods, bayesian methods and methods based on analysis of the model errors. Analytical methods compute probability distribution function of model outputs. They are applicable for simple models where the propagation of uncertainty through the model is straightforward and can be analytically expressed. Their applicability is generally limited to models with linear summation of independent inputs. Approximation methods provide only the moments of the distribution of the uncertainty output variable. Some of approximation based methods depend on the use of the Taylor series expansions to propagate the uncertainty through model.

Furthermore, the main advantage of approximation methods is that it is enough to propagate the moments of each probability distribution of the model inputs instead of the entire probability distributions. On the other side, the main disadvantages of these methods are, firstly, that they cannot be applied to problems with discrete or discontinuous behaviours because the model function should be differentiable. Secondly, it is computationally intensive as they typically require the evaluation of second order (may be higher) derivatives of the model. Thirdly, although these techniques are capable of propagating central moment of input distributions, information regarding the tails of the input distributions cannot be propagated. Sampling based methods are the most common techniques to study the propagation of uncertainty. These methods involve running a set of model simulations at a set of sampled points from probability distributions of inputs and establishing a relationship between inputs and outputs using the model results. There are two types of sampling, the first one is simple random sampling that depends on the entire population and second is stratified random samples, which separate the population elements into non overlapping groups called strata. These methods are able to work with complex and non-linear model and are capable of solving a great variety of problems. They do not require access to the model equations. Monte Carlo methods and Latin Hypercube Sampling methods are the most common sampling based uncertainty analysis methods. Bayesian methods utilize Bayes' theorem to estimate or update the probability distribution function of the parameters of the model and consequently estimate the uncertainty of model results. Generalized Likelihood Uncertainty Estimation [GLUE] is one of the sampling methods that is also based on Bayesian methods. The idea in the methods based on the analysis of model errors is to analyse the model residuals that occurred in reproducing the observed historical data (Devillers et al., 2010).

5. URBAN LAND USE SIMULATION MODELS

A model is a simplified representation of reality in either material form or symbolic form for purposes of description, explanation and forecasting of planning. GIS modelling involves symbolic representation of locational properties as well as thematic and temporal attributed describing characteristics

and conditions of space and time (Berry J. K., 1995). Urban models help scientists to analyse the form, speed and impact of urban growth in the past to assess current trends and to simulate scenarios of future development. A model maybe suitable for one area in terms of parameters and it has to be localized in another research area since the underlying factors of growth are different. An urban model has to be calibrated so as to 'learn' the endogenous characteristics of the particular environment that they explain and simulate (Silva & Clarke, 2002). However, in order to model urbanisation processes across different areas, it is vital to test the efficiency of the model's algorithms at capturing and simulating the land transformations that are specific to a place (Batty & Xie, 1994; Clarke & Hoppen, 1996; Li & Yeh, 2000). Cellular Automata (CA) models are ideal for modelling land use in areas where data is unavailable and experiencing rapid urban growth (Hill & Lindner, 2011).

5.1 Cellular Automata (CA) Model

Invented by John Von Neumann, cellular automata are discrete dynamical systems that model complex behaviour based on simple, local rules animating cells on a lattice. In CA, space is represented by a uniform M-dimensional grid of cells (M = 1,2,..) with each cell containing some data. Among all developed Urban Growth Models, CA models perform well in simulating urban development than conventional mathematical models. CA are able to predict urban growth based on the assumption that past urban development affects future patterns through local interactions among land uses (Sante, et al.,2010). CA simplifies the simulation of complex systems. CA are appropriate for urban modelling as the process of urban spread is entirely local in nature. Thus CA models are outstanding depicters of urban dynamics (Divigalpitiya, et al., 2007). Models based on CA are impressive due to their technological evolution in connection to urban application. Developing a CA model involves rule definition and calibration to produce results consistent with historical data and future prediction with set rules (Batty & Xie, 1994; Clarke & Gaydos, 1998; Yang & Lo, 2003).

Many CA based urban growth models are documented in the literatures (Triantakonstantis & Mountrakis, 2012). CA model reduces spaces into square grids (White & Engelen, 1993). CA implements the defined transition rules in recursive form to match the spatial pattern. CA models are usually designed based on individual preference and application requirements with transition rules being defined in an adhoc manner (Li & Yeh, 2003). Most of the developed CA models need intensive computation to select the best parameter values for accurate modelling.

