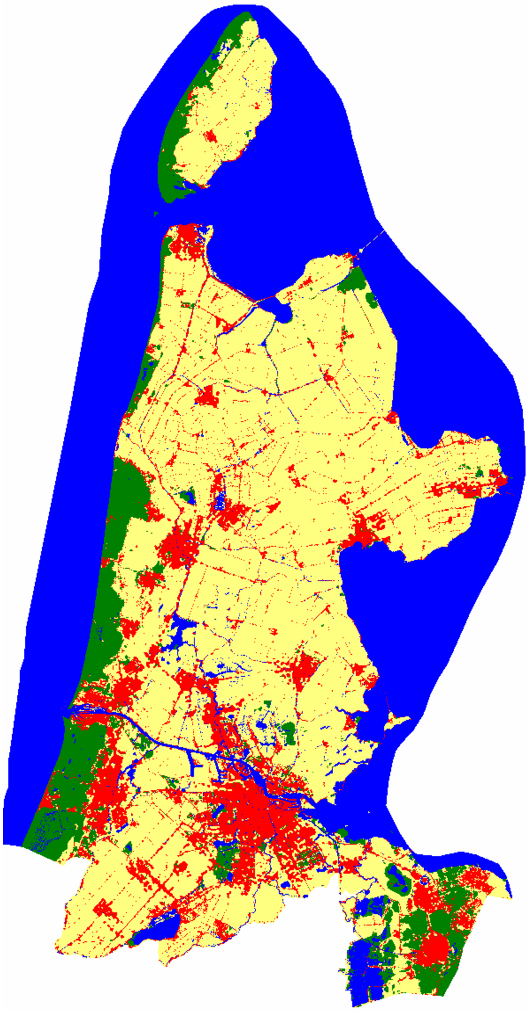
CENTRE FOR GEO-INFORMATION

THESIS REPORT GIRS-2006-03

VALIDATION OF LAND USE CHANGE MODELS

A CASE STUDY ON THE ENVIRONMENT EXPLORER

JASPER VAN VLIET



FEBRUARI 2006

WAGENINGEN UNIVERSITY
WAGENINGEN UR

VALIDATION OF LAND USE CHANGE MODELS

A case study on the Environment Explorer

JASPER VAN VLIET
REGISTRATION NUMBER 800723898130

SUPERVISORS:
DR. IR.RON VAN LAMMEREN
PROF. DR. IR. ARNOLD BREGT

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE DEGREE OF MASTER O
F SCIENCE AT WAGENINGEN UNIVERSITY AND RESEARCH CENTRE, THE
NETHERLANDS.

FEBRUARY 2006
WAGENINGEN, THE NETHERLANDS

Thesis code number: K075-704
Wageningen University and Research Centre
Laboratory of Geo-Information Science and Remote Sensing
Thesis Report: GIRS-2006-3

ACKNOWLEDGEMENTS

This report reflects about seven months of thesis work, or maybe even several years of study in geo-information science. Along these months I unexpectedly started to like scientific research. For that, as well as the help they gave me, I would like to thank the some people.

First of all I'd like to thank Ron van Lammeren a lot. Not only for the help during the period I was working, but especially when I wasn't, because of an aching shoulder. I'd also like to thank Arnold Bregt, most of all for introducing me to the right people, and directing me to some previously unknown sources of information. Finally I thank Henk Kramer for allowing me to use and providing me with the HGN datasets.

The thing that encouraged most was actually the fact that people outside the university were interested in my research and willing to help. Ton de Nijs, thank you for helping me with the Environment Explorer as well as giving feedback during the research. It speeded things up considerably. Finally I thank the people at RIKS and Alex Hagen in particular. I had a very good day at your office in Maastricht, and you were always prepared to answer the questions and solve the problems I was sending.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Land use and land cover change models are tools to analyze, simulate or predict the dynamics in land use and land cover. They are widely used in planning and policy making to explore scenarios. Technical issues of these models have been receiving much research effort since computer power has been increasing rapidly last decades. Model analysis and evaluation however received less attention. Due to a lack of proper techniques and methods, there is no standard or guideline to evaluate these models. In this thesis a guideline for operational validation of these models is proposed. A goodness of fit measure as well as a measure for patterns and complexity was applied on the case study model, the Environment Explorer.

The development of a model is an iterative process. In this process the following phases can be distinguished: the set-up phase, the conceptualization phase, the model construction phase, the evaluation phase, the use phase, and finally the reporting phase. While evaluation is a part of the model development cycle, distinct parts of the modeling process can again be evaluated. In this thesis only the validation of model results, operational validation, is considered. As proposed guideline, the model should perform better than the null model and a random model. As well, the sources of errors should be investigated. In what way it should perform better is depending on the desired functionality.

The Environment Explorer is a discrete model in both space and time. It aims at exploring possible future land use for a period up to 30 years ahead. For this research the model was calibrated in two ways, automated and by hand, with land use data of 1960 and 1980. To validate, the 2000 land use cover was simulated starting from the 1980 land use cover. Results were compared to the 2000 land use cover map.

Goodness of fit was measured with Kappa statistics. Because of the model's exploring character, two other methods were applied as well: Fuzzy Kappa statistics and landscape metrics. Four landscape metrics were selected: fractal dimension, shape index, perimeter area scaling and Shannon's diversity index. Metrics were compared globally and focally.

Fuzzy Kappa statistics showed only minor differences as compared to the standard Kappa statistic. Landscape metrics did reveal different results. It was concluded that landscape metrics are a useful addition to traditional validation based on accuracy measurements when applied on explorative models. Case study results on the Environment Explored showed as well, that is hard to outperform the null model when only a very limited amount of information was used. This was concluded for both simulations and both comparison

methods. The automatically calibrated model performed slightly better on the Kappa statistics, whereas the model calibrated by hand performed better on the landscape metrics.

1 INTRODUCTION

1.1 Context and background

1.1.1 Land use simulation models

Land use cover is constantly changing. These changes are the result of processes in a specific area and at a specific time, like social processes, natural processes or economic processes. Some are the result of human intervention and political decisions, as is clearly the case in urban planning but also in nature conservation. Others are the opposite, like urban sprawl or tidal sedimentation.

Land use and land cover change (LUCC) models are tools to analyze, simulate or predict dynamics in this land use cover. This is done to better understand land use changes and to support land use planning, environmental management and policy making. As well they are used for exploring future land use for different scenarios (Verburg et al., 2004). Since land use cover changes are caused by different processes, models simulate different processes and combinations of processes.

In recent years, several simulation models are developed to explore land use cover changes, most of which have a main focus on urbanization. Some examples are 'a self-modifying CA model' by Clarke et al. (1997), the land use scanner (Scholten et al., 2001), the MOLAND and MURBANDY models (Engelen et al., 2002), CLUE (Veldkamp and Fresco, 1996; Verburg et al., 1999) and the CLUE-S model (Verburg et al., 2002).

From a technical point of view these models have some parts in common. First of all they all incorporate time, since modeling time actually is modeling change (Kraak and Ormeling, 2003). The models mentioned above all have a discrete notion of time. They also have a discrete, cell based space. Land use change is simulated as a function of time steps and grid cells. The computational method used for this is a Cellular Automata, in some case extended with other factors of influence.

Differences are particularly found in the application of the models. These applications influence the number of classes, the cell size, the spatial domain, the temporal domain and transition rules.

1.1.2 Validation of land use simulation models

Technical issues, like incorporation of time have been a major research challenges in the field of geographical information systems. This research was mainly concerned with technical and implementation issues like the integration of space and time. Research on analysis and evaluation of these systems at the other side did not receive an equal amount of attention (Peuquet, 2000).

Land use simulation models in most cases are geographical information systems that incorporate both space and time. Due to a rapid increase in computer power, much research effort is put into technical and implementation issues. The speed with which model technology develops however is unmatched by the development of model testing and validation (Gardner and Urban, 2003). A problem they mention with respect to this is a lack of proper techniques and methods. As well they mention that it is a time consuming task.

A review of the models named above shows that some are validated, although these validations are quite dissimilar. Clarke's (1997) model was calibrated extensively, both visually and statistically. Though validation was not mentioned, the results of a Monte Carlo simulation were analyzed. The MOLAND and MURBANDY were also calibrated extensively, with respect to goodness of fit as well as spatial patterns, but no validation results were presented (Engelen et al., 2002). As the CLUE model originates from agronomy and was analyzed and described by de Koning (1998). The CLUE-S' result validation also included a measure for goodness of fit. Both were validated at different scales.

As can be seen from this overview, no standard model validation or result validation is used. Instead most models are analyzed or validated with respect to only some aspects. Most result validations include some measure of the result's goodness of fit. Few however consider spatial structures and patterns, whereas a lot of attention is given to this in model calibration especially in modeling urban areas. This urbanization is a complex process seems rather site-specific at first sight. However fractal analysis shows that cities are fundamentally similar in the way they evolve (Frankhauser, 2004; Tannier and Pumain, 2005). Variations on this are mainly due to local conditions like geography, transportation networks and local planning.

1.1.3 The Environment Explorer

The Environment Explorer is a model developed by the National Institute for Health and the Environment (RIVM) of the Netherlands to explore the effect of different spatial policy options on future land use. These spatial policy options are based on economical, sociological and ecological values, as well as the current land use (Engelen et al., 2003). The result is an indication of the effect on spatial development, land use transformations as well as the effect on a set of indicators under the given demographic and economic prognoses.

In de Nijs et al. (2001b) a case study was applied to examine the effect of three spatial policy options on the Netherlands until 2030. In this application a large number of improvements were indicated for further model development. As well it was mentioned that some model aspects needed further validation. A conclusion that was also drawn in de Nijs et al. (2001) is the need for techniques to validate patterns was specifically mentioned as well as the need to do this with a consistent geo data set.
A more extensive validation was performed for the period 1996 to 2000 (Engelen et al., 2005), after a calibration for the period 1989 to1996. The model was validated on the result's goodness of fit as well as on some spatial characteristics. The short calibration and validation period was indeed assigned as a drawback, since relatively few land use changed. In this thesis a simplified version for the province of Noord-Holland is used that does not use the economic indicators and different policy options.

1.2 Research objectives

As becomes clear from the introduction to the problem above, there is no standard procedure, nor are there any accepted standard guidelines for validating spatio temporal models. At the other side several authors have used different procedures and methods to test these models as well as their results. The objective of this thesis research is to find guidelines for validation specifically applicable to spatio temporal models. This objective is too broad to answer directly; therefore this will be obtained by answering the research questions as presented below.
This objective implies some restrictions to the type of models as well as validation topics considered. An important issue is the use of nominal data in LUCC models as opposed to for example some ecological or hydrological spatial models. Validation of this is not that straightforward (Pontius et al. 2001). Another issue related to the specific application in this thesis is the discretisation of space as well as time. The methods as presented in this thesis are thus specifically well applicable to LUCC models in general and the Environment Explorer in particular. However the answers to certain sub questions are more widely acknowledged in spatial and spatio temporal modeling.
To fulfill this research objective three research objectives are specified. Their answers will deal with the separate aspects of spatio temporal model validation. The main objective is to find guidelines for validating spatio-temporal models. Therefore it must be clear what spatio-temporal models are exactly, and what makes them distinct from other models. The answer is given as a description of the geocomputation as used in the Environment Explorer. For this

the aim here is not to give a complete overview and description of all spatio-temporal models. Instead the emphasis will be on the geocomputation as used in the case study. This result will be an answer to the following objective:

> *'How are space and time modelled in land-use simulation models, particularly the Environment Explorer?'* [Research objective 1]

Model validation itself is part of the modelling cycle, of which several parts can again be validated. A distinction here is often made between conceptual validation and validation of the model's result, here referred to as operational validation (see for example in Brimicombe, 2003). This validation types are applied in several ways by different modellers. An investigation of this, as part of the modelling cycle itself is the topic of the second research objective:

> *'Find or give validation guidelines for spatio-temporal models'* [research objective 2]

Within these guidelines, different methods can be applied. The result of this second research objective only forms the context in which the case study is placed. A precise description of the actual validation performed, as well as the validation itself, is therefore the third objective:

> *'Describe some methods to perform this validation'* [research objective 3a]

> *'Perform this validation on the study case'* [research objective 3b]

1.3 Structure of this thesis report

The outline of this thesis report will mainly follow the line of thought as introduced in the research objectives above. First some basic concepts about the research topic are explored. Then this thesis research' methodology is explained and finally results are presented and discussed and conclusions are drawn.

Chapter 2 starts with an investigation of spatial modeling and spatial data modeling in general. By that the Environment Explorer, the model used as a case study in this thesis, is put in a context. After that the Environment Explorer itself is investigated more thoroughly. By that this chapter is the answer to research question 1. It describes ways to model space and time as well as cellular automata as a case of spatio temporal models that are suitable for LUCC simulation.

The third chapter is a description of model evaluation. It starts with the model development cycle as a whole and then narrows to model evaluation and eventually model result validation. This chapter is essentially the answer to the second research question and in the meanwhile it is the context in which validation in the case study can be placed.

In chapter 4 the calibration as well as the validation as performed on the Environment Explorer is described. This comprises a description of the data sets used for this case study, as well as the exact procedures for both the calibration and validation.

The results of the calibration and validations performed in the case study are presented in chapter 5, together with a discussion. From this discussion eventually the conclusion is drawn and presented in chapter 6.

2 THE ENVIRONMENT EXPLORER EXPLORED

In this chapter I will first give an introduction to spatial models in general, and by that place the Environment Explorer in a context of models in geographic information science. This will include a description of different types of models, different levels of spatial data modeling, and the use of models in combination with GIS. After that, the representation of space and the representation of time is dealt with, both focusing on the concepts and methods that are used by the Environment Explorer.

It is not my purpose to give a complete overview of different kind of spatial models, neither is it to give an extensive descriptions of all the concepts that are used in the study case. Instead it is the intention to introduce some of the major concepts and clarify some terms that are used in this thesis.

2.1 Spatial modeling

Modeling is actually simplification of reality, or as Chorley and Haggett (1967) state: 'a simplified structuring of reality which presents supposedly significant features or relationships in a generalized form'. This reality is complex, and that complexity is reduced in the model. Not all aspects of reality can be included, and it is up to the modeler to decide which features or relationships are supposed to be significant. Models are thus a selective and subjective approximation. Naturally this selection of features and relationships is inspired by the intended functionality of the model and its application. The value of a model should therefore be considered within its application domain and with respect to its functionality.

Geospatial models then, are simplifications of processes spatially spread over the earth's surface, like population growth or groundwater flow. The prefix geo- indicates the topic being modeled, as opposed to spatial models of for example a human body or a molecular structure. Even within this range, a broad spectrum of models exist, ranging from qualitative to quantitative models and from physical processes to social phenomena. The topic being modeled with the Environment Explorer is land use.

Before the actual model is described, geospatial models are investigated a little closer, to put the Environment Explorer in a context. To classify models according to their functionality, a distinction can be made between predictive and explorative models (de Nijs, 2001; after Lowry 1964). Explorative models are used to generate possible or logical alternatives. Predictive models aim at explicitly making qualitative or quantitative predictions about the future. Both model types' result in assumed information about the future. The latter though aims at giving an explicit prediction, whereas the first only aims at giving a possible outcome. Though others (e.g. Brimicombe, 2003; after Chorley and Haggett, 1967) use somewhat different classifications for spatial models, they all agree that models have different functionalities, and therefore should be valued according to their supposed functionality. The Environment Explorer aims at being an explorative model. By that the results should be looked upon as possible directions for land use development.

2.2 Levels of spatial data modeling

Eventually digital spatial models are stored and structured physically by bits and bytes. Between the real world and this digital representation Molenaar (1998) distinguishes four levels of abstraction. The associated levels for data modeling are spatial modeling, conceptual modeling, logical modeling and physical modeling. The application domains of different disciplines in geographic information sciences are related to these levels, as shown in figure 2.1.

