Monitoring land use changes using geo-information

The presented research is funded by DWK program 358 of the Dutch Ministry of Agriculture, Nature and Fisheries

Monitoring land use changes using geo-information

Possibilities, methods and adapted techniques

C.J. de Zeeuw
G.W. Hazeu

Alterra-rapport 214
CGI-report 9

Alterra, Green World Research, Wageningen, 2001

ABSTRACT

Zeeuw, C.J. de and G.W. Hazeu, 2001. *Monitoring land use changes using geo-information; possibilities, methods and adapted techniques.* Wageningen, Alterra, Green World Research. Alterra-rapport 214/CGI-report 9. 84 blz., 18 fig.; 16 tab.; 46 ref.

Monitoring of land use with geographical databases is widely used in policy decision-making. This reports presents the possibilities, methods and adapted techniques using geo-information for monitoring land use changes. The municipality of Soest was chosen as study area and three national land use databases; Top10Vector, CBS-land use statistics and LGN, were used. The restrictions of geo-information for monitoring land use changes are indicated. New methods and adapted techniques improve the monitoring result considerably. However, providers of geo-information should co-ordinate on up-date frequencies, semantic content and spatial resolution to allow for better possibilities for monitoring land use by combining data sets.

Keywords: CBS-land use statistics, geo-information, LGN,.methodology, monitoring, land use, Soest, Top10vector, the Netherlands

ISBN: 90 327 0302 1
ISSN Reeks Alterra-rapporten: 1566-7197
ISSN Reeks CGI-rapporten: 1568-1874

This report can be ordered by paying 43,60 Dutch guilders into bank account number 36 70 54 612 in the name of Alterra, Wageningen, the Netherlands, with reference to rapport 214. This amount is inclusive of VAT and postage.

P.O. Box 47, NL-6700 AA Wageningen (The Netherlands).
Phone: +31 317 474700; fax: +31 317 419000; e-mail: postkamer@alterra.wag-ur.nl

Alterra is the amalgamation of the Institute for Forestry and Nature Research (IBN) and the Winand Staring Centre for Integrated Land, Soil and Water Research (SC). The merger took place on 1 January 2000.

Project 358-10069 [Alterra-rapport 214/HM/01-2001]

Contents

Preface

This report is an overview of the research on monitoring land use changes that has been carried out by the Centre for Geo-information (CGI) of Alterra over the past three years. It is an accumulation of the results of three research projects and contains the adapted versions of several articles, which were written to present the results of the first two projects.

In 1998 Rene Meijners started the first research project entitled "Monitoring Landgebruik Nederland" (MLN) in close collaboration with Kees de Zeeuw. The results were presented in various articles and a MSc. thesis. In 1999 a new project entitled "Ontwikkeling Strategie GIS Monitoring" (ORGISM) was started by Marthijn Sonneveld and Jeroen van den Brink under supervision of Kees de Zeeuw and Arnold Bregt. Results of this project were also presented in various articles. During this project a workshop was organised for Dutch GIS experts to identify the required follow-up of the monitoring research. The follow-up project (ORGISM II) started in the year 2000.

The Dutch Ministry of Agriculture, Nature Management and Fisheries (LNV) financed all three-research projects. This report and the third project "ORGISM II" were also financed by the DWK programme 358 "GIS and Remote Sensing" of LNV.

Besides those persons mentioned-above, Monica Wachowicz has also contributed to the ORGISM II project. She investigated the possibilities of applying knowledge discovery in databases (KDD) techniques for monitoring land use changes.

We would also like to thank Anna van Paddenburg for the revision of the English text and the municipality of Soest for their kind co-operation in the first project by facilitating all their data sources and knowledge of the area.

The editors,
Kees de Zeeuw and Gerard Hazeu.

Summary

The rapid development of national databases has introduced many new possibilities for decision-making and spatial analysis. The overlay of data sets has become standard GIS functionality and the updated versions of these data sets have been considered to be excellent data sources for multi-temporal analyses. However, until recently, little insight and concern has existed about the added value of the obtained monitoring results.