5.2 Cellular Automata Modelling

Earlier CA models viewed space as a lattice of identical cells in the 1960's and the 1970's in modelling urban land uses. Cities were considered as spatial distributed systems. Batty & Xie (1994) developed life CA models. The model was born on the idea of Conway's game of life. Clarke et al., (1997) developed self modifying CA. The model was called SLEUTH [Slope, Land cover, Exclusion, Urban, Transportation and Hill Shade] and explored complexities of urban cells and incorporated biophysical factors namely: urban, road, transportation, slope and exclusive layer. CA offers many advantages for modelling urban phenomena including representing urban dynamics and space time dynamics (Torrens & O'Sullivan, 2001). CA consists

of five basic elements namely, cell space, cell state, cell neighbourhood, transition rules and time. Several approaches had been identified in defining the basic elements of a CA to simulate the process of urban development (Torrens & O'Sullivan, 2001; Li & Yeh, 2000; Takeyama & Couelelis, 1997; White et al., 1997).

The space in which an automaton exists is called a *cell space or lattice*. *Cell states* most commonly represent land use but may be used to represent any spatially distributed variable for the purpose of modelling its spatial dynamics (White & Engelen, 2000). Each cell in CA can take only one state from a set of state at any one time and it can be a number that represents a property (Liu, 2008). *Transition rules* define how the state of one cell changes in response to its current state and the state of its neighbour (Liu, 2008). These rules determine how CA states adapt over time and can be designed using any combination of conditional statements or mathematic operators (Benenson & Torrens, 2004). The *cellular neighbourhood* of a cell consists of the surrounding (adjacent) cells. The neighbourhood is outlined by the grid (cell space) in which the CA are located. Liu (2008) defined *time* as the temporal dimension in which a CA exists. Thus the states of all cells are updated simultaneously at all iterations over time.

5.3 Uncertainties of a CA Model

Error Propagation assessment is important for understanding the results of simulation of urban land use models. Initial conditions, parameter values and stochastic factors influence the simulation results. Unexpected features may emerge during CA simulation due to interactions of various local actions. Error and uncertainty can propagate through modelling process. All the errors inherent in individual GIS layers can contribute to the final errors of the output during the overlay of different layers. To analyse error propagation, Monte Carlo simulation is used in many applications due to easy implementation and applicability. Error propagation in CA models is different from that of GIS overlay operations. In GIS operations, mathematical expressions can be given to calculate the errors presented in simple overlay using the logical AND and OR operators. But CA adopts complicated configurations using neighbourhood and iterations.

Urban CA models require the input of a large set of spatial data for realistic simulation. The outcome of CA models will be affected by a series of errors and uncertainties from data sources and model structures. Simulation is a dynamic process in which many complex features can arise according to transition rules. The conversion of the state of a central cell is influenced by the states of its neighbourhood. It is impossible to develop strict mathematical equations to represent the error propagation in dynamic process. The error problems of CA models are further exacerbated by taking into account model uncertainties. In CA simulation, not only input errors propagate through the simulation process, but model errors as well. Like any computer models, CA models could disagree with reality even when the inputs were completely error-free. CA models are only approximation to reality. Most of the existing CA models are just loosely defined and a unique model does not exist. Various types of CA models have been proposed according to individuals' perception and preference, and requirements of specific applications. The simulation results are hard to repeat when different CA models are used. A series of inherent model errors can be identified for CA models (A.Gar-On Yeh & X. Li, 2006).

6. HYDROLOGICAL MODELS

Hydrologic and water quality models are increasingly used to evaluate the impacts of climate, land use, and land and crop management practices on the quantity and quality of land and water resources. Calibration and validation of these models are necessary before using them in research or real-world applications. No universally accepted procedures or guidelines for calibration and validation currently exist in the literature. However, there are numerous viewpoints among model developers and model practitioners as to how calibration and validation should be implemented and reported to assist the peer-review process and to withstand legal scrutiny (Refsgaard & Storm, 1995; Refsgaard & Storm, 1996; Refsgaard, 1997; Santhi et al., 2001; Jakeman et al., 2006; Moriasi et al., 2007; Engel et al., 2007; Bennett et al., 2010).