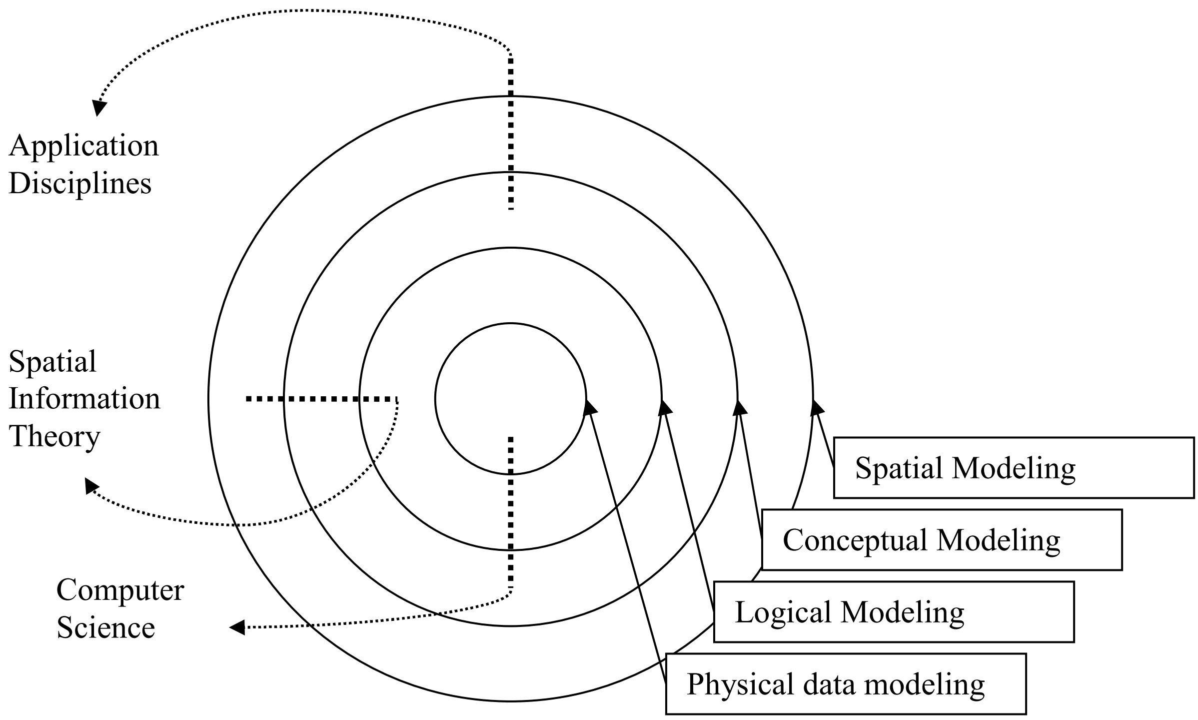


Figure 2.1 Levels of data modeling (after Molenaar, 1998, p7).

The spatial model formulates with parts of the real world are the topic of interest. It specifies the entities whose behavior and interrelationships are to be investigated. This spatial model is formulated in the context of an application domain. This phase is represented by the outer level of figure 2.1 above.

The first level of abstraction is the conceptual model. In conceptual modeling it is decided what aspects or objects of reality are topic of research, which thematic features are important and what relationships are among them. It will result in only a limited number of aspects or objects that are relevant in the application domain. In conceptual modeling it is also decided how this part of reality will be represented geometrically.

The next level of abstraction is the logical model, sometimes also referred to as the data structure of the model. The logical model indicates how the features of the conceptual model are organized and structured into tables, arrays or matrices. These data structures are dependent on to the software that is used for storage and data handling and only related with reality through this software. Examples are the hierarchical data model, the relational data model and more recently, the object oriented data model. The latter two, the most commonly used for geographical purposes, are described more extensively by Pilouk (1996).

The innermost circle finally represents the physical data modeling. At this level the data is physically stored on a medium by means of bits and bytes. This level is rather abstract and remote from a geographical way of thinking.

The logical and physical data modeling fall outside the scope of this thesis. Here the Environment Explorer will only be investigated at the more abstract levels of spatial modeling and conceptual modeling.

2.3 The representation of space

Representation of space as mentioned here deals with representation of spatial data. This data contains geometric information as well as thematic information at a specific moment, the 'what', 'where' and the 'when'. There are two general structures for linking thematic data to geometric data. These are the object based approach and the field based approach. In most applications the thematic aspects of the modeled space is of prime importance. Therefore the data processing and querying will be organized from a thematic perspective. The choice on how to structure the spatial data will thus primarily be based on the thematic data and its use, i.e. the purpose of the model, and only secondary on the geometric data.

The object based approach assumes that spatial objects can be defined as discrete entities that have a location, position and dimension as well as thematic characteristics. These geometric data and thematic data are connected by an identifier, see figure 2.2a. Space is viewed as a

container populated by these objects. Roads, buildings or land parcels are typical examples of entities that can be modeled using an object approach.
The field based approach, considers the real world as a spatial continuum. To attach thematic data, this continuum needs to be discretized in finite elements, cells. Thematic data again is attached to this cells, see figure 2.2b. For all modeled attributes a value is assigned to each cell. Typical entities modeled with the field approach are elevation or precipitation.

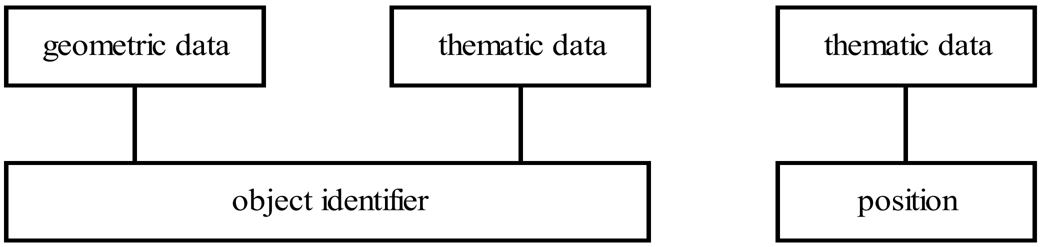


Figure 2.2: schematic representation of object and field based data structures.

Spatial data can be represented in two ways, the vector and the tessellation representation. The representation of space in the vector approach consists of points, lines and polygons. Space in the tessellation approach is represented as a space filling mesh, consisting of finite elements. These elements are usually regular squares (a raster or a grid), though other shapes do exist, like rectangles, hexagons or even triangular irregular networks (Laurini and Thompson, 1994). Since the micro model in the Environment Explorer uses a grid of regular squares, I will restrict myself to this.
This raster can be ordered in rows and columns. By numbering the rows (i) and columns (j) each grid cell can be identified by its position (i,j). These identifiers can be linked to coordinates by introducing a coordinate system. To introduce a coordinate system one needs to choose an origin, define two perpendicular axes and define the step size. The thematic data of an area will be directly expressed by the thematic attributes of the grid cells representing that area and are supposed to be representative for the whole cell, ignoring possible other values that also occur. A typical example is a land use data as being used in the Environment Explorer. Each cell will only represent the current major land use in a specific area, whereas other land uses that occur as well are not represented.
The topology of a raster is based upon the relationship between elements. Every grid cell has four full neighbors, and four diagonal neighbors. The full neighbors, also referred to as rook adjacent, together form a 4-neighborhood. This is also called a von Neumann neighborhood. Together with the four diagonal neighbors they form 8-neighborhood, also called a 3*3

Moore neighborhood. These neighbors are queen adjacent. Using this relations, windows can be specified. A window is the environment around a grid cell with a predefined size (see figure 2.X). This window is a subset of the whole raster on which a particular operation can be applied, like a moving average or a highest value. These operations applied on a window are called filters. The result of the operation is assigned to the central cell in the window. Both neighborhoods and moving window based structures are applied in the validation of the simulation results.

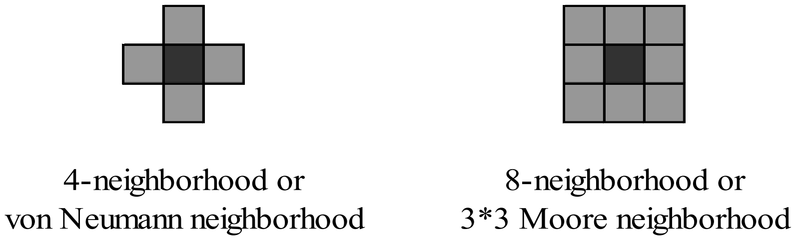


Figure 2.3: A 4-neighborhood and an 8-neighborhood.

This description of some basics of spatial modeling is far from a complete overview. Instead I introduced some of the main concepts as used in the Environment Explorer or the validation of its simulation results. Representation of space in this thesis for example is only considered with respect to 2 spatial dimensions, whereas 3 dimensional models do exist as well. For a more extensive overview of both spatial modeling in general as well as the object based and field based approaches see Laurini and Thompson (1994) or Molenaar (1998).

2.4 The representation of time

The spatial representation as discussed above is only capable of representing a static state situation. Some models however need to incorporate time as well, for example for representing events, states and episodes (Wachowicz, 2000). Modelling time in this context means modeling change; change in a features thematic aspects or its geometric aspect or in both.

An early attempt to visualize spatio-temporal information was Hagerstrand's Space-time cube. This static model is often seen as the start of time-space studies (Kraak, 2003). The earliest representation of time in GIS was the 'snapshot' or 'time slices' approach, were each layer represents a state at a certain point in time. Since snapshots represent states rather than changes, change is perceived by comparing different successive layers (Kraak and Ormeling, 2003). Recently attempts are made to fully integrate the spatial and the temporal dimensions.

This resulted in various attempts for event based data models, object oriented approaches and temporal databases. These models and databases focus on events instead of objects and therefore only register changes instead of states (Pang and Shi, 2002; Worboys, 2005).
The Environment Explorer computes land use change in discrete steps of one year. The result of this is a series of time slices or only the eventual situation, depending on users' preferences. This is done by means of a cellular automaton (CA). Other concepts of handling time are therefore not treated here. The next paragraph focuses on the implementation of time in spatio-temporal models by means of a CA. For a more conceptual discussion about the temporal dimension in spatial sciences see Wachowicz (1999) or Peuquet (2002).

2.5 Cellular Automata

The idea of cellular automata (CA) originates from the mathematicians John von Neuman and Stanislav Ulam who tried to express self reproducing systems mathematically. Their concept eventually found his way into many scientific disciplines like among others computer sciences, physics, ecology and geography. Probably the most well-known application is John Conway's the game of life, an attempt to simulate artificial life.
The objects or structure being modelled by cellular automata are in most cases described as complex systems, like for example non-linear chemical reactions, the growth of snowflakes and thermodynamics (Wolfram, 1994). Complex systems are systems that are "coherent in some recognizable way but whose elements, interactions, and dynamics generate structures admitting surprise and novelty which cannot be defined a priori" (Batty and Torrens, 2001). Geographic phenomena like urban growth as complex systems (Torrens and o'Sullivan, 2001), which makes them particularly suitable to model in a cellular automata.
Though cellular automata were not originally developed for geographic applications, they are widely used nowadays. It is especially the integration of cellular automata and GIS that extended the functionality of both (Wagner, 1997). This integration creates opportunities to simulate urban growth (Batty and Xie, 1994). More recently cellular automata are applied to larger areas as well, like in regional planning (White and Engelen, 1997) and in land use change models (Verburg et al., 2004). The use of CA in GIS however does not fulfil the strict requirements from the original CA. Instead a variety of focal functions are used which are referred to as 'relaxed CA' (Couclelis, 1997).

The Environment Explorer's micro model uses a CA for computing land use change over time. A CA is a discrete dynamic system, both in space and time. It basically exists of the

following elements: The space on which the automaton exists, represented as a lattice; the cell in which this automaton exists, which has a cell state; the neighborhood around the automaton that influences it and the transition rules that describe the influence of this neighborhood.
The lattice is the space on which the CA exists and evolves over time. Theoretically this space can have any number of dimensions. In practice the study area is a finite space usually restricted to two spatial dimensions. In most applications the lattices are filled with regular square cells, although other cell-shapes are possible as well, like triangles, hexagons or Thiessen polygons. Since the Environment Explorer uses regular square grid cells, I will restrict myself from now on to regular square cells.
Each cell in this lattice has a state that is only one of a finite number of states. In original CA's these were binary states. Either a cell was on, or it was off. The meaning of this on-off was depending on the application. The earlier land use models had the same options, they simulate urban growth representing cells that contain urban environment or not. More advanced land use simulations contain more different land uses, representing not only urban area, but for example nature, water or rural areas.
The neighborhood of a cell are the surrounding cells that have influence on it. This neighborhood can have any size and shape as long as it is uniform for all cells. The two basic types are the von Neumann and the Moore neighborhood. Since in land use applications a cell is usually being influenced by more then only its nearest neighbors, these neighborhoods are often bigger. Verburg et al. (2004b) indicates that the influence of neighboring cells is depending on both the land use type and the distance. In the Environment Explorer cells have a circular neighborhood with a diameter of 8 cells.
The transition rules eventually specify the behavior of cells over time, in discrete steps. In this way they are the equivalent of mathematics in rule based models or systems. These rules are fixed uniformly for all cells. A simple CA generally uses IF, THEN, ELSE statements with the predefined neighborhood as input (example from the Game of Life rules):

IF {two or three neighbors are alive}
THEN {target cell becomes alive}
ELSE {target cell becomes dead}

Originally those transition rules were decided more or less intuitively (Li and Yeh, 2004). More recently several techniques are applied to develop rules. These include the use of training data in neural networks (Li and Yeh, 2002), multi agent systems (Batty, 2003) or multi criteria evaluation (Wu, 1998).

In the Environment Explorer, first the amount of land use per class is computed in the macro model. Then these land uses are assigned to cells in order of decreasing potential per land use. The model uses four different land use classes: water, agriculture, forest & nature and urban. Of these only forest & nature and urban are dynamic and assigned as described above. Water is unable to change at all and agricultural land use is assigned to the remaining cells. Besides the neighboring land uses, the potential of a cell for a certain state is also dependent on the presence of a transport network and the size of a random parameter. The total potential formula for each cell is described by equation [1] below

$$P = v * R * A \qquad \text{Equation [1],}$$

Where P is the transition potential for a single cell at a single time for a specific land use, R is the contribution of the CA rules to the potential, v is a stochastic perturbation term, and A is the contribution of the accessibility to the potential. The stochastic perturbation term v is computed according to equation [2]

$$v = 1 + (-\log(random))^{alpha} \qquad \text{Equation [2],}$$

Where *random* is a uniform random function between 0 and 1 and *alpha* is the stochastic noise parameter (Engelen et al., 2002).

3 MODEL VALIDATION

3.1 The model development cycle

The development of a model is a process with several distinct phases. Though handbooks on this subject use somewhat different classifications and names, they generally identify the same phases. These are the setup phase, the conceptualization phase, the model construction phase, the evaluation phase, the use phase, and finally the reporting phase (van Waveren, 1999, Middlemis, 2000). Though modeling in this thesis refers to computer based Land Use Cover Change (LUCC) simulation models, many aspects are applicable to other models as well. Therefore not only principles from the LUCC modeling, but also ideas from other domains like ecology or computer simulation modeling are used.

The development of a model starts with a description of the problem and its context. From there the objectives and requirements of the model can be derived. By that the purpose and the intended functionality are defined. These steps together are the set-up phase. Already in this stage the requirements are defined by which the model will be evaluated eventually.

After it is agreed upon what functionality a model should have, the building of the conceptual model starts. This results in a model structure where the processes and the relationships are defined. It includes also the investigation of data availability and a description of the system. Based on this assumptions and simplifications can be made. Eventually the choice on which representation and data structures are used is being made in this phase.

Now the real construction of the model can start. This includes both the building of the (physical) model, and its calibration. Calibration is the estimation and adjustment of model parameters and constants. This means essentially making the model fit as good as possible with the data available.

