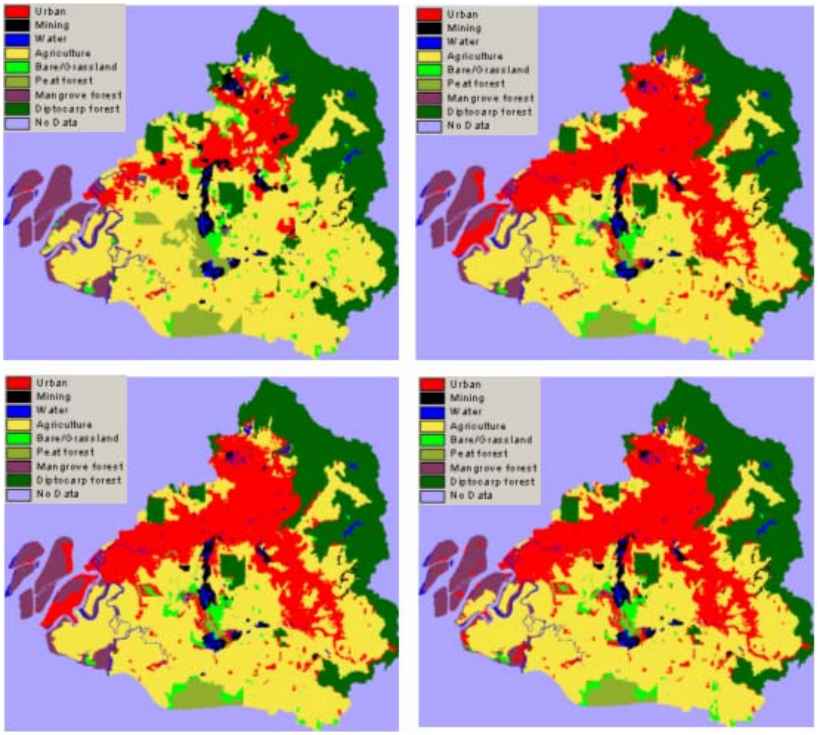
Centre for Geo-Information

Thesis Report GIRS-2005-06

Uncertainty Propagation in Predictive Modeling of Land Use Change in the Kuala Lumpur Region

Li Jie



March 2005

WAGENINGEN UNIVERSITY
WAGENINGEN UR

Uncertainty Propagation in Predictive Modeling of Land Use Change in the Kuala Lumpur Region

Li Jie

Registration number 79 11 06 515 030

Supervisors:

Dr. Gerard B.M. Heuvelink
Dr. ir. Sytze de Bruin

A thesis submitted in partial fulfillment of the degree of Master of Science at Wageningen University and Research Centre, The Netherlands.

March 2005
Wageningen, The Netherlands

Thesis code number: GRS-80337
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing
Thesis Report: GIRS-2005-06

Abstract

Predictive modelling of land use change combined with GIS comprise a set of indispensable tools to describe land use change processes, manage land use efficiently and predict possible land use changes. However, the inevitable uncertainties in spatial data and land use change model will propagate through the model and cause uncertainties in the model outputs. These uncertain results may be used to support decision making, which may cause unwanted results if the uncertainties in model inputs are significant to the application. Therefore, the accuracy of the model results need to be quantified and analysed to ensure that the uncertainties in the model outputs can be accounted for in further analysis and decision-making.

In this thesis research, spatial uncertainty propagation analysis was performed to evaluate the effect of uncertainties associated with input data on predictions of land use change made by the CLUE-S (Conversion of Land Use and its Effects at Small regional extent) model in the Kuala Lumpur region in Malaysia, which has been undergoing rapid developments since the early 1980's and where urbanization has dramatically changed land use and land cover in the region and the area around. The analysis included identification of important sources of input uncertainty, quantification of these uncertainties, propagation of the uncertainties to the land use change predictions, and analysis of the results of the uncertainty propagation.

Elevation (and the slope derived from it) and the travel time to sawmills, important towns and highways in general were identified as important uncertain inputs. Uncertainties in the Digital Elevation Model (DEM) were simulated using geostatistical methods that take spatial correlation into account. The required parameterization was done using the USGS 7.5 minute DEM documentation and using results from other studies. Uncertainties in travel times were also simulated using a pseudo random number generator. In this way normally distributed errors were generated with a mean equal to the average speed. The standard deviations of the errors in travel speed were specified using expert knowledge.

Next, quantified input uncertainties were propagated through the CLUE-S model to the predicted land use maps. Land use change prediction models are complex non-linear models and it is therefore not feasible to analyze error propagation with analytical methods. Instead, the Monte Carlo method was used. The idea of the Monte Carlo method is to compute the result of the model repeatedly, with uncertain input values randomly sampled from their joint probability distribution. The model results form a random sample from the distribution of the model output, so that parameters of the distribution, such as the mean and the variance, can be estimated from the sample.

Finally, the results of the spatial uncertainty propagation were analysed using Shannon's information entropy, confusion matrices and three land use fragmentation indices ('total area', 'clumpiness index' and 'perimeter to area ratio'). The results of the uncertainty analysis were

compared with a reference land use map and with the default result of the CLUE-S model, which was obtained using error-ignored inputs.

Overall, the propagated uncertainty in the predicted land use was small. DEM uncertainty caused substantial uncertainty in only a small part of the study area (i.e., the islands with mangrove forests that became partially urbanized). Elsewhere, the effects of DEM uncertainty were limited and the results were similar to the results obtained with the error-ignored DEM. Uncertainty in travel times affected a larger part of the study area. A comparison of the CLUE-S results with the reference land use map suggests that other sources of uncertainty also have an important contribution to the accuracy of the predicted land use. Uncertainty analysis of model settings and model parameters is therefore needed to gain a better understanding of the uncertainties involved with predicting land use change in the Kuala Lumpur region.

Keywords: Uncertainty propagation, CLUE-S model, Monte Carlo simulation, Digital Elevation Model, land use change, Shannon's information entropy, fragmentation index

Acknowledgements

I owe a lot to my supervisors Gerard Heuvelink and Sytze de Bruin. Without them, this thesis would not have been possible.

I am grateful to Gerard. He encouraged me to do the subject of uncertainty propagation. He also gave me many opportunities that I had never dreamed of and introduced me to some leading scientists in the field. His encouragement and trust stimulated me throughout the thesis. I just want to do my best and not to disappoint his good will.

I am indebted to Sytze. He patiently explained to me what uncertainty propagation is from the very beginning when I did not even know how to pronounce 'propagation'. He helped me throughout the thesis, gave me inspiration for my scripts (or sometimes, gave me scripts).

I am thankful to Peter Verburg, who provided the data and the CLUE-S model configurations for the Kuala Lumpur region, gave me a lot of help on the CLUE-S model and answered lots of questions at the beginning of the thesis research.

Thanks to Alias Mohd Sood in the Forestry Department Peninsular Malaysia, who provided the data for the Kuala Lumpur region.

I enjoyed the thesis work very much thanks to these three supervisors. Their comments and advice have always been very educational to me. I have learnt a great deal from the process, not only about uncertainty propagation but also about the attitude to conduct scientific research. The warm atmosphere and their good sense of humor also made every meeting an enjoyable experience. I hope that in the end this thesis is worthy of their efforts and trust.

Thanks to Willy ten Haaf, our course coordinator. He is always very kind to me and to all other students. Sometimes I just wonder how he could deal with so many things so well at the same time.

Thanks to Join Stuiver for his excellent teaching and patient help on Arc/Info and ArisFlow/ArisFlow Commander. After his class, I found programming in Arc/Info a pleasure rather than a task.

Thanks to all my colleague students. I will miss all the happy moments we spent in the thesis room and during the parties.

No word can express my gratitude to my parents, who put their child's welfare before all other considerations. I hope they are satisfied/proud of my work.

Table of Contents

List of Figures

List of Tables

List of Equations

List of abbreviations

The abbreviations listed are mainly used in the tables and captions of figures, tables and equations.

♦ A list of land use type in the Kuala Lumpur region case study and the abbreviation for each land use type (when applicable):

urban urban

mining mining

water water

agric agriculture

baregrs bare\grassland

peatfr peat forest

mgvfr mangrove forest

dipcfr diptocarp forest

♦ Other abbreviations used in tables and captions of tables and figures

dft The resulting land use map predicted by the CLUE-S model for the Kuala Lumpur region case study. It was produced using the default model settings presented in section 3.1 and error-ignored inputs. It is shown in Figure 3.3.

domi_uccd The dominant land use map for uccd_rst, i.e. a map displays the most frequently occurred land use type on every cell predicted by the CLUE-S model using uncertain travel speed, among all occurred land use type(s) for that cell predicted by the model using uncertain travel speed. The dominant land use map for uccd_rst shows the most probable land use type on each single cell that predicted by the CLUE-S model using uncertain travel speed.

domi_ucdem The dominant land use map for ucdem_rst, i.e. a map displays the most frequently occurred land use type on every cell predicted by the CLUE-S model using uncertain DEM, among all occurred land use type(s) for that cell predicted by the model using uncertain DEM. The dominant land use

map for ucdem_rst shows the most probable land use type on each single cell that predicted by the CLUE-S model using uncertain DEM.

entropy Specifically refers to Shannon's information entropy, when used in captions.

fi_uccd_rst The three land use fragmentation indices ('total area', 'clumpiness index' and 'perimeter to area ratio') of the model results using uccd_rst

fi_ucdem_rst The three land use fragmentation indices ('total area', 'clumpiness index' and 'perimeter to area ratio') of the model results using ucdem_rst

LU Land use

N.A. Not Analyzed

paccu The producer's accuracy

std Standard deviation

uaccu The user's accuracy

uccd Uncertain cost distance (as a result of uncertain travel speed). one hundred uncertain travel speed were simulated using the method described in the thesis.

uccd_rst Refers to a collection of the result (predicted land use map) of the CLUE-S model using simulated uncertain travel speed. Since one hundred uncertain travel speed were simulated, one hundred results of the CLUE-S model were obtained correspondingly. It refers to all the obtained results as a whole, not their average value or dominant value.

ucdem Uncertain DEM, created by adding uncertainties to the original error-ignored DEM. one hundred uncertain DEM were simulated using the method described in the thesis.

ucdem_rst Refers to a collection of the result (predicted land use map) of the CLUE-S model using simulated uncertain DEM. Since one hundred uncertain DEM were simulated, one hundred results of the CLUE-S model were obtained correspondingly. It refers to all the obtained results as a whole, not their average value or dominant value.

vli The validation map, i.e. the land use map of 1999 in the study area obtained from an independent source. It was used as a map to validate the CLUE-S model results using the uncertain inputs.

CHAPTER 1
INTRODUCTION

1.1 General background

As pressure on already limited land resources continues to grow, there is a growing demand to predict future land use change, so that measures can be taken to manage land use efficiently and prevent unwanted land use change or excessive use of land resources. However, land use change is a complex and dynamic process, whereby a large variety of driving factors influence the process. These factors not only include biophysical factors (such as elevation, temperature and soil type) but also include socio-economic factors (such as population density and market value of crops). Some factors act slowly and often obscurely over centuries (such as rock type), while other trigger events quickly and visibly (such as land use policy change and nature disaster). In any case, many driving factors are at work, sometimes operating independently but simultaneously, sometimes operating synergistically. The complexity makes predicting land use change a very difficult and demanding task and often inaccurate (Meyer and Turner 1994; Burrough and McDonnell 1998).

Following the development of geographical information systems (GIS) in the 1990s', along with the development of computer technology and mathematical tools for describing and resolving spatial problems, predictive modeling of land use change combined with GIS became a useful tool to help people managing, monitoring and predicting land use change (Burrough and McDonnell 1998).

The driving factors and land use maps are represented by two distinct formalized conceptual spatial data models in a GIS, object model and field model. In an object model, geographical world is represented as discrete entities. Although object model can be conveniently modeled by vector data using three basic geographical data primitives: point, line and polygon, and further specified by their attributes and geographical locations, it can also be modeled by raster data (as is the case for the road network in this thesis). The field model represents geographical phenomena by fields of attribute values. The attribute is usually assumed to vary continuously over space (Burrough and McDonnell 1998). Uncertainties are an integral part of geographical information and the way geographical information is modeled in GIS. A GIS built upon discrete objects is largely limited to primitive geometric operations that create new objects or compute relationships between objects while fields provide the alternative conceptualization of geographical phenomena, since they lend themselves naturally to the modeling of continuously varying, multivariate, and dynamic phenomena (Zhang and Goodchild 2002).

The rapid population increase and developments of the Kuala Lumpur region in the last two decades have caused dramatic land use and land cover changes in this region and the surrounding area. A predictive model of land use change -- the Conversion of Land Use and its Effects at Small regional extent (CLUE-S) modeling framework (Verburg and Veldkamp 2004) was used to simulate the land use change process and to predict the further land use change. The result of the model prediction can be used to facilitate land use management and support decision-making process.

1.1.1 Urbanization in the Kuala Lumpur region

The Federal Territory of Kuala Lumpur, located in the Selangor province in the mid-western part of Peninsular Malaysia, is 24,612 hectares in size. It has a population of 1.37 million people, which represent about 6% of the country's total population (in 1999). The importance of this region is reflected by its dominant position in the Malaysian economy. Together with its surrounding regions, they contribute about 28% of the gross domestic product (GDP) within an area occupying only 1.3% of the total area of Peninsular Malaysia (Kraim 1998).

The Kuala Lumpur region has been undergoing rapid developments since the early 1980's. The developments extended well beyond the boundary of the Territory, into the surrounding regions of the Selangor province. This caused intensive land use change, including merging several towns, development of new research institutions, construction of the new Kuala Lumpur International Airport, establishment of recreational facilities and building of a government administrative center (Kraim 1998).

The major land use types in the Kuala Lumpur region are agriculture, built-up areas and forest. The main agricultural crops are oil palm, natural rubber, coffee, cocoa and coconut (Kraim 1998).

Urbanization in the Kuala Lumpur region has dramatically changed land use and land cover in the region and the area around. To be able to utilize natural resources efficiently and mitigate the negative influences caused by urbanization to the local environment, it is important to understand the process of urbanization, and therefore be able to explain and predict future land use change. This knowledge is crucial to support decision-making. Due to the complex and dynamic nature of the urbanization process, a land use change model combined with GIS is an indispensable tool.

1.1.2 Land use change modeling with the CLUE-S model

The CLUE-S modeling framework is developed for spatially explicit simulation of land use change on a regional scale. The model is based on an empirical analysis of observed land use change trends combined with competition and interactions between the spatial and temporal dynamics of land use systems (Verburg, Soepboer et al. 2002). The land use change prediction procedures of the CLUE-S model are shown in Figure 1.1 (Verburg 2004).

The CLUE-S modeling framework consists of two independent models, a non-spatial demand model and a spatially explicit allocation model. The demand model is used to calculate the land use demands of all land use types in an aggregate level, on a yearly basis. The demands can be calculated using a simple extrapolation of historical data or complex socio-economic models. The spatial allocation procedure then allocates each land use type considering four categories of constraints (Verburg, Soepboer et al. 2002). The four categories of constraints are:

Constraint 1 Land use requirements (demands)
Demands are calculated in the non-spatial model. The total area (total number of cells * cell size) of a land use type must equal (within specified tolerance) to the demands for each time step.

Constraint 2 Spatial policies and restrictions
There are two types of spatial policies and restrictions considered in the CLUE-S model.

Constraint 2.1 Protected area where land use change is not allowed
This includes the establishment of protection zones where land use change is not allowed.

Constraint 2.2 Spatial policies that restrict certain types of land use conversions
The conversions restricted by a certain spatial policy are indicated in a land use conversion matrix. For example, residential constructions are not allowed to take place in designated agricultural areas or forest reserves.

Constraint 3 Land use type specific conversion settings
Two sets of parameters, conversion elasticity and land use conversion settings and transition sequence, are used together to model the real world land use conversion process.

Constraint 3.1 Conversion elasticity
Conversion elasticity represents the reversibility of land use change. Land use types that need higher capital investment have higher conversion elasticity. When the demands for those land use types that have higher conversion elasticity do not decrease, they are relatively more difficult to be converted into other land use types. Conversion elasticity ranges from zero (easy conversion) to one (irreversible change). Conversion elasticity for each land use type is specified in a land use conversion elasticity table.

Constraint 3.2 Land use conversion settings and transition sequence
Land use conversion settings control the availability of land use change. If a land use change is allowed, the transition sequence describes the temporal characteristics of land use conversions (how many time steps does it take for a land use type to be converted into other land use types, the maximum time steps that a land use type can stay the same, et cetera.). Land use transition sequences are specified in a conversion matrix.

Constraint 4. Location characteristics

Location characteristics ensure that land use conversions take place at the most preferable locations for a specific land use type at a time step. A binomial logit model is used to determine whether to convert a location into a certain land use type or not.

These four categories of constraints together create a set of conditions that model the real world land use change process. The model calculates the most probable locations for land use change through an iterative procedure.

Given an original land use map and a set of constraints, the most probable land use type for each grid cell is computed according to the allocation procedure (Verburg, Veldkamp et al. 2004).

First, all grid cells that are allowed to change are determined. Then, a preliminary allocation is made by allocating the land use types with the largest total probability using an equal value of the iteration variable. The total allocated area for each land use type is then compared with the demand, if it is not equal to the demand, the iteration variable is changed and cells are allocated again until the demand is finally met. Through this procedure, it is possible that the local suitability based on the location factors is overruled by the iteration variable due to the demands.

With the ability to incorporate both socio-economic and biophysical factors for spatial explicit prediction of land use change, the CLUE-S model is a valuable tool to support decision-making. However, spatial data and land use change models are simplifications of the real world phenomena; therefore, they inevitably contain errors. To ensure acceptable accuracy for the model predictions made by the CLUE-S model, it is important to analyze how errors propagate through the model and affect the predictions.

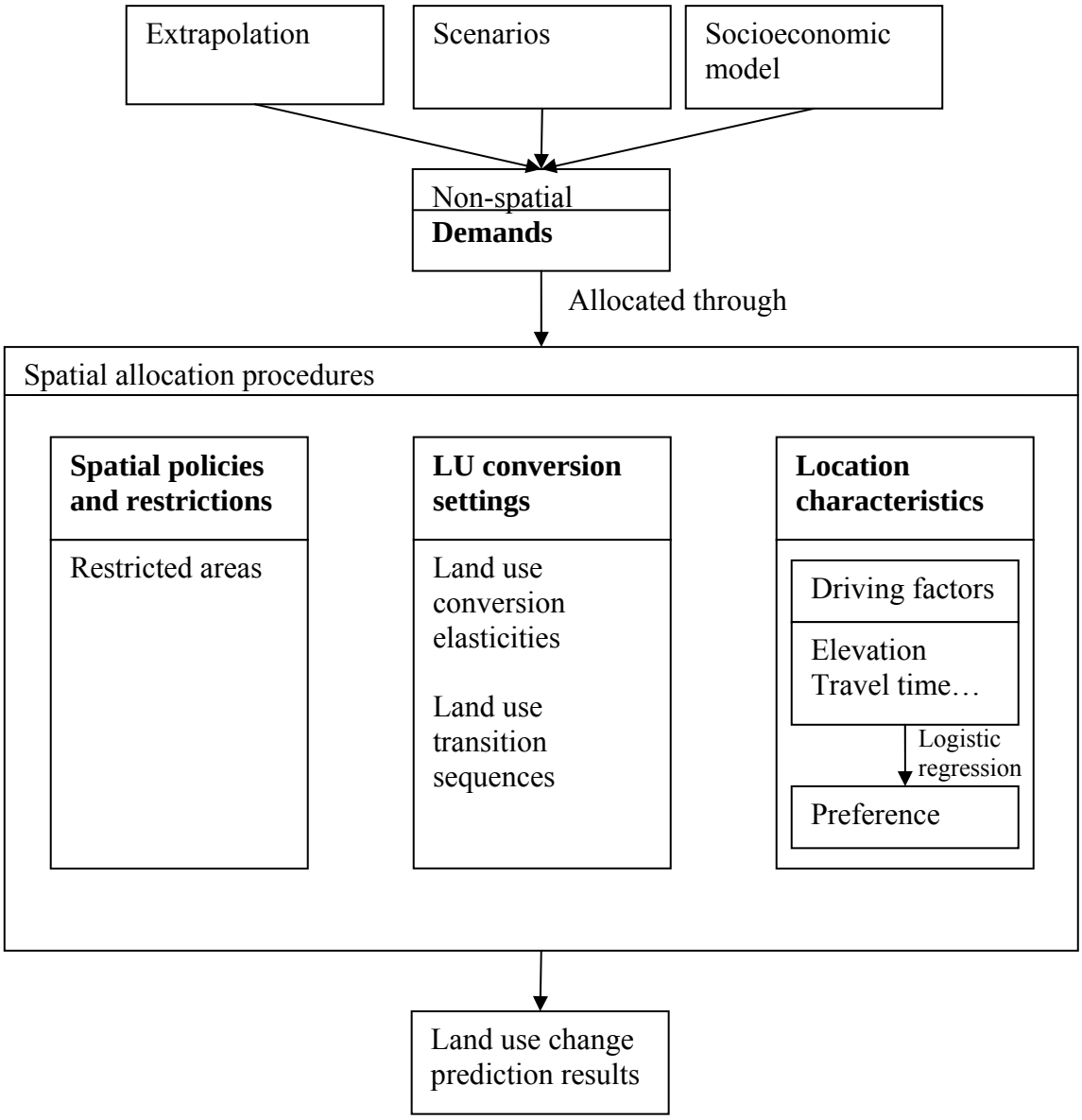


Figure 1.1 Land use change prediction procedures of the CLUE-S model

1.2 Problem definition

As the advance of computer technology and pervasive use of geographical information system, predictive modeling of land use change combined with a GIS has become a common practice to manage limited land resources, monitoring land use change process and predict further land use changes to facilitate decision-making on environmental policies. Spatial data are regularly used as inputs (either as original land use map or driving factors) of predictive modeling of land use change. Through a set of GIS operations, predictive modeling of land use change generates the result that shows the predicted land use distribution in the future. Such a result is often used by decision-makers and city-planners to support decision-making process. Clear-cut decisions often have to be made based on the result.

However, an important aspect of the result has often been overlooked -- the quality of the result made by predictive modeling of land use change. Spatial data contain inherent uncertainties due to for instance our vague definition of some of the geographical phenomena and measurement uncertainties. Predictive models of land use change themselves also contains inherent uncertainties such as those caused by the inevitable simplification and abstraction needed to represent the real world geographical phenomena. The combined effect of uncertainties in spatial data and models themselves will propagate through the model, and affect and accuracy and of the output. Yet, due to the (falsely) high precision and aesthetically pleasing graphics of modern GIS, it is easier to get a false illusion about the quality of the output. For some applications, the uncertainties in output are not significant enough to cause problem, but for other applications, uncertainties in output may have unexpected effects.