Our research on monitoring land use changes started early 1998 and can be divided into four steps:

- Step 1: Creation of awareness about the added value of using existing geo-information for monitoring purposes (clarification of the problem domain);
- Step 2: Investigation on what are the possibilities to improve monitoring results using available national databases and how to achieve a better application of GIS methods and functionalities in monitoring land use changes;
- Step 3: Recommendation of new methods to improve the geo-database structure for achieving better monitoring results;
- Step 4: Investigation on how the rapid technological development can provide us with new techniques in order to improve the monitoring of land use changes using geo-information.

Step 1

The research on the applicability of the available national databases for monitoring studies started early 1998 at the Centre for Geo-information (CGI) of Alterra. The national databases such as Top10Vector, land use statistics of Central Bureau of Statistics (CBS) and the National Land Use database (LGN) were used to obtain insight and awareness of the important value of the monitoring results. The municipality of Soest was chosen as a representative study area to perform a database analysis. In this analysis, a reference data set was constructed for the representative study area and the monitoring results were obtained using different national databases.

For the three multi-temporal national databases, land use changes were investigated using the simple overlay technique available in GIS. Data were analysed at three different hierarchical levels, i.e. superclass, class and sub-class levels as well as at regional and local scale levels. At the regional level, no importance was given to the geographical position of different types of land use. Only the statistical occurrences were considered at the regional level. However, at the local scale level, the geographical position was taken into account and has actually influenced the monitoring results. Land use changes that have occurred for every geographical point were analysed using the point wise approach.

The CBS-land use statistics were also used for investigating land use changes and they showed minor changes compared to the other databases. Although a direct comparison between the different national databases was not possible as time spans and legend contents vary, the magnitude of changes resulting from the CBS-land use statistics was most realistic when compared to the reference data set. Theoretically, an optimum combination of national databases could be established, but in this case, a closer co-ordination is required between the data providers of these databases for developing a systematic operationalisation of the update cycles and data characteristics.

The monitoring results showed that the changes in land use were over-estimated for the three multi-temporal national databases, including all hierarchical and scale levels. The over-estimation of changes (noise) was mainly due to the different recording techniques that have been applied to create the three databases. Every technique has brought improvements on the reliability and the accuracy of each database. The semantic and scale characteristics have also been modified for each database. These were the main factors that have caused the noise during the analysis. Moreover, the difficulties in defining a common ontology and the conversion of land cover to land use classes have introduced an additional error. One of the main conclusions found in this step was that the national databases have not been created for monitoring purposes, and as a result, they are not expected to reach a satisfactory level in the short term.

Step 2

In 1999 a new project was started entitled "Ontwikkeling Strategie GIS Monitoring" (ORGISM) focusing on the possibilities of using GIS to optimise monitoring results. The methodology was based on the reduction of unreal changes using the Top10Vector database. The municipality of Soest was used as a case study area to test the methodology.

The approach developed in this phase dealt with the problem of noise reduction at the level of noise expression. In this approach, a data set was defined to have the lowest level of uncertainty once it was created in the most recent point in time. This definition was derived from the perspective that most of the current data sets are being produced using improved techniques and concepts than before. Therefore, the most recent data set (from t=2) was used as a *reference source* while the earliest data set (from t=1) was used as the *test source*. The reference source was considered to have the lowest level of uncertainty while the test source was used to determine the positional accuracy of a spatial object. The approach considers not only individual objects, but also the relationships between them. The changes generated by overlaying multi-temporal versions of the Top10 vector database were identified and analysed.

Several criteria (geometric as well as thematic criteria) were defined to evaluate whether observed changes could be linked to object related uncertainties and cartographic errors, or whether they exceeded the limits for similarity. For example, we have used relative geometric criteria (reliable change, significant overlap and shape factor) that were directly linked to polygon characteristics in order to determine the degree of spatial isomorphism of a spatial object. Instead of using these criteria to determine only the degrees of spatial isomorphism (Holt et al., 1998), we have also used them as Boolean operators, since we have considered the situations of 'no-change' (or similarity) and 'change'. The thematic criterion has been defined as class characteristics to distinguish between most likely and less likely classifications. The magnitude of noise was filtered out using a quantified stepwise approach that was fully dependent on the chosen criteria.