The model is now ready for use, but the value of the results is not yet determined. That happens in the evaluation phase. This evaluation should be done with the functionality and objectives as determined in the set-up of the model in mind. Passing the evaluation in this context essentially means that a model has an acceptable fit for a desired level of utility. The validation is part of, but not equal to this evaluation of the model.

Finally, the model is ready for use and can be described in a report. It should be noted however that modeling is not a linear process, but iterative instead. Every step as described above can result in a revision of an earlier step. If the evaluation test fails for example,

calibration might need to be improved. The whole process including feedback is shown in appendix C.
In the next part of this chapter the evaluation phase be investigated more thoroughly, and some terms will be defined (3.2). From this evaluation phase again, the validation of the model results (3.3) and specific problems for LUCC simulation models (3.4) are treated further. The actual methods to perform this validation are described in the methodology chapter (4).

3.2 Model evaluation

In the above a short outline of the development of a model is given. This thesis focuses on the validation of a model, which is a part of the evaluation phase. In literature however it is not generally agreed upon what exactly belongs to the evaluation of a model. An important issue here is the terminology used in this discussion (Refsgaard and Henriksen, 2004). Therefore I will first give an overview of different components of the evaluation phase and their meaning in this thesis' context. For example validation itself is often confused with or treated similar to calibration as well as to verification.

First of all the term calibration sometimes appears when discussing model evaluation. This however is not a part of model evaluation, and should be done prior to and independent of the evaluation. Calibration is the procedure in which model parameters are optimized to get the best output result. The evaluation phase again investigates how well this optimized model is eventually.
Verification and validation then are often treated as synonyms in everyday language, though in a modeling context a differentiation between them is necessary. Verification is an establishment of truth. To verify a model would thus mean to demonstrate its correctness. This verification can be mechanical or logical (Rykiel, 1996). The logical verification, referring to a conceptual model, would mean that the conceptual model is a correct representation of the real world. This is impossible, both from a philosophical and a practical point of view (Oreskes et al., 1994). Verifying the mechanics in the computer model means to demonstrate that a model is correctly implemented. This is therefore quite different from logical verification and this is possible.
Validation itself is an investigation of the quality of the model or its results. Passing the validation is therefore a demonstration that a model within its domain of applicability possesses a satisfactory range of accuracy. Thus it is acceptable for use (Sargent, 2001). It

does not mean it represents truth, neither is it the best possible model. Kok et al. (2001) come to the conclusion that most land use models lack proper validation, for several reasons. The necessity of proper validation however is indicated by several authors (Power, 1993; Pontius and Schneider, 2001; Pontius et al., 2004).
This verification and validation together is what is referred to as model evaluation in this thesis. The original (non-simplified) Environment Explorer is evaluated quite thoroughly (Engelen et al., 2005).

While evaluation is a part of the model development cycle, distinct parts of the modeling process can again be evaluated (Caughlin, 2000). Sargent (1998) classifies four model evaluation activities, which are connected to a part of this modeling cycle (see figure 3.2 below). These are: data validation, conceptual validation, computer model verification and operational validation.

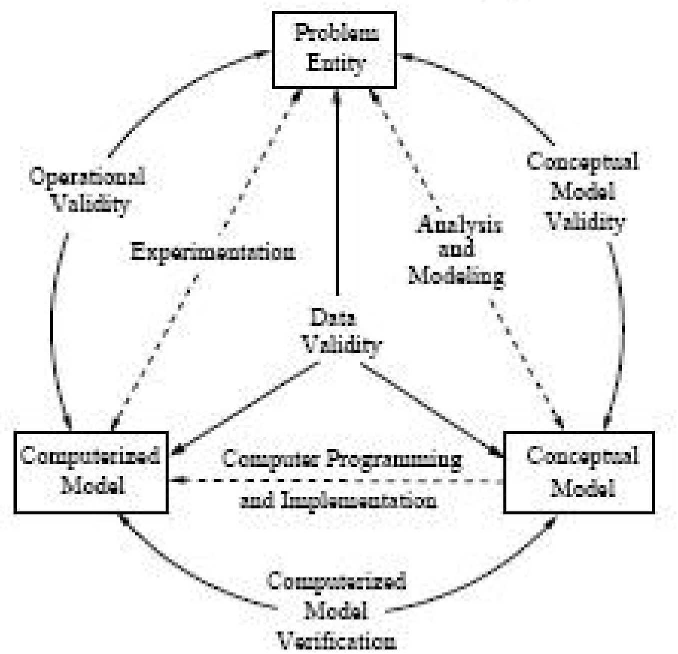


Figure 3.1: Evaluation of modeling phases (after Sargent, 1998).

Spatial data are generalizations of the real world, as are models. This does mean that data never represent the real system accurately. Although data validation is usually not considered to be part of the model validation, it is discussed here, because it is important to be aware of. Data is used both in the construction and calibration as well as the evaluation of the model. The model and its result will never be more accurate then the data used as input. Therefore it is important to know the accuracy and quality of the data used for evaluating a model. Besides the accuracy it is also the interpretation of the data that is important. The same is true for data manipulation, since every data handling step can cause inaccuracies. Although the validation of data is not treated any further in this thesis, it will be considered once more for the data used in the case study.

In the phase of conceptual modeling assumptions are made and relations are defined or neglected as well as selections of what is relevant and what not in that specific model. The result, the conceptual model, will be a representation of the real world with these assumptions and simplifications. Validating this means determinating whether these assumptions and used theories are acceptable. It also means determining that the structure, logic and relationships used are reasonable for its intended functionality (Sargent, 1998).

The assumptions and theories mentioned refer to linearity, independence, stationarity and the like. These properties can be tested with statistical analysis. Two methods to validate the structure logics and relationships are face validation and the use of traces through the model. Face validation requires experts to examine the flowchart or a graphical representation of the model. The use of traces is the tracking of entities through the model to determine correct logics

In conceptual modeling also the necessary simplifications are assumed. Occam's razor tells us that "entities are not to be multiplied without necessity", or that if two models have equal predictive power, the simpler is preferred. Einstein stated about the same subject that "models should be as simple as possible, but no simpler" (Clarke, 2003). Therefore the level of simplicity is referred to as the Occam- or Occam-Einstein dimension (Brimicombe, 2003). Although it is obvious that a model should not oversimplify or at the other side be too complex, it is hard to determine what exactly is simple enough.

Since the decision about the representation of space and time is taken in the conceptual modeling phase, it should be evaluated here as well. Spatial and temporal resolutions are related to each other as well as to the phenomena being modeled. Events affecting large areas generally take more time then events effecting smaller areas (Delcourt and Delcourt, 1988). As well Kavouras (2001) states that the temporal as well as the spatial resolution are related to the temporal and spatial extend of a model. The larger the total time span of a model, the larger the temporal resolution. The same goes for the spatial resolution.

Still, no clear guidelines or rules-of-thumb are present. Just like the Occam-Einstein dimension for simplification, resolutions need not to be unnecessarily high. LUCC models predicting land use on a daily basis for each square meter for the next 50 years in the

Netherlands is obviously nonsense. In practice spatial as well as temporal resolution is usually a trade-off between data availability, available computer power, desired accuracy, the entities being modeled and of course finances.
The only part of a model that can be verified is the implementation of the conceptual model in a computer program. Verification in this context only means that the mathematical expression of the conceptual model is correct in a technical sense. It is thus a technical matter about how faithful and accurate ideas are translated into computer code or mathematical formulas (Sargent, 1998). For relatively large models this is of course extremely difficult to verify that a model implementation is entirely correct (Rykiel, 1996).

Operational validation or result validation finally validates whether the result of the whole model meets the pre-set accuracy. This validation is primarily concerned with how well the model simulates the real world representation. Since models are only calibrated for a specific place or time span, this validation is site and time specific as well (Refsgaard and Henriksen, 2004). Often statistical tests are used here and comparisons between real and simulated data. Correct model results however do not guarantee correct mechanisms. It is very well possible to generate good results without, as is the case in neural networks.
Operation validity tests should be independent of the model construction and calibration (Refsgaard and Henriksen, 2004; Pontius and Malanson, 2005). An appropriate method for that is called data splitting. In this method the available data is split in two independent parts. The first part of the data is then used to calibrate the model, whereas the second part can be used afterwards to validate the calibrated model. Usually the split is temporal, spatial or both, meaning the model is calibrated and validated over different (but comparable) periods in time, or areas in space or both.

3.3 Operational validation

As mentioned above, two criteria are commonly accepted in operational validation. The first is that the requirements to pass should be pre-set and defined in the set-up phase of the project already. These requirements are specified for the intended functionality of the model. The second prerequisite for good validation is independency between the construction and calibration at the one side and validation at the other. This independency includes especially the data being used, though some even argue that managerial and financial independency are required as well (Arthur and Nance 2000).

Pontius et al. (2004) propose four standard components of validating LUCC models: budgets of sources of errors, comparison against a null model, comparison against a random model and analysis at different scales. By these tests it can be checked whether a simulation model has any forecasting power.

Budgets of sources of error refer to the cause of errors in these simulation models. For example a low agreement between two land use maps can be caused by a wrong number of pixels per class, the wrong location mechanism or both (Pontius, 2002). In this thesis' application the number of pixels is predefined. By that the errors in the Environment Explorer are mainly due to the location mechanism.

A null model is a model that generates no change at all, both in location and in quantity. For validation both the null model and the simulation results can be compared with the real situation. At least simulation models should perform better then the null model to have any predictive power.

A second validation comparison is that of the simulation against a random model. This random model predicts the same amount of land use change, but it has no spatial preference for locating this change. There are several random models possible. The random model used in this thesis is the random constraint match, which is described in chapter 4. Again, to have any predictive power the model should have better results that this random model.

Finally Pontius et al. (2004) mention that the model should be tested at multiple (spatial) resolutions. They show that the budgets of sources of error as well as agreement due to chance change with the resolution. A conclusion that is shared in Jantz and Goetz (2005) and Pontius and Malandson (2005) as they conclude that spatial resolution is of significant influence in simulation model results. Kok and Veldkamp (2001) however come to the conclusion that not spatial resolution, but the spatial extent is of greater influence. Due to time limitations this validation is not performed in this thesis.

3.4 Difficulties in validating LUCC models

When evaluating LUCC models some specific problems appear that make this not a very straightforward procedure. The first problem is that much land use changes are very place and time specific. Local situations and policies can have great influence on spatial development, whereas they are hard to capture in rules and functions. A related aspect is time dependency. Government policies can change rapidly in a few years, which make it hard to model locational as well as quantitative changes over longer periods.

A second problem is that LUCC models usually have nominal classes, sometimes combined with ordinal classes like sparse and dense residential. Much validation methods used in other disciplines, like hydrology (i.e. Andersen and Bates, 2001) or Ecology (i.e. Power, 1993) assume quantitative results. This implies that more standard statistics can be applied to them. An additional difficulty is that some models use a hierarchical classification, where superclasses like agriculture have several subclasses, like arable land, greenhouses or horticulture. In LUCC this is therefore less straightforward, and alternative statistical tests need to be performed. Several solutions for this, most of them based on a contingency table, are proposed (for example Pontius and Schneider, 2001; Pontius et al, 2001). Those that are used in this thesis are described in chapter 4.

Finally, LUCC simulation models, and the Environment Explorer in particular, do not aim at predicting one future situation, instead, they aim at exploring possible directions of development (de Nijs et al., 2001). Results should therefore be seen as possibilities. This makes a crisp pixel by pixel comparison less logical. Therefore methods that take this consideration into account are needed to validate exploratory models.

3.5 A note on philosophy

So far I investigated first what model evaluation and eventually model validation is. This is logically followed by a description how this is performed in chapter 4. Both are described from a rather practical point of view. More philosophically however it can be questioned whether LUCC models can be validated at all.

Oreskes et al. (1994) states that models, in particular numerical models of natural systems, cannot be validated or verified by definition. The main argument they use is that natural systems are never closed, so the input data is inherently incomplete. She also mentions the scaling up of non additive properties, assumption ladenness of data and underdetermination as reasons that there is no logical proof a model is true. From her point of view a model can only be partially confirmed by demonstrating agreement between observation and prediction. Due to this problem model results will always be topic of question, and their value is mainly heuristic.

Oreskes' view on validation and verification is too strict according to Suppe (1998). He agrees that it is not possible to logically prove that a model is true, but that rigorous view would make all empirical knowledge impossible. He states that simulation models can be validated in some way and therefore some information about the real world can be obtained.

He also indicates that the other objections Oreskes et al. mention are true but should not be overestimated.

This was investigated once more by Rykiel (1996) specifically for (ecological) simulation models. Since their aim is to predict or forecast some future situation the problem arises that there are no observations to compare model results with. The future has not yet arrived. He concludes that instead of verifying the truth, validation should be approached as confirming acceptance for its intended use. Power (1993) states that indeed a model can never be fully validated, but each validation test passed will increase confidence in the model and its result. This will lead to the conclusion that models can only be not-invalidated, instead of validated.

Without going any deeper into philosophical backgrounds of science or epistemology in particular, this thesis will follow Suppe's line of thought. Validation is therefore not intended as verification of truth, but as a confirmation of reasonable outcomes instead. This is in line with the aim of the model, which is to explore possible or reasonable land use cover, instead of forecasting some truth.

4 METHODOLOGY FOR OPERATIONAL VALIDATION

In this thesis the Environment Explorer is validated with the use of historic data. The model was calibrated for the province of Noord-Holland, the Netherlands, for the period from 1960 to 1980. With this calibrated model the land use of 2000 was simulated starting from 1980. That result was validated by comparing it to the available land use data from 2000. The land use data used for this is the HGN SE dataset as described below.

In this chapter first the data (4.1) and the calibration of the model (4.2) are described. After that the methods applied to perform the operational validation are treated (4.3 to 4.4).

4.1 The HGN SE datasets

The data used for the calibration and validation is derived from the historic land use dataset of the Netherlands (HGN). This dataset comprises a series of maps containing the land use in the province Noord-Holland, the Netherlands for the years 1850, 1900, 1930, 1960, 1980 and 2000 in a 50*50 meter resolution.

These maps are derived from old topographic maps that were scanned and reclassified in 10 land use classes. These years however are not exact. The data is derived from several topographic maps originating from a period around this specific year. Validation of this data showed that the classifications have accuracy between 95% and 98% (Knol et al. 2003). An exception to this is the 2000 land use map, which is derived from the LGN4 dataset which is again derived from remote sensing imagery (personal communication with Henk Kramer).

This difference in source and hence classification methods will cause noise in the comparisons of simulation results. This noise is originating from different classification methods, instead of actual land use change. Since all simulations start with the same land use map (the 1980 map) and all results are compared to the same land use map (the 2000 map) this noise will be the same in all comparisons and no systematic errors will appear.

For the actual calibration only the 1960 and 1980 data was used. The reason for this is that land use changes are more or less consistent in the period from 1960 to present, w1hereas it was not compared to earlier periods (personal communication with Ton de Nijs). For this thesis the data was reclassified to a 100*100 meter resolution. This reclassification is done by applying a majority aggregation on the original HGN data. This reclassification does have effect on the maps, since single pixels patches disappear. This especially affects the class 'urban land use' where single pixels represent a few houses in between agricultural land.

These new datasets are called HGNSE_YYYY. Where SE indicates 'Special Edition' since the data was reclassified for this purpose to a 100 meter resolution. The YYYY indicates the year the data represents. So HGNSE_1980 is the land use data for the province of Noord-Holland in the year 1980 on a 100 meter resolution.
A second, thematic reclassification was applied to this HGNSE datasets to reduce the number of classes. The original data contained ten land use classes, among which two rural classes and five different nature types. This was reduced to the following four classes: rural areas, nature, urban areas and water. The reason for this reclassification is twofold. At first it results in a simpler model, with only a limited number of transition rules. Second, this simplified version of the Environment Explorer does not aim to simulating these specific land use classes. For example the change in specific nature types as used in the HGN data, is not cause by its neighboring land use or accessibility, but by soil characteristics instead (personal communication with Ton de Nijs). The result of this reclassification is showed in appendix A.