6.1 Soil and Water Assessment Tool (SWAT) Model

A hydrologic simulation model is composed of three basic elements, which are: (1) Equations that govern the hydrologic processes, (2) Maps that define the study area and (3) Database tables that numerically describe the study area and model parameters. The SWAT model (Arnold et al., 1998a) is a hydrologic/water quality tool developed by the United States Department of Agriculture-Agriculture research Service [USDAARS]. The SWAT model, as a physically based model could simulate most of the key processes at basin scale (Arabi et al., 2007). SWAT allows the use of GIS inputs like DEM, land use and soil maps. Sherman (1932) developed the unit hydrograph concept of modelling on the basis of superposition. These models are based on complex physical theory and require large amount of data and computational time. Usually, the hydrological processes of SWAT model are divided into two phases: the land phase and the channel phase. The Modified Universal Soil Loss Equation [MUSLE] was used to estimate sediment yield at the hydrologic research unit (HRU) level. More information about SWAT model was discussed by Douglas-Mankin et al. (2010).

6.2 SWAT Modelling

Watershed hydrology and river water quality models are important tools for watershed management for both operational and research programs (Quilb'e & Rousseau, 2007; Van et al., 2008; Sudheer & Lakshmi, 2011). However, due to spatial variability in the processes, many of the physical models are highly complex and generally characterized by a multitude of parameters (Xuan et al., 2009). The SWAT model parameters can be divided into the conceptual group and the physical group (Gong et al., 2011). A typical hydrological model generally consists of a large number of equations describing the hydrological processes and a small number of parameters describing the watershed properties. Without a realistic assessment of parameter uncertainty, it is difficult to have confidence in tasks, such as evaluating prediction limits on future hydrological responses and assessing the value of regional relationships between model parameters and watershed characteristics (Kuczera & Parent, 1998).

6.3 SWAT Model Uncertainties

Much attention has been paid to uncertainty issues in hydrological modelling due to their great effects on prediction and further on decision making (Van et al., 2008; Sudheer & Lakshmi, 2011). Usually, the uncertainty in hydrological

modelling is from model structures, input data and parameters (Lindenschmidt et al., 2007). Currently, parameter uncertainty is a hot topic in the uncertainty research field (Shen et al., 2008; Sudheer at al. 2011). Several calibration and uncertainty analysis techniques have been applied in previous research work, such as the first order error analysis [FOEA] (Melching & Yoon, 1996), the Monte Carlo method (Kao & Hong, 1996) and the Generalized Likelihood Uncertainty Estimation Method [GLUE] (Beven & Binley, 1992). The GLUE method can also handle the parameter interactions and non-linearity implicitly through the likelihood measure (Vazquz et al., 2009). In addition, GLUE is a simple concept and is relatively easy to implement. The GLUE method (Beven & Freer, 2001) is an uncertainty analysis technique inspired by importance sampling and regional sensitivity analysis (Hornberger & Spear, 1981). In GLUE, parameter uncertainty accounts for all sources of uncertainty; i.e., input, structural, parameter and response uncertainties. Hence, this method is widely used in many areas as an effective and general strategy for model calibration and uncertainty estimation associated with complex models.

With the advances in computing technology, Monte Carlo based methods are popular , with strengths in dealing with the non-linearity and interdependency of parameters in complex hydrological models. Two Monte Carlo based approaches are commonly seen in literatures: importance sampling and Markov Chain Monte Carlo (MCMC) simulation (e.g. Beven & Binley 1992; Smith and Roberts, 1993; Kuczera & Parent, 1998; Bates & Campbell, 2001; Gallagher & Doherty, 2007). Importance sampling is a technique for randomly sampling from a probability distribution and was implemented in Generalised Likelihood Uncertainty Estimation (GLUE) by Beven & Binley (1992). The efficiency of this algorithm depends strongly on the choice of the importance distribution. If one or more importance weights dominate, the algorithm can produce unreliable inferences (Gelman et al., 1995). MCMC is one of the most important numerical technique for creating a sample from the posterior distribution, which has been widely used in hydrological modelling to quantify parameter uncertainties (e.g. Kuczera & Parent, 1998; Campbell et al., 1999; Makowski et al., 2002; Vrugt et al., 2003). Its underlying rationale is to set up a Markov chain to simulate the true posterior distribution by generating samples from a random walk. An obvious advantage of this method is that it does not require linearity assumptions in model or even that model outputs do not need to be differentiable with respect to parameter values (Gallagher & Doherty, 2007). Because of its robust performance, MCMC is often used to assess parameter uncertainties in combination with GLUE or Bayesian inference by estimating a probability density for model parameters conditioned on observations. Blasone et al. (2008) recently found that using a MCMC sampling scheme coupled with GLUE significantly improves the efficiency and effectiveness of the methodology of GLUE. Another widely used method for parameter estimation and uncertainty analysis is the Bayesian method which provide more information than single-point estimates (Bates & Campbell, 2001; Engeland et al., 2005; Gallagher & Doherty, 2007; Yang et al., 2007). The posterior distribution could be obtained by applying Bayes' theorem based on a prior distribution and observed data (Gelman et al., 1995). The Monte Carlo based approaches are often used to generate a large-enough sample from the posterior distribution so that desired features of the posterior distribution may be summarized.