Since land use change prediction models are complex non-linear models, the quantification of uncertainties in the model output is not a simple adding of the uncertainties in the inputs. Some of the questions involved that need to be considered are listed. Firstly, the causes of uncertainties in spatial data and models are so numerous; it is not straightforward to identify the sources of uncertainty that are essential for the application. Secondly, correlations may exist between uncertainties in different inputs. Last by not least, predictive models of land use change are seldom liner and involve many computations; the traditional analytical uncertainty propagation methods are not suitable to propagate the uncertainties to the output due to the assumptions involved for the analytical methods and numerical complexity.

On the other hand, there are growing demands to have an understanding of the level of uncertainties in the model output, giving the uncertainties in the inputs and model itself. Given such information on, for example, the most important and uncertain inputs to the model, the sensitivity of the model to the uncertainties in the model inputs and the amount of uncertainties in the model output, et cetera, the uncertainties in the model outputs can be accounted for in decision-making process and further analysis, and efforts could be directed to improve the accuracy that are most important and uncertain to the model in case the uncertainties in the model outputs are significant for the application and the accuracy need to be improved.

To evaluate the uncertainties in the model result, uncertainty propagation analysis can be performed to identify sources of the uncertainties, quantify the uncertainties, propagate the uncertainties to the result, and analysis the result of the uncertainty propagation.

1.2.1 Identification of error sources

Errors in model output can originate from two sources: errors in inputs and errors in model.

Errors in model

There are inevitable errors exist in model since model is a simplification and abstraction of the real world land use change process, which is a complex process and involves many driving factors (both socioeconomic and biophysical). Which aspect is simplified, abstracted, and represented by the model is dependent on the focus of the model and application. In practice, only the key processes that are important to the application are modeled by a set of GIS operations. Model uncertainty includes uncertainties in the structure of the model (conceptual or logical uncertainties), uncertainties in model parameters and uncertainties in the solution of the model(Brown and Heuvelink 2004a).

Errors in model parameters

Model parameters are specified by users according to the application, so that they fit in a particular situation. However, in practice it is difficult to specify correct values because models are approximation of reality. Individuals may also have different perceptions on what are the appropriate values according to their discipline and focus.

Errors in model structure

Uncertainties in model structure may originate from a lack of knowledge about the real processes of land use change. Important processes may be overlooked. Errors in model structure may also be caused by the simplification and abstraction of the land use change process. When less important processes and driving factors are ignored, they become a form of uncertainty. Model structure uncertainties are more difficult to quantify than uncertainties in inputs and model parameters, although in practice they may be more important than parameter uncertainty in evaluating the uncertainty in the model output (Brown and Heuvelink 2004a).

The combined effects of errors in model parameters and model structure will propagate through the model and affect the accuracy of the model results (Brown and Heuvelink 2004a).

Errors in inputs

Various errors exist in inputs, these errors are introduced during the process the real world geographical phenomena are abstracted and simplified by spatial data. Our vague definition on the real world geographical phenomena, the limitation of measurement instruments and human factors can also cause errors.

Conceptual level

At a conceptual level, errors occur from our definition of the geographical phenomena. Many commonly used geographical terms such as 'mountain', 'river', 'towns' and 'lake' are loosely defined. Different terms can be used for different levels of resolution, and their exact form and extent may also change with time (Burrough and McDonnell 1998).

Temporal accuracy

The spatial and thematic component of the real world phenomena change with time (Veregin 1999). However, geographical data only record them for a certain period at particular locations. When time passes by and the spatial and thematic information of the real world phenomena change, errors will arise on the spatial data that are originally recorded. For example, road map is often used as input for predictive modeling of land use change, when its spatial and thematic value changed during the prediction period while not updated on time, errors arise in the input and propagate through the model causing uncertainties in the output.

Measurement, sampling and interpolation error

All spatial data inevitably contain measurement errors due to the limitation of the measuring instruments, methods used for measuring (direct measurement or secondary source) and experience of the personnel who conduct the measurement, et cetera.

For practical reasons such as the cost and time for collecting data, data for large areas or of a long time-span are often obtained by measuring at selected locations and interpolating to the entire area. The errors introduced in this process depend on the sampling density, the algorithms used to interpolate the data and the degree of spatial variability in the interpolated variable.

Classification error

The initial land use map is essential input for predictive modeling of land use change. The initial land use map is usually obtained by remote sensing techniques. A typical type of error for remotely sensed images is classification error. The magnitude and character of the classification error is related to the methods, algorithms used and the characteristics (e.g. spectral signature, pattern) of the land use under examination.

Scale issue

Errors in input data are strongly related to the scale of the application. Detailed information may be lost if data are aggregated to a more general level.

1.3 Research objectives

The objective of this thesis work is to analyze how the most important uncertainties in input data propagate to the CLUE-S model predictions of land use change in the Kuala Lumpur region.

To achieve the objective, several research questions should be answered:

1. What are the most important uncertain inputs in the CLUE-S model when applied to the Kuala Lumpur region?
2. How to identify and quantify the uncertainty in the selected uncertain inputs?
3. What is the contribution of the individual selected uncertain inputs to the output uncertainty?
4. Which input uncertainty should be reduced in order to improve the accuracy of the land use change predictions made by the CLUE-S model?

1.4 Thesis outline

Chapter 2 introduces the methods used to select uncertain inputs and to quantify the uncertainties. The Monte Carlo method was used to propagate the uncertainties to the output.

In chapter 3, the methods described in chapter 2 were applied to the Kuala Lumpur region case study and the results are presented. Shannon's information entropy map, confusion matrices and fragmentation indices were used to analyze the results of the uncertainty propagation.

Chapter 4 discusses and compares the results of analysis presented in chapter 3. The results using uncertain DEM and uncertain travel speed are discussed and compared with the default result of the CLUE-S model using error-ignore inputs, with the validation map and with each other. The difference between the model outputs using error-ignored and error-perturbed inputs and using different types of uncertain inputs are interpreted and explained.

Chapter 5 presents the conclusions of the thesis research.

CHAPTER 2
METHODOLOGY OF SPATIAL UNCERTAINTY PROPAGATION ANALYSIS

At the beginning of the uncertainty analysis problem is to effectively identify the uncertain inputs that are important to the application. The criteria used for selection are presented in section 2.1.

2.1 Selection of uncertain inputs

Selection of uncertain inputs was based on the following criteria:

1. Magnitude of uncertainty

   The inputs selected should have relatively large uncertainties. The assessment can be based on available data, literature review of similar studies or expert judgment.

2. Model sensitivity

   The model should be sensitive to the uncertain inputs selected. The selection could be based on a sensitivity analysis. Sensitivity analysis was performed by changing the values of the inputs (with relatively the same amount) one at a time while keeping others constant and observing the corresponding change in the model output. The most sensitive input is the one that causes the largest change in the model output.

2.2 Quantification of uncertainties

After uncertain inputs are identified, the errors associated with them need to be quantitatively defined.

2.2.1 Definition of error model

Error is defined as the discrepancy between the 'true' value of the real world phenomena and the representation of it, and expressed in Equation 2.1 (Heuvelink 1998):

$$e(x) = t(x) - r(x) \qquad (2.1)$$

where t(x) is the 'true' value of a spatial attribute

r(x) is the representation of the 'truth'

e(x) is the error

Since in practice, we are uncertain about the true value, so it is represented by a random stochastic variable, which is characterized by a probability distribution (Heuvelink 1998).

In practice, e(x) in Equation 2.1 is assumed to follow a particular distribution, such as Gaussian distribution, which can be completely characterized by mean and standard deviation.

For earth science data, values at neighboring locations are often correlated, and so are their errors. The spatial correlated errors at all locations D can be modeled by a spatial random field $E(x)$, following Equation 2.2 (Aerts, Heuvelink et al. 2003):

$E(x) = \mu(x) + \varepsilon(x)$ for all $x \in D$ (2.2)

where $\mu(x)$ is the mean value of E(x) and represent the systematic error or bias, which is often assumed to be zero.

The random field $\varepsilon(x)$ represents the random error. It has zero mean and variance $\sigma(x)^2$. Its spatial autocorrelation is characterized by variance. If we assume the spatial autocorrelation of the variance do not depend on the absolute locations of the values, but only dependent on lag $|h|$, the distance between them (i.e. second-order stationary (Isaaks and Srivastava 1990)), the spatial autocorrelation of the variance can be characterized by a variogram. A variogram describes the spatial continuity of a variable as a function of the distance between them (Isaaks and Srivastava 1990) and is defined in Equation 2.3 (Aerts, Heuvelink et al. 2003).

$$\gamma_z(|h|) = \frac{1}{2} E\left[(E(x) - E(x+h))^2\right] \qquad (2.3)$$

where E stands for mathematical expectation

$E(x)$ stands for spatial random field of errors

$E(x+h)$ stands for spatial random field of errors that are separated by the lag $|h|$

There are three parameters that are used to describe the important features of a variogram: range, sill and nugget effect. As the separation distance between pairs of values increases, the corresponding variogram value will also generally increase. Eventually, when an increase in the separation distance no longer causes a corresponding increase in the average squared difference between pairs of values and the variogram reaches a plateau. The distance at which the variogram reaches this plateau is called the range. Sill is the plateau that the variogram reaches at the range. Nugget effect is the vertical jump from the value of zero at the origin to the value of the variogram at extremely small separation distances.

To ensure that error variance is always positive, positive definite variogram must be used (Isaaks and Srivastava 1990).There are three commonly used positive definite variograms: the spherical model, the exponential model and the Gaussian model. The exponential model is suitable to describe a random function that varies largely over a short distance. The spherical model is suitable to describe a random function that is less erratic than the exponential model. Gaussian model is suitable to describe a random function that is extremely continuous; the variogram is tangential to the x-axis at the origin, and rises very slowly (Isaaks and Srivastava 1990).

2.2.2 Quantification of uncertainties in spatial objects

With an object model, geographical reality is represented in the form of static geometric objects and associated attributes. Many geographical data are naturally modeled as objects, such as road networks, buildings, pipelines, et cetera (Zhang and Goodchild 2002).

Positional uncertainties

For spatial data represented by an object model, the position of an object is described by a set of coordinates. Positional data are subject to gross, systematic, and random measurement errors. Gross errors may be identified by inspection or by repeated measurement. Systematic and random errors are traditionally analyzed through well-established statistical approaches. Error ellipses (for point features) and epsilon error bands (for line features) are widely used to model the positional uncertainty (Zhang and Goodchild 2002). Positional uncertainties are not considered in this thesis work.

Attribute uncertainties

An attribute value is a specific quality or quantity assigned to an object (Zhang and Goodchild 2002). The uncertainties associated with quantitative data can be quantified by using more accurate data or expert knowledge. The quantities determined via expert knowledge are subject to large amount of uncertainties due to human factors and are subjective. Different experts (as well as different groups of experts) may come up with different values of how uncertain something is according to their world-view, background, information available and focus on the application (Brown and Heuvelink 2004b). However, when more accurate data are not available, one has to rely on this approach.

To evaluate how uncertainties in quantitative attribute values influence the accuracy of the model results, the uncertainties can be modeled by random variables. Note that this does not mean that the real attribute values themselves are random variables. Rather, it is a reflection of our inability to specify the objective 'true' value. A common approach to quantify the uncertainties is to use a probability model to describe the spatial data that recognizes the uncertainties. For errors in continuous variable at a single location, the probability that the value of uncertainties below a specified value can be expressed by Equation 2.4 (Brown and Heuvelink 2004b).

$$E(z) = \text{Prob}(Z \le z) \quad -\infty < z < \infty \qquad (2.4)$$

In practice, the uncertainties associated with quantitative attribute are usually assumed to follow an idealized distribution. One commonly used distribution is Gaussian distribution. For a Gaussian distribution, only mean and standard deviation are required to characterize the entire distribution of the random variable.

2.2.3 Quantification of uncertainties in spatial fields

There are two types of variables to consider for spatial field model, namely continuous variable and categorical variable. Only uncertainties in continuous variables were considered in this thesis.

Identification of error models for continuous spatial field

In section 2.2.1, error model is defined, yielding a number of parameters (mean, variance, variogram) that characterize the uncertainties. To identify the error models, the value of the parameters need to be estimated.

If independent validation data are available, the standard deviation (square root of variance $\sigma(x)^2$) can be estimated by Root Mean Squared Error when the error mean $\mu(x)$ is zero. RMSE is defined in Equation 2.5.

$$\mathrm{RMSE} = \sqrt{\frac{\sum (r(x) - t(x))^2}{n}} \quad (2.5)$$

where r(x) is the sampled value (the representation)

t(x) is the 'true' value, or more accurate data in practice

n is the number of samples

The parameters for the variogram model can be estimated by fitting a variogram to the errors derived from validation data and the data used.

2.3 Error propagation methods

After uncertainties in the inputs and model have been quantified, their effects on model output can be analyzed using uncertainty propagation methods. The problem of uncertainty propagation can be formulated generically as in Equation 2.6 (Heuvelink 1998):

$$y = g(u_1, ..., u_m) \quad (2.6)$$

where y is the output of a model.

g is any computational model that incorporates any number of certain elements.

u_i represent input uncertainties as well as model uncertainties.

With this definition, the aim of the uncertainty propagation method is to determine the uncertainty in the output y, given the operation g and the uncertainties in the inputs. The variance of y is a measure for the propagated uncertainty in y. When g is linear and all uncertainties are quantitative and quantified, the variance of y can be derived analytically. Some commonly employed pseudo-analytical methods include first order Taylor series method, second order Taylor series method and Rosenblueth's method (Heuvelink 1998). The Taylor series methods can only be used when g is continuously differentiable. When g is not continuously differentiable, Rosenblueth's method could be used. The advantages of a pseudo-analytical method are that it yields an analytical expression for the variance of the output error, but the solution is approximate only. However, land use change prediction models are complex models and rarely linear, so that analytical models therefore are not suitable for analyzing the propagation of errors due to the numerical complexity.

2.3.1 The Monte Carlo method

The Monte Carlo method (Heuvelink 1998) is more suitable to perform error propagation for a complex non-linear model. The idea of the method is to compute the result of the model repeatedly, with uncertain input values randomly sampled from their joint distribution (realizations). The model results form a random sample from the distribution of the model output, so that parameters of the distribution, such as the mean and the variance, can be estimated from the sample (Heuvelink 1998). To perform the Monte Carlo method, first

generate many set of realizations of the uncertain inputs (and uncertain model parameters, structure and solution, if applicable) that have the same probability of being selected (note this does not mean they have the same probability), for each set of realizations, run the model and store the model output. Then, compute and store sample statistics from the generated outputs.
The realizations that are needed by the Monte Carlo simulation can be simulated by various stochastic simulation techniques (Goovaerts 1998). Among them, unconditional sequential Gaussian simulation (sGs) (Deutsch and Journel 1992) is a commonly used method. sGs proceeds as follows (Goovaerts 1998):

sGs visits each node to be simulated using a random path. At each node, the parameters (mean and variance) of the Gaussian cumulative conditional distribution function (ccdf) are determined using simple Kriging with the variogram model. The parameters are calculated using all the previously simulated nodes to preserve the spatial variability as modeled in the variogram.

Then, a random number is drawn from the distribution that has a variance equivalent to the kriged variance and a mean equivalent to the kriged value. The random number is drawn using a Pseudo-Random Number Generator (PRNG) (Niel and Laffan 2003). Theoretically, the value drawn should have an equal chance of being selected. In practice, since all PRNG create sequences using a deterministic algorithm, the numbers created are pseudo-random. The first value created by PRNG is from a seed value, within the algorithm, each value from a random sequence is calculated from the previous one. Given a good algorithm, the statistical dependence of sequential values can be relatively insignificant (Niel and Laffan 2003).

After the Pseudo-Random Number (PRN) is drawn, it is added to the conditional data set. Afterwards, proceed to the next node along the random path and repeat the procedures until every location is visited once. If a simulation is conditional, it honors the observed value (ground control points), while the unconditional does not honor the observed value.

The most important advantage of the Monte Carlo method, when compared to analytical error propagation methods is that it can yield the entire distribution of the output at an arbitrary level of accuracy. The accuracy of the Monte Carlo method is inversely related to the square root of the number of runs N. As N increases, the accuracy will slowly improve. Other important advantages are that the method is easily implemented and generally applicable. The major disadvantage of the Monte Carlo method is the computational load. In order to make sure that the accuracy of the result is within acceptable limits, the operation has to be executed many times. It becomes extremely time-consuming if the operation itself is a complex model that takes a long time to run. Another disadvantage is that the result of the Monte Carlo method does not come in a nice analytical form, as is the case for the Taylor method. To see how a reduction of input error will have an effect on the output, the entire simulation has to be executed again (Heuvelink 1998).

2.4 Analysis of the results of uncertainty propagation

Results of the Monte Carlo simulation method are a set of random samples drawn from the population of the model prediction. Each of the samples drawn has the same probability of

being selected. There are several methods to analyze the results of the uncertainty propagation.

2.4.1 Shannon's information entropy

Shannon's information entropy was used to measure uncertainties associated with results of predictive models of land use change. The entropy of a discrete probability set is described as follows (Dominich, Goth et al. 2004):

Given events (alternatives) E_j, j = 1, . . ., m; let p_j denote the probability to select alternative (probability of occurrence of event) E_j. The entropy *H* for information is expressed in Equation 2.7.

$$H = -\sum_{j=1}^{m} p_j \log_2 p_j \quad (2.7)$$

The information entropy H satisfies the following properties:

- H is nonnegative because discrete probabilities cannot be greater than one. If and only if exactly one alternative is selected, H reaches its lowest value, being zero.
- H is maximal and equal to $\log_2 m$ if all p_j are equal to 1/m.

For land use allocation models, the events (alternatives) are land use types and p_j is the probability that a particular land use type occurs at specific locations. If n realizations of the model result have been simulated, then p_j at location u is calculated as the number of times a particular land use type occurs divided by n. To the extreme situation, if H is equal to the upper bound $\log_2 m$, then every land use type has the same probability of occurrence (hence model allocation is very uncertain). On the other hand, if H is equal to zero, then only one particular land use type occurred during the model runs (therefore model prediction is very certain). The rest of the values range between these two extremes, when H is close to the upper bound, model predictions are uncertain, and vice versa.

2.4.2 Confusion matrix

A common method to assess the accuracy of a land use classification is to draw a confusion matrix (or a contingency table, or an error matrix). The confusion matrix can also be adapted to evaluate the differences between the results of predictive models of land use change made using error-ignored inputs and error-disturbed inputs, but the meaning of the confusion matrix and its several derived measures (such as producer's accuracy, user's accuracy and overall accuracy) change. These characteristics are no longer measures of the classification accuracy, but become merely measures of how different the two types of model results are.

A confusion matrix is derived by comparing a classification against known reference data. In this thesis work, confusion matrixes were derived by comparing one of the realizations of the model predictions using error-perturbed inputs (on the rows) against the model prediction using error- ignored inputs (on the columns). To summarize the results of the confusion matrix, the 5^{th}, 50^{th} (or median) and 95^{th} percentile of the derived confusion matrices were tabulated.

Several characteristics of the difference between two types of model outputs can be expressed by a confusion matrix. For example, the diagonal of the confusion matrix records the grid cells that stayed the same using error-ignored and error-perturbed inputs. The sum of the diagonal elements divided by the total number of grid cells is called the overall accuracy. The overall accuracy represents the probability that predicted land use at a randomly selected location corresponds with the reference land use.

Producer's accuracy is calculated by dividing the number of cells that stayed the same in each category (on the major diagonal) by the total number of cells in that category in the reference data (the model result using error-ignored inputs). This characteristic is a measure of omission error.

User's accuracy is computed by dividing the number of cells that stayed the same in each category (on the major diagonal) by the total number of cells in that category in a realization (the model result using error-perturbed inputs). The user's accuracy is a measure of commission error and indicates the probability that predicted cells for a land use type correspond with the reference land use.

A point should be made about interpreting classification accuracies. This is the fact that even a complete random assignment of cells will result in correct values in the confusion matrix. To eliminate the effect of the change agreement, the Kappa coefficient was used. The Kappa coefficient is computed in Equation 2.6 (Lillesand and Kiefer 2000).

$$K = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^{r}(x_{i+} \cdot x_{+i})} \quad (2.6)$$

where r = number of rows in the confusion matrix

x_{ii} = the number of cells in row i and column i (on the major diagonal)

x_{i+} = total of observations in row i

x_{+i} = total of observations in column i

N = total number of observations included in the matrix

This statistic indicates the percentage of correctly classified values of a confusion matrix that are due to true agreement versus chance agreement. As the true agreement approaches 1 and the chance agreement approaches 0, K approaches 1. In reality, K usually ranges between 0 and 1. A K of 0 suggests that a given classification is no better than a random assignment of cells.

In this thesis work, confusion matrices were calculated to compare the differences of the model results obtained using error-ignored inputs with those obtained with error-perturbed inputs. Confusion matrices were also calculated to compare the differences of the model results using error-ignored inputs and the validation map.

The statistics calculated from the confusion matrices (e.g. producer's accuracy and user's accuracy) reveal the differences of land use allocation between each land use type. In the

cases of comparing default land use map predicted by the CLUE-S model and the results using uncertain inputs, the differences indicate how uncertainties in the inputs influence the model results. If the producer's and user's accuracy are large (close to 1) for a land use type, then the influence of uncertainties is small, and vice versa. In the extreme situation, if a producer's accuracy or user's accuracy for a land use type is 1, it denotes that there was almost no influence for that land use type whether the errors in the inputs were ignored or not.

2.4.3 Dominant map

The dominant land use map is a map that shows the most frequently occurred land use type on every cell predicted by the CLUE-S model using uncertain travel speed and uncertain DEM. The most frequently occurred land use type is determined among all occurred land use type(s) on each cell predicted by the model using uncertain travel speed and uncertain DEM. The dominant land use map shows the most probable land use type on every cell that predicted by the CLUE-S model using uncertain travel speed and uncertain DEM.

2.4.4 Land use fragmentation index

Although the preciously used coefficients and statistics gave good indications of the difference between the model outputs using error-ignored inputs and error-perturbed inputs, they fail to express the difference of the patterns of the model outputs. There are numerous indicators of patterns of land use maps; some of them communicate the same information. Perimeter to area ratio (PERIAREA), clumpiness index (CLUMPINESS) (McGarigal, Cushman et al. 2002) and total area per class (CAREATOL) were considered relevant and complementary indicators to express the differences in land use patterns in this thesis. All the land use fragmentation indices were on class level (McGarigal, Ene et al.).

Perimeter to area ratio (per class level)

The Perimeter to area ratio is calculated by dividing the total perimeter for a land use type by its corresponding total area per land use class. It is a measurement of the complexity of the shapes of patches and an expression of the spatial heterogeneity of a landscape mosaic (Eiden, Kayadjanian et al. 2000).

Clumpiness index (per class level)

Clumpiness index measures the level of aggregation for a land use type. Its value ranges from minus one to positive one. CLUMPINESS equals minus one when the patches of a land use type are maximally disaggregated, equals zero when the focal patches of a land use type are distributed randomly, and approaches one when the patches are maximally aggregated (McGarigal, Cushman et al. 2002).