4.2 Model calibration

For this thesis the model was adjusted to be used with the HGN SE data as described above. This means that as opposed to the original Environment Explorer, the model uses a 100 meter cell size and the macro model has only one region. For the calibration of this model the land use maps of 1960 and 1980 as well as some information about the transport networks was available.
The Environment Explorer works with dynamic land uses and features. Features are land uses that do not change during the simulation. In this model water is the only feature. The other three land uses are dynamic, meaning that they can increase or decrease at each other's cost. In this case the urban and nature land use classes are truly dynamic, whereas the rural land use is a vacant class. This means that the increase or displacement of urban or nature pixels are computed and located first and rural land use fills the remaining pixels.
To calibrate the model, three factors can be adjusted. These are the influence of the different land uses on the transition rules, the influence of the transport network on the different land uses and the total potential formula. By this the 1980 land use map was simulated starting with the 1960 land use map.

The influence from the different land uses on each other is represented by the transition rules. A transition rule is defined by a spline indicating the attraction of one land use on the other as a function of the distance. This spline is defined by points indicating the attraction between two land uses at a certain distance. Intermediate points are interpolated. The range of influence is fixed at 8 cells, were it equals zero. The influence table keeps track of the transition rules from each land use on itself and the other land uses. In the Environment Explorer these transition rules can be calibrated automatically.

The second thing that can be adjusted for calibration is the attraction of the transport networks on different land uses. Three networks are implemented: roads, railroads and waterways. Together with these networks some nodes are introduced: highway junctions, as well as intercity and normal railway stations. In this model it is assumed the transportation network only influences urban land use. For the other land uses the attraction is equal to 0. The attraction of these networks and nodes is defined by a weighting factor and a halving distance, in cells.

Finally the algorithm to compute the transition potential is adjustable. You can either choose to make a new one specifically for this model, or use and adjust an existing rule. For this application the algorithm of the original Environment Explorer model was used, as stated in chapter 2. The parameter alpha that indicates the amount of randomness in the simulation is the actual parameter that is calibrated for this application.

4.3 Model simulations

Starting with the HGNSE_1980 land use map the following simulations are performed: a null simulation, four random simulations and simulations with the calibrated models.

The null simulation is the simulation where there is no change at all. The result of this simulation is equal to the starting situation, HGNSE_1980.

The random simulations are the simulations where the location of land use change is located randomly in the map, whereas the amount of change is equal to the calibrated models. As a random model the random constraint match (RCM) is used from a tool called the Map Comparison Kit (MCK). This tool is described by Hagen-Zanker et al. (2005).

The random constraint match is a method that computes the amount of change between two maps and locates them randomly in the reference map of the two. In this case the amount of change between the land use classes of HGNSE_1980 and HGNSE_2000 was computed and located randomly on the HGNSE_1980 map. Because this was not performed within the Environment Explorer itself, water is no longer a feature in the RCM. Since the amount and location of water differs between the two maps, some random pixels are located at sea. The RCM results are presented in appendix A. The results of the RCM are further referred to as rcm1, rcm2, rcm3 and rcm4

The calibrated simulations finally are the simulation runs from the two calibrated models. The result maps are sim_au and sim_hm, with au indicating 'automated' calibration and hm 'handmade' calibrated

4.4 Map comparison methods

Intuitively the human eye can very well compare maps and identify similarities and dissimilarities. However this 'visual map comparison' is highly subjective. Results differ from person to person and even from time to time. Therefore the need for objective and repeatable measurements of map similarity is widely acknowledged (Winter 2000; Power et al. 2001; Brown et al. 2005).

Recently various different methods for automated map comparison are developed. A map in this discussion is a 2 dimensional grid of cells containing nominal data. These methods compare maps with respect to different criteria. These criteria are predefined and not adjustable to a specific application. The reason they are applied anyhow is their objectivity as well as their repeatability (Hagen 2003).

In this thesis several methods are used, that compare maps with respect to both thematic aspects and geometric characteristics. The thematic aspects are compared with Kappa statistics and a variation on this, Fuzzy Kappa statistics. The latter takes into account fuzziness with respect to location only. The geometric characteristics are compared by looking at several landscape metrics. These metrics include diversity, patch shape complexity and patch area-perimeter ratio. All computations are executed by the 'Map Comparison Kit' (Hagen-Zanker et al., 2005b).

4.4.1 Kappa statistics

The most straightforward way of comparing two maps is by assessing similarity pixel by pixel. Every pixel is either equal to its counterpart on the other map or not. The goodness of fit is then expressed in the percentage of agreement, which is the number of similar pixels divided by the total number of pixels. The visual result of this is a map containing two categories: equal and unequal. An example of this is given in figure 4.1 below. The figure shows two maps being dissimilar in some way and the third shows the result map of this comparison.

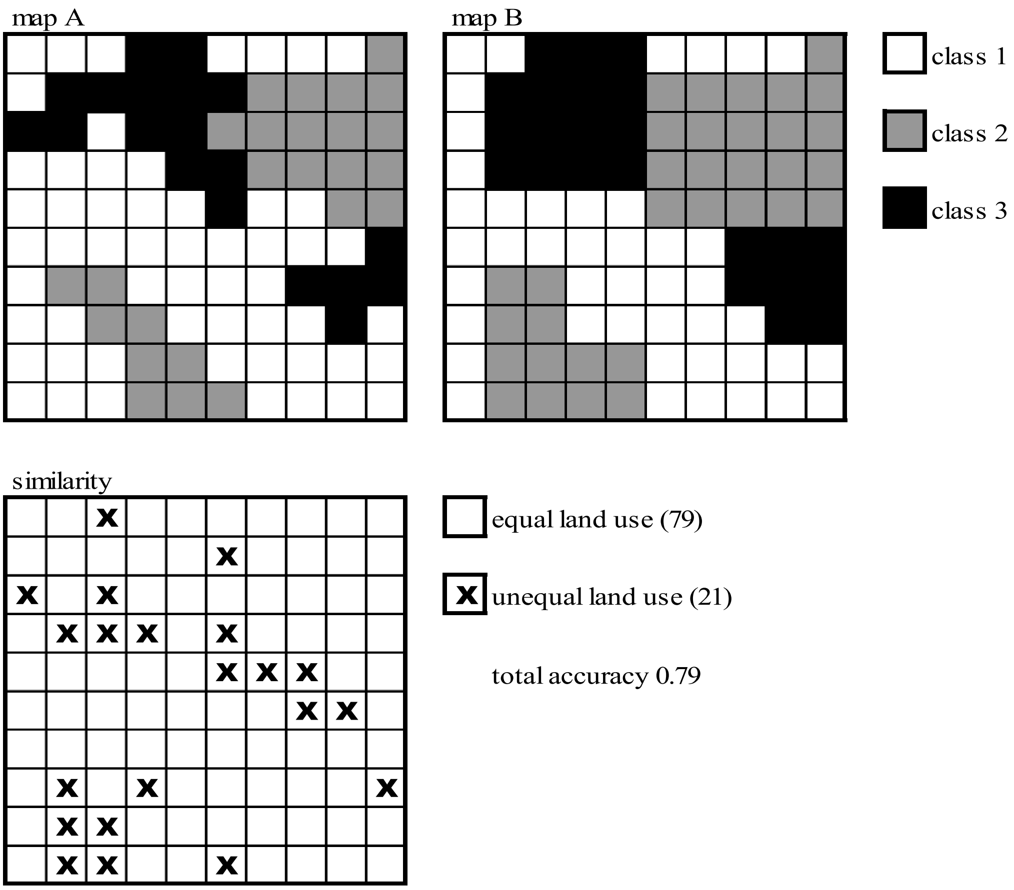


Figure 4.1: Maps A and B show two classification results with the same land use classes. Map C shows which pixels are classified similar and which not.

This comparison can be enhanced by investigating each land use class separately. By comparing two maps per class, a contingency table or error matrix can be calculated. Table 4.1 below shows this schematically. Table 4.2 is the numeric result of this example.

Table 4.1: Schematic example of a contingency table (after Bishop 1975, p394)

	Map B				
Map A	Class 1	Class 2	..	Class I	Totals
Class 1	x_{11}	x_{12}	..	x_{1I}	x_{1+}
Class 2	x_{21}			x_{2I}	x_{2+}
.	.			.	.
.	.			.	.
Class I	x_{I1}	x_{I2}	..	x_{II}	x_{I+}
Totals	x_{+1}	x_{+2}	..	x_{+I}	N

Table 4.2: a numeric example of a contingency table derived from the comparison between maps A and B in figure 4.1.

	Map B			
Map A	Class 1	Class 2	Class 3	Totals
Class 1	41	8	7	56
Class 2	2	23	0	25
Class 3	1	3	15	19
Totals	44	34	22	100

From the obtained contingency table some statistics can be calculated. These are the observed fraction of agreement, and the expected fraction of agreement. These can be calculated according to equations [3] and [4] below respectively.

$$P(O) = \sum_{i=1}^{c} x_{ii} / N$$ Equation [3]

$$P(E) = \sum_{i=1}^{c} x_{+i} * x_{+i} / N^2$$ Equation [4]

Here P(O) is the observed fraction of agreement, P(E) the expected fraction of agreement, i the row or column number and c the total number of columns. N is the total number of observations, x_{ii} is the number of observations in row i, column i and x_{i+} and x_{+i} are the totals of row i and column i respectively as is shown in table 4.2.

To compare results more easily, it is preferable to have the map similarity expressed in one single number. This can be done with Kappa statistic (Congalton, 1991), which is calculated according to equation [5]. The Kappa statistic actually computes the goodness of fit as a fraction of agreement corrected for the expected fraction of agreement.

$$Kappa = \frac{P(O) - P(E)}{1 - P(E)}$$ Equation [5]

Since this Kappa is standardized, it can be used to compare similarities between maps. Values for Kappa are between -1 and 1. Kappa greater then 0 indicates more similarity than random, a Kappa smaller than 0 indicates less similarity then random. However, in order to compare similarities an equal number of classes are needed. The comparison between maps A and B in figure 4.1 results in a Kappa value of 0.665. The formulas are presented in Bishop et al. (1975) and introduced to the field of geo-information by Congalton et al. (1983).

4.4.2 Fuzzy Kappa statistics

The Kappa statistics and its adjustments as described above have some drawbacks. For example small displacements are characterized as unequal while intuitive they can be quite alike. For this kind of unsure or vague information fuzzy set theory was developed. The usefulness of fuzzy map comparison methods is already indicated by several authors and in several disciplines (Fritz and See, 2005; Wealands et al., 2005). Reasons they mention are among others the appearance of mixed pixels (Power et al., 2001), classification inaccuracies (Foody, 2002) and the fact that geographic entity boundaries and classes are inherently fuzzy (Fisher, 2000).

Fuzziness in maps occurs with respect to both classification categories and location. Fuzziness due to classification categories implies that some categories are more similar then others. For example in hierarchical classes; horticulture is more similar to other agricultural land uses then it is to residential zones. It also occurs when ordinal scales are used; dense and light residential are more similar to each other then they are to nature. Fuzziness of location means that a category that in a map is positioned at a specific location can be interpreted as being somewhere in the proximity of that location.

The full derivation of Fuzzy Kappa is given in Hagen-Zanker et al. (2005). Here only a short introduction is given for fuzziness with respect to location. Fuzzy classification is not considered in this thesis.

Fuzzy similarity with respect to location for a pixel on two maps is depending on the distance. It is expressed as a value between 0 and 1. For example if the pixel has value 'nature' and exactly the same pixel on the other map has a value 'nature' as well, its similarity is one. If the pixel 'nature' however is displaced by one cell, its value is less then one, and can be calculated according to some distance decay function. This will result in the 'one-way' similarity between two maps.

Once this fuzzy similarity is calculated twice (map A compared to map B as well as map B compared to map A) the two way similarity can be calculated. The eventual similarity between two pixels is the lower of two one-way comparisons. Two-way comparison prevents two pixels being similar only because their neighbors are similar. The overall map similarity P(O) can be computed as an average of pixel similarity over the whole map. The calculation of the expected similarity differs also from the normal Kappa and is as well described in Hagen-Zanker et al. (2005). The final statistic to compute is the Fuzzy Kappa, or Kfuzzy. This can be calculated similar to the normal Kappa by equation [5] and does not involve fuzzy set theory. An example computation is shown in figure 4.2. The one-way and two-way similarity is computed for the centre pixel. Pixels displaced by one cell are counted as half similar.

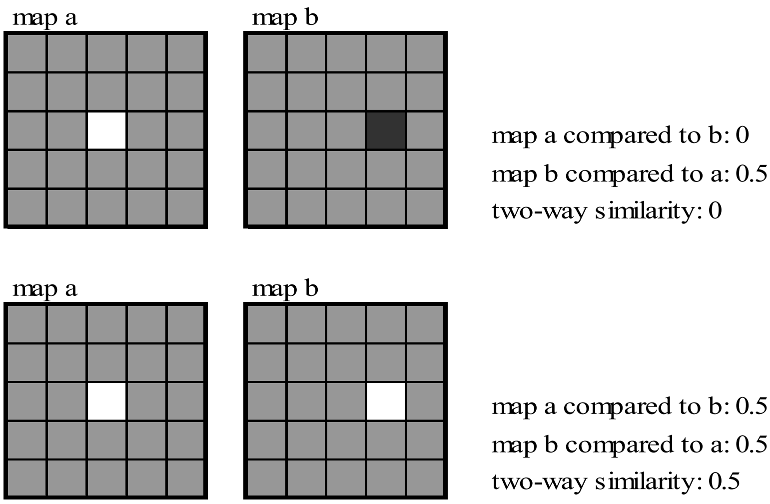


Figure 4.2 Example of fuzzy similarity with respect to location

4.4.3 Landscape metrics

Kappa statistics and its varieties as described above deal with the thematic aspects of data, like land use. It lacks however the notion of geometry that land use maps do contain (Herzog and Lausch, 2001). Some attempts that also consider geometry are the use of landscape metrics (Barredo and Demichelli, 2003) and a calculation of the diversity in a pixels' neighborhood (Verburg et al., 2003). In this thesis' approach a comparison with the use of landscape metrics is presented as well. As opposed to the method used by Barredo and Demichelli (2003) here the notion of location is considered since metrics are compared focally instead of globally.

Landscape metrics originate from and are widely used in landscape ecology applications (for example McGarigal and Marks, 1994). They are measurements of patch geometry, shape, diversity, complexity et cetera. An application landscape metrics in raster maps is given in Brown et al. (2005). An application in an economic driven land use change model is presented by Parker and Meretsky (2004).

Among the possible metrics there is much redundancy. Riiters et al. (1995) found that no more then six univariate metrics explain about 87% of the characteristics. This result was

confirmed by Cain et al. (1997) when tested at different resolutions, though others (Giles and Trani, 1999; Remmel and Csillag, 2003) found different results. Five of them are: patch compaction, image texture, patch shape, patch perimeter-area scaling and the number of classes. The sixth, ‘large patch area density scaling’ was actually only useful in one specific case. It was noted however that it is unsure whether these metrics also indicate changes well.
Since the number of classes in this thesis’ case study is no greater then four, the metric ‘number of classes’ was not taken into account. These are patch perimeter-area ratio (PA), Shannon’s diversity index (SHDI), shape index normalized to a square with the same perimeter (Shape) and fractal dimension (D) respectively. The correlation of these four metrics is shown in table 4.3. Absolute values > 0.27 are significantly correlated and therefore underlined. Computation of significantly correlated metrics means the results do not reveal entirely different information.

Table 4.3: Correlation between landscape metrics (after Ritters, 1995).