7. CONCLUSIONS

Uncertainty analysis, a prerequisite for model building, aims at quantifying the overall uncertainties associated with the model response as a result of uncertainties in the model input. This paper reviewed the quality components and uncertainties of spatial data with detailed emphasis on urban growth simulation and hydrological models and their model uncertainties.

REFERENCES

Anthony Gar-On Yeh and Xia Li, 2006. Errors and uncertainties in urban cellular automata, *Computers, Environment and Urban Systems* (30), pp. 10-28.

Arabi, M., Govindaraju, R.S and Hantush M. M., 2007. A probabilistic approach for analysis of uncertainty in the evaluation of watershed management practices. *Journal of Hydrology* (333), pp. 459–471.

Arnold, J. G., Srinivasan, R., Muttiah, R.S. And Williams, J.R., 1998. Large area hydrologic modeling and assessment Part I: model development. *Journal of the American Water Resources Association* 34(1), pp. 73-89.

Bakhtiar Feizizadeh, Piotr Jankowski and Thomas Blaschke, 2014. A GIS based spatially-explicit sensitivity and uncertainty analysis approach for multi-criteria decision analysis, *Computers & Geosciences*, (64), pp.81–95.

Bates, B.C. And Campbell, E.P., 2001. A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modelling. *Water Resour. Res*. (37), pp. 937–947.

Batty, M., and Y. Xie. 1994. From cells to cities. Environment and Planning B-Planning & Design 21: pp. 31-38.

Benenson, I., and Torrens, P. M., 2004. Geosimulation: Automata-Based Modeling of Urban Phenomena (1 Ausg.). London: Wiley.

Bennett, N. D., B. F. W. Croke, A. J. Jakeman, L. T. H, Newham, and J. P. Norton. 2010. Performance evaluation of environmental models. In Proc. Intl. Congress on Environmental Modelling and Software (iEMSc 2010), 1703-1711. D. A. Swayne, W. Yang, A. A. Voinov, A. Rizzoli, and T. Filatova, eds. Manno, Switzerland.

Berry, J. K. 1995. Spatial Reasoning for Effective GIS. New York: John Wiley & Sons.

Beven, K. and Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction, *Hydrol. Process*, (6), pp. 279-298.

Beven, K. J. and Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems, *J. Hydrol*., (249), pp. 11-29.

Blasone, R.S., Vrugt, J.A., Madsen, H., Rosbjerg, D., Robinson, B.A., and Zyvoloski, G.A.,2008. Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov chain Monte Carlo sampling. Adv. Water Resour.

Buckley, D.J., 1997. "The GIS Primer, The Nature of Geographic Information." Innovative GIS Solutions, Inc., An

Introduction to Geographic Information Systems. Website: http://www.innovativegis.com/education/primer/nature.html.

Campbell, E.P., Fox, D.R. And Bates, B.C., 1999. A Bayesian approach to parameter estimation and pooling in nonlinear flood event models. *Water Resour. Res.* 35 (1), pp. 211–220.

Chong, A. K., 1997. "A Field Check Sampling Procedure to Evaluate the Positional Accuracy of Digital Landbases," Proceedings of ASPRS/ACSM, Volume 2, Annual Convention & Exposition of Technical Papers, Seattle, Washington, pp. 1.

Clarke, K., Hoppen, S., and Gaydos, L., 1997. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. Environment and Planning B: *Planning and Design*, 24(2), pp. 247-261.

Clarke , K. and Gaydos, L. J. 1998. Loose coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographical Information Science*, 12(7), pp. 699-714.