The monitoring results have shown that the land use changes have reduced at around 30% at the sub-class level for the Soest municipality. At both class level and superclass level the reduction was less than 10%. The remaining noise was due mainly to the complexity of certain objects and the fuzzy nature of some classes. Therefore, the final outcome was considered a noise-reduced monitoring result rather than a noise-free monitoring result. The main advantages of the method are related to the fact that detailed knowledge on the data quality is not required and the method can be applied to different databases at different hierarchical levels of legends.

Step 3

The ORGISM II project started in the year 2000. For this project, the decision was taken to focus on the process of data capturing used for the creation of a land use database. The objective was to investigate if the problems encountered in the steps 1 and 2 could be avoided by developing a new database structure in conjunction with operational procedures for constructing this database. The multi-temporal versions of the National Land Use (LGN) database were used in this investigation.

A new data layer was constructed as a database structure for storing real changes. The LGN3 data set was taken as the reference source. The real changes were obtained by comparing satellite images at different points in time (t and t+1) and they were stored in a change data layer. Subsequently, the adapted LGN2 and LGN4 data sets were created using the information available from this change data layer. As a result, the LGN3 data were considered as current (true) characteristics for areas where changes have not been found. The adapted LGN2 and LGN4 data sets were used for monitoring purposes. A syntax scheme for land use classes was developed for providing a common legend structure for all three data sets.

The monitoring results have shown the occurrence of under-estimations of real changes in land use at both superclass and class level. One or a combination of the following factors can explain the under-estimation of changes and the impossibility to monitor at the sub-class level:

- Subjectivity of the method and level of expert knowledge;
- Spatial and temporal scale;
- Relation between land cover and land use.

Step 4

Simultaneously with step 3, new techniques were investigated for monitoring land use changes using information on land use classes (i.e. class type, location, and classification accuracy) as well as quantitative information on land surface properties such as spectral and spatial resolution. Knowledge discovery methods and their implemented tools were selected to uncover spatio-temporal relationships from the multi-temporal versions of the National Land Use database. The main objective was to extract from them useful patterns, objects, events, categories and structures with which to construct a model of the relationships between the different land use classes over time. Knowledge discovery subsumes the field of data mining, which can be viewed as comprising the techniques of pattern recognition and change detection. The MineSet System of Silicon Graphics Inc (SGI) was used for mining the LGN1, LGN2, and LGN3 data sets. Association rule generators have been used as the data mining technique for searching these data sets, revealing the nature and frequency of relationships or associations between different land use classes.

The data mining results have confirmed that the data quality of the LGN1, LGN2, and LGN3 data sets is significantly poor for monitoring purposes due to the noise already identified in step 1. The results have also shown that the three hierarchical levels (i.e. sub-classes, classes and superclasses) will be particularly useful for monitoring thematic changes, such as from one land use class to another. (e.g. from agriculture to continuous urban area at the superclass level). However, it is prohibitive to use these hierarchical levels for monitoring other changes, especially at the class and superclass levels. In fact, the regional and local scale levels play a more significant role as the main spatial component to be used for revealing land use changes. The scale level being used, particularly for spatial changes of land use classes with a well-defined spatial distribution such as the grassland class will affect the monitoring results.

1 Introduction

1.1 Background

In the Netherlands the human interaction with the environment plays an important role in the change of the actual land use. Yet, human interaction is a complex process to model. As a consequence there is a clear shift among policy makers and researchers from the use of modelling techniques towards the development of monitoring tools.

Monitoring can be defined as 'periodical measurements of a particular phenomenon in time and the recording of observed changes'. Although some researchers do not explicitly distinguish between monitoring land use and monitoring land cover (Aspinall et al., 1993), the two are not the same. Following the definition provided by Lindgren (1985), land use includes all land that is used by residents of the country while land cover refers to the vegetational and artificial constructions covering the land surface. Land use is considered as an attribute value corresponding to the function of the land.

Geographical Information Systems (GIS) have become a leading tool in the development and application of contemporary research in land use planning (Kent et al., 1993). The increased use of GIS in land use planning is due largely to the development of computer technology, improved methods of data acquisition, growing data availability and declining costs of computer hardware and software (Michalak, 1993; Openshaw, 1998).

1.2 Problem definition and objectives

In 1998, the Centre for Geo-information of Alterra acknowledged the rapid development of national databases and the need to introduce new possibilities for decision making en spatial analysis. The overlay of data sets has become standard GIS functionality and the updated versions of these data sets have been considered to be excellent data sources for multi-temporal analyses. However, little insight and concern has existed about the added value of the obtained monitoring results.