SHDI	1	.	.	.
D	<u>-0.47</u>	1	.	.
ShapeIndex	<u>-0.32</u>	<u>-0.47</u>	1	.
P/A ratio	0.04	<u>0.51</u>	-0.18	1
	SHDI	D	ShapeIndex	P/A ratio

Units used for computing are cells and cell-units. Since all maps used in this thesis use the same resolution, the cell sizes are equal and this is no problem for similarity comparison. The Patch perimeter-area ratio can simply be calculated by dividing the patch perimeter by the patch area as in equation [6].

$$PA = \frac{p}{a} \qquad \text{Equation [6]}$$

Here *p* is the perimeter of the patch, and *a* the area. Bregt and Wopereis (1992) found that this was a good metric for measuring map complexity.
Shannon’s diversity index is a measure that originates from information theory and is also in use in landscape ecology. It is combination of the relative proportion of each patch type (evenness) and the number of different patch types present (richness). The result ranges from 0 to infinity, where 0 is the value for only one patch comprising the whole map, and increasing value for more patch types or a more even distribution of patches. The index is

found to be more sensitive to richness then it is to evenness (McGarigal and Marks, 1994). Shannon's diversity index is given by equation [7]:

$$SHDI = -\sum_{i=1}^{m} P_i \ln P_i \qquad \text{Equation [7]}$$

Here m is the number of patch types and P_i is the proportion of the landscape occupied by each patch type i (Read and Lam, 2002).

Shape index is another measurement of the complexity of patch shapes. It represents the complexity of a patch compared to a square (McGarigal and Marks, 1994). A square has a value of 1 in this index. More complex shapes have increasing values. It can be computed according to equation [8].

$$Shape = \frac{p/4}{\sqrt{a}} \qquad \text{Equation [8]}$$

Here p is again the patch perimeter and a the patch area.

Fractal dimension (D) finally is a third measurement for the complexity of shapes. Small values indicate a more simple shape; large values indicate a more convoluted shape, like a lint. The results range is between 1 and 2. The fractal dimension can be calculated according to equation [9].

$$D = \frac{\ln \frac{p}{4}}{\ln \sqrt{a}} \qquad \text{Equation [9]}$$

Here p is the patch perimeter and a the patch area (Hagen-Zanker et al. 2005b).

Except Shannon's diversity index, these metrics can be calculated per patch. These calculated values will then be assigned to the pixels belonging to this patch. By this a weighted average can be computed as the overall mean. For Shannon's diversity index it is necessary to define the neighborhood for which it is calculated. For all metrics this will result in a map with numerical values for each geometric characteristic. All metrics will suffer from edge effects were the boundary of a map does not coincide with the boundary of a patch.

As for the Kappa statistics these results are computed for the calibrated, the null and the random simulations. Results are again compared to the values obtained from the HGNSE_2000 map. An example, based on the maps from figure 4, is given in figure 4.2. First the patches are shown. From each map one patch was chosen to calculate the landscape metrics used in this thesis' case study. The SHDI values in this map are computed for the map as a whole.

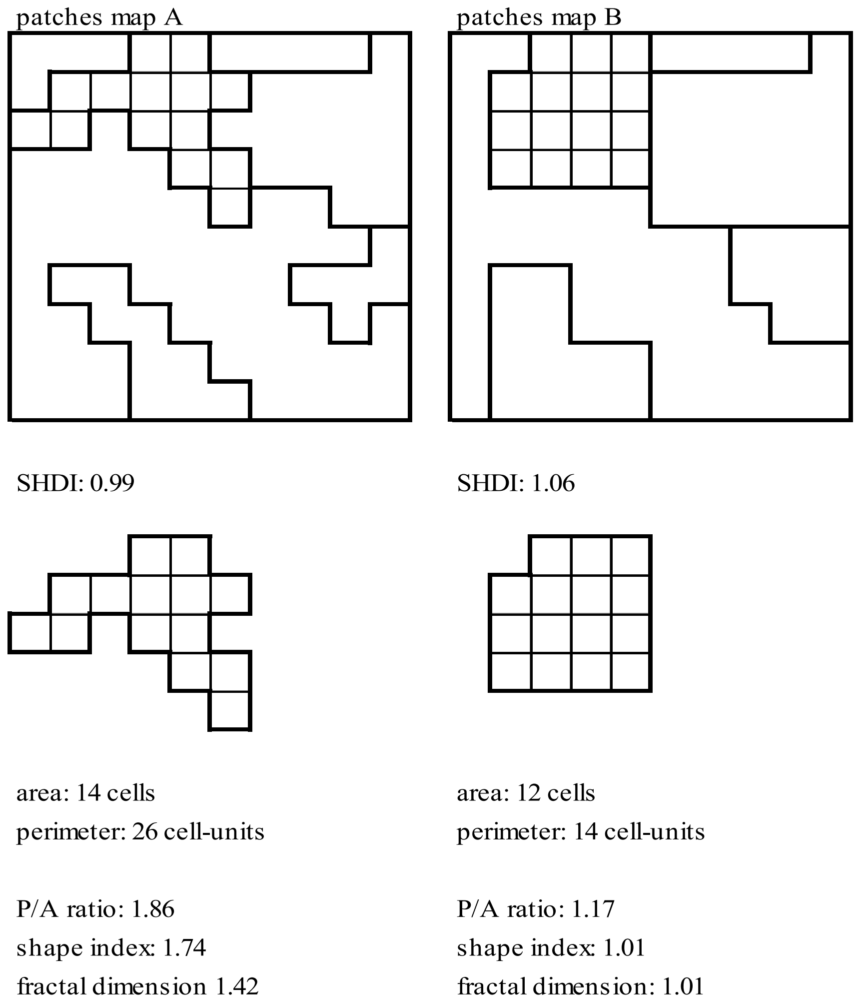


Figure 4.3: numeric examples of landscape metrics

Two ways of comparing these maps are used here. At first the global average value for each landscape metric is computed. The difference of that value with the global average of the HGNSE_2000 metrics is an indication of map similarity.

Global average however does not consider local similarities and dissimilarities. Therefore a second comparison was computed, based on a moving average. In this method a moving average was applied on the numeric values. After that the total map similarity can be computed by taking the global averages of the absolute differences per moving average. Since Shannon's diversity index is already computed for an area instead of a single pixel, no moving average was applied to this metric.

5 RESULTS AND DISCUSSION

5.1 Calibration results

Two models were calibrated for simulation. First, cal_au was calibrated entirely to reach the highest Kappa value. Second, cal_hm, was adjusted to look more realistic. This more realistic appearance was not measured in a way, and therefore highly subjective. Both models were calibrated to simulate HGNSE_1980 from the HGNSE_1960 map. For this calibration the amount of pixels with urban or natural land use were predefined, so errors in these classes are the result of misallocations.

The cal_au simulation was created by first adjusting the attraction of the transport networks by hand. This was repeated until the simulation resulted in clusters of urban land use around this transport network of about the right proportion as compared to the HGNSE_1980 result map. After that the automatic calibration adjusted the attraction of different land uses on each other such that the Kappa was highest. The adjustment of the transport network and automatic calibration of the transition rules was iterated a few times. Finally the random parameter alpha was adjusted so that the result appeared as scattered as it appeared in the HGNSE_1980 map.

The cal_hm simulation largely followed the same procedure. Only after the last iteration, the transition rules were again adjusted manually. This resulted in a lower Kappa than the cal_au model. At the same time the shape of especially the city patches looked more realistic. Cities were no longer growing in circles, but in more star shaped instead. This attempt to make the result look more realistic was purely done by trial and error and not measured by a result statistic.

Both simulations were compared to the null simulation (the 1960 map) and four random simulations. The result of this is presented in table 5.1 below. The Fuzzy Kappa was computed with a distance decay function, with a radius of 4 cells and a halving distance of 2. The result maps of these calibration runs are presented in appendix B: Calibration results.

Table 5.1: Calibration results compared to HGNSE_1980. Underlined results are those indicating the most similarity to the reference map.

			Kappa per category			Fuzzy
	Kappa	Rural	Nature	Urban	Water	Kappa
HGNSE_1980	1	1	1	1	1	1
cal_au	0,781	0,776	0,823	0,553	0,897	0,649
cal_hm	0,753	0,740	0,828	0,473	0,897	0,663
null	0,777	0,772	0,832	0,437	0,897	0,696
RCMa	0,680	0,659	0,812	0,295	0,877	0,574
RCMb	0,682	0,661	0,812	0,299	0,878	0,578
RCMc	0,681	0,660	0,812	0,298	0,878	0,575
RCMd	0,682	0,661	0,812	0,299	0,878	0,577

5.2 Discussion of calibration results

The cal_au model has a slightly better goodness-of-fit than the null model, and much better than the random model. If however Fuzziness with respect to location is considered, the null model performs best over the calibration period. Since the amount of change for the variable classes was predefined, this performance was mainly due to the allocation of the changed cells.

The cal_hm model was outperformed by the null model and the cal_au model in the Kappa and the Fuzzy Kappa statistic over the calibration period. The results were still better than the results from the random model. Also in this model the amount of change for the variable classes was predefined. Since the statistics were computed for the whole map, all differences are significant. However since the model implies a stochastic process, a Monte Carlo analysis would give more insights in the spread of the values.

Measurements for the goodness-of-fit show that performance of both models is quite low over the calibration period, 1960 to 1980. Several reasons can be indicated as a cause for this low performance. These are the simplicity of the model, changes due to planning decisions and the fact that water cannot change into land and vice versa.

First the model is kept very simple. Only very limited amount information was used for computing land use change. At the other hand a high spatial resolution was used. Figure 5.1 shows the result of the automatic calibration and the reference map for 1980. It shows that it is hard to accurately simulate detailed land use changes with only this limited amount of information. Since the model was this simple, I argue that it is not useful to use spatial resolution as detailed as 100 meter.

Second, much land use changes as a result of local planning policies. These changes are hard to simulate by transition rules mainly based on the neighboring land use, since they do not depend on this. In some case the appearance of nodes in the transport network, like highway junctions and railway station indicates locations of change. Still they do not indicate the exact location. An obvious example is the village of Heerhugowaard, which was hardly recognizable in the 1960 map. Due to local planning activity however increased much in size until 1980. Also additional information that is of importance for land use change is not considered. Examples are open spaces that indicate parks in cities

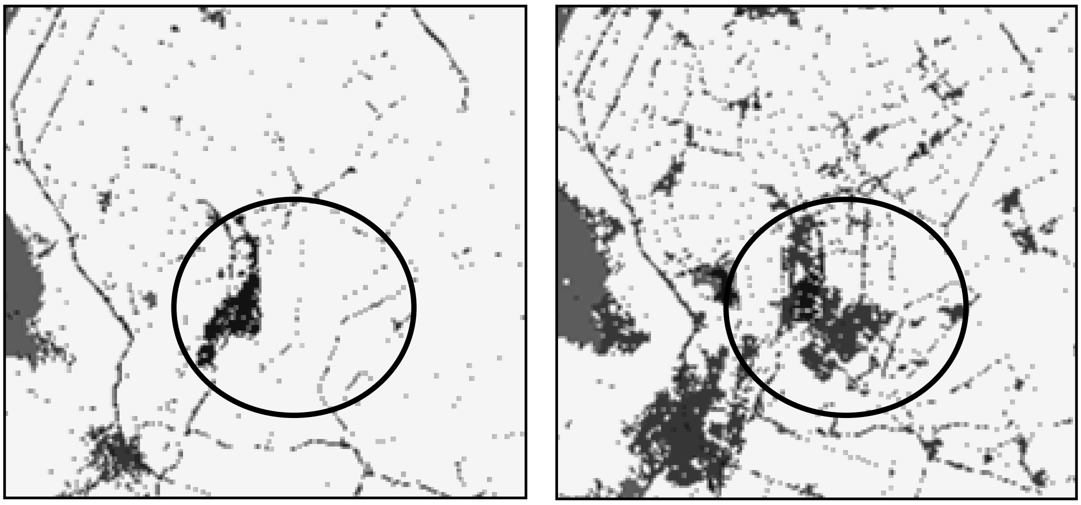


Figure 5.1: Cutout of the HGN SE 1960 and HGN SE 1980 land use maps around Heerhugowaard. Dark gray indicates urban area.

A third cause for dissimilarities in comparing results is the rule that water cannot change in this model. In reality however land is changing into water and vice versa. This is either caused by political decision making and human action or by the force of nature. An example of the first are the new harbors in Amsterdam, an example of the latter is the genesis of the island Noorderhaaks, close to Texel. Both changes cause dissimilarities in the result comparison since they are not simulated by the model. The model however was not meant to simulate these changes.

Besides it generates disagreement in all comparisons, since all models in the comparison started with the same 1960 land use map. The effect is thus a lower overall percentage of agreement, but no bias towards one simulations or the other. The RCM models are able to change water into land, since they are not simulated in the Environment Explorer, but in the Map Comparison Kit instead. However the simulated change is located randomly here, so they are very unlikely to produce a harbor or an island.

5.3 Validation results

The models derived in the calibration were only changed in two ways for the simulations. First the starting year was changed from 1960 to 1980. Second the starting map was changed from HGNSE_1960 to HGNSE_1980. Thus the transition rules, attraction of the transport network and random parameter remained the same. The HGNSE_1980 map is now treated as the null model, and the random simulation results were computed as a random constraint match between the HGNSE_1980 and the HGNSE_2000 maps.

At first, these results maps were compared to the HGNSE_2000 reference map. The results of which are presented in table 5.3 below. The Fuzzy Kappa was computed with a distance decay function, with a radius of 4 cells and a halving distance of 2. The result maps of the validation runs are presented in appendix C: Validation result maps.

Table 5.2: Validation results compared to HGNSE_2000. Underlined results indicate the most similarity with the reference map.

		Kappa per category				Fuzzy
	Kappa	Rural	Nature	Urban	Water	Kappa
HGNSE_2000	1	1	1	1	1	1
HGNSE_1980	0.800	0.800	<u>0.764</u>	0.646	<u>0.913</u>	<u>0.737</u>
sim_au	<u>0.801</u>	<u>0.804</u>	0.745	<u>0.665</u>	<u>0.913</u>	0.726
sim_hm	0.795	0.793	0.754	0.650	<u>0.913</u>	0.733
RCM1	0.760	0.759	0.713	0.594	0.895	0.666
RCM2	0.760	0.758	0.713	0.595	0.896	0.667
RCM3	0.760	0.759	0.714	0.594	0.895	0.667
RCM4	0.760	0.759	0.714	0.595	0.896	0.666

Besides a comparison of the thematic aspects with the Kappa and Fuzzy Kappa statistic, a comparison of the geometric aspects was performed. For this comparison four landscape metrics were computed for all result maps, Shannon's diversity index, fractal dimension, shape index and perimeter/area ratio. For each result the global average was computed and compared with the value obtained from the reference map. The results of this are presented in table 5.3.

Table 5.3: Absolute difference in global average values for landscape metrics compared to HGNSE_2000. Underlined results indicate the most similarity with the reference map.

	SHDI	D	ShapeIndex	P/A ratio
HGNSE_2000	0.000	0.000	0.000	0.000
HGNSE_1980	0.008	0.033	2.941	0.053
sim_au	0.019	0.009	<u>0.509</u>	0.037
sim_hm	<u>0.001</u>	<u>0.006</u>	1.225	<u>0.003</u>
RCM1	0.085	0.098	9.705	0.262
RCM2	0.085	0.098	9.783	0.261
RCM3	0.085	0.098	9.694	0.261
RCM4	0.085	0.096	9.062	0.262

Second a comparison of the landscape metrics was made based on a moving average. For this comparison first a global average with a radius of 10 pixels, 1 kilometer, was computed. With these results the same method as for the pixel by pixel comparison was used to compute the overall result. This result is presented in table 5.5.

Table 5.4: global average of pixel by pixel absolute differences after applying a moving window. Underlined results indicate the most similarity with the reference map.