Clarke, K., Hoppen, S. and Gaydos, L., 1996. Methods and techniques for rigorous calibration of cellular automaton model of urban growth. Third international conference/workshop on integrating GIS and environmental modeling. Santa Fe, New Mexico.

Cukier, R.I., C.M. Fortuin, K.E. Shuler, A.G. Petschek, and J.H. Schaibly, 1973. Study of the Sensitivity of coupled reaction systems to uncertainties in rate coefficients, I: Theory, *Journal of Chemical Physics*, (59), pp. 3873–3878.

Devillers, R, Stein, A, Bédard, Y, Chrisman, N, Fisher, P, and Shi, W. 2010. Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities. *Transactions in GIS*, 14(4), pp. 387–400.

Divigalpitiya, P., Ohgai, A., Tani, T., Watanabe, K. and Gohnai, Y. 2007. Modeling Land Conversion in the Colombo Metropolitan Area Using Cellular Automata. *Journal of Asian Architecture and Building Engineering*, 6(2), pp. 291-298.

Douglas-Mankin, K. R., R. Srinivasan and J. G. Arnold. 2010. Soil and Water Assessment Tool (SWAT) model: Current development and applications. Trans. ASABE 53(5), pp. 1423-1431.

Engel, B., D. Storm, M. White, J. Arnold, and M. Arabi. 2007. A hydrologic/water quality model application protocol. *J. American Water Resour. Assoc.* 43(5), pp. 1223-1236.

Engeland, K., Xu and C.Y., Gottschalk, L., 2005. Assessing uncertainties in a conceptual water balance model using Bayesian methodology. *Hydrol. Sci. J*.(50), pp. 45–63.

Frank Forier and Frank Canters., 1996. A User-friendly Tool for Error Modelling and Error Propagation in a GIS Environment, In Mowrer, H.T., Czaplewski, R.L. & Hamre, R.H. (eds), *Spatial Accuracy Assessment in Natural resources and Environmental Sciences*. USDA Forest Service Technical Report RM-GTR-227, pp.225-234.

Gallagher, M. and Doherty, J., 2007. Parameter estimation and uncertainty analysis for a watershed model. *Environ. Modell. Softw.* (22), pp. 1000–1020.

Gelman, A., Carlin, J.B., Stren, H.S. And Rubin, D.B., 1995. Bayesian Data Analysis. Chapman and Hall, London. pp. 526.

Gertner, G.Z., Parysow, P. and Guan B, 1996. Projection variance partitioning of a conceptual forest growth model with orthogonal polynomials, *Forest Science*, 42(4), pp. 474-486.

Gertner, G.Z., X. Cao, and H. Zhu, 1995. A quality assessment of a Weibull based growth projection system, *Forest Ecology and Management*, (71), pp. 235–250.

Gong, Y. W., Shen, Z. Y., Hong, Q., Liu, R. M. and Liao, Q., 2011. Parameter uncertainty analysis in watershed total phosphorus modeling using the GLUE methodology, *Agr. Ecosyst. Environ.*, (142), pp. 246-255.

Guangxing Wang, George Z. Gertner, Shoufan Fang, and Alan B. Anderson., 2005. *Photogrammetric Engineering & Remote Sensing* 71(12), pp. 1423–1432.

Heuvelink, G.B.M., 1998. Error Propagation in Environmental Modeling with GIS, Taylor & Francis Ltd., London.

Hill, A., and Lindner, C. 2011. Simulation informal urban growth in Dar es Salaam, Tanzania - A CA-based land-use simulation model supporting strategic urban planning. In P. Mandl, & A. Koch, Modeling and Simulating Urban Processes (S. 77-98). Munster: LIT-Verlag.

Hornberger, G. M. and Spear, R. C., 1981. An approach to the preliminary analysis of environmental systems, *J. Environ. Manage.*, (12), pp. 7-18.

Jakeman, A. J., R. A. Letcher, and J. P. Norton. 2006. Ten iterative steps in development and evaluation of environmental models. *Environ. Modelling Software* 21(5), pp. 602-614.

Kao, J. J. and Hong, H. J., 1996. NPS model parameter uncertainty analysis for an off-stream reservoir, *J. Hydrol.*, (32), pp. 1067-1079.

Kuczera, G., and E. Parent., 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm, *J. Hydrol.*,(211), pp. 69–85.

Lanter, D.P. and H. Veregin, 1992. A research paradigm for propagating error in layer-based GIS, *Photogrammetric Engineering & Remote Sensing*, 58 (6), pp. 825- 833.