Total area (per class level)

Total (Class) Area equals the sum of the areas of all patches of the corresponding patch type, that is, total class area. It is a measure of landscape composition; specifically, how much of the landscape is comprised of a particular patch type (land use type). In addition to its direct interpretive value, class area was used in the computations for perimeter to area ratio (McGarigal, Cushman et al. 2002).

CHAPTER 3
THE KUALA LUMPUR REGION CASE STUDY

The methodology presented in chapter 2 was applied to the Kuala Lumpur region case. The CLUE-S model was used to predict the land use change.

3.1 The default model specifications and result

The eight land use types defined for the CLUE-S model in the Kuala Lumpur region are shown in Figure 3.1.

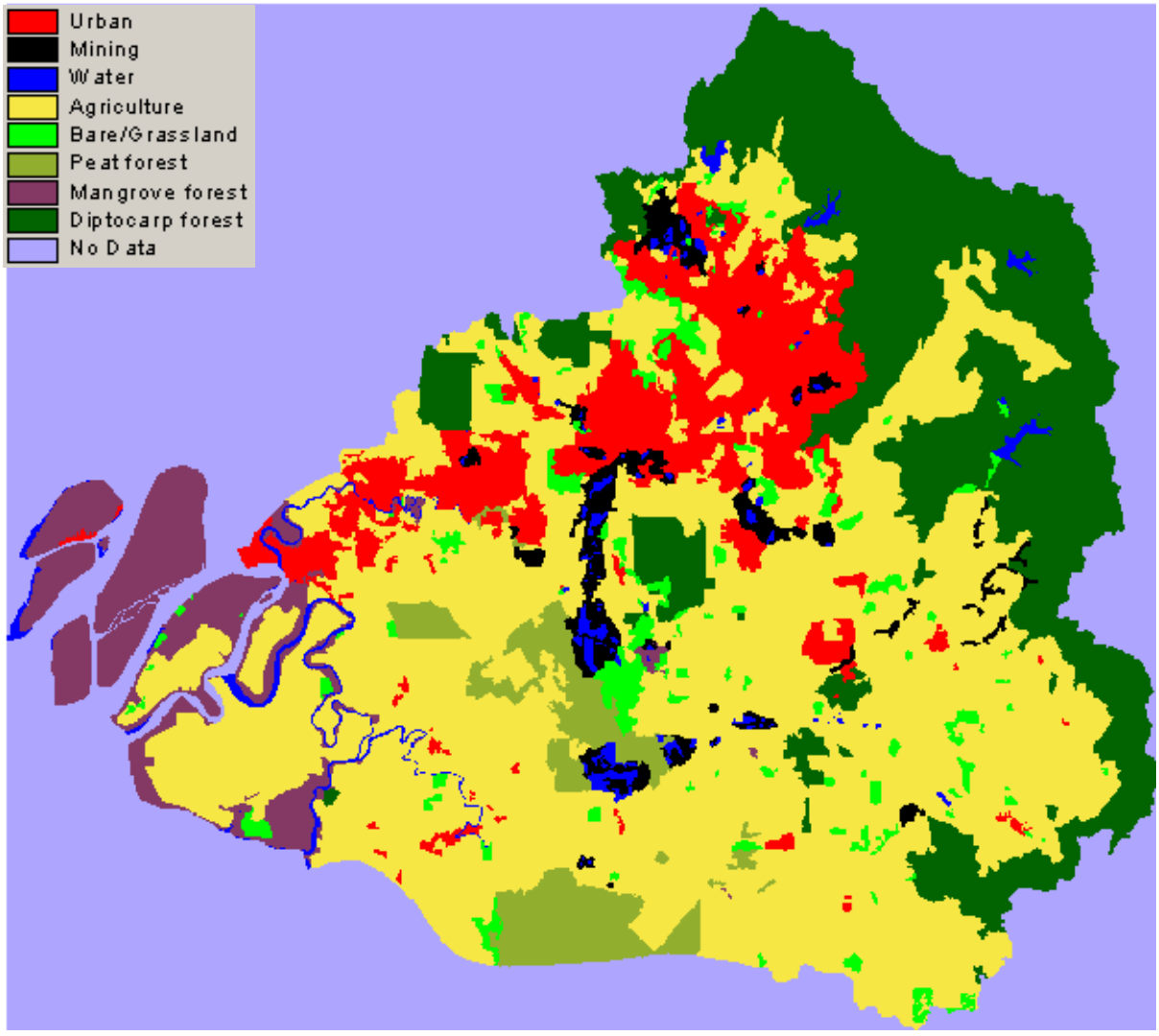


Figure 3.1 Original land use map of the Kuala Lumpur region

The CLUE-S model allocates the land use change through four categories of constraints.

Constraint 1 Land use requirements (demands)

The CLUE-S model allocates the land use change according to the demands calculated in the non-spatial part of the model. The predefined demands for the Kuala Lumpur region case study are listed in Table 3.1.

Table 3.1 Land use demands for the Kuala Lumpur region

LU / Year	**urban**	**mining**	**water**	**agric**	**baregrs**	**peatfr**	**mgvfr**	**dipcfr**
1989	39433	9544	6828	191724	11256	15675	22866	83628
1990	45810	9113	6870	188244	10964	14815	22387	82757
1991	52188	8682	6912	184765	10671	13955	21907	81886
1992	58565	8250	6954	181285	10379	13095	21428	81014
1993	64942	7819	6995	177805	10086	12234	20948	80143
1994	71320	7388	7037	174326	9794	11374	20469	79272
1995	77697	6957	7079	170846	9502	10514	19990	78401
1996	84074	6526	7121	167366	9209	9654	19510	77529
1997	90451	6094	7163	163887	8917	8794	19031	76658
1998	96829	5663	7205	160407	8624	7934	18551	75787
Annual increment	6377	-431	42	-3480	-292	-860	-479	-871

Constraint 2 Spatial policies and restrictions

The restricted areas defined in the study are shown in Figure 3.2.

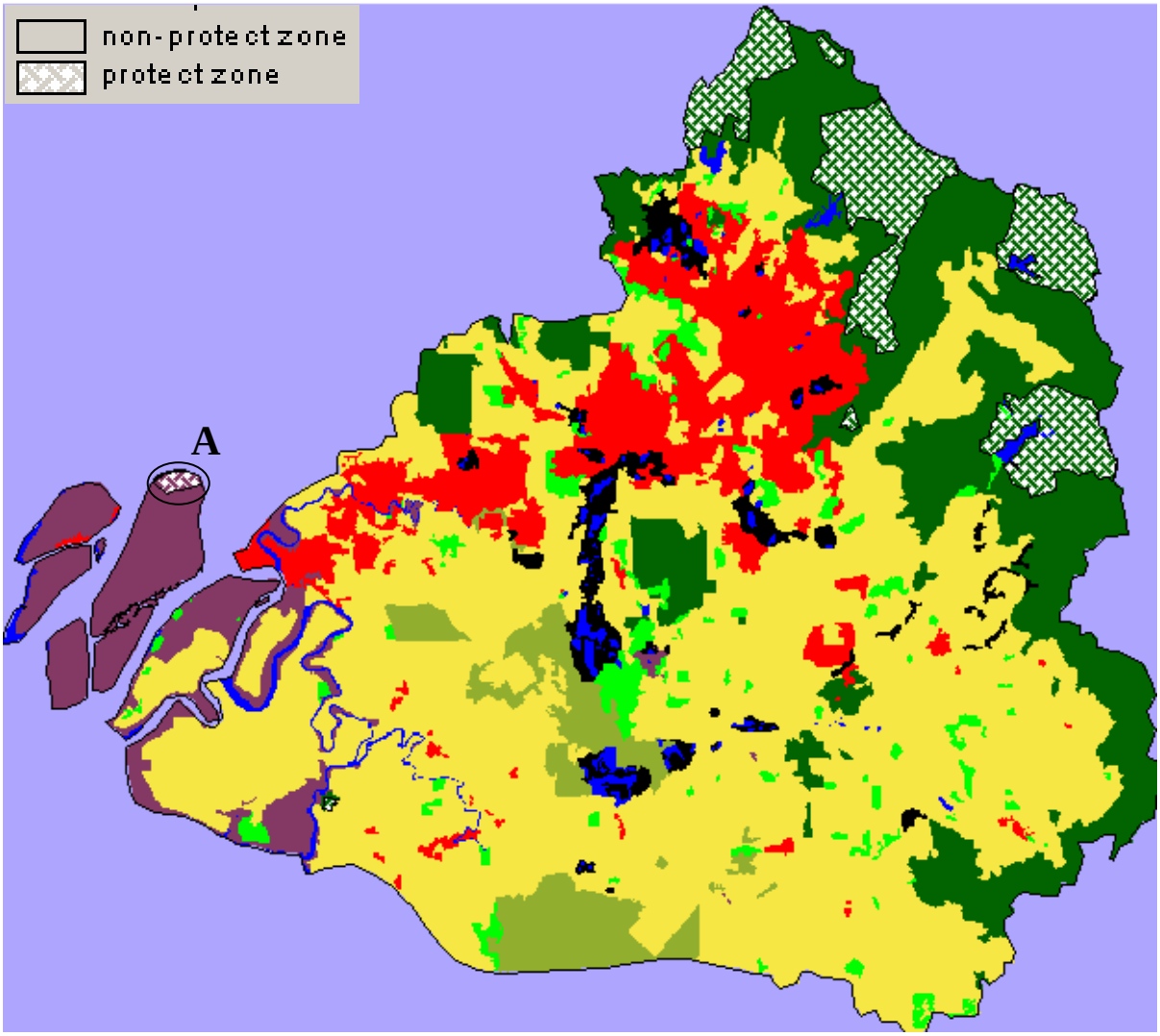


Figure 3.2 Land use protection zone of the Kuala Lumpur region

The restriction zone is mainly located in ‘diptocarp forest’. There is also a small patch of restriction zone in ‘mangrove forest’ in area A of Figure 3.2.

Constraint 3 Land use type specific conversion settings

Constraint 3.1 The conversion elasticity

The conversion elasticity for each land use type for the Kuala Lumpur region case study (Verburg 2004) is listed in Table 3.2.

Table 3.2 Land use conversion elasticity (CE) for the Kuala Lumpur region

LU \ CE	
0 urban	1.0
1 mining	1.0
2 water	1.0
3 agriculture	0.2
4 bare/grassland	0.0
5 peat forest	1.0
6 mangrove forest	1.0
7 diptocarp forest	1.0

Land use type 'bare/grassland' had a conversion elasticity of zero, which means that it is easy for this type of land use to convert to other land use types. 'Agriculture' had a conversion elasticity of 0.2; it is relatively more difficult to convert 'agriculture' to other land use types. The rest of land use types had conversion elasticity equal to one, which means irreversible change.

The conversion elasticity is set in the 'main parameter' file of the model.

Constraint 3.2 Land use transition sequences

Land use transition sequences for the Kuala Lumpur region case study are given in Table 3.3.

Table 3.3 Land use conversion settings for the Kuala Lumpur region

Present LU \ Future LU	0	1	2	3	4	5	6	7
0 urban	1	0	0	0	0	0	0	0
1 mining	1	1	1	1	1	0	0	0
2 water	1	0	1	0	0	0	0	0
3 agriculture	1	1	1	1	1	0	0	0
4 bare/grassland	1	1	1	1	1	0	0	0
5 peat forest	0	1	1	1	1	1	0	0
6 mangrove forest	1	0	1	1	1	0	1	0
7 diptocarp forest	1	1	1	1	1	0	0	1

The rows in Table 3.3 represent the present land use (LU) types, while the columns represent the future land use types. A value one in

Table 3.3 denotes that land use conversion is allowed, while a value zero denotes that land use conversion is not allowed.

Constraint 4 Location characteristics

The 25 driving factors defined for the CLUE-S model in the Kuala Lumpur region are listed in Table 3.4. Some are biophysical factors (e.g. elevation, slope, soil suitability), some are socioeconomic factors (e.g. forest reserves and travel time). All driving factors were assumed constant during the prediction period.

The model predicted land use change from 1989 to 1998. The resulting map of the CLUE-S model prediction 1998 (the default run using error-ignored inputs) is shown in Figure 3.3.

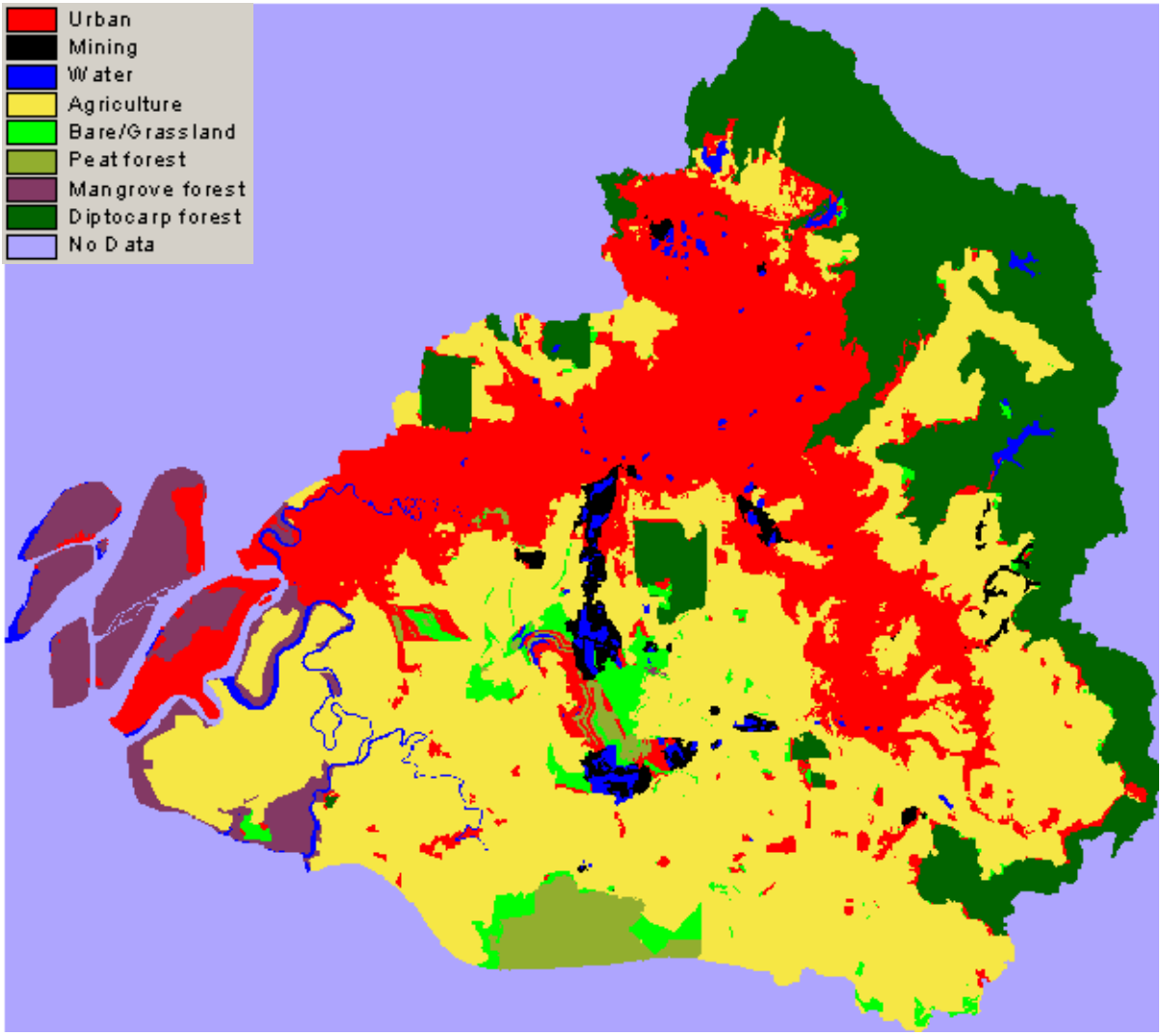


Figure 3.3 Default result of the CLUE-S model using error-ignored inputs

3.2 Selection of uncertain inputs

Table 3.4 was made containing the result of sensitivity analysis and the uncertainty level of all continuous inputs was made. The inputs that have both high level of sensitivity and uncertainty were selected to be analyzed.

To select the inputs that the model is sensitive to, a sensitivity analysis was performed by adding and subtracting 20% of the original value of a driving factor, while kept other driving factors unchanged, run the CLUE-S model and observe the resulting change. If the value change in a driving factor caused a relatively large change in the model result, then the driving factor was given a large sensitivity rank and vice versa. The result of the sensitivity analysis is listed in Table 3.4. The level of uncertainty associated with each of the driving factor was determined via expert knowledge and is also listed in Table 3.4. The scale is from 1 to 5, the larger the number the larger the sensitivity or uncertainty.

Table 3.4 Result of sensitivity analysis and uncertainty assessment of the driving factors

code	Driving factors	Uncertainty	Sensitivity
0	**elevation**	**3**	**4**
1	**slope**	**3**	**5**
2	travel time (tt) to all towns 97	4	3
3	tt to important towns 97	4	3
4	tt to highway 97	4	3
5	distance to sawmill	2	2
6	tt to all roads 97	4	3
7	distance to coast	2	3
8	protection zone	1	N A
9	alluvial	N A	N A
10	igneous	N A	N A
11	limestone	N A	N A
12	metamorphic	N A	N A
13	high erosion	N A	N A
14	moderate high erosion	N A	N A
15	low erosion	N A	N A
16	soil suitability low	N A	N A
17	soil suitability moderately low	N A	N A
18	soil suitability reasonable	N A	N A
19	tt to all towns	4	2
20	**tt to important towns 89**	**4**	**3**
21	**tt to highway 89**	**4**	**3**
22	**tt to sawmill 89**	**4**	**2**
23	tt to all roads 89	3	3
24	forest reserves 1990	2	N A

The driving factors that have both high uncertainty and high sensitivity were selected as uncertain inputs to be analyzed. The categorical variables were not analyzed (N A). Note that if the driving factor were allowed to change during the prediction period, driving factors two to six are used in year 1997 instead of driving factors 19 to 23, but since all the driving factors were assumed to be constant for the Kuala Lumpur region case study, driving factor two to five were actually not used by the model, therefore they were not selected although they were also uncertain and the model was also sensitive to them. The results of the sensitivity analysis showed that the CLUE-S model was very sensitive to the DEM and its derivative, slope. The model sensitivities for the rest of the analyzed driving factors were generally smaller and at the same level.

According to the expert knowledge, travel time to highway, sawmills, all towns and important towns were the most uncertain driving factors for the Kuala Lumpur region.

Combining the results of the sensitivity analysis and expert knowledge on the uncertainties of the inputs, DEM (and slope) and travel time were selected as uncertain inputs.

3.3 Quantification of uncertainties in travel speed

According to the expert knowledge on local road network, the uncertainties in travel speed in the Kuala Lumpur region was assumed to be normally distributed.

The mean value of the normal distribution for each road type was equal to the average speed available for that road type. The standard deviation was related to the road types. Road type 'others' had the largest uncertainty, while 'railway line' had the smallest uncertainty. The average value for each road type and the standard deviation associated with it are specified in Table 3.5. All road segments of the same type were assumed to have the same travel speed in a single realization.

Table 3.5 Average travel speed for each road type and the corresponding standard deviation

Road type	Speed (km/h)	Std. (km/h)
Dual Carriage Road	60	9
Highway	110	8
Motorable Track	40	7
Railway Line	50	6
Others	40	10

3.4 Quantification of uncertainties in DEM

To quantify the uncertainties in DEM, first the error sources in DEM have to be identified.

3.4.1 Identification of error sources in DEM

DEM generally contains three types of errors: blunders, systematic errors and random errors. A blunder is a vertical error that usually exceeds three standard deviations from the mean. When the standard deviation of the error is known, blunders can be identified. Blunders are cause by mistakes such as misleading contours, erroneous correlations or careless observations, et cetera, and are removed prior to entry in the database (USGS 1997).

Systematic errors are those errors that follow some fixed pattern or rule and are generally of constant magnitude or sign. They are introduced by procedures or systems, and are typically predictable. Systematic errors can be eliminated or substantially reduced when the error source is known (USGS 1997).

Random errors arise from accidental and unknown combinations of sources beyond the control of the observer and remain after blunders and systematic errors have been eliminated (USGS 1997). In this thesis work, it was assumed that blunders and system errors had been eliminated from the DEM, thus errors in DEM only refer to the random errors. Random errors were assumed as normally distributed with zero mean (USGS 1997).

3.4.2 Parameterization of error model

Typically, errors in DEM's are reported by Root Mean Squared Error. However, RMSE is a global average value for the whole DEM; it does not describe the local variability and spatial correlation of the DEM errors. To simulate the spatial correlation of the DEM errors, a variogram model was used.

To estimate the parameters (sill, range and nugget effect) of a variogram model, an independent DEM data set with higher accuracy can be used. The standard deviation can be estimated by RMSE when error mean is zero, which is generally the case for U.S. Geological Survey (USGS) DEM products (USGS 1997). An experimental variogram can be fitted to the errors (difference between more accurate DEM and the one used in the application). Nugget effect and sill can be estimated from the experimental variogram.

To parameterize a variogram model, expert judgments will always be required, since the variogram model must be interpreted and processed by people (Brown and Heuvelink 2004a). Nugget effect (sudden value changes at extremely small distances) is not likely to happen in the Kuala Lumpur region; therefore, it is set to zero. Sill equals to the standard deviation of the DEM. Since USGS DEM products generally have a zero mean, RMSE can be used to estimate the standard deviation. The DEM used in the study area has a cell size of 225 m^2, which is 15 meter in length and width. The resolution for the DEM used is comparable with USGS 7.5-minute DEM, which has a resolution of 10 or 30 meter (USGS 1997). Therefore, the parameterization of the error model for the DEM used was performed by consulting USGS documentation (USGS 1997) and studies on DEM's when independent validation data is available, combined with the expert knowledge of DEM in the study area.

According to USGS documentation (USGS 1997), for 7.5-minute DEM's derived from a photogrammetric source a vertical RMSE of 7 m is the desired accuracy standard. A RMSE of 15 m is the maximum permitted; 90 percent have a vertical accuracy of seven-meter RMSE or better and ten percent are in the eight to fifteen meter range. According to expert knowledge on the DEM used, range 7-10 meter is reasonable range for the DEM error in the mountain part of the study area. Other studies on DEM's in mountainous region used (Aerts, Heuvelink et al. 2003) 65 to one hundred as sill value. Therefore, eight m was chosen as the RMSE for the mountain part of the study area. On the other hand, errors in plain part of the study area are much smaller according to the expert knowledge, therefore three m was chosen as the RMSE for the plain part of the study area. Since elevation in the Kuala Lumpur region varies extremely continuous in short distances, the errors associated with

elevation are also expected to vary continuously (Kyriakidis, Shortridge et al. 1999; Aerts, Heuvelink et al. 2003). Therefore, a Gaussian variogram was used to model the errors in elevation. In this study I assumed that there are no abrupt changes within a short distance for DEM in the Kuala Lumpur region, so zero nugget was used. I also assumed that DEM errors are no longer autocorrelated if the distance between two locations is longer than 1200 m. therefore, 1200 was chosen as the range value for the error model of DEM.