The research questions posed in early 1998 were related to concerns such as: Are the available national land use databases applicable for monitoring studies?; What if the monitoring results obtained using the overlay technique in GIS do not reflect the real-world changes with a sufficient degree of accuracy?; What are the improved GIS methods?; and Should we adapt the data structure of our national databases for monitoring purposes?

The objectives of this research can be described according to the following four steps:

- Step 1: Creation of awareness about the added value of using existing geo-information for monitoring purposes (clarification of the problem domain);
- Step 2: Investigation on what are the possibilities to improve monitoring results using available national databases and how to achieve a better application of GIS methods and functionalities in monitoring land use changes;
- Step 3: Recommendation of new methods to improve the geo-database structure for achieving better monitoring results;
- Step 4: Investigation on how the rapid technological development can provide us with new techniques in order to improve the monitoring of land use changes using geo-information.

This report presents the methodology, the application, and the results obtained for each of the steps described above. The research was carried out within a three-year project in which four steps were realised during 1998-2000. The results of step 1 and 2 have also been published in various articles (de Zeeuw, 1998; de Zeeuw et al., 1999a; de Zeeuw et al., 1999b; de Zeeuw et al., 2000; Sonneveld et al., 2000). The reason to summarise the results of the four research steps into one report is mainly to provide GIS users with a complete overview of the three-year research project. The target audience for this report is researchers, policy makers and GIS users in the Netherlands, mainly those readers who have used Dutch national databases in their analysis. However, the results described in this report are expected to be of interest and of use to the international GI community as well.

1.3 The research approach in brief

In 1998 CGI started a research project entitled "Monitoring Landgebruik Nederland" (MLN) in order to qualify and quantify the problem of using existing geo-information for monitoring purposes. In 1999 a new project was started entitled "Ontwikkeling Strategie GIS Monitoring" (ORGISM) focussing on the possibilities of GIS to optimise monitoring results. During this project a workshop was organised for Dutch GIS experts to identify the required follow-up of the research until then. The follow-up project (ORGISM II) started in 2000 and focussed on the need to construct databases in such a way that problems encountered in the first two projects could be avoided. Figure 1 represents a schematic overview of the project history, dividing the process into four steps.

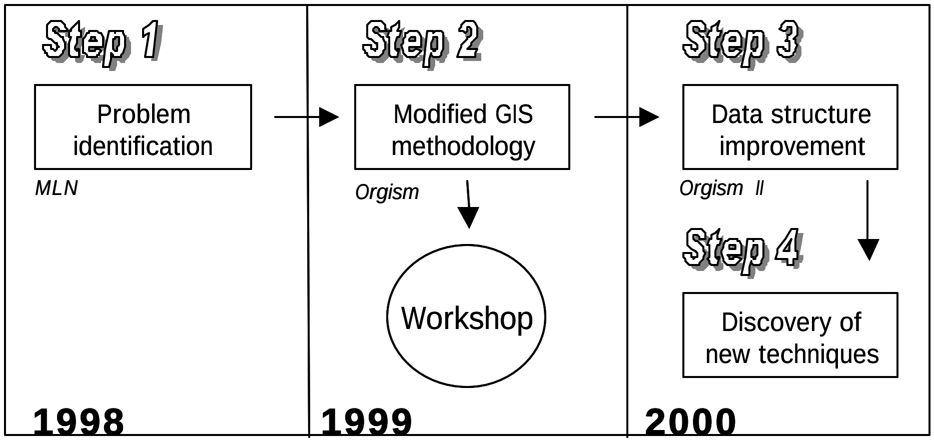


Figure 1. Schematic overview of the project

Step 1. Qualify and quantify the problem

For the databases available at the national scale it has been investigated till what extent these databases are suitable for the monitoring of land use. The objective can be subdivided in the following research questions:

- At what level in classification hierarchy are the databases suitable for monitoring?
- At what scale or at what statistical aggregation level are the databases suitable for a well thought use?
- How often are the databases updated (multi-temporality) and what is the consequence for the suitability for monitoring practices?
- Is there any possibility for improving monitoring results by combining the different national databases?