Li, X. and Yeh, A. 2000. Modelling sustainable urban development by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science,* 14(2), pp. 131-152.

Li, X. and Yeh, A. G., 2003. Error propagation and model uncertainties of cellular automata in urban simulation with GIS. 7th International Conference on GeoComputation. Southampton: University of Southampton.

Lindenschmidt, K. E., Fleischbein, K. and Baborowski, M., 2007. Structural uncertainty in a river water quality modelling system. *Ecol. Model.*, (204), pp. 289-300.

Liu, K. F. and Shi, W. Z., 2006. Computing the fuzzy topological relations of spatial objects based on induced fuzzy topology. *Int. J. Geog. Inform. Sci*. 20(8): pp. 857–883.

Liu,Y.,2008. Modelling Urban Development with Geographical Information Systems and Cellular Automata (1 ed.). Florida: CRC Press.

Lodwick, W. A., Munson W., and Svoboda, L., 1988. Sensitivity Analysis in Geographic Information Systems, Part 1: Suitability Analysis. Research Paper RP8861, University of Colorado–Denver.

Lunetta, R.S., Congalton R. G., Fenst.ermaker L.K, Jensen J. R.,McGwire K. C., and Tinney L.R., 1991. Remote sensing and geographic information system data integration: Error sources and research issues, *Photogrammetric Engineering & Remote Sensing,* 57(6), pp. 677–687.

Makowski, D., Wallach, D. and Tremblay, M., 2002. Using a Bayesian approach to parameter estimation, comparison of the GLUE and MCMC methods. *Agronomie* (22), pp. 191–203.

Marble D.F. and Peuquet D.J., 1983. Geographic Information System and Remote Sensing: Manual of Remote Sensing, 2nd ed. Falls church, VA: *American Society of Photogrammetry and Remote Sensing*, pp. 923-958.

Melching, C. S. and Yoon, C. G., 1996. Key sources of uncertainty in QUAL2E model of Passaic River, *J Water Res. Pl.-ASCE*, (122), pp. 105 -113.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50(3), pp. 885-900.

Newcomer, J. A., and J. Szajgin. 1984. ''Accumulation of Thematic Map Errors in Digital Overlay Analysis.'' *The American Cartographer* (11), pp. 58–62.

Openshaw S., 1992. Learning to live with errors in spatial databases, Accuracy of spatial databases [M. Goodchild and S. Gopal, editors] Taylor & Francis Ltd., pp. 1294-1298.

Quilbe, R., and Rousseau, A. N., 2007. An integrated modelling system for watershed management – sample applications and current developments, *Hydrol. Earth Syst. Sc.,* (11), pp. 1785-1795.

Rahim Ali. Abbaspour, Mahmoud Reza Delavar and Reihaneh Batouli., 2003. The Issue of Uncertainty Propagation in Spatial Decision Making. The 9th Scandinavian Research Conference on Geographical Information Science, Espoo, Finland – Proceedings.

Refsgaard, J. C., and B. Storm. 1995. MIKE SHE. In Computer Models of Watershed Hydrology, V. Singh, ed. Highlands Ranch, Colo.: Water Resources Publications, pp. 809-846.

Refsgaard, J. C., and B. Storm. 1996. Construction, calibration,and validation of hydrological models. In Distributed Hydrologic Modeling, 41-54. M. B. Abbot and J. C. Refsgaard, eds. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Refsgaard, J. C. 1997. Parameterization, calibration, and validation of distributed hydrological models. *J. Hydrol*.198(14): pp. 69-97.

Ries, T., 1993. "Design Requirements for Location as a Foundation for Transportation Information Systems," GIS-T '93 Geographic Information Systems for Transportation Symposium, Albuquerque, New Mexico, pp.48.

Santé, I., García, A. M., Miranda, D., and Crecente, R. 2010. Cellular automata models for the simulation of real-world urban processes: A review and analysis *Landscape and Urban Planning*, 96(2), 108-122.

Santhi, C., J. G. Arnold, J. R. Williams, W. A. Dugas, R. Srinivasan, and L. M. Hauck. 2001. Validation of the SWAT model on a large river basin with point and nonpoint sources. *J. American Water Resour. Assoc.* 37(5), pp. 1169-1188.