To avoid abrupt value changes along the border of the plain and mountain area (see Figure 3.4), a unit variogram was used to model the errors in the entire area. The value simulated were then multiplied by the corresponding sill in the mountain and plain area and added to the original DEM.

The original error-ignored DEM is shown in Figure 3.4.

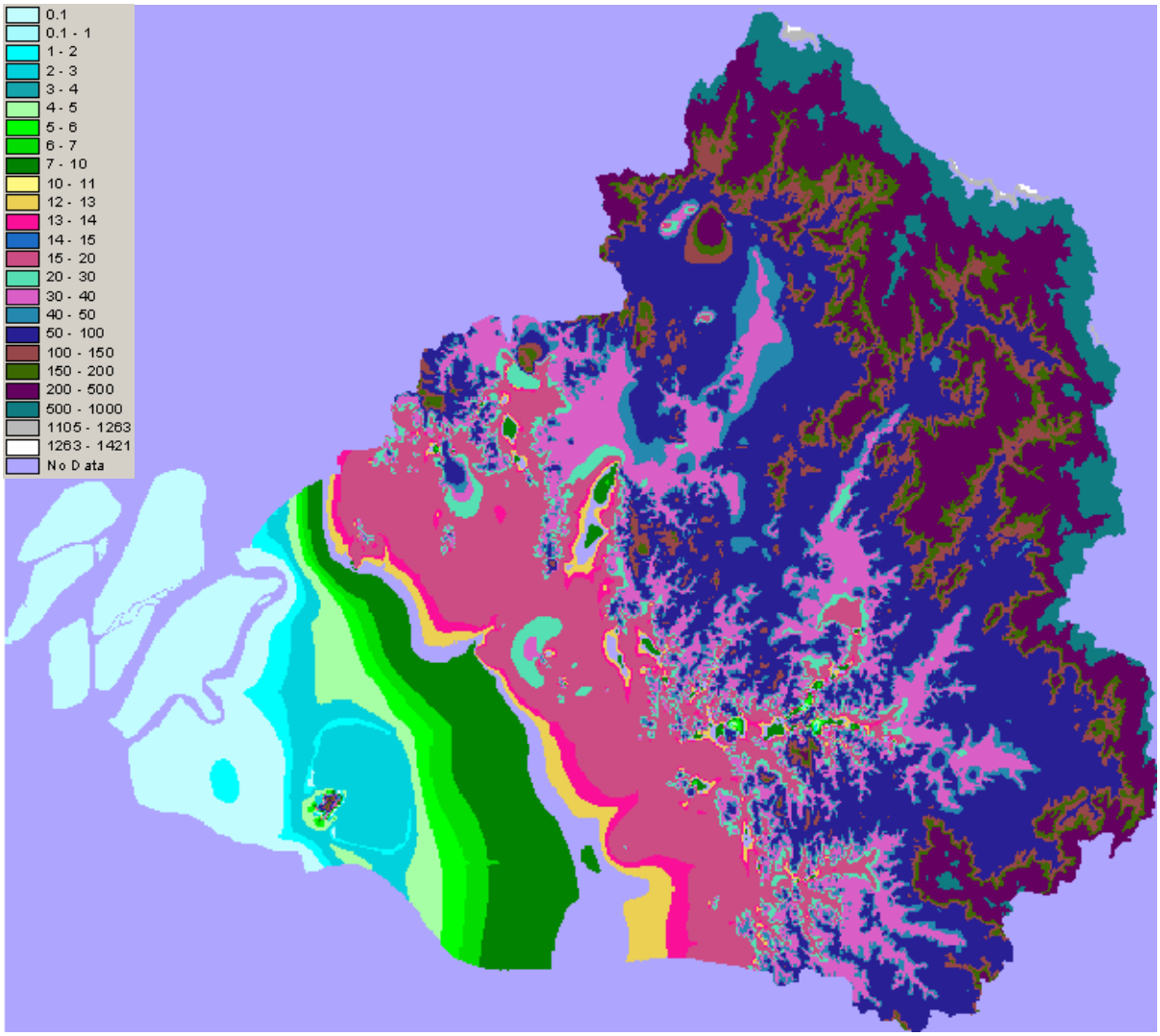


Figure 3.4 Original error-ignored DEM

One realization (the first one) of the error-perturbed DEM is shown in Figure 3.5.

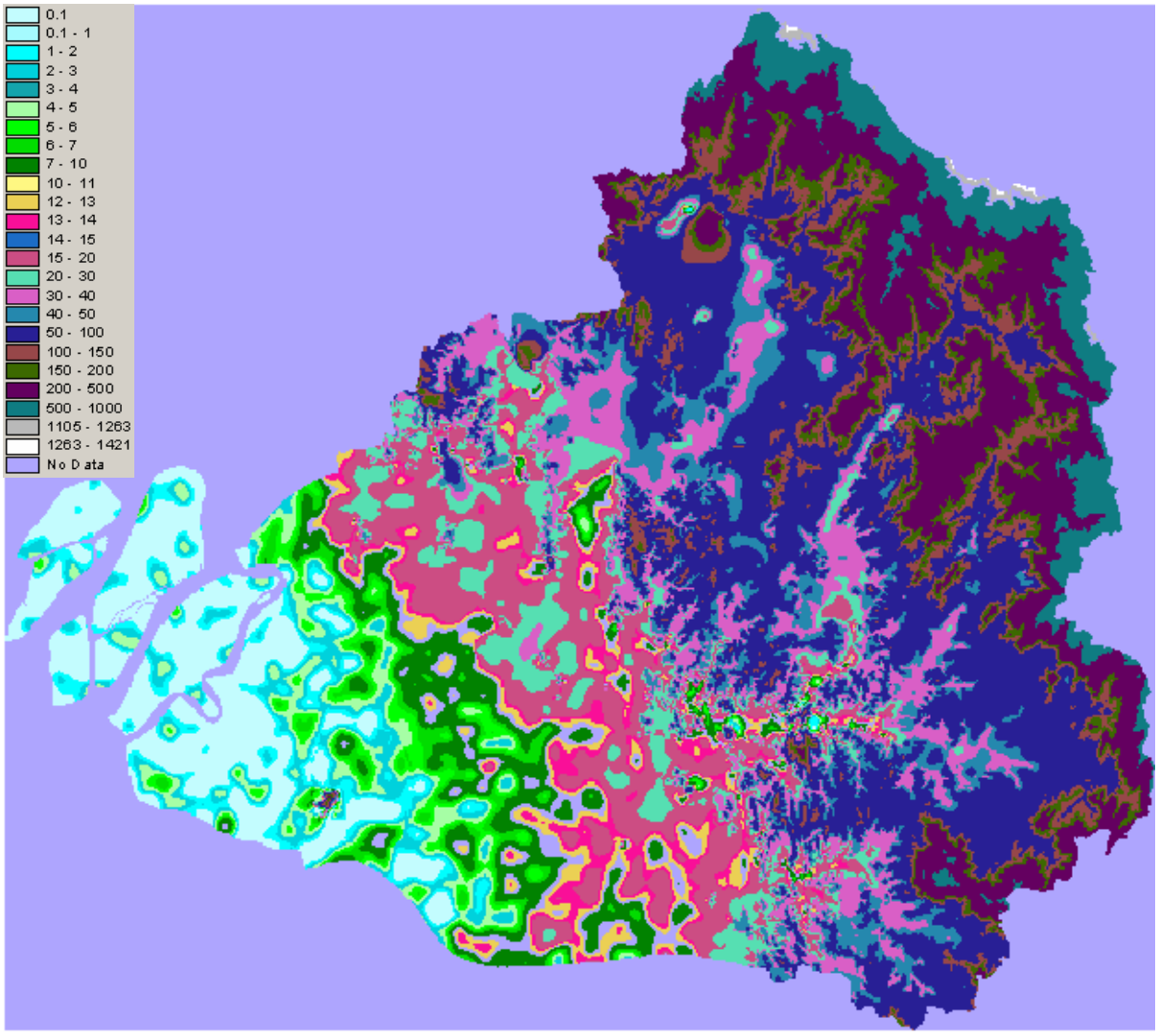


Figure 3.5 The first simulated uncertain DEM

3.5 Applying the Monte Carlo method

The Monte Carlo method was used to perform the error propagation due to the complex nature of the predictive land use change model.

3.5.1 The Monte Carlo simulation for uncertain travel speed

One hundred realizations for uncertainties in travel speed were simulated using PRNG. This was realized using the PRNG in Excel. First, for each road type, one hundred numbers with normal distribution were simulated with zero mean and standard deviation equal to the standard deviation specified in Table 3.5. The resultant numbers were then added to the average speed for each road type (see Table 3.5) correspondingly to derive the error-perturbed travel speed.

3.5.2 The Monte Carlo simulation for uncertain DEM and slope

One hundred realizations for uncertainties in DEM were simulated using unconditional sequential Gaussian simulation (sGs) (Deutsch and Journel 1992). The unconditional sGs was perform using the unconditional sGs algorithm in gstat (Pebesma 1999) with search radius equal to 2000, omax (maximum number of observations per quadrant) equal to four, sk_mean (simple Kriging mean) equal to zero. One hundred slopes were subsequently derived from the simulated DEM. The derivations of slopes were performed in Arc/Info using the SLOPE function (ESRI 2001). For each set of simulated uncertain DEM and slope, the CLUE-S model was executed and the result was saved.

3.6 The CLUE-S model results using simulated uncertain DEM and travel speed

The CLUE-S model was executed for each set of the one hundred simulations of uncertain inputs. When using uncertain DEM, travel speed was kept constant and vice versa. The CLUE-S model results using one realization of uncertain travel speed and DEM are shown in Figure 3.6 and Figure 3.7, respectively.

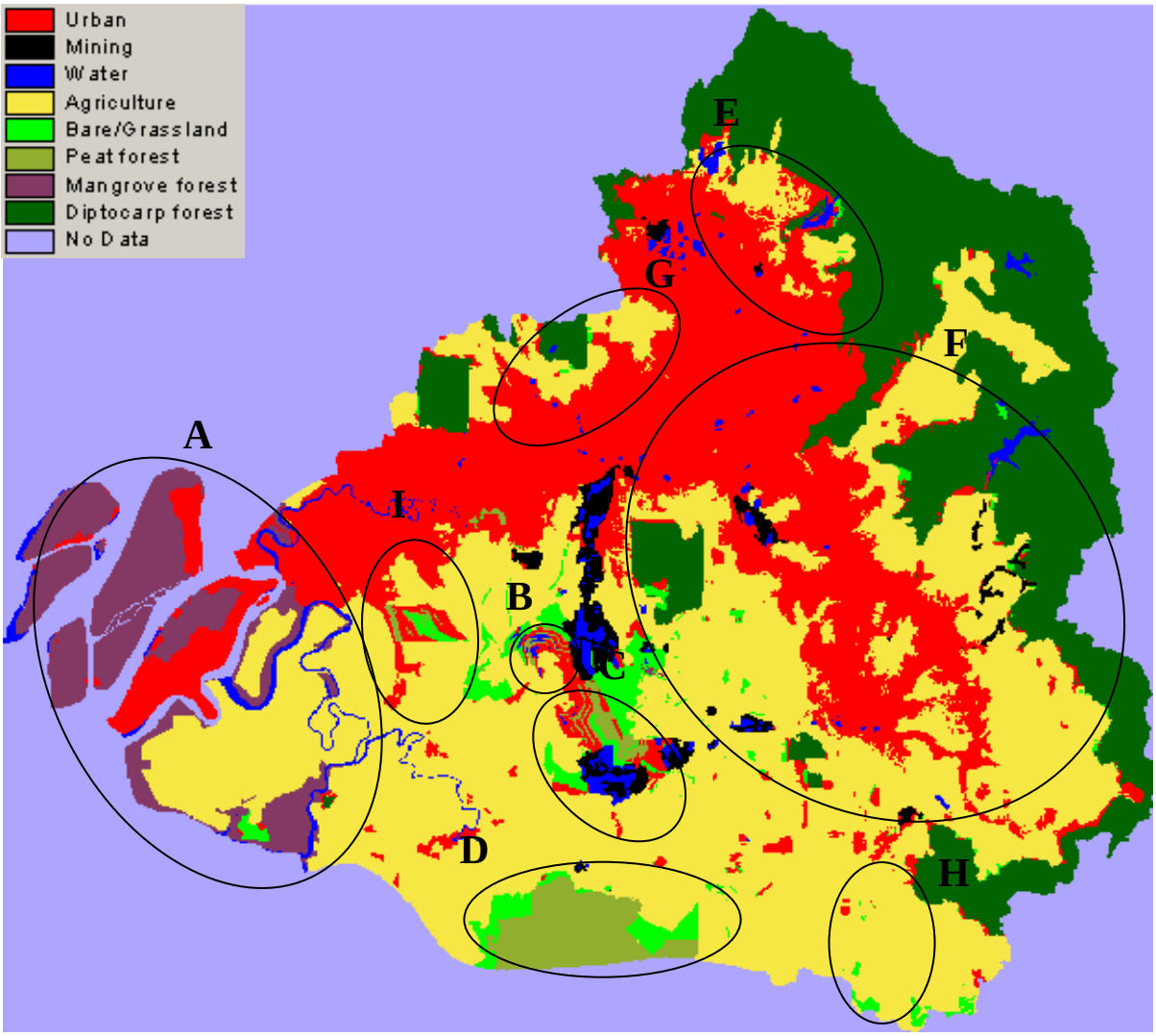


Figure 3.6 The model result using one set of simulated uncertain travel speed

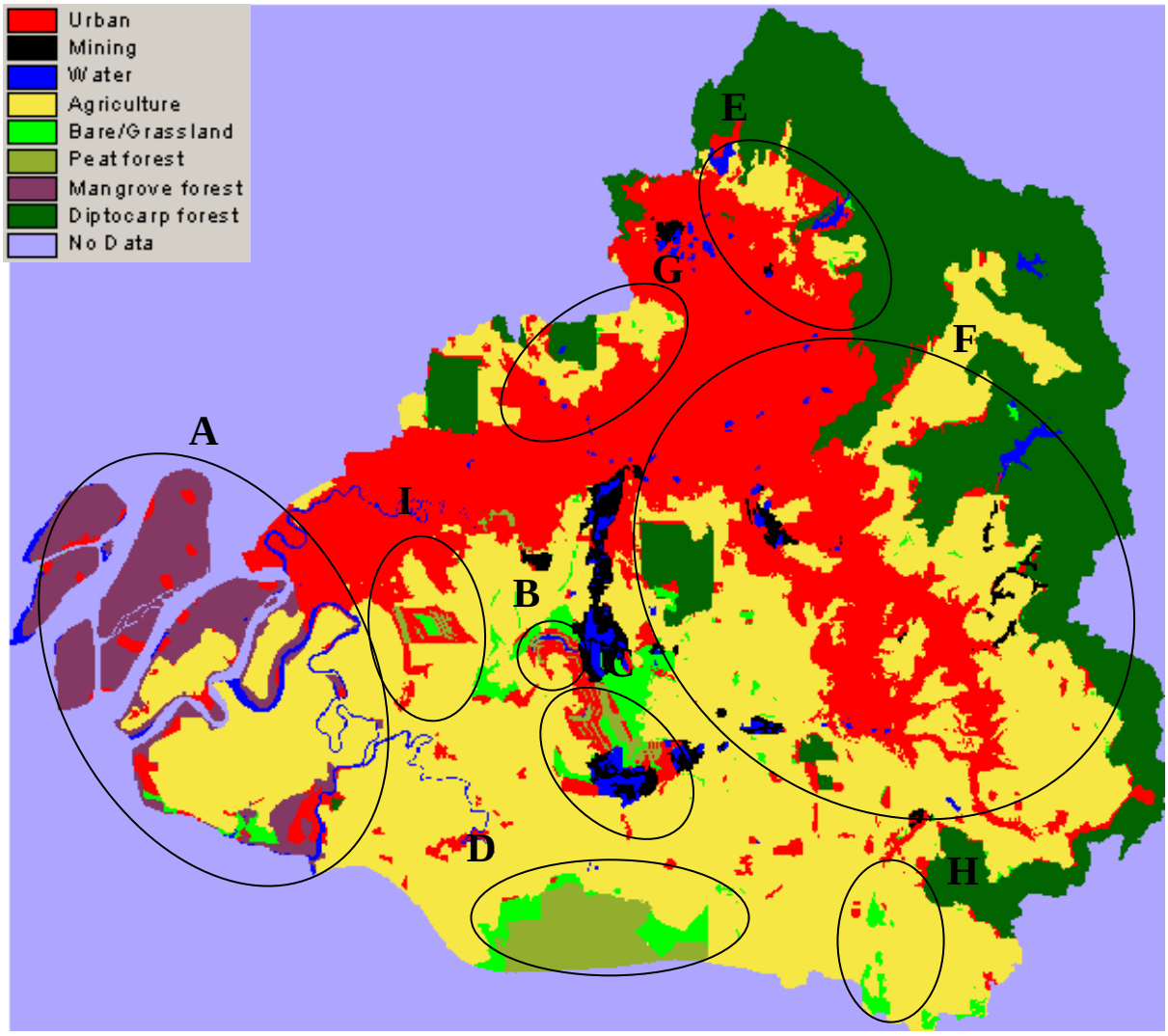


Figure 3.7 The model result using the first simulated uncertain DEM

In general, Figure 3.6 was not much different from Figure 3.7. The most noticeable different is in area A of there two figures. Other differences are surrounded by ellipse A to ellipse I. the differences between the CLUE-S model results using uncertain DEM and uncertain travel speed are discussed later in the discussion part.

3.7 Analysis of the results of the uncertainty propagation

The results of the uncertainty propagation were analyzed using Shannon's information, confusion matrices and fragmentation indices. The differences between the model results using default error-ignore inputs and simulated uncertain inputs were compared, and they were both compared with the validation map in Figure 3.8.

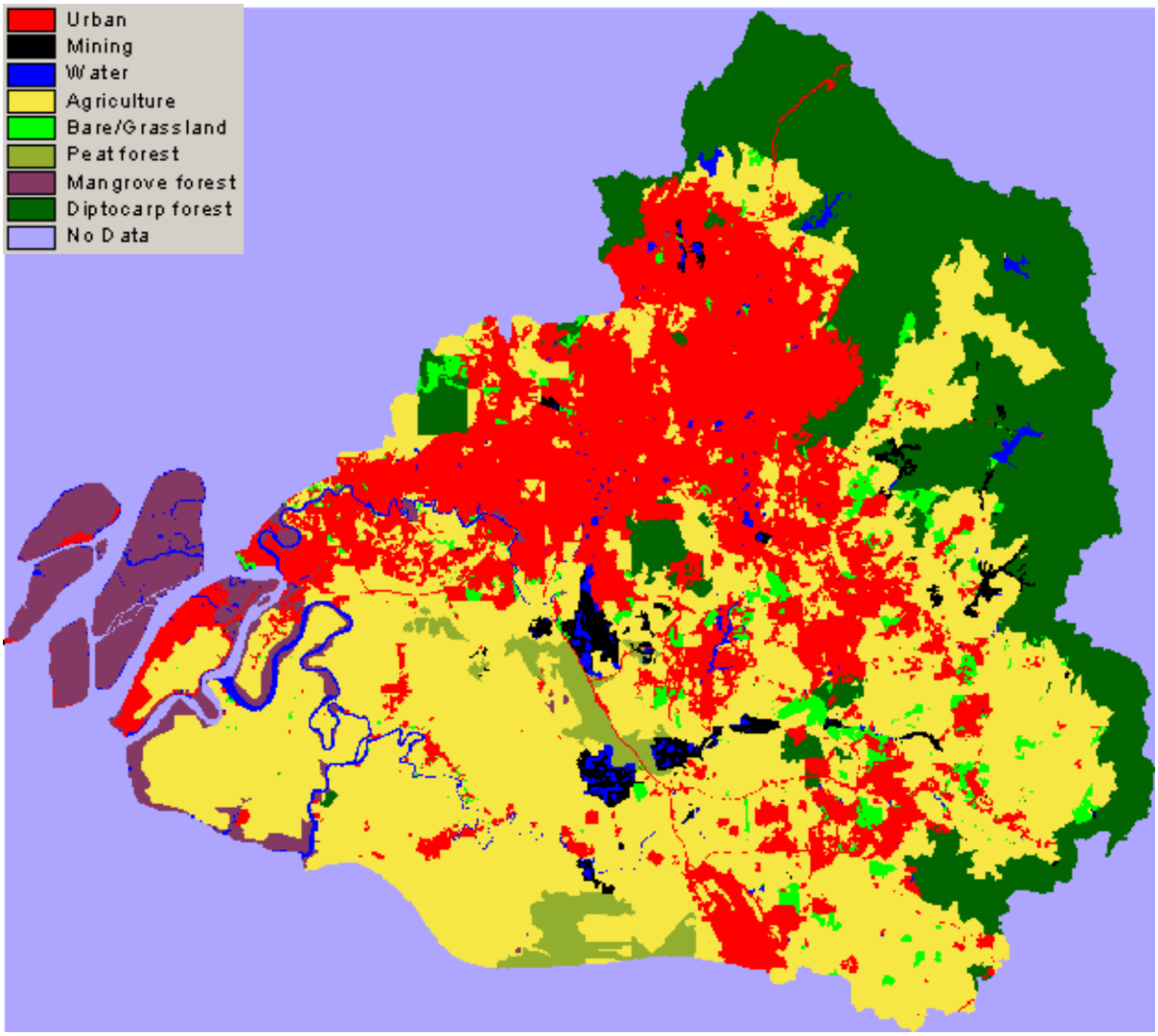


Figure 3.8 The validation map of the Kuala Lumpur region

3.7.1 Analysis of the entropy maps

According to Equation 2.5, for eight possible land use types, Shannon's information entropy range from three (all the land use types have an equal probability of occurrence) to zero (it is certain that a land use type will occur).

The entropy map for the CLUE-S model using uncertain DEM is shown in Figure 3.9 Entropy map ucdem_rst. The entropy map for the CLUE-S model using uncertain travel speed is shown in Figure 3.10 Entropy map uccd_rst.