Step 2. Try to improve monitoring results using existing data and tools

Vector-based geo-data sets that are available for detecting land use changes often do not include quality descriptions of their fitness for monitoring (see also Frank, 1998). There are several examples of error source classification in the literature (e.g. Burrough, 1986; Walsh et al., 1987; Benhardsen, 1999) but these cannot serve as a basis for noise reduction if the relation between the expression of noise and its origin is unknown. It is therefore useful to make a distinction between the causes of noise and the way noise is expressed within a multi-temporal data set. The approach followed in this phase of the research project tackles the problem of noise reduction at the level of noise expression. A geo-data set was defined from a more recent point in time (T=2) as the data set with the lowest level of uncertainty. This definition is derived from the view that recent geo-data sets are being produced by means of improved techniques and concepts. The most recent representation served as a reference source while the earliest representation (from T=1) served as the tested source to determine the positional accuracy of a spatial object. Several criteria (geometric as well as thematic) could then be defined to evaluate whether observed changes could be linked to object related uncertainties and cartographic errors, or whether they exceeded the limits for similarity.

We used relative geometric criteria that are directly linked to polygon characteristics to determine spatial isomorphism. Instead of using these criteria to determine the degree of spatial isomorphism (Holt et al., 1998), we used them as boolean operators, since we considered only the situations 'no-change' (or similarity) and 'change'. As thematic criterion we defined class characteristics to distinguish between likely and less likely classifications.

Evaluation moment
A workshop was organised to assimilate results and future requirements of the monitoring research. Based on presentations, discussion and a monitoring game (see http://www.geo-informatie.nl) the research results were presented to the GIS community in the Netherlands and input was obtained for the desired follow-up of the project.

Step 3. Make data suitable for monitoring in an early stage during data capturing process
Rather then to deepen the experiences obtained during step 2 of the research, it was decided to step back to the process of data capturing in order to try to improve monitoring results based on geo-information. The objective was to investigate if the problems encountered in step 1 and 2 could be avoided by structuring geo-databases in another way and to adapt the operators' procedures while constructing the database.

Step 4. Discover new techniques
Simultaneously to step 3 it was investigated if new techniques developed in the IT/database environment could have an added value to the monitoring process with geo- information. We focussed mainly on 'Knowledge based Knowledge Discovery Techniques' (KDD) in order to derive time dominant rules from the data itself.

The use of a case study area and available national databases
A case study area in the Netherlands was selected meeting the following criteria:

- Multi-temporal availability in relevant national spatial databases;
- Representative in it's different types of land use;
- Large enough to be representative, small enough to collect field data for a reference data set;
- Availability of additional (field) data.

Topographic map 32 West, covering mainly the municipality of Soest, was selected as the area meeting these requirements.

1.4 Content of this report

After the background of the study has been clarified in chapter 1, in chapter 2 the theoretical background of adding temporal aspects to geo-databases is being presented. This chapter is conceptual in its' approach and forms the basis of the research approach as presented in this report. In chapter 3 a description is given of the case study area and the used geo-databases. The chapters are written by Kees de Zeeuw, except for section 2.3 which was written by Monica Wachowicz. Chapter 3 is based on articles published in congress proceedings and scientific papers (de Zeeuw, 1998; de Zeeuw et al., 1999a; de Zeeuw et al., 2000).

Chapter 4 focuses on the first step of the research process, the multi-temporal overlay technique, and is based on work done by Rene Meijners and Kees de Zeeuw (Meijners, 1998; de Zeeuw et al., 1999a)

In Chapter 5 the research on improving the GIS methodology (step 2) is presented as done by Marthijn Sonneveld and Jeroen van den Brink in 1999. The chapter is based on previously written articles under supervision of Kees de Zeeuw and Arnold Bregt (de Zeeuw et al., 1999b; Sonneveld et al., 2000).

Chapter 6, written by Gerard Hazeu, deals with the process of building a geo-database (step 3). The possibility of making layers of changes in a geo-database during the data capturing process is investigated.

Chapter 7 was written by Monica Wachowicz and it describes the results of applying knowledge discovery techniques for mining association rules in geo-databases. The advantages and disadvantages of using data mining for the purpose of monitoring land use changes is investigated.