	SHDI	FD	Shape	P/A ratio
HGNSE_2000	0.000	0.000	0.000	0.000
HGNSE_1980	<u>0.091</u>	0.037	3.345	0.097
sim_au	0.104	0.035	<u>2.022</u>	0.120
sim_hm	0.097	<u>0.029</u>	2.179	<u>0.095</u>
RCM1	0.137	0.098	9.882	0.271
RCM2	0.137	0.098	9.938	0.272
RCM3	0.136	0.098	9.869	0.271
RCM4	0.137	0.096	9.258	0.273

5.4 Discussion of validation results

The goodness-of-fit validation results show the same preference as the calibration results. Based on the Kappa statistic, the sim_au model shows a slightly better result than the null model which again performs better than the sim_hm model. When measured with the Fuzzy Kappa statistic, the null model is still the best prediction of the future. With both statistics the models perform much better than the random allocation model. Since the map comprises the population again, all differences are significant.

For this quite low performance as compared to the null model the same causes can be identified as in the calibration phase. These are simplicity of the model, changes due to planning decisions and the fact that water is a feature. In fact it shows an invalidation of the sim_hm model, and hardly a validation for the sim_au. Because of these limitations of the model as described in the discussion of the calibration, it is likely that another calibration will hardly improve these results. Probably decision in the conceptual model, like resolution, simplifications and assumptions, are of much bigger influence on the eventual result.

However, landscape metrics show a different result. In fact it shows that the simulated models produce more realistic results with respect to patterns of land uses and complexity. At least both calibrated models generally perform better than the null and the reference model. The exception of this being Shannon's diversity index that still shows a preference for the null model. This indicates that the model algorithm and transition rules, as well as the amount of randomness indeed are capable of simulating more or less realistic land use changes after calibration.

5.5 Discussion of the use of landscape metrics in operational validation

Validation of simulation result with landscape metrics yields different results than validation based on goodness-of-fit measures. This indicates that Landscape metrics indeed are an addition in measuring these results. Here comes the purpose of a model into picture. For this case study it was explicitly stated (de Nijs et al., 2001) that the Environment Explorer aims at exploring possible future land use. This makes it reasonable not to validate the result on the level of single pixels. Instead two criteria remain. The first is whether the placement of the land uses about right (Fuzzy Kappa). The second whether the patterns and complexity generated are realistic. Since thematic information is of prime importance in simulation models, metrics cannot be a stand alone replacement for thematic validation. However, they are a good addition.

Two different ways to compute map similarities with landscape metrics indeed show different results. This implies that regional differences are identified and simulated as such in the simulation result maps. The preference of one computation method above the other is then dependent on the homogeneity of the area simulated. For example the Netherlands as a whole show quite some difference between regions. Whereas the province of Flevoland has big clusters of uniform land use, 'de Achterhoek' shows a lot of diversity within small distances. The larger the regional differences the less favorable a global average becomes.

Since these landscape metrics indicate certain patterns, single pixel differences are not a good indication for similarity as well. Therefore an intermediate method was applied: differences based on moving windows. This moving window should then have a size that successfully measures patterns and complexity. In this thesis a 10 pixel radius was used, which is equal to a 1000 meter radius. Total area for the moving window is therefore 317 pixels or 3.17 km2.

6 CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

6.1.1 Guidelines for validating simulation models

Development of simulation models roughly goes through the following phases: project set up, conceptual modeling, implementation and calibration, validation and finally the use and report of the model. Separate steps should be evaluated, specifically the conceptual model, the implementation of the conceptual model, the result of the calibrated model and the input data. In this thesis the validation of model results was considered.

Two standard validation comparisons are proposed here. Simulation models should perform better than the null model and a random model. Besides that, sources of disagreement should be identified. In what way they should perform better is depending on the purpose of the model.

In this case study the purpose was to generate possible outcomes for future land use. This means that a traditional measure of pixel based goodness-of-fit would not be appropriate. Instead the model was validated in two other ways. First a goodness-of-fit based on fuzzy location rules was applied, the fuzzy kappa. Second, the model was validated with respect to the generated geometry, by calculating landscape metrics. This latter method seems to be a good indication for the realness of the generated patterns

6.1.2 The use of landscape metrics for validation

Since simulation models do not aim at forecasting the exact future situation, goodness-of-fit measures, like Kappa statistics, are not an appropriate method to validate results. Methods that take into account fuzziness, like the Fuzzy Kappa, are an improvement of these traditional goodness-of-fit measures. In this way pixels that are slightly displaced, are not entirely wrong. An additional way to validate results is to take into consideration spatial patterns.

Landscape metrics are measurements that consider patterns instead of goodness-of-fit. By that they measure geometric aspects of simulation results instead of thematic aspects. For this thesis four metrics were selected and investigated in three comparison methods. These comparisons were based on a global average, a pixel by pixel comparison and a moving average.

From these comparisons it can be concluded that metrics indeed had a different preference for simulation results than the traditional Kappa comparisons. Since landscape metrics indicate patterns rather than pixel characteristics, it is not useful to compare them on a pixel by pixel basis. Whether the global average is preferable, or the moving window based comparison depends on the homogeneity in the area of interest. For a homogeneous area a global average will suffice. For areas that are spatially diverse a moving window based comparison gives more detailed results.

6.1.3 Calibration and validation of the Environment Explorer

Two versions of the model were calibrated with respect to goodness-of-fit for the period 1960 to 1980, one automatically and one by hand. Based on the Kappa statistic only the first model performed slightly better than the null model over this period. A comparison with the Fuzzy Kappa showed that none of the models performed better than the null model. In both cases the results from the random models were far worse.

Over the validation period, 1980 to 2000, both models were again compared with the null model and a random model. Based on Kappa and Fuzzy Kappa statistics this validation yielded the same results as obtained from the calibration. The automatically calibrated model performed better on the Kappa statistic, whereas the null model performed best on the Fuzzy Kappa statistic.

A validation based on the landscape metrics was performed as well. This showed that both models have indeed some forecasting power. PA ratio, fractal dimension and Shape index showed a preference for the calibrated models above the null and random models. Shannon's diversity index was a notable exception to this, since its preference was depending on the comparison method.

Overall it was obvious that the calibrated models hardly performed better than the null model if at all. The random models performed far worse. A major cause of this it's the simplicity of the model.

6.2 Recommendations for further research

6.2.1 Recommendations for model evaluation

Evaluation comprises more than only operational validation. The conceptual model and assumptions made in the conceptualization are not considered here, but they offer some challenging problems. For example the simplifications made for this version of the model might affect the final simulation result much more than a parameter adjustment. The most interesting problems are the validation of the resolution and the simplifications and assumptions made.

The resolution, both spatial and temporal, depends among others on the spatial extend and the topic being modeled. But it is not clear what range or resolution to extend ratio would be acceptable. In most cases still computation time required or finances are limiting.

The topic being modeled, in this case land use, changes on different spatial scales, ranging from a single house to whole new areas of land. An investigation of a characteristic scale for a certain modeling exercise would be appropriate. For signal modeling the Nyquist rate was defined. This rate states that the sample frequency should be at least twice the frequency of the signal. A free translation to spatial modeling would mean a spatial resolution of no more than half the defined characteristic scale.

Another problem is to justify simplifications and assumptions in a model and confirm they are correct in an application. For example the appearance of Amsterdam Airport Schiphol was neglected in this thesis' case study, whereas its influence on land use change might be larger than the neighborhood effect of water on nature. The difficulty here lies in the fact that these problems are highly site specific.

6.2.2 Recommendations for investigating landscape metrics

Validation with the use of landscape metrics showed that good agreement with respect to nominal data does not indicate an equally well similarity with respect to patterns. Which of both is preferable is hard to say and at least dependent on the purpose of a simulation. However it would be useful to see how people value both characteristics. For this application I showed the maps to some friends to get a clue whether I was on the right track. Most of them intuitively shared my opinion that the metrics were a useful addition to the goodness of fit measurements like Kappa and Fuzzy Kappa. However, it would be interesting, and a stronger proof if this was confirmed by a large group of interpreters. Be it professionals or not. This thesis investigated the credibility of model results, whereas its acceptability by the audience is not treated.

Also it is not clear from this research whether a difference in landscape metrics indicates a significant difference in simulation results. Since the Environment Explorer and many other models comprise an amount of randomness in a simulation, a Monte Carlo analysis would be applicable. By this it can be investigated whether a difference as measured in this way could indeed be identified as significant or not. At least it gives more insight in the spread of values for different metrics.

Finally, the landscape metrics were compared after applying a moving window filter. Also the Fuzzy Kappa used a moving window. The size of this window is actually a topic of discussion. It is not investigated whether the 10 cell radius is a good environment for calculating metrics.

REFERENCES

Anderson, M. G. and P. D. Bates (2001). Model validation: perspectives in hydrological science. Chichester, John Wiley and Sons.

Arthur, J. D. and R. E. Nance (2000). Verification and validation without independence: A recipe for failure. 2000 Winter Simulation Conference.

Barredo, J. I. and L. Demicheli (2003). "Urban sustainability in developing countries' megacities: modelling and predicting future urban growth in Lagos." Cities **20**(5): 297-310.

Batty, M. (2003). Agents, cells and cities: new represntational models for simulating multi-scale urban dynamics. CASA working paper series. London, University College London**:** 43.

Batty, M. and P. M. Torrens (2001). Modeling Complexity: the limits to prediction. CASA working paper series. London, University College London**:** 37.

Batty, M. and Y. Xie (1994). "From cells to cities." Environment and Planning B **21**(1): 31-48.

Bishop, Y. M. M., S. E. Fienberg and P. W. Holland (1975). Agreement as a special case of association. Discrete multivariate analysis: theory and practice. Cambridge MA, MIT press**:** 393-400.

Bregt, A. K. and M. C. S. Wopereis (1990). "Comparison of complexity measures for chloropleth maps." The Cartographic Journal **27**: 85-91.

Brimicombe, A. (2003). GIS, environmental modelling and engineering. New York, Taylor & Francis.

Brown, D. G., S. Page, R. Riolo, M. Zellner and W. Rand (2005). "Path dependence and the validation of agent based spatial models of land use." International Journal of Geographic Information Science **19**(2): 153-174.

Cain, D. H., K. Riiters and K. Orvis (1997). "A multi-scale analysis of landscape statistics." Landscape Ecology **12**: 199-212.

Caughlin, D. (2000). An integrated approach to verification, validation and accreditation of models and simulations. 2000 Winter Simulation Conference.

Chorley, R. C. and Haggett, P. (1967). Models in geography. London, Methuen & Co.

Clarke, K. C. (2003). The limits of simplicity: Toward geocomputational honesty in urban modeling. 7th International Conference on GeoComputation, Southampton.

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

Congalton, R. G. (1991). "A review of assessing the accuracy of classifications of remotely sensed data." Remote Sensing of Environment **37**: 35-46.

Congalton, R. G., R. G. Oderwald and R. A. Mead (1983). "Assessing Landsat classification accuracy using discrete multivariate statistical techniques." Photogrammetric Engineering and Remote Sensing **49**(12): 1671-1678.

Couclelis, H. (1997). "From cellular automata to urban models: new priciples for model devellopment and implementation." Environment and Planning B **24**(2): 165-174.

Delcourt, H. R. and P. Delcourt (1988). "Quarternary landscape ecology: Relevant scales in space and time." Landscape Ecology **2**: 23-44.

Engelen, G., A. Hagen-Zanker, A. C. M. de Nijs, A. Maas, J. Van Loon, B. Straatman, R. White, I. Uljee, M. van der Meulen and J. Hurkens (2005). Kalibratie en validatie van de LeefOmgevingsVerkenner. Bilthoven, MNP**:** 160.

Engelen, G., R. White and I. Uljee (2002). The MURBANDY and MOLAND models for Dublin. Maastricht, RIKS**:** 172.

Engelen, G., R. White and T. de Nijs (2003). "Environment Explorer: a spatial support system for the integrated asessment of socio-economic and environmental policies in the Netherlands." Integrated Assessment **4**(2): 97-105.

Fisher, P. (2000). "Sorites paradox and vague geographies." Fuzzy Sets and Systems **113**: 7-18.

Foody, G. M. (2002). "Status of land cover classification accuracy assessment." Remote Sensing of Environment **80**: 185-201.

Frankhauser, P. (2004). Comparing the morphology of urban patterns in Europe - a fractal approach. European Cities - Insights and Outskirts. A. Borsdorf and P. Zembri. Brussels**:** 79-105.

Fritz, S. and L. See (2005). "Comparison of land cover maps using fuzzy agreement." International Journal of Geographic Information Science **19**(7): 787-807.

Gardner, R. H. and D. L. Urban (2005). Model validation and testing: past lessons, present concerns, future prospects. The role of models in ecosystem science. C. D. Canham, J. C. Cole and W. K. Lauenroth. Princeton, Princeton University Press**:** 186-205.

Giles, R. H. J. and M. K. Trani (1999). "Key elements of landscape pattern measures." Environmental Management **23**(4): 477-481.

Hagen-Zanker, A., B. Straatman and I. Uljee (2005). "Further developments of a fuzzy set map comparison approach." International Journal of Geographic Information Science **19**(7): 769-785.

Hagen-Zanker, A., G. Engelen, J. Hurkens, R. Vanhout and I. Uljee (2005b). Map comparison kit 2: user manual. Maastricht, RIKS**:** 120.

Hagen, A. (2003). "Fuzzy set approach to assessing similarity of categorical maps." International Journal of Geographic Information Science **17**(3): 235-249.

Harmoniqua. 2005, from http://harmoniqua.wau.nl/public/most/flowchart.htm.

Herzog, F. and A. Lausch (2001). "Supplementing lans-use statistics with landscape metrics: some methodological considerations." Environmental Monitoring and Assessment **72**(37-50).

Jantz, C. A. and S. J. Goetz (2005). "Analysis of scale dependencies in an urban land-use-change model." International Journal of Geographical Information Science **19**(2): 217-241.

Kavouras, M. (2001). Understanding and modelling spatial change. Life and motion of socio-economic units. F. A. Raper, and J. P. Cheylan, Taylor & Francis.

Knol, W. C., H. Kramer, G. J. van Dorland and H. Gijsbertse (2003). Historisch Grondgebruik Nederland: tijdreeksen grondgebruik Noord-Holland van 1850 tot 1980. Wageningen, Alterra**:** 50.

Kok, K., A. Farrow, A. Veldkamp, P.H. Verburg (2001). "A method and application of multi-scale validation in spatial land use models." Agriculture, Ecosystems and Environment **85**(3): 223-238.

Kok, K. and A. Veldkamp (2001). "Evaluating impact of spatial scales on land use pattern analysis in Central America." Agriculture, Ecosystems and Environment **85**: 205-221.

Koning, G. H. J. de, A. Veldkamp and L. O. Fresco (1998). "Land use in Ecuador: a statistical analysis at different aggregation levels." Agriculture, Ecosystems and Environment **70**: 231-247.

Kraak, M. J. (2003). The space-time cube revisited from a geovisualization perspective. 21st International Cartographic Conference, Durban.

Kraak, M. J. and F. J. Ormeling (2003). Cartography, visualization of geospatial data. London, Prentice Hall.

Laurini, R. and D. Thompson (1994). Fundamentals of spatial information systems. London, Academic Press.

Li, X. and A.G. Yeh (2002). "Network based cellular automata for simiulating multiple land use changes using GIS." International Journal of Geographic Information Science **16**(4): 323-343.

Li, X. and A.G. Yeh (2004). "Data mining of cellular automata's transition rules." International Journal of Geographic Information Science **18**(8): 723-744.

Lowry, I. S. (1964). A model of metropolis. Santa Monica, California, The Rand Corporation.

Mcgarigal, K. and B. J. Marks (1994). FRAGSTATS - Spatial pattern analysis program for quantifying landscape structure - user's manual version 2.0.

Middlemis, H., N. Merrick and J. Ross (2000). Groundwaterflow modelling guideline. South Perth, Aquaterra Consulting Pty Ltd.

Molenaar, M. (1998). An introduction to the theory of spatial object modelling for GIS. London, Taylor & Francis.