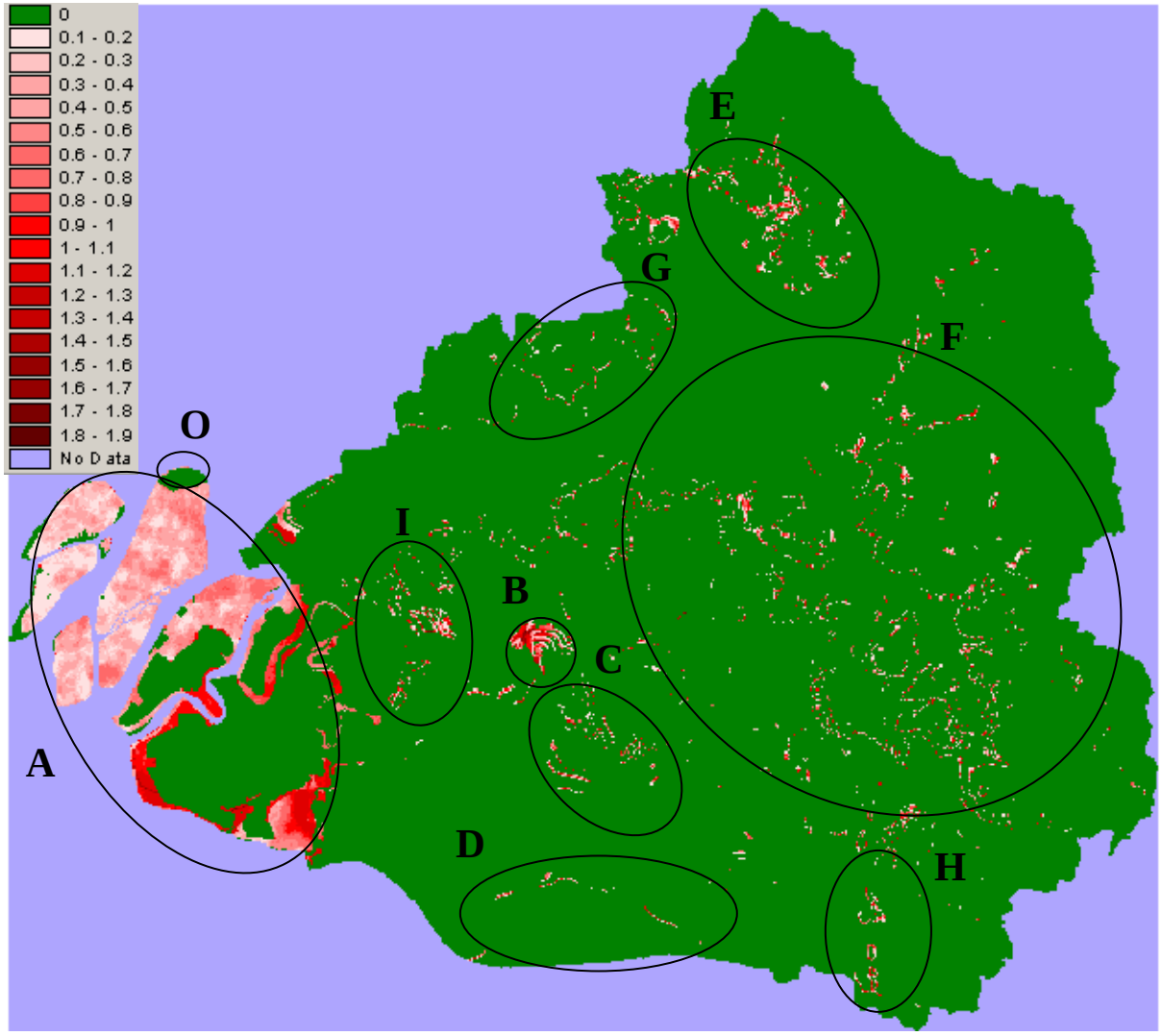


Figure 3.9 Entropy map ucdem_rst

In Figure 3.9, the most obvious uncertainties are located in area A and B, there are also other relatively small uncertainties scattered across the Kuala Lumpur region. The causes of the uncertainties are discussed in chapter four.

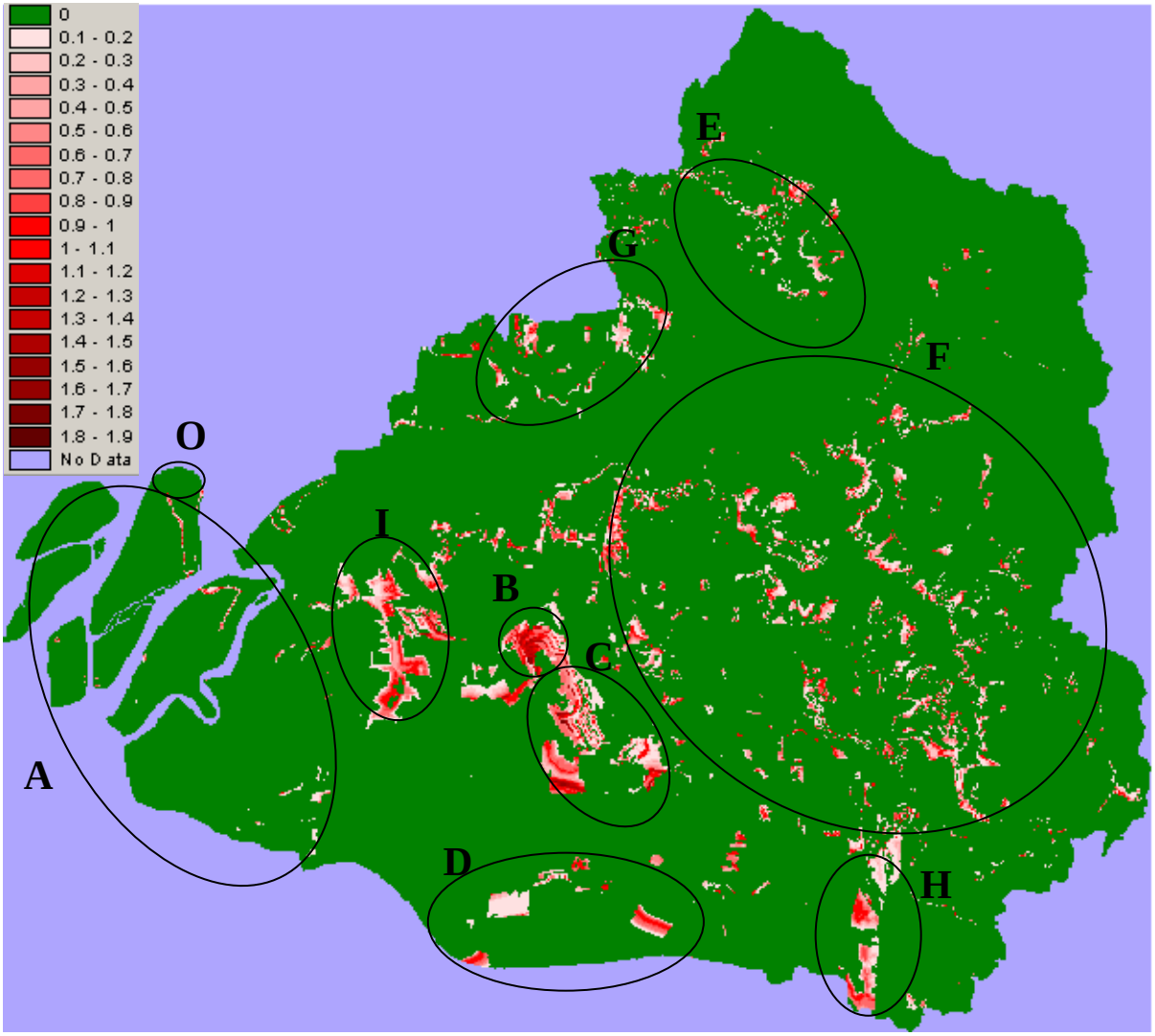


Figure 3.10 Entropy map uccd_rst

Figure 3.10 does not has the large uncertainties in area A of Figure 3.9. While for other parts, the uncertainties are relatively larger. The causes of uncertainties and the difference between Figure 3.9 and Figure 3.10 are discussed in chapter four.

3.7.2 Analysis of the confusion matrices

Since one hundred realizations were obtained for the CLUE-S model results using uncertain DEM and travel speed, one hundred confusion matrices were made respectively, comparing each realization with the default map.

The percentiles of the derived confusion matrices comparing the CLUE-S model results using uncertain DEM and the default map predicted by the model (Figure 3.3) are listed in Table 3.6. The median (the 50th percentile) of the derived confusion matrixes is displayed in bold.

Table 3.6 Percentiles of confusion matrices (dft vs ucdem_rst)

dft \ ucdem_rst		**urban**	**mining**	**water**	**agric**	**baregrs**	**peatfr**	**mgvfr**	**dipcfr**	**paccu**
	5	39672	52	14	1705	18	179	1025	64	0.92
urban	**50**	**39791**	**63**	**28**	**1724**	**23**	**195**	**1142**	**69**	**0.92**
	95	39913	76	44	1753	29	209	1255	73	0.93
	5	30	2430	0	18	5	0	0	0	0.97
mining	**50**	**42**	**2444**	**0**	**19**	**9**	**0**	**0**	**0**	**0.97**
	95	56	2454	0	19	13	0	0	0	0.98
	5	19	0	3115	0	0	20	0	0	0.97
water	**50**	**28**	**0**	**3138**	**0**	**0**	**26**	**0**	**0**	**0.98**
	95	47	0	3154	0	18	35	0	0	0.99
	5	1318	0	0	69458	377	2	5	0	0.97
agric	**50**	**1362**	**0**	**0**	**69485**	**399**	**2**	**40**	**0**	**0.97**
	95	1398	0	0	69509	425	2	85	0	0.98
	5	502	0	0	32	3109	63	0	2	0.81
baregrs	**50**	**554**	**0**	**11**	**35**	**3153**	**75**	**0**	**4**	**0.82**
	95	592	0	18	39	3211	91	0	8	0.84
	5	141	0	2	0	128	3194	0	0	0.91
peatfr	**50**	**154**	**0**	**6**	**0**	**143**	**3218**	**0**	**0**	**0.91**
	95	163	0	23	0	162	3241	0	0	0.92
	5	951	0	0	2	16	0	6949	0	0.84
mgvfr	**50**	**1085**	**0**	**0**	**10**	**88**	**0**	**7058**	**0**	**0.86**
	95	1234	0	0	33	148	0	7196	0	0.87
	5	4	0	0	0	0	0	0	33692	1.00
dipcfr	**50**	**7**	**0**	**0**	**0**	**0**	**0**	**0**	**33697**	**1.00**
	95	13	0	0	0	2	0	0	33701	1.00
	5	0.92	0.97	0.98	0.97	0.81	0.91	0.84	1.00	
uaccu	**50**	**0.92**	**0.97**	**0.99**	**0.97**	**0.82**	**0.91**	**0.86**	**1.00**	
	95	0.93	0.98	0.99	0.98	0.84	0.92	0.87	1.00	

The median of the overall accuracy for Table 3.6 was 0.96.

The percentiles of the derived confusion matrices comparing the CLUE-S model results using uncertain travel speed and the default map predicted by the model are listed in Table 3.7.

Table 3.7 Percentiles of confusion matrices (dft vs uccd_rst)

dft \ uccd_rst		**urban**	**mining**	**water**	**agric**	**baregrs**	**peatfr**	**mgvfr**	**dipcfr**	**paccu**
	5	40774	11	2	731	5	45	0	66	0.95
urban	**50**	**41931**	**20**	**26**	**920**	**13**	**80**	**7**	**72**	**0.97**
	95	42114	97	57	1808	89	127	31	111	0.98
	5	7	2406	0	0	0	0	0	0	0.96
mining	**50**	**19**	**2473**	**0**	**3**	**3**	**0**	**0**	**0**	**0.98**
	95	88	2497	0	32	10	0	0	0	0.99
	5	3	0	3118	0	0	1	0	0	0.97
water	**50**	**15**	**0**	**3163**	**0**	**0**	**19**	**0**	**0**	**0.99**
	95	37	0	3193	49	21	32	0	0	1.00
	5	569	0	0	69130	46	0	2	0	0.97
agric	**50**	**731**	**0**	**0**	**70308**	**187**	**0**	**26**	**0**	**0.99**
	95	1541	43	0	70551	880	0	38	0	0.99
	5	23	0	0	2	2824	21	0	3	0.74
baregrs	**50**	**158**	**0**	**4**	**47**	**3574**	**29**	**4**	**3**	**0.93**
	95	535	21	9	371	3731	73	47	8	0.97
	5	34	0	0	0	19	3300	0	0	0.94
peatfr	**50**	**68**	**0**	**3**	**0**	**49**	**3399**	**0**	**0**	**0.97**
	95	105	0	20	5	83	3447	0	0	0.98
	5	11	0	0	0	0	0	8136	0	0.99
mgvfr	**50**	**37**	**0**	**0**	**0**	**0**	**0**	**8207**	**0**	**1.00**
	95	105	0	0	3	0	0	8231	0	1.00
	5	67	0	0	0	3	0	0	33586	1.00
dipcfr	**50**	**72**	**0**	**0**	**0**	**4**	**0**	**0**	**33630**	**1.00**
	95	114	0	0	0	5	0	0	33634	1.00
	5	0.95	0.96	0.98	0.97	0.74	0.94	0.99	1.00	
uaccu	**50**	**0.97**	**0.98**	**0.99**	**0.99**	**0.93**	**0.97**	**1.00**	**1.00**	
	95	0.98	0.99	1.00	0.99	0.97	0.98	1.00	1.00	

The median of the overall accuracy for Table 3.7 was 0.98.

The percentiles of the derived confusion matrices comparing the CLUE-S model results using uncertain DEM and the validation map is listed in Table 3.8.

Table 3.8 Percentiles of confusion matrices (vli vs ucdem_rst)

vli \ ucdem_rst		**urban**	**mining**	**water**	**agric**	**baregrs**	**peatfr**	**mgvfr**	**dipcfr**	**paccu**
	5	28138	1042	348	11193	284	67	690	1015	0.65
urban	**50**	**28216**	**1049**	**348**	**11217**	**295**	**68**	**823**	**1016**	**0.66**
	95	28344	1059	348	11243	313	70	889	1017	0.66
	5	286	908	371	587	134	8	1	178	0.36
mining	**50**	**299**	**915**	**380**	**592**	**141**	**12**	**2**	**178**	**0.36**
	95	311	919	388	596	147	16	2	178	0.37
	5	303	194	1869	423	21	19	306	18	0.58
water	**50**	**320**	**196**	**1871**	**430**	**24**	**20**	**325**	**18**	**0.58**
	95	338	198	1872	433	28	21	341	18	0.59
	5	9359	203	122	55622	2681	1270	346	1324	0.78
agric	**50**	**9484**	**207**	**139**	**55656**	**2712**	**1276**	**487**	**1332**	**0.78**
	95	9623	211	152	55689	2749	1286	627	1341	0.78
	5	985	31	12	2159	39	4	1	566	0.01
baregrs	**50**	**999**	**31**	**19**	**2170**	**41**	**4**	**2**	**569**	**0.01**
	95	1011	32	21	2179	43	4	3	572	0.01
	5	662	67	31	160	445	2102	0	0	0.60
peatfr	**50**	**678**	**67**	**50**	**160**	**458**	**2112**	**0**	**0**	**0.60**
	95	694	67	72	160	477	2119	0	0	0.60
	5	910	8	280	176	25	24	6481	0	0.79
mgvfr	**50**	**1042**	**8**	**280**	**211**	**76**	**24**	**6605**	**0**	**0.80**
	95	1165	8	280	237	118	24	6755	0	0.82
	5	1962	33	96	838	66	0	0	30649	0.91
dipcfr	**50**	**1976**	**34**	**97**	**843**	**74**	**0**	**1**	**30658**	**0.91**
	95	1988	34	97	849	88	0	1	30667	0.91
	5	0.65	0.36	0.59	0.78	0.01	0.60	0.79	0.91	
uaccu	**50**	**0.66**	**0.36**	**0.59**	**0.78**	**0.01**	**0.60**	**0.80**	**0.91**	
	95	0.66	0.37	0.59	0.78	0.01	0.60	0.82	0.91	

The median of the overall accuracy for Table 3.8 was 0.74.

The percentiles of the derived confusion matrices comparing the CLUE-S model results using uncertain travel speed and the validation map is listed in Table 3.9.

Table 3.9 Percentiles of confusion matrices (vli vs uccd_rst)

vli \ uccd_rst		**urban**	**mining**	**water**	**agric**	**baregrs**	**peatfr**	**mgvfr**	**dipcfr**	**paccu**
	5	28046	971	341	11212	259	61	695	1016	0.65
urban	**50**	**28181**	**1034**	**345**	**11398**	**264**	**64**	**705**	**1017**	**0.65**
	95	28425	1090	347	11450	411	66	721	1019	0.66
	5	289	897	371	551	115	10	1	176	0.36
mining	**50**	**301**	**924**	**377**	**575**	**144**	**15**	**2**	**178**	**0.37**
	95	319	962	385	608	148	19	8	178	0.38
	5	358	197	1869	424	20	16	275	18	0.58
water	**50**	**369**	**200**	**1871**	**427**	**22**	**18**	**278**	**18**	**0.58**
	95	371	202	1872	438	27	22	280	18	0.59
	5	9399	197	130	55233	2680	1246	714	1297	0.77
agric	**50**	**9493**	**218**	**150**	**55401**	**2715**	**1267**	**726**	**1310**	**0.78**
	95	9684	235	162	55482	2776	1310	754	1329	0.78
	5	987	31	11	2144	44	4	1	565	0.01
baregrs	**50**	**1001**	**31**	**19**	**2154**	**47**	**4**	**1**	**569**	**0.01**
	95	1019	31	22	2178	62	4	1	571	0.02
	5	374	53	47	155	407	2092	0	0	0.59
peatfr	**50**	**626**	**67**	**57**	**161**	**481**	**2130**	**0**	**0**	**0.60**
	95	675	68	77	413	516	2148	6	0	0.61
	5	1162	8	280	199	13	24	6506	0	0.79
mgvfr	**50**	**1177**	**8**	**280**	**211**	**13**	**24**	**6529**	**0**	**0.79**
	95	1215	8	280	231	13	24	6534	0	0.79
	5	1746	34	93	877	63	0	0	30595	0.91
dipcfr	**50**	**1875**	**34**	**97**	**938**	**128**	**0**	**0**	**30612**	**0.91**
	95	2024	34	97	1032	146	0	1	30628	0.91
	5	0.65	0.36	0.58	0.77	0.01	0.59	0.79	0.91	
uaccu	**50**	**0.65**	**0.37**	**0.59**	**0.78**	**0.01**	**0.60**	**0.79**	**0.91**	
	95	0.66	0.38	0.59	0.78	0.02	0.61	0.79	0.91	

The median of the overall accuracy for Table 3.9 was 0.74.

The confusion matrix for the default map predicted by the CLUE-S model and the dominant map using uncertain DEM is listed in Table 3.10.

Table 3.10 Confusion matrix of domi_ucdem vs dft

domi_ucdem_rst / dft	**urban**	**mining**	**water**	**agric**	**baregrs**	**peatfr**	**mgvfr**	**dipcfr**	**paccu**
urban	39667	53	21	1706	19	195	1314	58	0.92
mining	33	2452	0	19	9	0	0	0	0.98
water	28	0	3146	0	0	25	0	0	0.98
agric	1338	0	0	69539	406	2	0	0	0.98
baregrs	555	0	11	34	3153	74	0	5	0.82
peatfr	153	0	2	0	145	3223	0	0	0.91
mgvfr	429	0	0	4	6	0	7804	0	0.95
dipcfr	1	0	0	0	0	0	0	33704	1.00
uaccu	0.94	0.98	0.99	0.98	0.84	0.92	0.86	1.00	

The overall accuracy for Table 3.10 was 0.96.

The confusion matrix for the default map predicted by the CLUE-S model and the dominant map using uncertain travel speed is listed in Table 3.11.

Table 3.11 Confusion matrix of domi_uccd vs dft

domi_uccd_rst / dft	**urban**	**mining**	**water**	**agric**	**baregrs**	**peatfr**	**mgvfr**	**dipcfr**	**paccu**
urban	42146	14	19	715	10	63	0	66	0.98
mining	13	2496	0	2	2	0	0	0	0.99
water	8	0	3173	0	0	18	0	0	0.99
agric	612	0	0	70538	111	0	24	0	0.99
baregrs	139	0	4	4	3650	27	4	4	0.95
peatfr	59	0	2	0	49	3413	0	0	0.97
mgvfr	30	0	0	0	0	0	8213	0	1.00
dipcfr	68	0	0	0	4	0	0	33633	1.00
uaccu	0.98	0.99	0.99	0.99	0.95	0.97	1.00	1.00	

The overall accuracy was for the dominant map using uncertain travel speed versus the reference map 0.99.

3.7.3 Analysis of the dominant maps

The dominant map for the CLUE-S model results using uncertain DEM is listed in Figure 3.11.

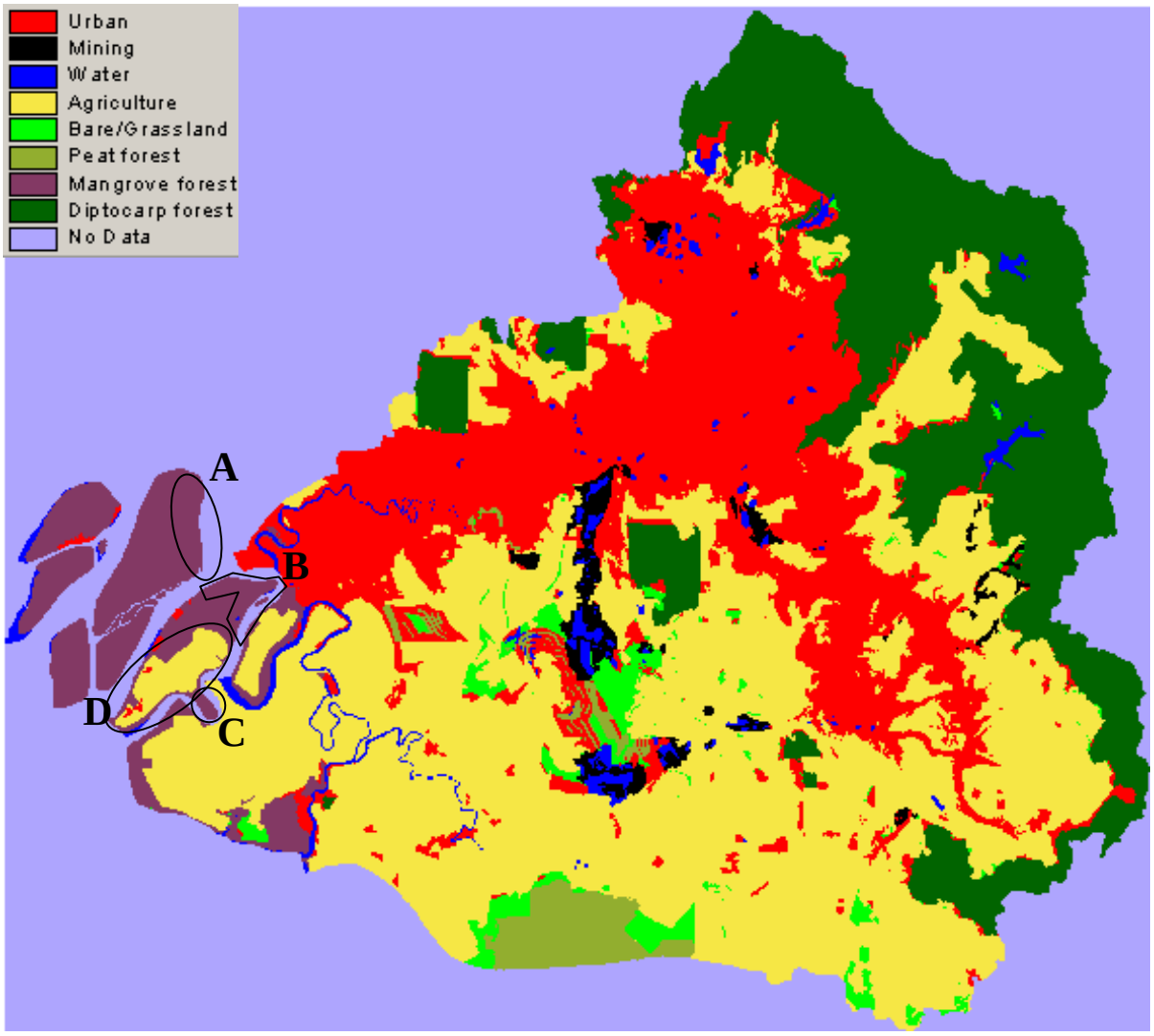


Figure 3.11 Domi_ucdem_rst

The dominant map for the CLUE-S model results using uncertain travel speed is listed in Figure 3.12.