In chapter 8 conclusions are presented and recommendations are given, focusing on both considered use of geo-information for monitoring land use and requirements for further research and methodological developments.

2 Theoretical background of multi-temporal GIS

The purpose of this chapter is to charcoal the available conceptual GIS theories with respect to the use of geo-information (GI) in multi-temporal analysis and to explain the terminology used in the chapters clarifying the different research steps.

2.1 Representing the real world in geo-databases

The success of using a multi-temporal geo-data set for monitoring depends on the monitoring requirements on the one hand and the suitability of the data set for monitoring on the other. Aalders (1998) provides a framework for the nominal ground in the creation and use of geo-data sets as represented in Figure 2. The concept of nominal ground is introduced to avoid confusion between real world objects and their data representations. It is based on the set of specifications used for the data capture and on the region of the world that must be captured (Vaughlin, 1998).
The nominal ground corresponds with "what should have been produced" while the geo-data set relates to "what has actually been produced". Data quality can be measured by assessing the discrepancy between two data sets: the set to be used and a second set of higher accuracy (Vaughlin, 1998).

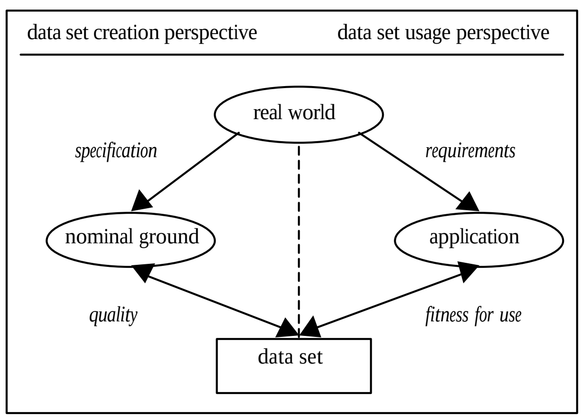


Figure 2: Framework for using the nominal ground (after Aalders, 1998).

Real world entities are represented as internal objects in a database (Bishr, 1997; Fritsch and Anders, 1996) and this representation is performed in either a field approach or an object approach (Molenaar, 1998). The present study made use of an object approach to describe phenomena in the real world and considered real world entities as area objects (polygons) with attribute values and vector geometry. Although the assumption of a one-to-one correspondence between real world objects and their symbolic representations in the database environment is weaker in relational databases than in truly object-based systems (Kent, 1991; in Bishr, 1997), objects can still be represented by means of an identifier to which thematic and

geometric data are linked. Following Molenaar (1998), we regard our approach as object-structured instead of object-oriented, to avoid confusion with modern concepts and techniques in computer science (see also Fritsch and Anders, 1996).
Thematic information is linked to the objects in the form of a list of attributes. An attribute specifies the thematic class to which an object belongs. By convention, all objects belong to some (thematic) classes and each object belongs to exactly one class. The relationships between the classes and the geometric elements of the map can thus be found through the objects. The relationship between classes can be found through a class hierarchy, on which classes are sub-sets of superclasses (Molenaar, 1998).

Monitoring does not necessarily have to focus on the class level but may also concentrate on the superclass level. The choice for the thematic level for monitoring is mainly a consequence of user demands.

Attribute values can also be used as selection criteria for the importance of objects (Molenaar 1996, 1998). A rank order selection of objects may thus help to differentiate between objects of various significance.

The poor quality of multi-temporal geo-data sets for monitoring is also partly due to uncertainty over the characteristics of land use objects. Some land use objects in geo-data sets are fuzzy in the sense that it is not absolutely sure whether they are correctly mapped. Three types of uncertainty are involved when dealing with spatial objects; uncertainty in the assignment of objects to object classes, of attribute values to objects and of spatial description to objects (Molenaar, 1998). Uncertainty in the assignment of object to object classes is largely related to class membership functions: an object may belong to a class with a fuzzy class membership function M, where $0 \leq M[0,C] \leq 1$ (Molenaar, 1998). Fuzzy object geometry is expressed in uncertain spatial boundaries of an object, resulting in a large number of smaller polygons ('slivers') when a topological overlay is applied.