Shen, Z.Y., Hong, Q. and Yu, H., 2008. Parameter uncertainty analysis of the non-point source pollution in the Daning River watershed of the Three Gorges Reservoir Region, China, *Sci. Total Environ.*, (405), pp. 195-205.

Sherman, L. K., 1932. Streamflow from Rainfall by the Unit Graph Method. *Eng. News Rec.*, (108), pp. 501-505.

Shi, W. Z., Fisher, P. F. and Goodchild, M. F. (ed.), 2002. Spatial Data Quality. London and New York: Taylor & Francis.

Shi W, 2010. Modeling Uncertainties in Spatial Data and Spatial Analyses, CRC press, Taylor and Francis Group.

Shrestha, D.L. and Solomatine, D. 2008. Data-driven approaches for estimating uncertainty in rainfall-runoff modelling. *Intl. J. River Basin Management,* 6(2), pp. 109-122.

Silva, E. and Clarke, K. C., 2002. Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal. *Computers, Environment and Urban Systems*, 26(6),pp. 525–552.

Smith, A.F.M. and Roberts, G.O., 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. Roy. Stat. Soc., Ser.* B (55), pp. 3–23.

Sobol, I.M., 1993. Sensitivity estimates for nonlinear mathematical models, *Mathematical Modeling and Computational Experiments*, 1(4), pp. 407–414.

Sudheer, K. P., Lakshmi, G. and Chaubey, I., 2011. Application of a pseudo simulator to evaluate the sensitivity of parameters in complex watershed models, *Environ. Modell.Softw.,* (26), pp. 135- 143.

Takeyama, M. and Couclelis, H., 1997. Map dynamics: integrating cellular automata and GIS through Geo-Algebra. International Journal Geographical Information Sciences, 11(1), pp. 73-91.

Torrens , P. M. and O'Sullivan , D. 2001. Cellular automata and urban simulation: where do we go from here? Environment and Planning B: *Planning and Design*, 28(2), pp. 163-168.

Triantakonstantis, D. and Mountrakis, G. 2012. Urban Growth Prediction: A Review of Computational Models and Human Perceptions. *Journal of Geographic Information System*, 4(6), pp. 555-587.

Van, G. A., Meixner, T., Srinivasan, R. and Grunwals, S, 2008. Fit for-purpose analysis of uncertainty using split-sampling evaluations, *Hydrolog. Sci. J.*, (53), pp. 1090-1103.

Vazquez, R. F., Beven, K. and Feyen, J., 2009. GLUE based assessment on the overall pre-dictions of a MIKE SHE application, *Water Resour. Manage.*, (23), pp. 1325-1349.

Vitek, J.D., Walsh S.J. and Gregory M.S., 1984. Accuracy in geographic information systems: an assessment of inherent and operational errors. Proceedings PECORA IX Symposium, pp. 296-302.

Vrugt, J.A., Gupta, H.V., Bouten, and W., Soorooshian, S., 2003. A shuffled complex evolution metrolopolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour. Res*. 39 (8), pp. 1201–1215.

White, R. and Engelen, G., 2000. High-resolution integrated modelling of the spatial dynamics of urban and regional systems. *Computers, Environment and Urban Systems*, 24(5), pp. 383-400.

White, R., Engelen, G. and Uljee, I., 1997. The use of constrained cellular automata for high resolution modelling of urban land-use dynamics. *Environment and Planning B*, 24(3), pp. 323-343.

White, R. and Engelen, G., 1993. Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use. *Environment and Planning A,* 25(8), pp. 1175-1199.

William Rasdorf., 2000. A GIS Data Quality Case Study: Accuracy Comparison of Length Measurements Using GIS/NED and DMI, Technical Report, Department of Civil Engineering, North Carolina State University.

Xuan, Y., Cluckie, I. D. and Wang. Y., 2009. Uncertainty analysis of hydrological ensemble forecasts in a distributed model utilising short-range rainfall prediction, *Hydrol. Earth Syst. Sci.,* (13), pp. 293-303.

Yang, J., Reichert P. and Abbaspour, K.C., 2007. Bayesian uncertainty analysis in distributed hydrological modelling: a case study in the Thur River basin (Switzerland). *Water Resour. Res*. (43), W10401.

Yang , X. and Lo, C. P., 2003. Modelling urban growth and landscape changes in the Atlanta metropolitan area. International Journal of Geographical Information Science, (17), pp. 463-488.