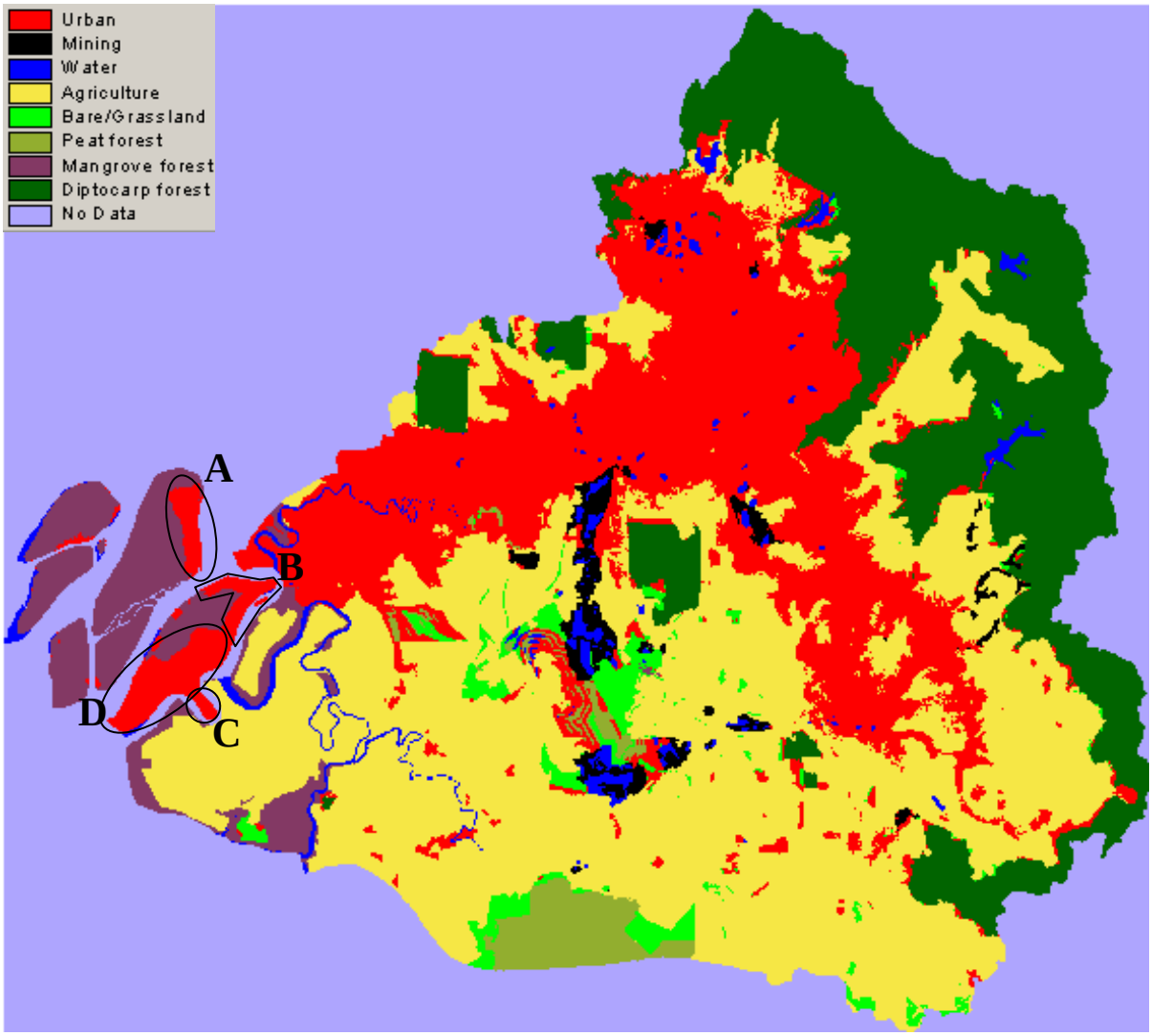


Figure 3.12 Domi_uccd_rst

The differences between Figure 3.11 (domi_ucdem) and Figure 3.3 (the default prediction) are shown in Figure 3.13.

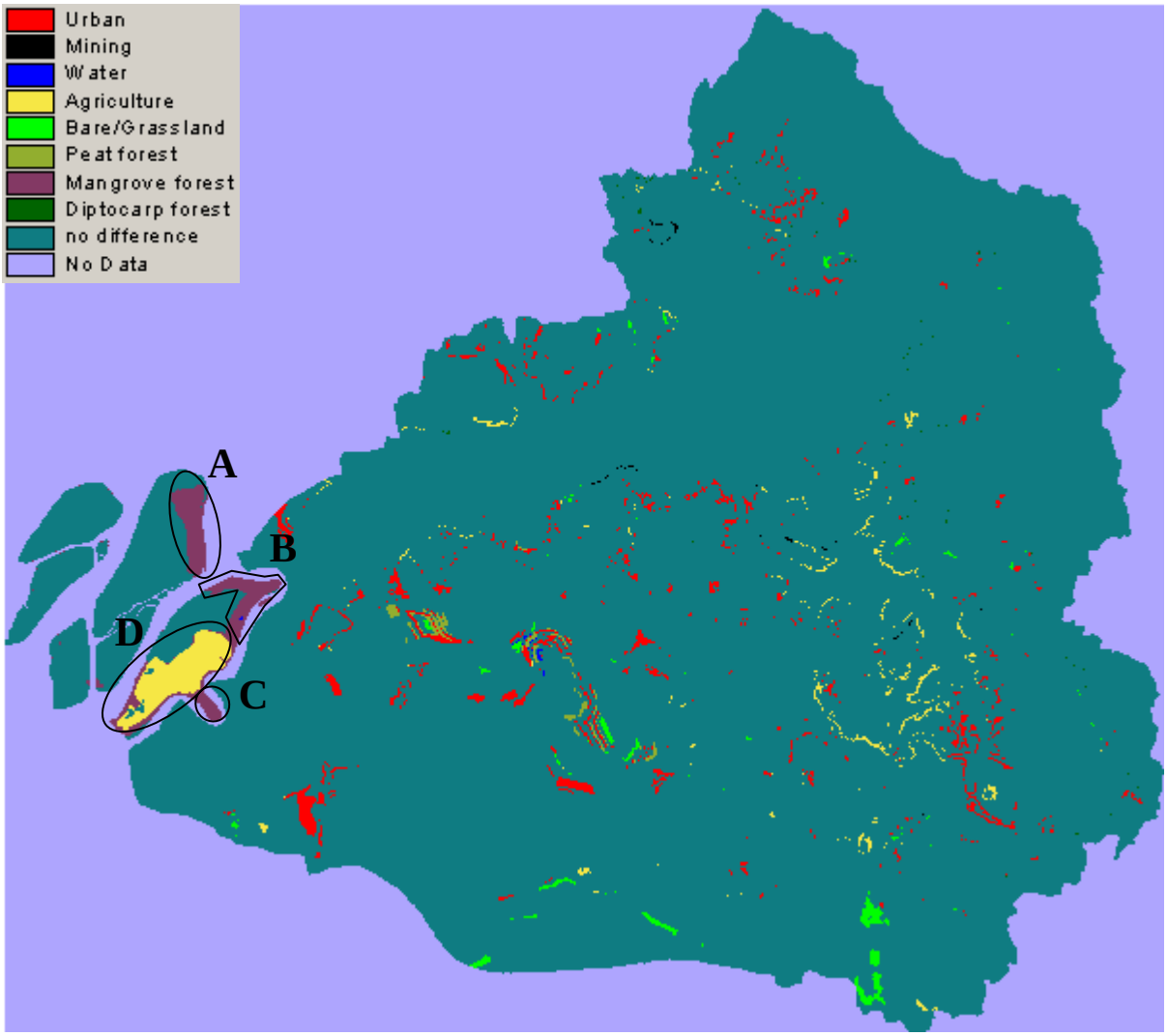


Figure 3.13 Difference between domi_ucdem and dft

The land use types shown in Figure 3.13 are from Figure 3.11 (domi_ucdem).

The differences between Figure 3.12 (domi_uccd) and Figure 3.3 (the default prediction) are shown in Figure 3.14.

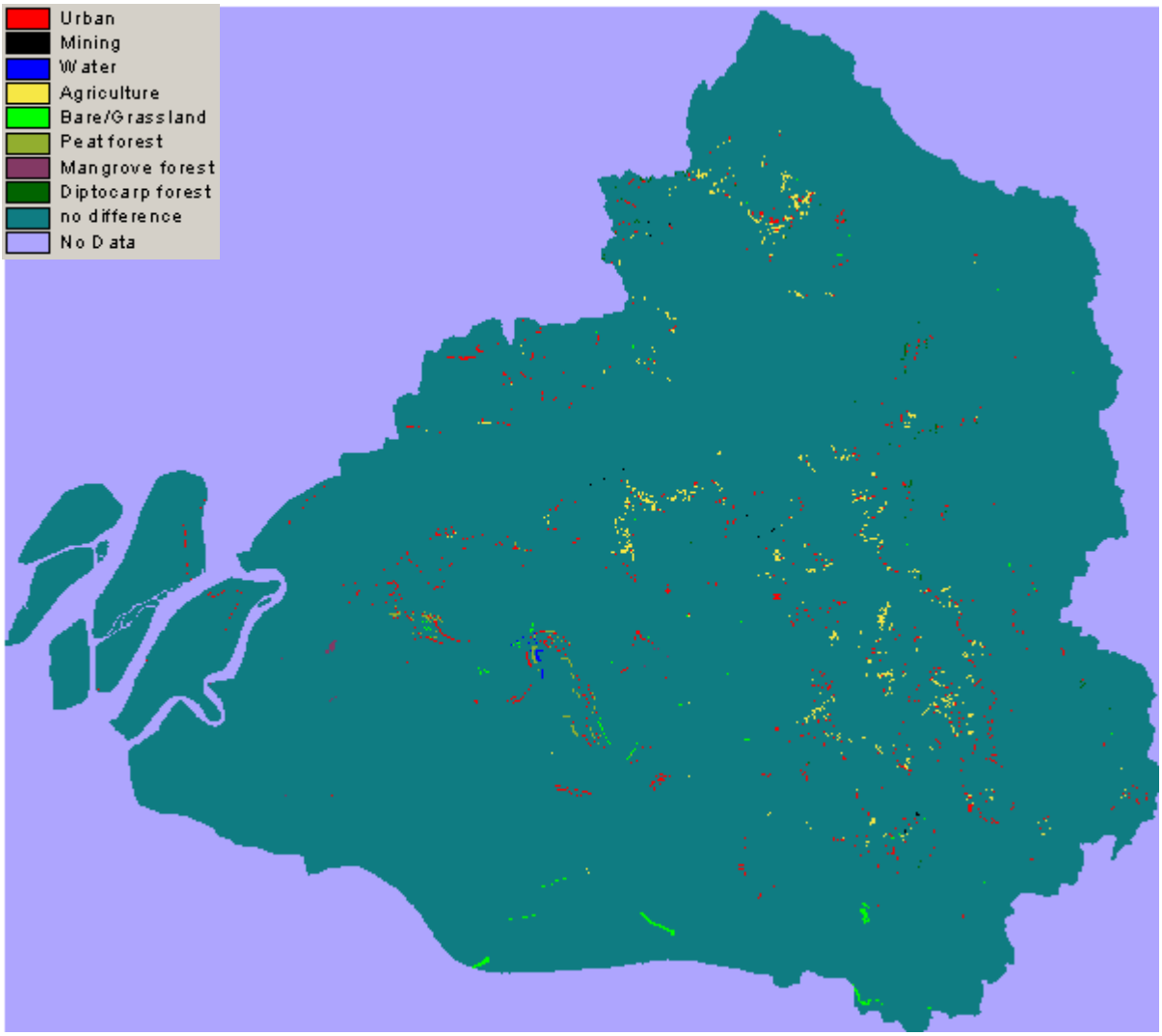


Figure 3.14 Difference between domi_uccd and dft

The land use types shown in Figure 3.14 are from Figure 3.12 (domi_uccd).

The differences between Figure 3.11(domi_ucdem) and Figure 3.12 (domi_uccd) are shown in Figure 3.15.

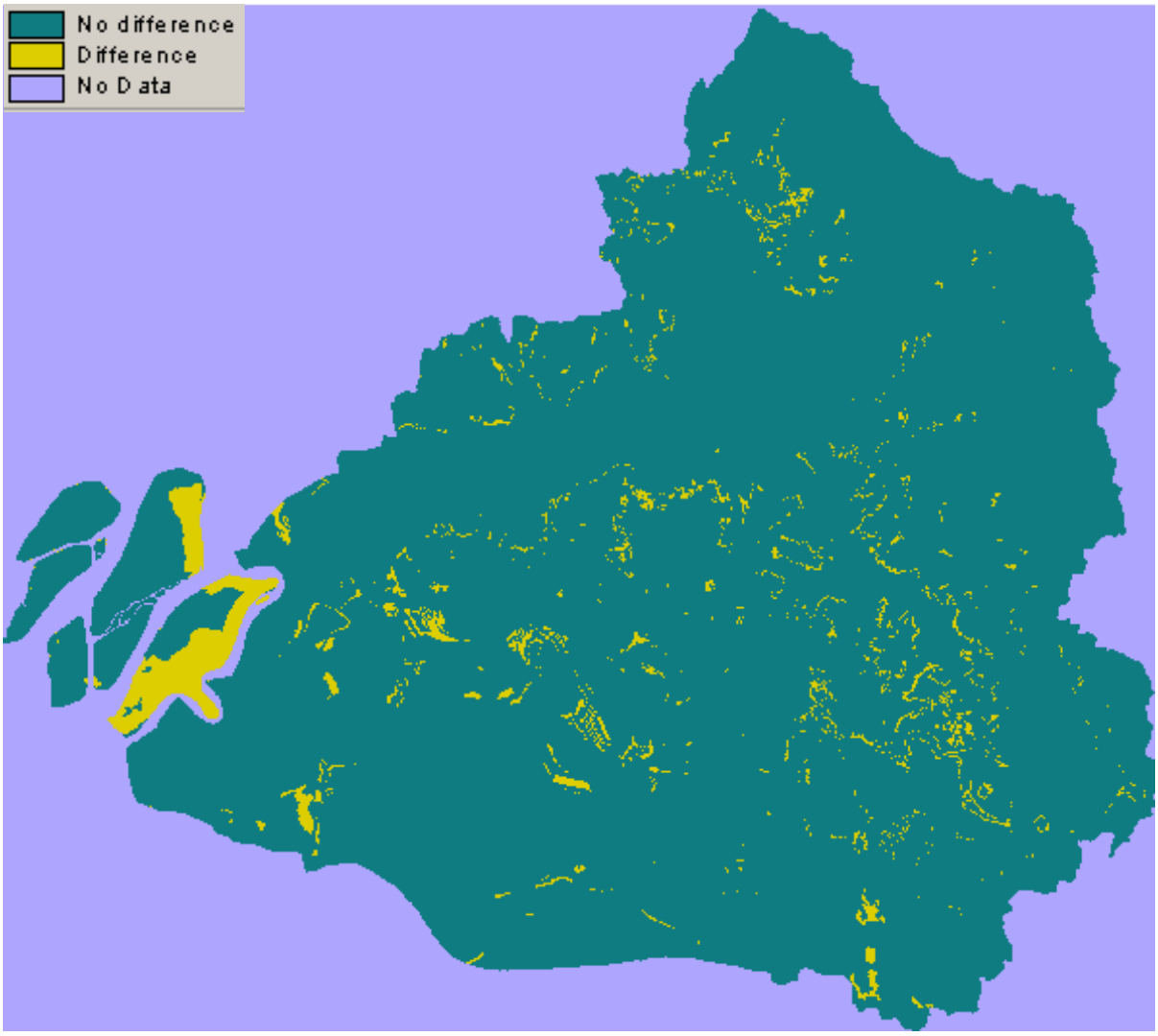


Figure 3.15 Difference between domi_ucdem and domi_uccd

3.7.4 Analysis of the land use fragmentation indices

Statistics of the land use fragmentation indices

The mean, median and standard deviations of the fragmentation indices of the default map and the model results using uncertain DEM are listed in Table 3.12.

Table 3.12 The mean, median and standard deviation of the fi_ucdem_rst

fi_ucdem_rst \ LU		**urban**	**mining**	**water**	**agric**	**baregrs**	**peatfr**	**mgvfr**	**dipcfr**
Total Area (per class)	Mean	0.25	0.01	0.02	0.42	0.02	0.02	0.05	0.20
	Median	0.25	0.01	0.02	0.42	0.02	0.02	0.05	0.20
	Std	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Dft	0.25	0.02	0.02	0.42	0.02	0.02	0.05	0.20
CLUMPINESS Index (per class)	Mean	0.90	0.79	0.64	0.92	0.78	0.91	0.91	0.96
	Median	0.90	0.79	0.64	0.92	0.78	0.91	0.91	0.96
	Std	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	Dft	0.91	0.79	0.64	0.92	0.77	0.92	0.92	0.96
Perimeter to Area Ratio (per class)	Mean	21.24	60.22	92.31	13.13	59.47	25.19	13.79	7.26
	Median	21.01	59.99	92.34	13.02	59.53	25.18	13.83	7.23
	Std	1.41	1.61	0.28	0.66	0.60	0.33	1.06	0.20
	Dft	18.53	58.89	92.36	12.44	61.55	22.58	14.16	7.20

The mean, median and standard deviations of the fragmentation indices of the default map and the model results using uncertain DEM are listed in Table 3.13.

Table 3.13 The mean, median and standard deviation of the fi_uccd_rst

fi_uccd_rst \ LU		**urban**	**mining**	**water**	**agric**	**baregrs**	**peatfr**	**mgvfr**	**dipcfr**
Total Area (per class)	Mean	0.25	0.01	0.02	0.42	0.02	0.02	0.05	0.20
	Median	0.25	0.01	0.02	0.42	0.02	0.02	0.05	0.20
	Std	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Dft	0.25	0.02	0.02	0.42	0.02	0.02	0.05	0.20
CLUMPINESS Index (per class)	Mean	0.90	0.79	0.64	0.92	0.77	0.92	0.92	0.96
	Median	0.90	0.79	0.64	0.92	0.77	0.91	0.92	0.96
	Std	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
	Dft	0.91	0.79	0.64	0.92	0.77	0.92	0.92	0.96
Perimeter to Area Ratio (per class)	Mean	20.01	59.39	92.32	13.13	61.31	23.67	14.54	7.34
	Median	20.04	59.11	92.41	13.09	61.30	24.31	14.50	7.34
	Std	0.14	0.74	0.36	0.20	0.69	1.56	0.36	0.02
	Dft	18.53	58.89	92.36	12.44	61.55	22.58	14.16	7.20

Error bar plots for the land use fragmentation indices

The error bar plots of the land use fragmentation indices 'total area', 'clumpiness index' and 'perimeter to area ratio' for the CLUE-S model results using uncertain DEM and travel speed are shown in Figure 3.16, Figure 3.17 and Figure 3.18, respectively.