2.2 Setting up a spatial data set

To set up a spatial data set, usually three steps are made (consciously or unconsciously):

- Choose a spatial data model;
- Map the land cover;
- Document characteristics.

2.2.1 Spatial data modelling

A static (mono-temporal) geodata set is being defined using four conceptual choices: the model, the data structure, the classification method and the classification hierarchy (de Bruin, 2000).

The conceptual model for geo-information can be twofold according to Molenaar (1998). It can be a discrete object model where space is composed of well-defined spatial homogeneous entities or it can be a continuous field model where space is considered a continuum. De Bruin (2000) adds a third possible model called the fuzzy sets, in which classes have no sharply defined boundaries. The spatial data structure can be defined as raster data or vector data. For raster data points of lattice or cells can be used. For vector based structures point, line and polygon features can be defined which can be used in a combined way. The classification method can result in two types of classes: Crisp classes or fuzzy classes, dependent on the applied data model and spatial data structure. Finally the classification hierarchy can be a class driven aggregation ("is a …" qualification) or a functional aggregation ("part of …" qualification). Figure 3 shows a diagram of the possible choices for setting up a spatial data set. As an example the spatial data model as applied for the LGN database of Alterra is indicated in this figure.

Conceptual model *__D__iscrete* *Continuous* *Fuzzy*
/ \
Data Structure *__R__aster* *Vector* … …
/ \
Classification method *__C__risp* *Fuzzy*
/ \
Classification hierarchy *__C__lass driven* *Functional aggregation*

Figure 3. The process of spatial data modelling. As example, in bold face the combination of choices as applied for the national land use database LGN are indicated (D-R-C-C). The hierarchy tree for continuous and fuzzy models is similar tot the discrete conceptual model, but not showed in this figure.

2.2.2 Land cover mapping

Once the conceptual framework is established the actual land cover mapping consists of the data capture and land cover classification (de Bruin, 2000).

The data capture can be realised doing a field inventory or survey. In the case of land use inventories this is normally done by fieldwork in combination with a base geometry. The actual land use is classified by adding attribute values to mapped entities. At this stage the distinction between land use and land cover is an important issue, effecting the database significantly (Lindgren, 1985).

On the other hand land cover boundaries and land cover types can be derived from remotely sensed data (aerial photography or satellite imagery). Land cover classification derived from spectral multi-band information is proven technology and common practice (e.g. Noordman et al., 1997). However, better classification results can be obtained by applying stratification, multi-temporal analysis within one growing season and the combination with ancillary geo-information (de Wit et al., 1999).

2.2.3 Documentation of characteristics (metadata)

Spatial information without additional information documenting the data characteristics (metadata) is difficult to re-use. The metadata should give details on:

- Data set identification;
- Administrative information;
- Quality parameters (uncertainty, inaccuracy and precision).

For the format of metadata various standards have been developed. However, even in most recent standards (e.g. ISO, CEN, NEN, OpenGIS), little attention is given to temporal characteristics of data and versioning attributes of single objects.

2.3 Adding the temporal dimension to a spatial data base

2.3.1 Spatio-temporal data modelling

One of the most common findings in literature is that time is just one more dimension to be added to the spatial dimension. This perspective is indeed the underlying rationale behind most of the implementations of spatial and temporal database models. However, the synergy of space and time requires "*spatio-temporal concepts*" that represent space-time dynamics (for example, patterns and process of change) and the "*human cognition of a knowledge domain*" (for example, distinction between observed spatio-temporal patterns and derived knowledge). Adding the time dimension to a spatial data model is inadequate for representing space and time in databases. Mainly because this will result in a database model that represents the time dimension in the same manner as the spatial dimension, and as a result, it may only capture time-referenced sequences (snapshots) of spatial data.

Spatio-temporal database models usually support the representation of *classes, states, events, and episodes* within a database (Wachowicz, 1999; Kucera, 1996; Peuquet, 1994). A class is a set of objects containing similar properties. A state represents a version of what we know about an object in a given moment. An event is the moment in time an occurrence, action, or observation takes place. Events and states are also part of a process of change caused by the passage of time. In this process, an episode is the length of time during which change occurs, a state exists, or an event lasts. Consequently, spatio-temporal database modelling is about using modelling abstractions such as classes, states, events, and episodes to explain changes that have occurred in the real-world (Kucera, 1996; Frank, 1994; Egenhoffer and Al-Taha, 1992; Langran, 1989). This requires a good understanding of how certain processes are related to an observed pattern. Generally, there is a better understanding of the range of types of patterns (e.g. types of land use) rather than the process responsible for their generation (e.g. fire, urbanization, climate change). Classes, states, events, and episodes are modelling abstractions that can be used for representing both patterns and processes in a spatio-temporal database model.