Nijs, T. de, G. Engelen, R. White, H. van Delden and I. Uljee (2001). De LeefOmgevingsVerkenner - technische documentatie. Bilthoven, RIVM.

Nijs, T. de, R. de Niet, G. de Hollander, F. Filius and J. Groen (2001). De LeefOmgevingsVerkener: Kaartbeelden van 2030 - Een verkenning van de inzet bij beleidsondersteuning. Bilthoven, RIVM.

Oreskes, N., K.Shrader-Frechette and K. Belitz (1994). "Verification, validation and confirmation of numerical models in the earth sciences." Science **263**: 641-646.

Pang, M. Y. C. and W. Shi (2002). "Development of a process-based model for dynamic interaction in spatio-temporal GIS." GeoInformatica **6**(4): 323-344.

Parker, D. C. and V. Meretsky (2004). "Measring pattern outcomes in an agent-based model of edge-effect externalities using spatial metrics." Agriculture, Ecosystems and Environment **101**: 233-250.

Peuquet, D. J. (2000). Space-time representation: an overview. Time in GIS: Issues in spatio-temporal modelling. L. Heres. Delft. **47:** pp 3 - 13.

Peuquet, D. J. (2002). Representations of space and time. New York, The Guilford Press.

Pilouk, M. (1996). Integrated modelling for 3D GIS. Wageningen, Wageningen Agricultural University**:** 201.

Pontius, R. G. and J. Malanson (2005). "Comparison of the structure and accuracy of two land change models." International Journal of Geographic Information Science **19**(2): 243-265.

Pontius, R. G. and L. C. Schneider (2001). "Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA." Agriculture, Ecosystems and Environment **85**(3): 239-248.

Pontius, R. G., D. Huffaker and K. Denman (2004). "Usefull techniques for spatially explicit land-change models." Ecological Modelling **179**: 445-461.

Pontius, R. G., J. D. Cornell and C. A. S. Hall (2001). "Modeling the spatial pattern of land-use change with GEOMOD2: Application and validation for Costa Rica." Agriculture, Ecosystems and Environment **85**: 191-203.

Power, M. (1993). "The predictive validation of ecological and environmental models." Ecological Modelling **68**(1): 33-50.

Power, M., A. Simms and R. White (2001). "Hierarchical fuzzy pattern matching for the regional comparison of land use maps." International Journal of Geographic Information Science **15**(1): 77-100.

Read, J. M. and N. S. N. Lam (2002). "Spatial methods for characterising land cover and detecting land-cover changes for the tropics." International Journal of Remote Sensing

23(12): 2457-2474.

Refsgaard, J. C. and H. J. Henriksen (2004). "modelling guidelines - terminology and guiding principles." Advances in water resources **27**(1): 71-82.

Remmel, T. K. and F. Csillag (2003). "When are two landscape pattern indices significantly different?" Journal of Geographical Systems **5**: 331-351.

Riiters, K. H., R. V. O'neill, C. T. Hunsaker, J. D. Wickham, D. H. Yankee, S. P. Timmins, K. B. Jones and B. L. Jackson (1995). "A factor analysis of landscape pattern and structure metrics." Landscape Ecology **10**(1): 23-39.

Rykiel, E. J. (1996). "Testing ecological models: the meaning of validation." Ecological Modelling **90**(3): 229-244.

Sargent, R. G. (1998). Verification and validation of simulation models. 1998 Winter Simulation Conference.

Sargent, R. G. (2001). Some aproaches and paradigms for verifying and validating simulation models. 2001 Winter Simulation Conference.

Scholten, H. J., R. J. van de Velde and J. A. M. Borsboom van Beurden (2001). RuimteScanner: Informatiesysteem voor de lange termijnverkenning van ruimtegebruik. Nederlandse Geografische studies 242. Amsterdam, KNAG/Faculteit der economische wetenschappen en econometrie, VU Amsterdam: 152.

Straatman, B., R. White and G. Engelen (2004). "Towards an automatic calibration procedure for constrained cellular automata." Computers, Environment and Urban Systems **28**: 149-170.

Suppe, F. (1998). "Verification, validation and confirmation of simulation models: Suppe vs. Oreskes et al." from http://carnap.umd.edu/phil250/lectures/Suppe_Oreskes2.html.

Tannier, C. and D. Pumain (2005). "Fractals in urban geography: a theoretical outline and an empirical example." Cybergeo **307**.

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

Veldkamp, A. and L.O. Fresco (1996). "CLUE-R: an integrated multiscale model to simulate land use change scenarios in Costa Rica." Ecological Modelling **91**: 231-248.

Verburg, P. H., G. H. J. de Koning, K. Kok, A. Veldkamp and J. Bouma (1999). "A spatial explicit allocation procedure for modelling the pattern of land use change based upon actual land use." Ecological Modelling **116**: 45-61.

Verburg, P. H., P. Schot, M. Dijst and A. Veldkamp (2004). "Land use change modelling: current practice and research priorities." Geojournal **61**(4): 309-324.

Verburg, P. H., T. C. M. de Nijs, J. R. van Eck, H. Visser and K. de Jong (2004b). "A method to analyse neighborhood characteristics of land use patterns." Computers, Environment and Urban Systems **28**: 667-690.

Verburg, P. H., W. Soepboer, A. Veldkamp, R. Limpiada, V. Espaldon and S. S. A. Mastura (2002). "Modeling the spatial dynamics of regional land use: the CLUE-S model." Environmental Management **30**(3): 391-405.

Wachowicz, M. (1999). Object-oriented design for temporal GIS. London, Taylor and Francis.

Wachowicz, M. (2000). The role of geographic visualisation and knowledge discovery in spatio-temporal data modelling. Time in GIS: Issues in spatio-temporal modelling. L. Heres. Delft**:** 13-26.

Wagner (1997). "Cellular automata and geographic information systems." Environment and Planning B **24**(2): 219-234.

Waveren, R. H. van., S. Groot, H. Scholten, F. C. van Geer, J. H. M. Wosten, R. D. Keoze and J. J. Noort (1999). Good modelling practice handbook. Lelystad, STOWA/RIZA.

Wealands, S. R., R. B. Grayson and J. P. Walker (2005). "Quantitative comparison of spatial fields for hydrological model assessment - some promissing approaches." Advances in Water Resources **28**: 15-32.

White, R. and G. Engelen (1997). "Cellular automata as the basis of integrated dynamic regional modelling." Environment and Planning B **24**(2): 234-246.

Winter, S. (2000). "Location similarity of regions." ISPRS Journal of Photogrammetry & Remore Sensing **55**: 189-200.

Wolfram, S. (1994). Cellular automata as models of complexity, Addison-Wesley.

Worboys, M. (2005). "Event-oriented approaches to geographic phenomena." International Journal of Geographic Information Systems **19**(1): 1-28.

Wu, F. (1998). "SimLand: a prototype to simulate land conversation through the integrated GIS and CA with AHP-derived transition rules." International Journal of Geographical Information Science **12**(1): 63-82.

APPENDIX A: CALIBRATION RESULT MAPS

The numbered maps in the following pages represent respectively:

1: HGNSE1960 – the begin map for the calibration period

2: HGNSE1980 – the end map for the calibration period

3: cal_au – the calibration result of the automatic calibrated model

4: cal_hm – the calibration result of the model adjusted by hand

5: RCMa – result a of the random constraint match

6: RCMb – result b of the random constraint match

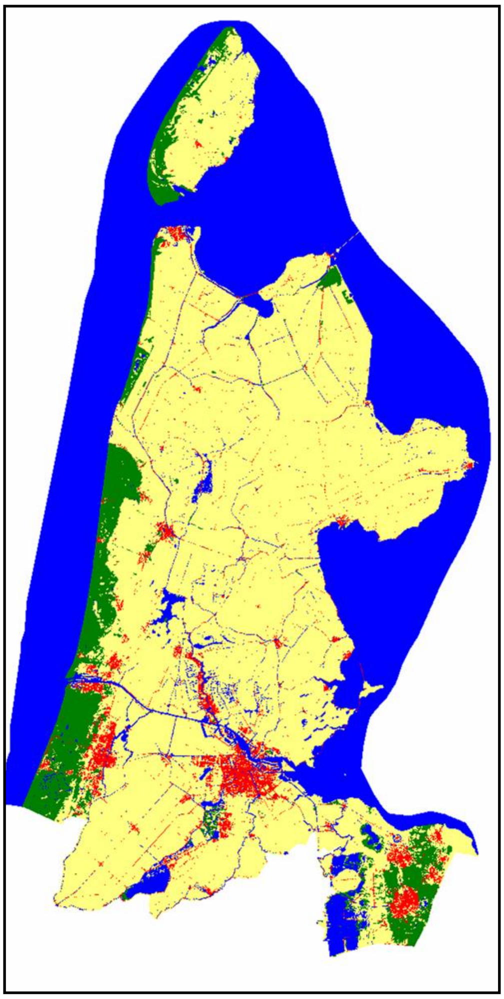
7: RCMc – result c of the random constraint match

8: RCMd – result d of the random constraint match

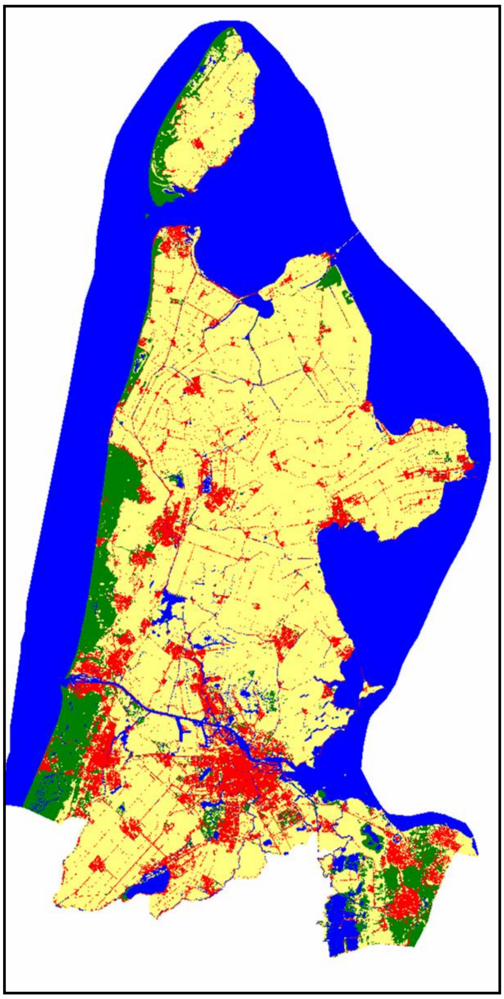
Legend:

Rural land use

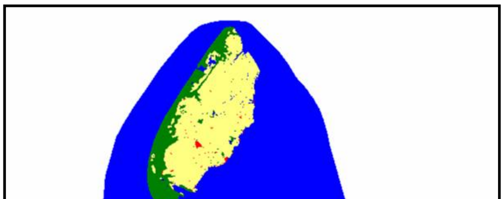
Nature



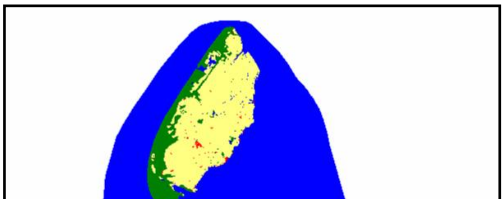
1



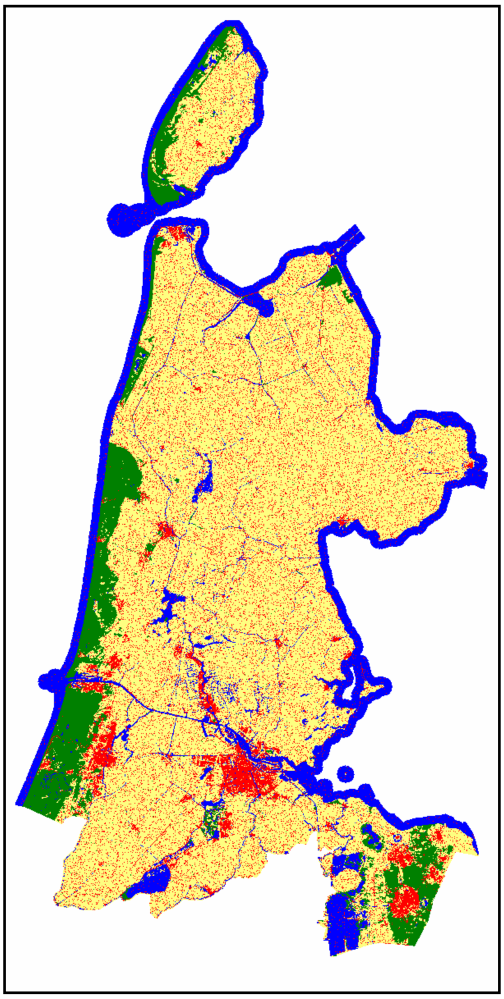
2



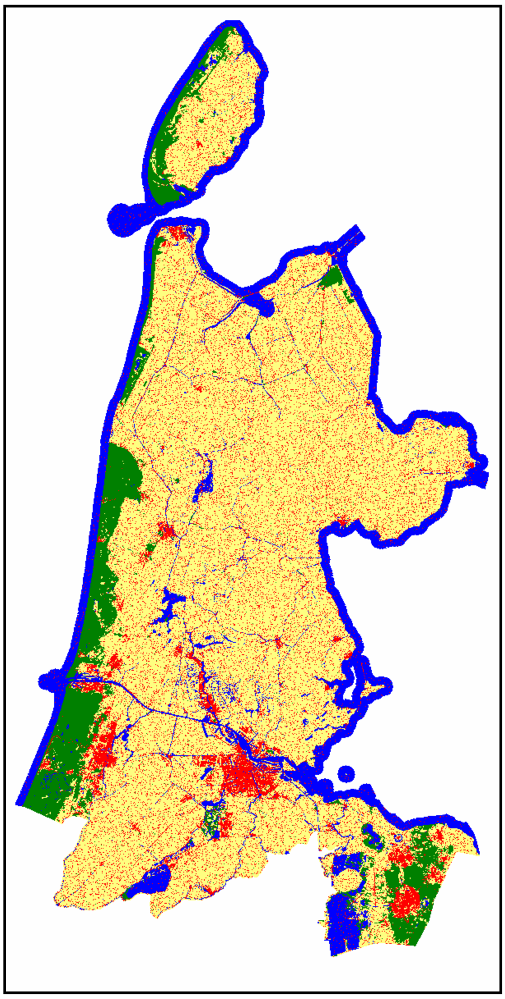
3



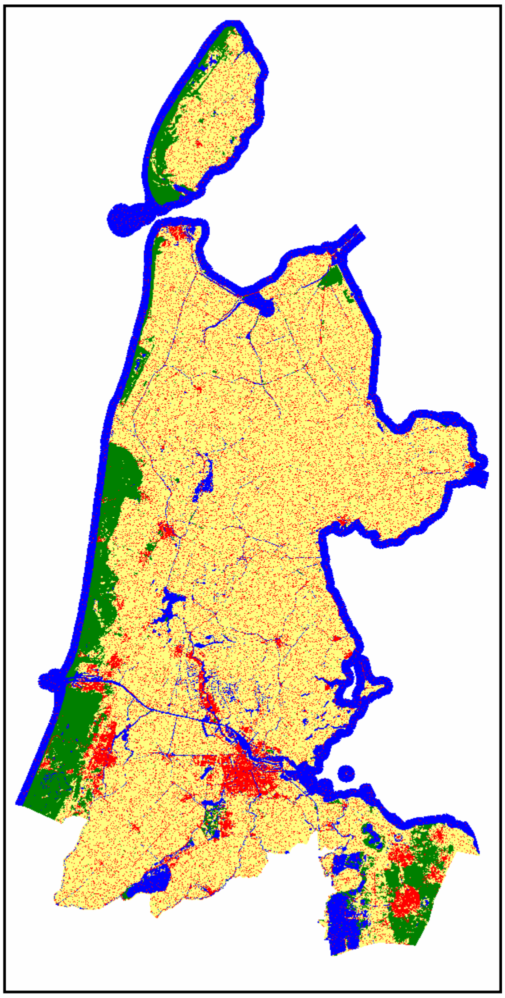
4



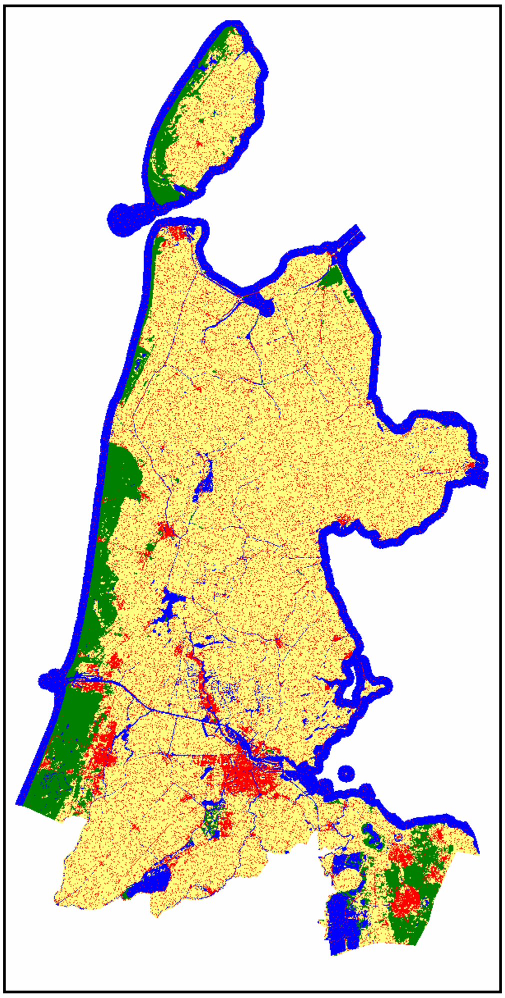
5



6



7



8

APPENDIX B: VALIDATION RESULT MAPS

The numbered maps in the following pages represent respectively:

1: HGNSE1980 – the starting map for the validation period

2: HGNSE2000 – the 2000 land use map to compare results with

3: sim_au – the simulation result of the automatic calibrated model

4: sim_hm – the simulation result of the model adjusted by hand

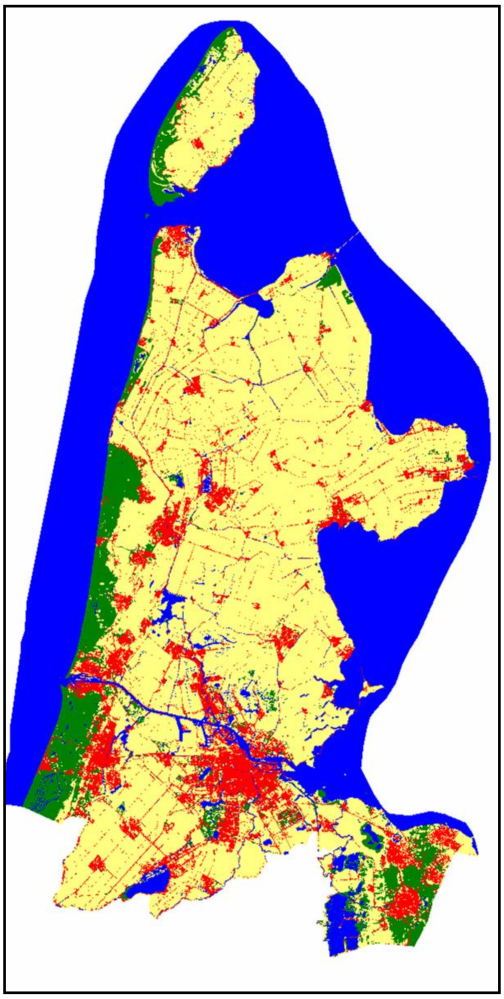
5: RCM1 – result 1 of the random constraint match

6: RCM2 – result 2 of the random constraint match

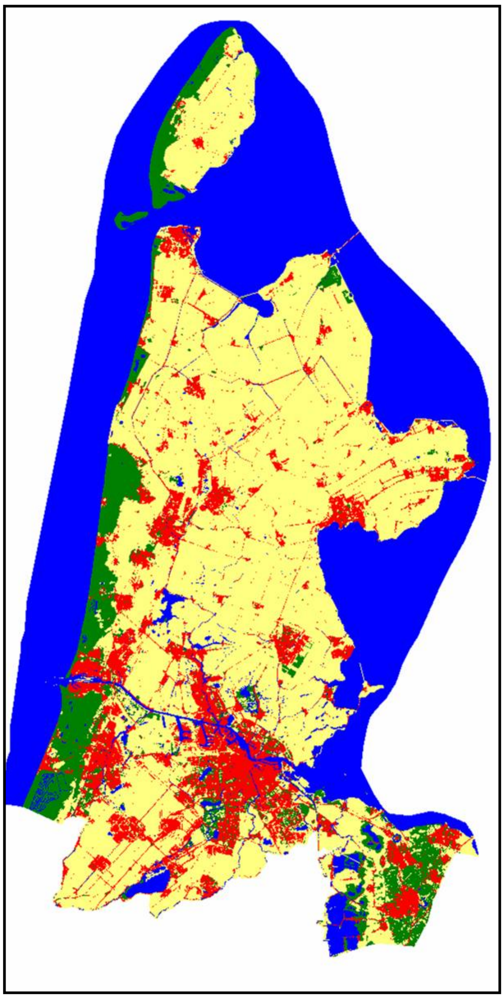
7: RCM3 – result 3 of the random constraint match

8: RCM4 – result 4 of the random constraint match

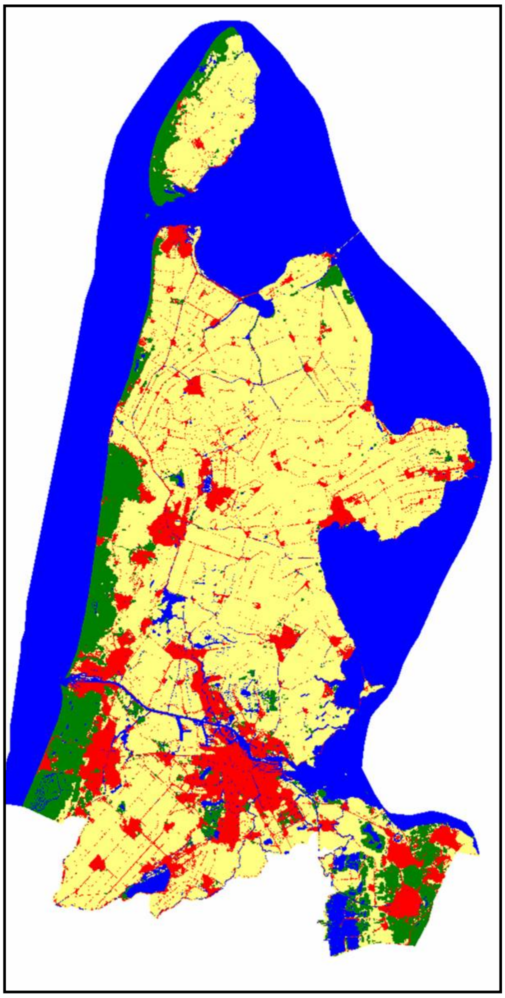
The legend is equal to that in Appendix A.



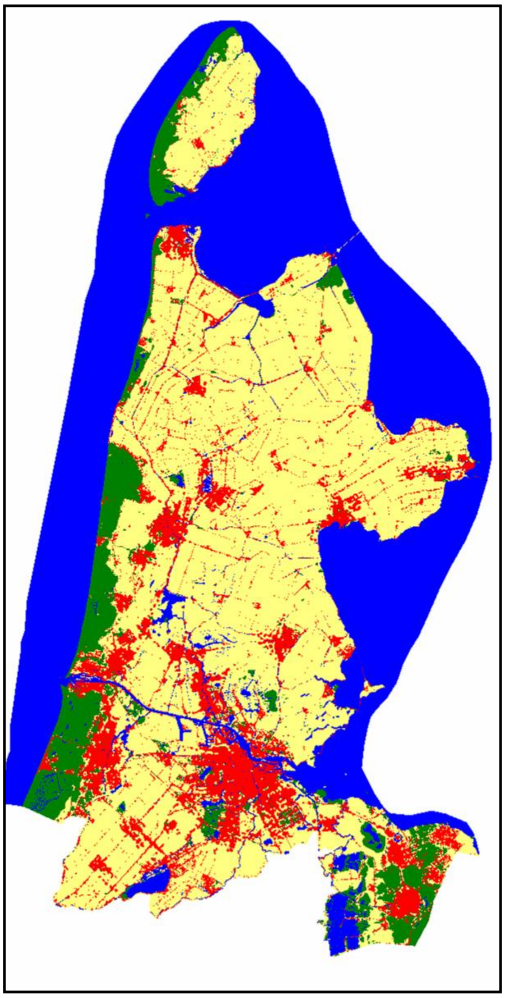
1



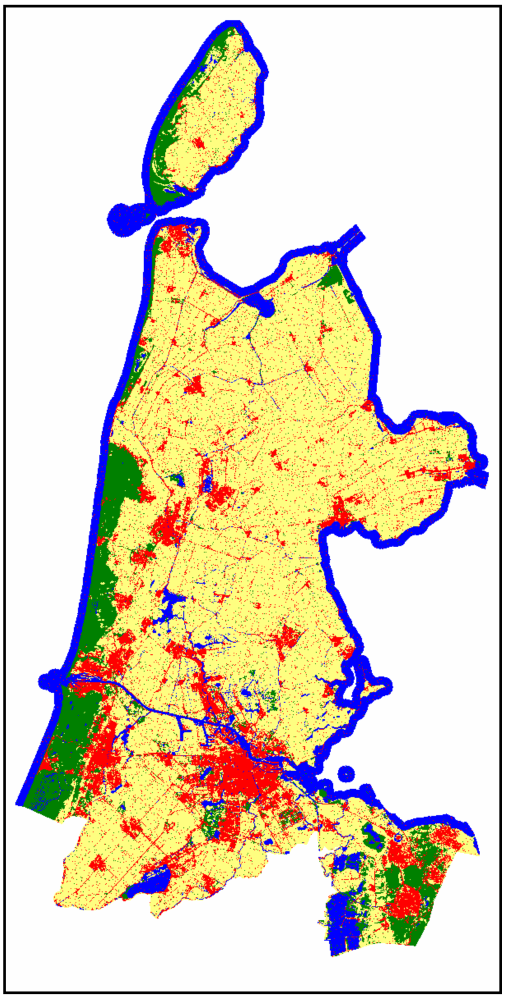
2



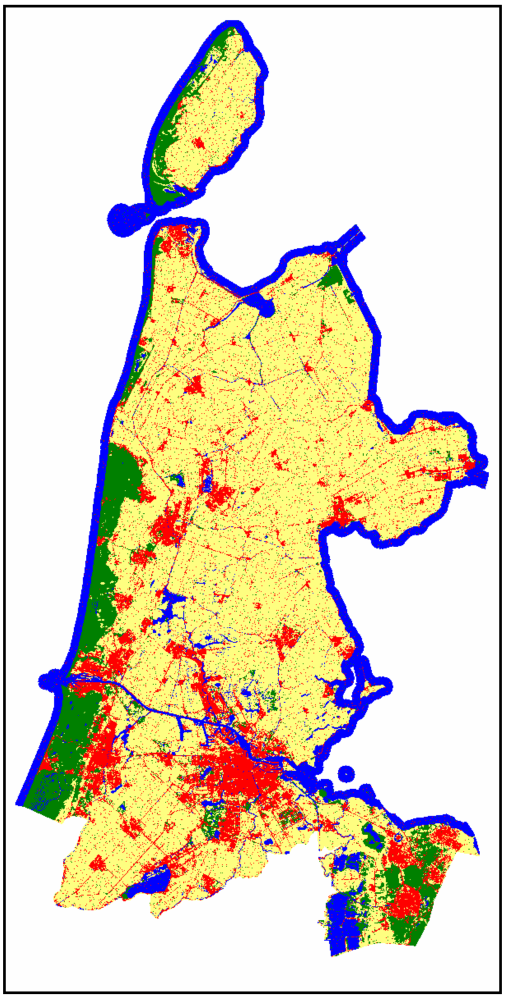
3



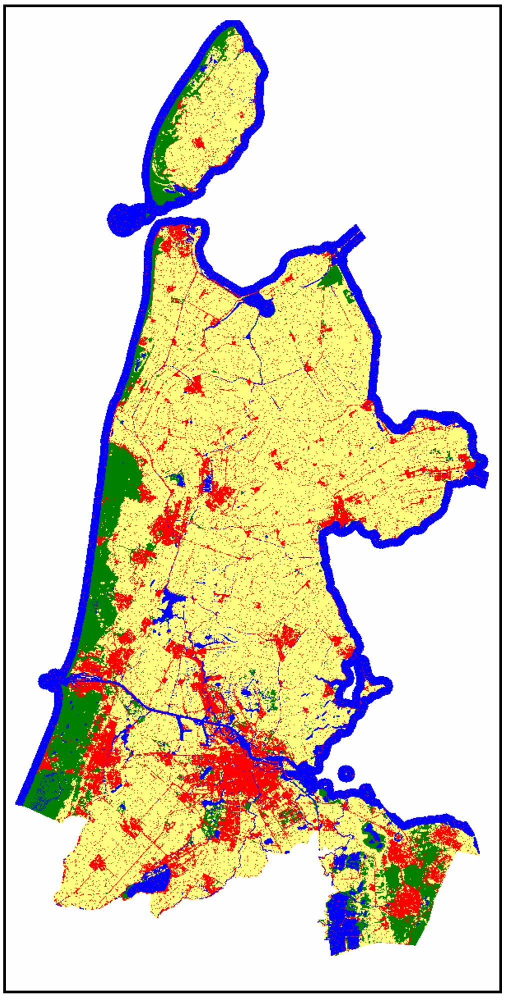
4



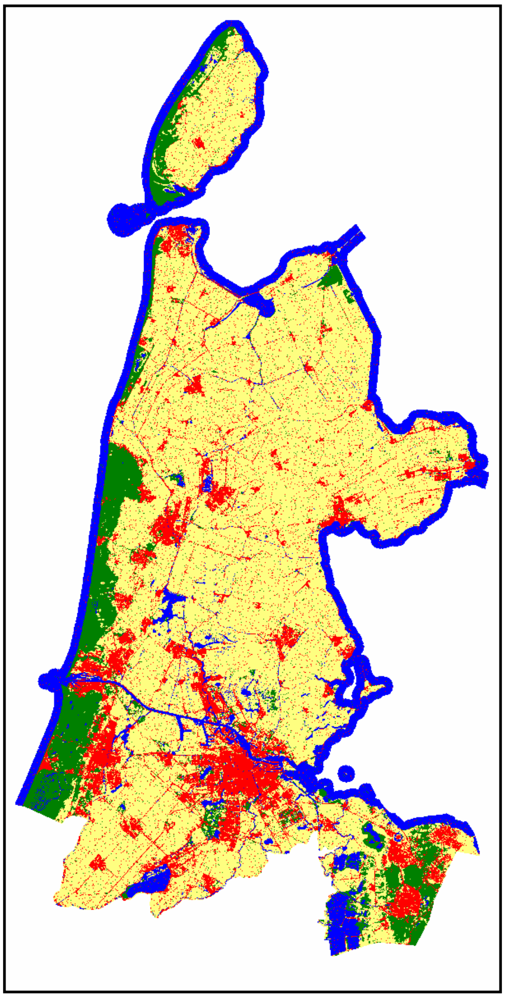
5



6



7



8

APPENDIX C: MODELLING CYCLE FLOWCHART

(After Harmoniqua)