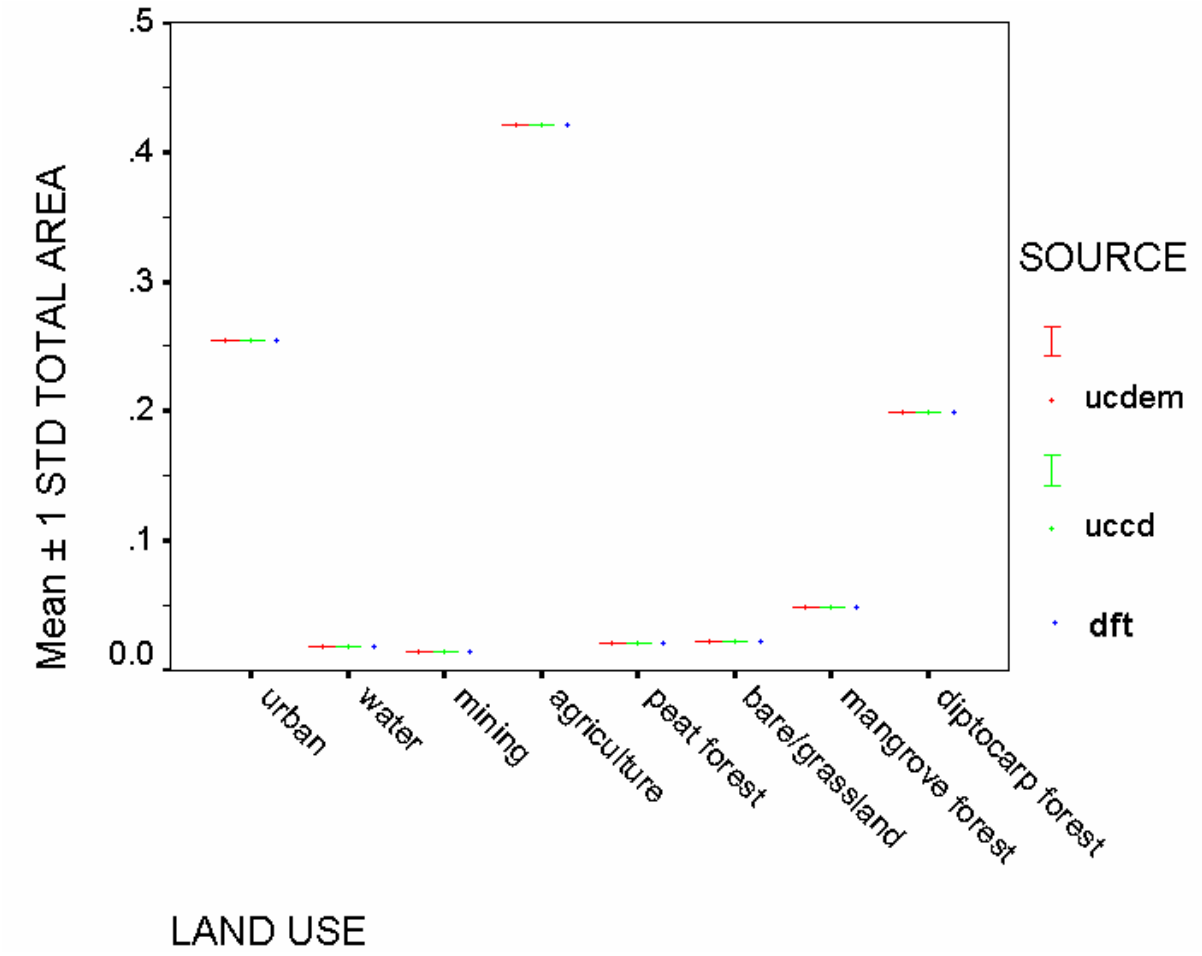


Figure 3.16 Mean ± Std of the fragmentation index 'Total Area' for ucdem_rst, uccd_rst and dft

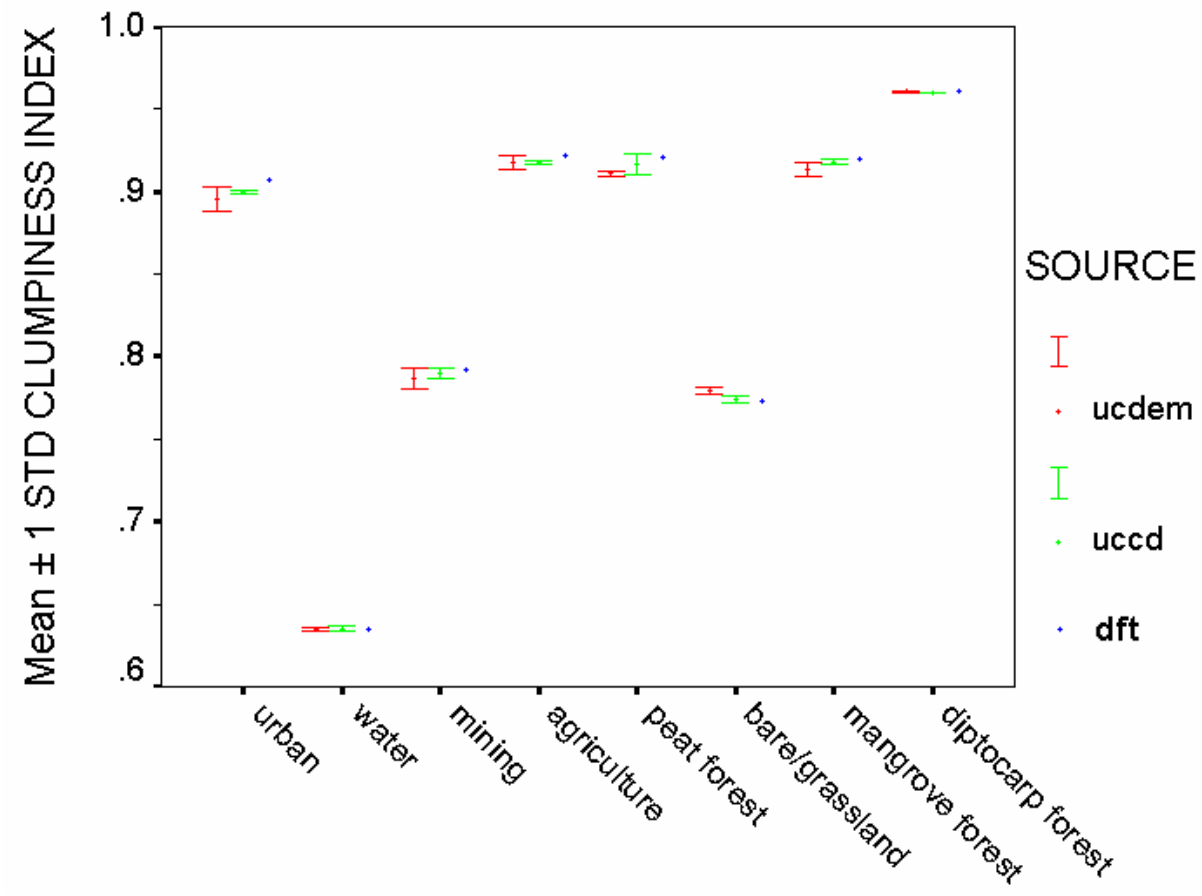


Figure 3.17 Mean ± Std of the fragmentation index 'Clumpiness Index' for ucdem_rst, uccd_rst and dft

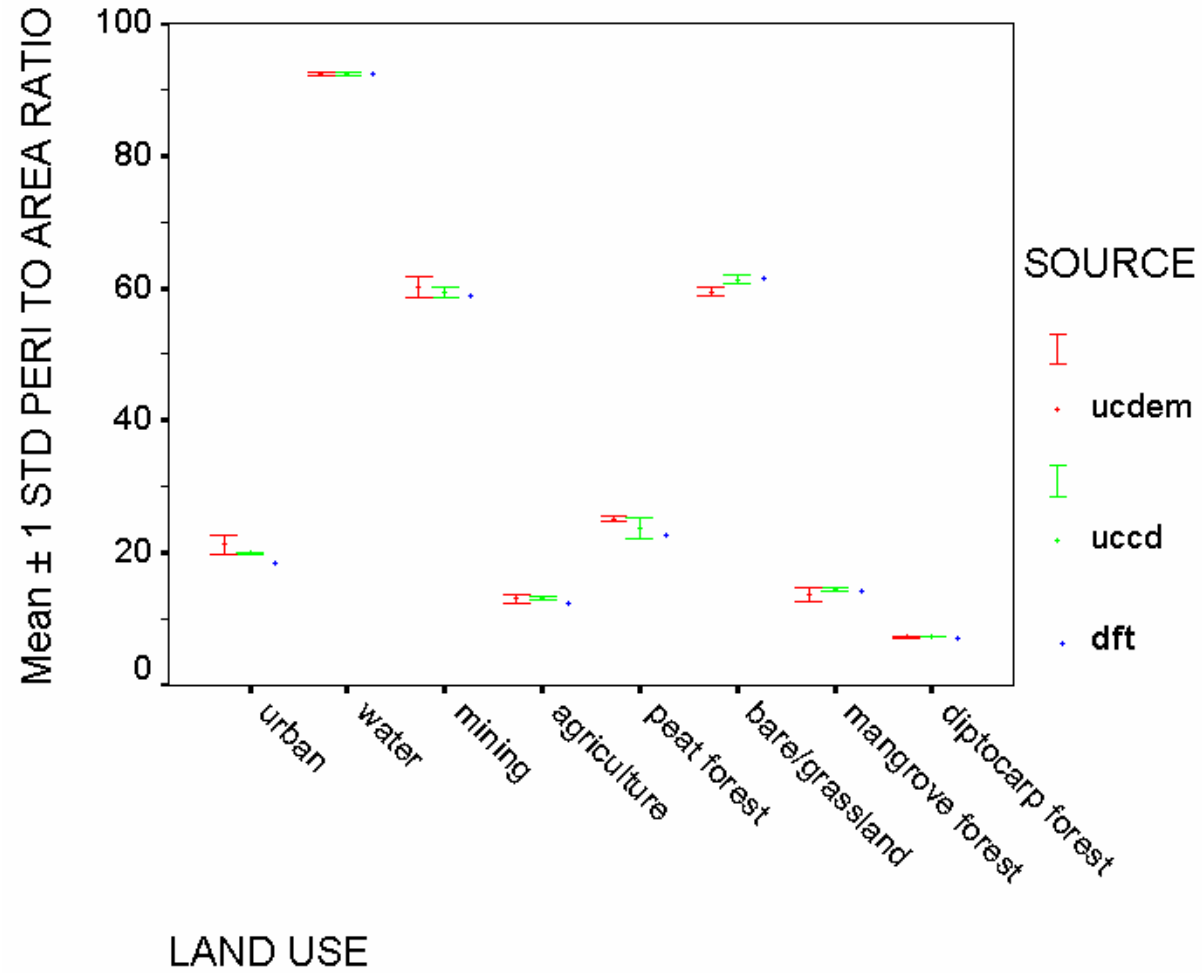


Figure 3.18 Mean ± Std of the 'Perimeter to Area Ratio' for ucdem_rst, uccd_rst and dft

CHAPTER 4
DISCUSSION OF THE CASE STUDY

4.1 Discussion of the entropy maps

Figure 3.9 shows uncertainties in the CLUE-S model outputs caused by uncertainties in DEM; since all other inputs were kept to their original value. Figure 3.10 shows uncertainties in the CLUE-S model outputs caused by uncertainties in travel speed. From Figure 3.9 and Figure 3.10 it can be observed that most parts of the study area have entropy equal to zero, indicates that the model prediction is stable in most part of the study area; the uncertainties in DEM and travel speed did not cause much uncertainty in the model outputs.

4.1.1 Large areas of uncertainties in circle A of Figure 3.9

In Figure 3.9, the most obvious uncertainties are in area A. There are also other smaller areas of uncertainties scatter across the study area. To further investigate the reason of this pattern, Figure 3.5 and Figure 3.7 are zoomed in to the area A in Figure 3.9, and are shown in Figure 4.1 and Figure 4.2, respectively.

Comparing Figure 4.2 with Figure 4.1, it shows that those cells that are originally predicted as ‘mangrove forest’ in Figure 3.3 have been changed to ‘urban’ in Figure 4.2, if the elevation in Figure 4.1 at these grid cell locations are larger than four meter. These grid cells are surrounded by circle A to circle N of Figure 4.1 and Figure 4.2.

In contrast, gird cells in circle O of Figure 3.9 has an entropy value equal to zero, this means they have constantly been predicted by the model as 'mangrove forest' in all the model realizations using uncertain DEM, even when elevation in circle O is larger than four meter. This is because these grid cells are in protection areas (Figure 3.2, circle A) in which land use change is not allowed.

Similar findings are discovered when comparing Figure 3.9 with the 10^{th} simulated DEM and the model result using it (not shown in this thesis work). It shows that the pattern is not random, the similar pattern occurred in all the model results using uncertain DEM.

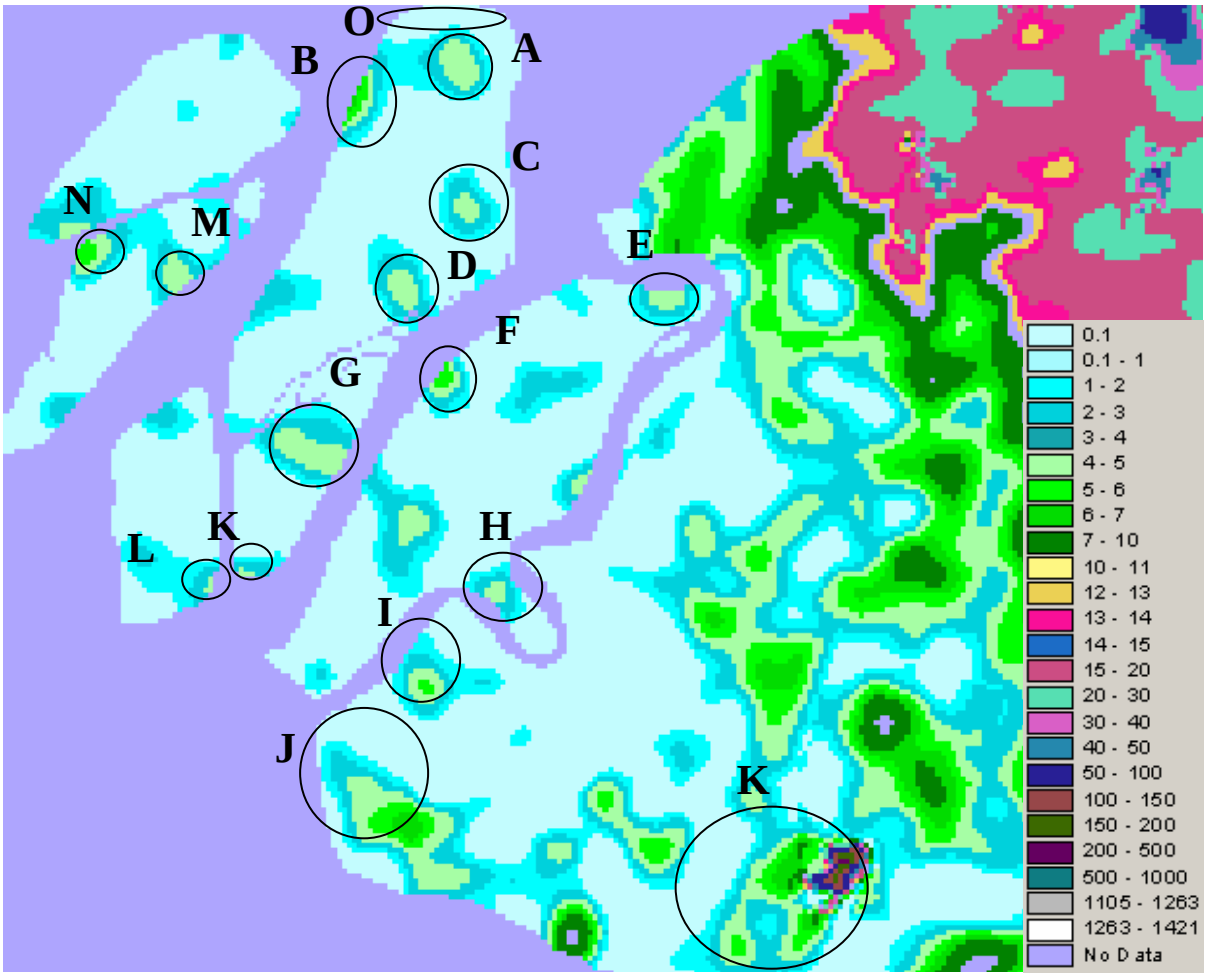


Figure 4.1 The first simulated uncertain DEM (zoomed in)

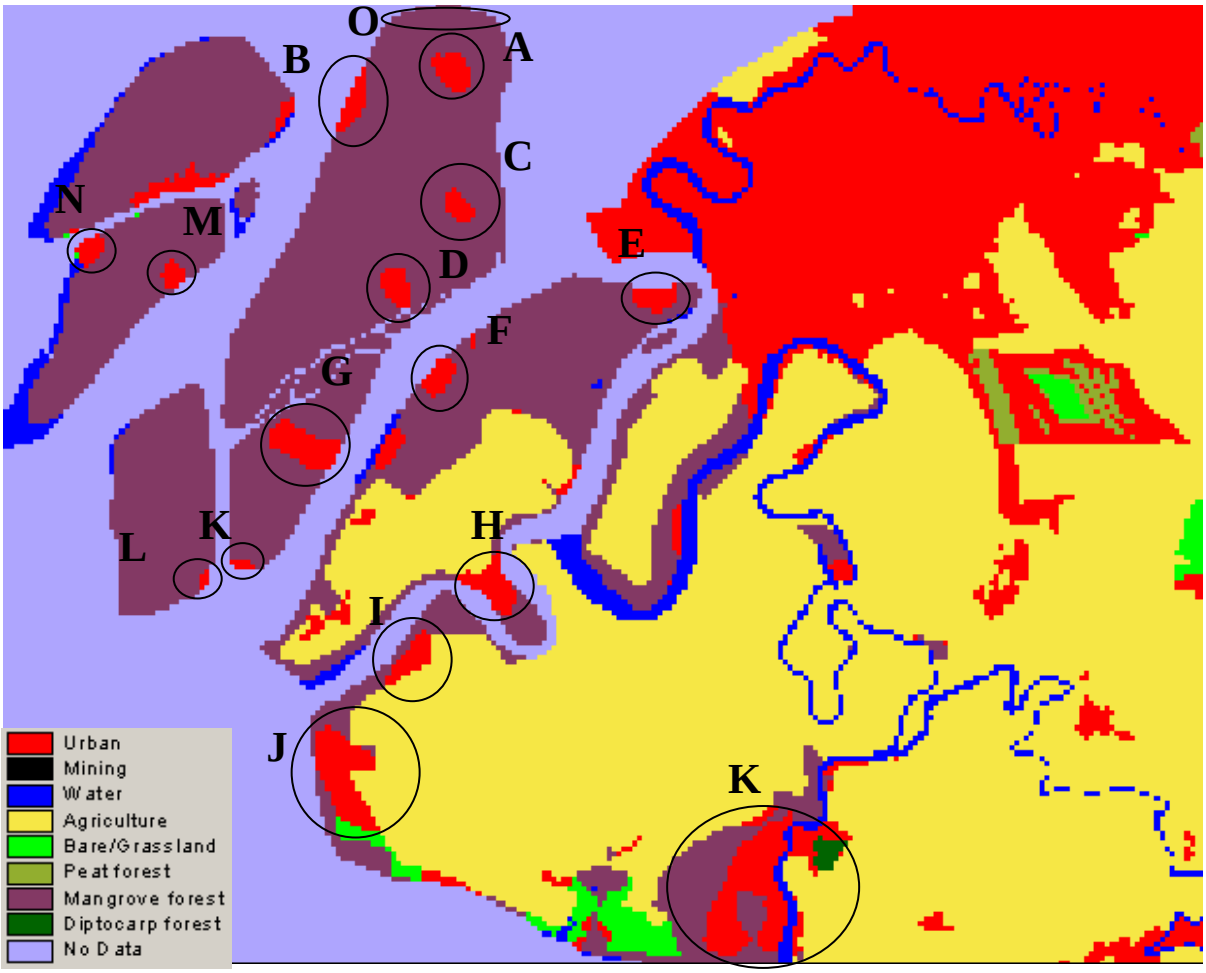


Figure 4.2 The model result using the first simulated DEM (zoomed in)

4.1.2 Large uncertainties in circle B of both Figure 3.9 and Figure 3.10

Circle B of Figure 3.9 and Figure 3.10 shows large uncertainties. Comparing Figure 3.9 and Figure 3.10 with the model results using uncertain DEM and travel speed (see circle B in Figure 3.6 and Figure 3.7), it shows that grid cells in circle B of Figure 3.9 are located in land use transitional zones (the border where one or several land use types change to other land use types and the surrounding area). For circle B, this is particularly the case, because this is the area where 'urban', 'water', 'agriculture', 'bare/grassland', 'mining' and 'peat forest' change land use types.

4.1.3 Large uncertainties in circle C of Figure 3.10 compared with small uncertainties in circle C of Figure 3.9

Although both areas in circle C are located in land use transitional zones, circle C in Figure 3.10 shows much larger uncertainties than in Figure 3.9. Being in land use transitional zones is not the only reason for circle C in Figure 3.10 to show large uncertainties. To further investigate the reason, towns are overlaid on Figure 3.3 and shown in Figure 4.3.

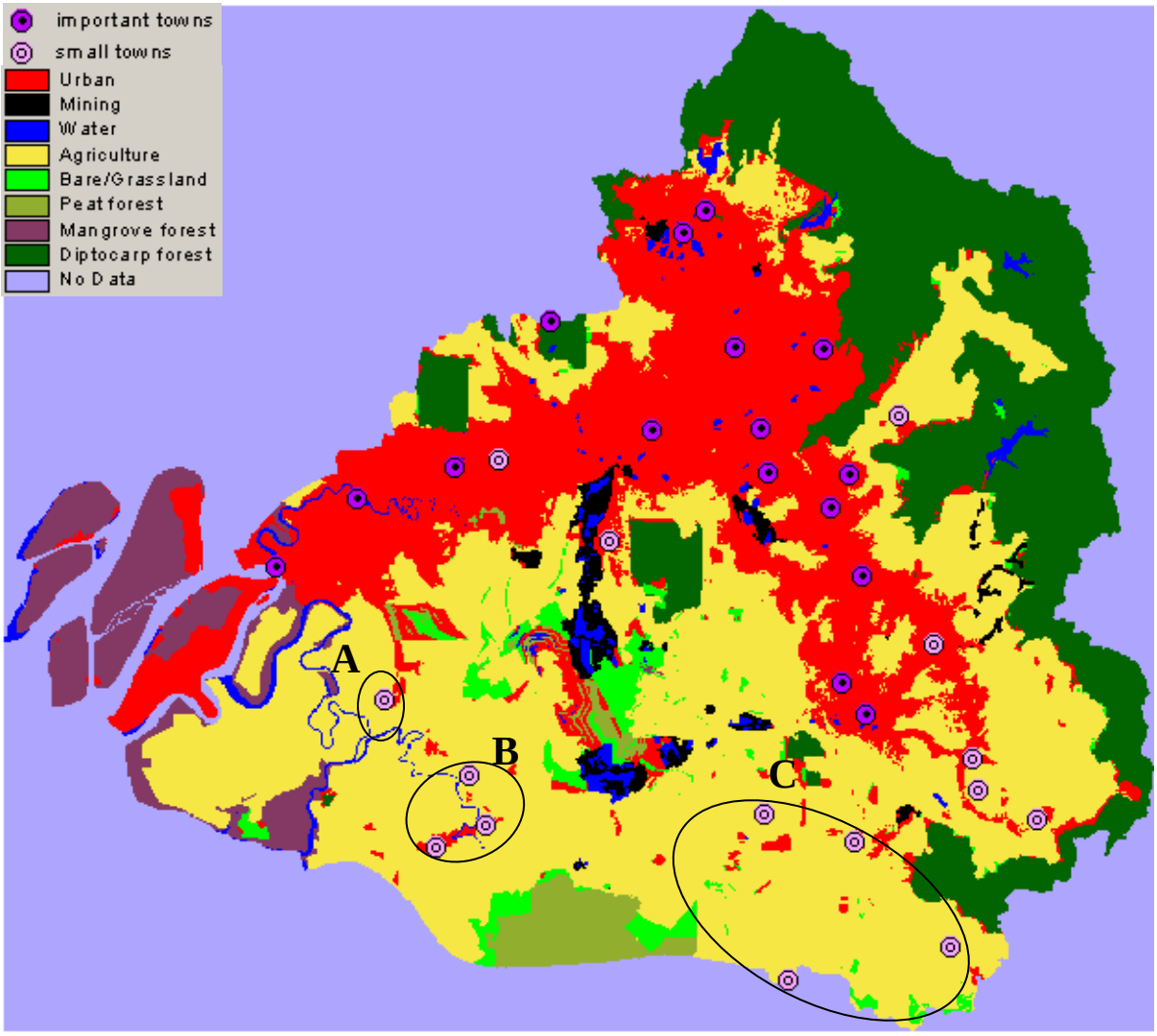


Figure 4.3 Towns overlaid on Figure 3.3

Figure 4.3 shows that circle C in Figure 3.10 is close to towns in circle B in Figure 4.3, the uncertainties in travel speed caused accessibility to these towns to change (in contrast, uncertainties in DEM did not much changes with respect to accessibility) and therefore triggered large area of uncertainties in circle C of Figure 3.10.

4.1.4 Large uncertainties in circle I and H in Figure 3.10 compared with small uncertainties in the same area in Figure 3.9

The large area of uncertainties in circle I and H in Figure 3.10 are also caused by uncertain accessibility to nearby towns (as a result of uncertain travel speed, see circle A and B in Figure 4.3).

This effect is especially obvious in circle H in Figure 3.10 where it is not located in land use transitional zones but surrounded by towns (see circle C in Figure 4.3).

4.1.5 Large uncertainties in circle D of Figure 3.10 compared with small uncertainties in the same area of Figure 3.9

Although grid cells in circle D in both located in Figure 3.9 and Figure 3.10 are located in land use transitional zones, the uncertainties in Figure 3.10 are much larger. This is because uncertainties in travel speed caused accessibility of 'peat forest' to nearby towns and sawmill to change, consequently caused large uncertainties in the location of 'peat forest'. On the other hand, uncertainties in DEM did not cause large changes in term of accessibility to towns and sawmills. This is also the reason for larger area of uncertainties in circle E and circle G in Figure 3.10 than the same area in Figure 3.9.

4.1.6 Uncertainties in circle F of Figure 3.9 and Figure 3.10

The large area of uncertainties in circle F of Figure 3.9 and Figure 3.10 are located in the area where urban expansion took place. Large 'urban' area expanded into 'agriculture'. The large uncertainties follow the pattern of the edge of the transitional zone between 'urban' and 'agriculture'. The low conversion elasticity of 'agriculture' (Table 3.2) also contributed to the uncertainties.

From both Figure 3.9 and Figure 3.10, it can be observed that although the exact locations of the urban expansion remain uncertain, the general pattern of the urban expansion follows the location of towns, especially important towns. Besides, 'urban' tends to aggregate together with the existing 'urban' areas.

The general pattern of uncertainties in circle F of Figure 3.9 and Figure 3.10 is similar, but model result using uncertain travel speed have larger areas of uncertainties.

4.1.7 Difference of uncertainties in mountain and plain area of Figure 3.9

In spite of the larger DEM errors in the mountain area, in Figure 3.9 the mountain area is relatively certain compared with plain area. This is because the demand for 'peat forest' (see Table 3.1) is relatively stable compared with the total area of 'diptocarp forest', and the model setting that no other land use types can be converted to 'diptocarp forest' (see Table 3.3) (therefore 'peat forest' has to stay to the same location to meet the demands).

It can also be observed in Figure 3.9 that the manually digitized line to separate mountain area and plain area did not cause sudden changes of uncertainties along the line.

It should be noted that the causes of uncertainties in the model output are caused by the combined effects of uncertainties in inputs; and influenced by other factors (such as the distance to towns and sawmills). Therefore when evaluating the uncertainties in the model outputs, many factors should be considered comprehensively.

4.2 Discussion of the results of the confusion matrices

For the CLUE-S model, the value of the producer's accuracy and user's accuracy are the same for realizations using uncertain DEM and travel speed versus the default map. This is because the demands remain the same when the uncertain inputs were used. The CLUE-S model in effect re-allocated the same amount of cells into different locations.

The number of grid cells that have changed as a result of uncertain DEM and travel speed for each land use type can be observed in the confusion matrices (Table 3.6 and Table 3.7),

while the location of the changes can be observed in the entropy maps (Figure 3.9, Figure 3.10), respectively.

4.2.1 Discussion of Table 3.6 and Table 3.7

The producer's accuracies vary considerably among different land use types. 'Diptocarp forest' has the largest producer's accuracy of one (discussed in section 4.1.7).

On the other hand, 'bare/grassland' has the lowest producer's accuracy of 0.82. This is due to its low conversion elasticity of 0.2, as specified in Table 3.2. this low conversion elasticity means that 'bare/grassland' is relatively easy be converted to other land use types when there is a demand, therefore relatively easy to shift location.

The producer's accuracy for 'mangrove forest' is particularly low in Table 3.6 (only 0.86), when compared with the producer's accuracy for 'mangrove forest' in Table 3.7 (almost one). This is because the uncertainties in DEM have caused the original 'mangrove forest' to be converted to 'urban', as have already been discussed in section 4.1.1.

'Urban' in the default has been converted to many other land use types (especially 'agriculture' and 'mangrove forest'). In the model results using uncertain DEM, 'urban' also contained many cells converted from other land use types (especially 'agriculture' and 'mangrove forest'). The lower producer's accuracy for 'urban' in Table 3.6 than in Table 3.7 is mainly caused by this 'mixing' of 'mangrove forest' and 'urban' (discussed in section 4.1.7). The producer's accuracy is non-spatial measurement; it does not measure spatial distribution of classification accuracy (in this thesis work, the percentage of cells that stayed the same). Apart from the area in circle A of Figure 3.9, the producer's accuracy is higher for 'urban' in Table 3.6 than in Table 3.7.

From the column 'urban' in Table 3.6, we can observe where the grid cells predicted as 'urban' in the model output using uncertain DEM came from. For example, from 50% of all model results using uncertain DEM, 39791 (92%, the producer's accuracy) came from the original 'urban', 1362 grid cells were converted from 'agriculture', 1085 grid cells were converted from 'mangrove forest', et cetera. The other land use types can be interpreted in the same way.

On the column 'dipcfr' in Table 3.6, only a small number of grid cells are converted from 'urban' and 'bare/grassland' in the default map. The high producer's accuracy of 'diptocarp forest' confirms the small uncertainties where the cell location is 'diptocarp forest', as shown in Figure 3.9.

For the model results using uncertain travel speed (Table 3.7), the percentiles of the producer's accuracy generally (except for 'mangrove forest') have a larger variation than the model results using uncertain DEM (Table 3.6). It indicates that generally the model results using uncertain travel speed have larger uncertainties. The variation of the producer's accuracy for 'bare/grassland' in Table 3.6 is particularly large, among 90% of all the model results using uncertain travel speed, the producer's accuracy ranges from (0.74 to 0.97). This is because many cells that were originally 'bare/grassland' in the default map have been converted to 'urban' and 'agriculture' when uncertain travel speed was used.

The variations of producer's accuracy and user's accuracy in Table 3.7 are generally larger than in Table 3.6. This means that the model results using travel speed are more uncertain than using uncertain DEM.

4.2.2 Discussion of Table 3.8 and Table 3.9

Overall, the producer's accuracy and user's accuracy in Table 3.8 and Table 3.9 are relatively low. This means the model results using uncertain DEM and travel speed are quite different from the validation map.

In Table 3.8 and Table 3.9, 'diptocarp forest' has the largest producer's accuracy (0.91), the cause of it has been discussed in section 4.1.7.

On the other hand, 'bare/grassland' has the lowest producer's accuracy (0.01); most of the cells that are 'bare/grassland' in the validation map were predicted as 'urban' and 'agriculture' in the model results using uncertain DEM.

The producer's accuracy and user's accuracy for other land use type are in the middle of these two extremes, and can be interpreted in a similar way as for ‘bare\grassland’ in the last paragraph.

The variations of producer's accuracy and user's accuracy in Table 3.8 and Table 3.9 are relatively low. This means that the model prediction is stable.

4.2.3 The dominant map using uncertain DEM versus the default model result

In general, the producer's accuracy and user's accuracy in Table 3.10 are relatively high. This indicates that the dominant maps using uncertain DEM and travel speed are quite similar to the default model result.

From the row 'urban' in Table 3.10, it can be observed that grid cells that are originally 'urban' in the default model result have been predicted as many other land use types, especially as 'agriculture' (1706 cells) and 'mangrove forest' (1314 cells). From the row 'agriculture' in Table 3.10, it shows that 1338 cells that are original 'agriculture' in the default model result have been predicted as 'urban' when using uncertain DEM. In other words, 1338 cells that are originally 'agriculture' in the default model result shifted location when uncertain DEM are used, from the mainland to the island, as can be seen from circle D in Figure 3.13.

The column 'urban' in Table 3.10 shows how many cells from each land use type that were originally not 'urban' in the default model result have been converted to 'urban' in the dominant map using uncertain DEM. The locations of these cells are shown in Figure 3.13.

It can be observed that the producer's accuracy of 'mangrove forest' in Table 3.10 is much larger than in Table 3.6. This shows that the isolated patches of 'urban' in Figure 3.7 caused by uncertain DEM are randomly distributed, no particular locations are dominated by 'urban' constantly (because no particular locations have high elevation constantly in the simulated DEM). This finding confirms the pattern in circle A in Figure 3.9, where large areas of uncertainties occur, but the magnitude of uncertainties themselves are not high. Moreover, the pattern of DEM is expected using the unconditional sequential Gaussian simulation presented in chapter 2.

The producer's accuracy for 'mangrove forest' shows that there is a 95% probability that 'mangrove forest' stays the same location when uncertain DEM are used. On the other hand, the user's accuracy for 'mangrove forest' is relative low (0.86), it shows that there is a 86% probability that 'mangrove forest' in the model results using uncertain DEM actually belonged to 'mangrove forest' in the default model result.

In Table 3.10, it can be observed that the producer's accuracy of 'bare/grassland' is quite low (0.82). This is because many cells in the default model result (555) have been converted to 'urban' in the dominant map using uncertain DEM. The user's accuracy of 'bare/grassland' is also quite low, this is because among all the cells that are predicted as 'bare/grassland' in the dominant map using uncertain DEM, 406 cells and 145 cells are originally belong to 'agriculture' and 'peat forest' in the default model result, respectively. This shifting of location for 'bare/grassland' is caused by the zero conversion elasticity specified in Table 3.2.

4.2.4 The dominant map using uncertain travel speed versus the default model result

In general, the producer's accuracy in Table 3.11 is very high, and higher than the producer's accuracy in Table 3.10, it means there is a very high probability that land use types will stay the same in the dominant map using uncertain travel speed. The user's accuracy in Table 3.11 is also very high, it means that for the cells that are in the dominant map using uncertain travel speed, there is a very high probability that they came from the same land use type in the default land use map.

The columns 'urban' in Table 3.11 shows 'urban' in the dominant map using uncertain travel speed. The column also shows which land use type these cells came from. The locations of these cells that originally not 'urban' in the default model result are shown in Figure 3.14.

4.2.5 Comparison of dominant map using uncertain travel speed and uncertain DEM

In general, the producer's accuracy and user's accuracy in Table 3.11 are larger than in Table 3.10.

The producer's accuracy and user's accuracy of 'bare/grassland' is much larger in Table 3.11 than that in Table 3.10.

The user's accuracy of 'mangrove forest' in Table 3.11 is much larger than that in Table 3.10.

The difference between the dominant map using uncertain travel speed and uncertain DEM is shown in Figure 3.15.

4.3 Discussion of the results of the land use fragmentation indices

The value changes for the land use fragmentation indices as a result of using uncertain DEM and travel speed can be observed in Table 3.12 and Table 3.13, respectively. The mean values for 'total area', 'clumpiness index', and 'perimeter to area ratio' and its corresponding standard deviations are plotted on the error bar plots in Figure 3.16, Figure 3.17 and Figure 3.18.

4.3.1 Discussion of the index 'total area'

For a visual presentation of the mean and standard deviation of the 'clumpiness index' of the model results using uncertain DEM and uncertain travel speed see the red dots/bars and green dots/bars, respectively in Figure 3.17.

The values of the land use fragmentation index ‘total area’ are the same for model results using uncertain DEM (see Table 3.12) and uncertain travel speed (Table 3.13), and both equal to the total area for each land use type of the default model result.(dft in Table 3.12 or Table 3.13). Moreover, the standard deviation of 'total area' for each land use type is zero (see the red dots/bars in Figure 3.16). This result is expected since the demands are the same for both error-ignored input and error-perturbed inputs (therefore only the cell location changes, the number of cells for each land use type do not change, thus the total area for each land use type is stable). The slightly different value for 'mining' in Table 3.12 is due to the tolerance settings for model convergence and rounding errors.

'Agriculture' has the largest area of 0.42 hectare. 'Urban' and ‘diptocarp forest’ has the area of 0.25 hectare and 0.2 hectare. The main parts of the study area are dominated by these three land use types. The 'total area' of 'mining', 'water', 'bare/grassland' and 'peat forest' are quite small, they are around 0.02.

4.3.2 Discussion of the index 'clumpiness index'

For a visual presentation of the mean and standard deviation of the 'clumpiness index' of the model results using uncertain DEM and uncertain travel speed see the red dots/bars and green dots/bars, respectively in Figure 3.18.

'Diptocarp forest' had the largest 'clumpiness index' of 0.96 in both Table 3.12 and Table 3.13 and standard deviation of zero. This result indicates that most parts of 'diptocarp forest' are aggregated together into large homogeneous patches. The values are equal to the value in default model result, this confirms the results derived previously that the location of 'diptocarp forest' is very stable.

'Water', on the other hand, had the lowest value of 0.64 in both Table 3.12 and Table 3.13 and equal to the value for default model result. This finding indicates that there are many small patches of 'water' scattered across the study area in the default model result (Figure 3.3) and the model results using uncertain DEM (Figure 3.6) and travel speed (Figure 3.7). The standard deviation of the index 'clumpiness index' is zero, denotes that 'water' are constantly been predicted as many patches scattered across the study area.

the 'clumpiness index' for 'urban' in Table 3.12 shows that the model results using uncertain DEM are slightly less aggregated compared to the default model result. The 'clumpiness index' for 'urban' in Table 3.12 is equal to Table 3.13 but with a slightly larger standard deviation, it shows that although the pattern of the model prediction using uncertain DEM (see Figure 3.7) is different from using uncertain travel speed (see Figure 3.6), there is no difference in term of clumsiness. However, the variation for the 'clumpiness index' is slightly higher for the model results using uncertain DEM than using uncertain travel speed. This larger variation for 'urban' confirms the large area of uncertainties in circle A in Figure 3.9.

The 'clumpiness index' for 'mangrove forest' in Table 3.12 shows that the model results using uncertain DEM are slightly less aggregated than the model results using uncertain travel

speed and the default model result. This smaller value for 'mangrove forest' confirms the large area of uncertainties in circle A in Figure 3.9.

The 'clumpiness index' of 'peat forest' in Table 3.13 is slightly larger than that in Table 3.12. This is because uncertain travel speed caused more changes in the accessibility of 'peat forest' to nearby towns and sawmill (modeled by cost distance) than caused by uncertain DEM.

In general, 'clumpiness index' values in both Table 3.12 and Table 3.13 were very similar to the default model result, and the variations are almost equal to zero. This indicates that uncertain DEM and travel speed have very small effects on the clumpiness of the model results.

4.3.3 Discussion of the index 'perimeter to area ratio'

For a visual presentation of the mean and standard deviation of the 'clumpiness index' of the model results using uncertain DEM and uncertain travel speed see the red dots/bars and green dots/bars, respectively in Figure 3.18.

From Table 3.12 and Table 3.13 it can be observed that 'water' has the largest 'perimeter to area ratio' but its standard deviation is small. This result means 'water' has complicated border lines (for the default model result and the model results using uncertain travel speed and DEM the complicated border lines means many small patches, see Figure 3.3, Figure 3.6 and Figure 3.7, respectively) while the uncertainties in inputs have little influence on the shape of the borders in the model output. This result confirms the result of 'clumpiness index' that 'water' has the smallest clumpiness.

'Diptocarp forest' has the lowest 'perimeter to area ratio' and standard deviation in Table 3.12 and Table 3.13. This is because 'diptocarp forest' forms large homogeneous patches.

'Perimeter to area ratio' of 'peat forest' in Table 3.13 is much larger than that in Table 3.12. This is because uncertainties in travel speed changed the accessibility of 'peat forest' to towns and sawmills, and therefore cause the location for 'peat forest' to be uncertain. On the contrary, uncertain DEM did not cause the accessibility of 'peat forest' to change in a large degree.

The 'perimeter to area ratio' of 'bare/grassland' and the corresponding standard deviation is very similar in Table 3.12 and Table 3.13. however, when compare the standard deviation for 'perimeter to area ratio' with the standard deviation for 'clumpiness index', it reveals that the uncertainties in uncertain DEM and travel speed only decreased the complexity of the borders to a small degree, but has no effects on the clumpiness of the land use type.

In general, the standard deviation of 'perimeter to area ratio' for each land use type in Table 3.12 is larger than those in Table 3.13, except for 'peat forest' and 'bare/grassland', which have been discussed above. This means the changes of border complexity of the model outputs using uncertain DEM are slightly larger than that using uncertain travel speed, but the differences are not large.

All in all, the land use fragmentation indices are very similar; the uncertainties in DEM and travel speed do not cause large changes to the output.

4.4 Issues of model execution time and model convergence

For the Kuala Lumpur region case, the default run of the CLUE-S model takes about 1 hour to finish (dependent on the hardware used and the application). When using error-disturbed travel speeds, the execution time increased slightly, to about 1 hour and a half. Thus, for one hundred runs, in principle it will take about 150 hours to finish. In practice, the execution time was longer than expected because 10 runs did not converge. In order for the model to converge, two criteria have to be fulfilled. First, the area allocated for each of the land use type has to be close enough to the demands calculated in the non-spatial model. This parameter was set to be 15 in the default settings. Second, when the first criterion has been met, the total difference of each land use type and its corresponding demand has to be within a certain limit. This limit was set to 50 in the default settings. If the model could not find a solution that fulfills the two criteria within 20000 iterations, it stops and reports an error message. To obtain one hundred realizations of the model results using uncertain travel speeds, 10 extra runs using the uncertain travel speeds were executed.

The time needed for the model to finish one run using an uncertain DEM increased to about 2 hours per run. The time increased more dramatically than using uncertain travel speed. The chance that a model run will converge is also considerably decreased. From the execution of 30 runs, only 10% converged. To obtain one hundred realizations of the model results using uncertain DEM, the default setting of convergence criteria was changed. The maximum difference allowed for a particular land use type allocated and the demand for it was increased to 50, the maximum difference allowed for all land use types allocated and the total demand was increased to 200. Using this setting, one hundred runs were executed, 10 runs did not converge. Therefore, 10 more runs were executed using the uncertain DEMs drawn from the same probability distribution.

CHAPTER 5
CONCLUSIONS

After performing the uncertainty propagation on the selected uncertain inputs, several conclusions can be drawn from the analysis. The research questions listed at the beginning of the thesis can now be answered.

The research questions are answered as follows.

1. What are the most important uncertain inputs in the CLUE-S model when applied to the Kuala Lumpur region?

   According to the selection criteria listed in chapter 2, the most important model inputs are DEM (and its derivative slope) and travel speed to highway, sawmill and important towns. The uncertainty in model parameters was not analyzed.

2. How to identify and quantify the uncertainty in the selected uncertain inputs?

   For continuous spatial data, uncertainties can be identified and quantified by comparing them with more accurate data. When such more accurate data are not available, one has to rely on expert knowledge and literatures of similar studies.

3. What is the contribution of the individual selected uncertain inputs to the output uncertainty?

   Taking the entire study area as a whole, uncertain DEM contributes more uncertainties to the output uncertainty. 'Mangrove forest' is very sensitive to uncertain DEM.

   Considering the spatial distribution of output uncertainties, uncertain DEM contributes more uncertainties to the output in circle A of Figure 3.9. For the rest of the study area, uncertain travel speed contributes more uncertainties to the model output.

   The most probable areas where uncertainties will occur are in land use transitional zones when uncertain DEM and travel speed are used. However, situated in land use transitional zones is not the only reason for the model prediction to be uncertain (and sometimes also not the main reason). Uncertainties in the model output are caused by the combined effects of many factors, such as the uncertainties in inputs, land use transitional zones and the distance to towns and sawmills, et cetera.

4. Which input uncertainty should be reduced in order to improve the accuracy of the land use change predictions made by the CLUE-S model?

   The uncertainties in DEM and travel speeds do not cause large uncertainties in the output of the CLUE-S model in the Kuala Lumpur region. Uncertainty analysis of model settings and parameters is therefore needed to have a better understanding of the uncertainties involved with the model predictions.

References

Aerts, J. C. J. H., G. B. M. Heuvelink, et al. (2003). "Accounting for spatial uncertainty in optimization with spatial decision support systems." Transactions in GIS **7**(2): 211-230.

Brown, J. D. and G. B. M. Heuvelink (2004a). Assessing uncertainty propagation through physically based models of soil water flow and solute transport. Encyclopedia of Hydrological Sciences. A. M.G. Chicester, UK, Wiley.

Brown, J. D. and G. B. M. Heuvelink (2004b). On the identification of uncertainties in spatial data and their quantification through probabilistic means. Handbook of GIS. S. Fotheringham and J. Wilson. Chicester, UK, Wiley.

Burrough, P. A. and R. A. McDonnell (1998). Principles of Geographical Information Systems. New York, Oxford University Press Inc.

Deutsch, C. V. and A. G. Journel (1992). GSLIB: geostatistical software library and user's guide. New York, Oxford University Press.

Dominich, S., J. Goth, et al. (2004). "An entropy-based interpretation of retrieval status value-based retrieval, and its application to the computation of term and query discrimination value." Journal of the American Society for Information Science and Technology: 613-627.

Eiden, G., M. Kayadjanian, et al. (2000). FROM LAND COVER TO LANDSCAPE DIVERSITY
IN THE EUROPEAN UNION.

ESRI (2001). ARC/INFO help.

Goovaerts, P. (1998). impact of the simulation algorithm, magnitude of ergodic fluctuations and number of realizations on the spaces of uncertainty of flow properties, Stanford Center for Reservoir Forecasting, Stanford University, Unpublished annual report No 11.

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

Isaaks, E. H. and R. M. Srivastava (1990). An Introduction to Applied Geostatistics. New York, Oxford University Press.

Kraim, O. A. (1998). http://www.eng.ukm.my/~othman/klang_langat_info.html.

Kyriakidis, P. C., A. M. Shortridge, et al. (1999). "Geostatistics for conflation and accuracy assessment of digital elevation models." International Journal of Geographical Information Science **13**(7): 677-707.

Lillesand, T. M. and R. W. Kiefer (2000). Remote Sensing And Image Interpretation. New York, John Wiley & Sons, Inc.

McGarigal, K., E. Ene, et al. FRAGSTATS USER GUIDELINES.

Meyer, W. B. and B. L. Turner, Eds. (1994). Changes in land use and land cover: A global perspective. Cambridge, UK, Cambridge University Press.

Niel, K. V. and S. W. Laffan (2003). "Gambling with randomness: the use of pseudo-random number generators in GIS." International Journal of Geographical Information Science **17**(1): 49 - 68.

Pebesma, E. J. (1999). GSTAT USER'S MANUAL. http://www.gstat.org/manual/gstat.html.

USGS (1997). "Standards for Digital Elevation Models."

Verburg, P. (2004). "Projecting land use transitions at forest fringes in the Philippines at two spatial scales." Landscape Ecology **19**(1): 77-98.

Verburg, P. and A. Veldkamp (2004). "Projecting land use transitions at forest fringes in the Philippines at two spatial scales." Landscape Ecology **19**(1): 77-98.

Verburg, P. H., W. Soepboer, et al. (2002). "Modeling the Spatial Dynamics of Regional Land Use: the CLUE-S Model." Environmental Management **30**(3): 391-405.

Verburg, P. H., A. Veldkamp, et al. (2004). Landscape Level Analysis of the Spatial and Temporal Complexity of Land-Use Change. Ecosystems and Land Use. R. DeFries, G. Asner and R. A. Houghton, AGU. **153:** 217-230.

Veregin, H. (1999). Data quality parameters. Geographical Information Systems Principles and Technical Issues. P. A. Longley, M. F. Goodchild, D. J. Maguire and D. W. Rhind. New York, John Wiley & Sons, Inc. **1**.

Zhang, J. and M. Goodchild (2002). Uncertainty in Geographical Information. London, Taylor & Francis Inc.