The key issue here is to develop a methodology, which can help us to understand the spatial, temporal, and thematic aspects of the pattern-process paradigm. There is not always a 1:1 mapping of a given pattern to a given process. Different processes can produce the same pattern. From a pragmatic perspective, it is essential to be able to explore databases for finding patterns and processes of change in such a way that we can dynamically define classes, states, events, and episodes within a spatio-temporal database model.

However, current databases only support the *absolute view of space*, which considers space as finite, homogeneous, and isotropic, with an existence fully independent of any object it might contain. Time is implicitly incorporated into the spatial data model every time some sort of change occurs. As a result, a snapshot of a layer is created every time an update occurs. A sequence of snapshots describes the passage of time. In this approach, it is not possible to know how an updated layer might affect other associated layers of the same geographic space. Current databases support two types of a layer based model, i.e. raster (field) or vector (object) models. These models present spatially depicted classifications of objects grouped into layers or sets in time. The geographic space is grouped along the spatial dimension after some sort of categorisation, and time is grouped along the time dimension after some sort of periodisation. The analysis is carried out based on similarity or dissimilarity between layers (aggregations) at different points in time. The LGN database is one example of this approach.

Peuquet (1994) points out that absolute space is objective since it gives us an immutable structure that is rigid, purely geometric and serves as the framework in which objects may or may not change (change- or update-based scenario). This is probably the reason why most of the GIS products adopted the space-dominant view within their database models.

2.3.2 Multi-temporal land cover mapping

Currently in GIS, for each possible update procedure, a change is associated with the geometry, topology, and thematic properties of an object in space. Kucera (1996) has also advocated the need for developing data-driven update procedures in GIS; procedures based on where and when the changes occur.

However, it is the *relative view of space and time* that is the most fundamental importance for representing space and time in database models. The concept of relative space is more general and empirically more useful than the concept of absolute space. Jammer (1969, p.23) defines relative space as "an ordering relation that holds between bodies and determines their relative positions… a system of interconnected relations.' The profound implication is that any relation defined on a set of entities creates space. In other words, defining a relation automatically defines a space. Harvey (1969) provides an excellent review of the two perspectives, absolute and relative space. The concept of absolute space overemphasises the absolute location of entities within a spatial data model. In contrast, relative space focuses on the relative location among objects and events. This relativistic point of view is usually associated with studies of forms, patterns, functions, rates, and diffusion processes.

A complementary concept is relative time - time measured in relation to something, not constrained to a single dimensional axis. Cyclical time - the repeating of a day, week, or year - is an example of relative time. In absolute time, 15 June 2000 cannot be repeated. But in relative time, Thursdays keep returning. Most questions about change will be understood from this perspective (Ornstein, 1969). Relative time is subjective since it assumes a flexible structure that is more topological in the sense that is defined in terms of relationships between events. For example, Frank (1994) suggested an ordinal model of time in which an episode is defined according to relativity among events of a time-line rather than attaching precise dates for these events.

The relative space-time view embraces human activity over the landscape that results from studying land use changes and processes within a knowledge domain. A *process study* seeks to identify the rules that govern spatio-temporal sequences, in such a form that the patterns are interpretable in terms of the results of the sequence. This is particularly relevant when applying knowledge discovery methods for monitoring land use changes (Wachowicz, 2000c).

Unfortunately, we have frequently taken a particular conceptual view (i.e. absolute view or relative view) for constructing a data model without examining the rationale for such a choice. After all, we should not discriminate one from the other. They are complementary. The absolute view requires some sort of measurements referenced to a constant base, implying non-judgmental observation. The relative view, on the other hand, involves interpretation of processes and the flux of changing patterns within a knowledge domain. However, a question still remains about integrating absolute and relative views. How can we have both perspectives placed in the same spatio-temporal database model?

3 Characteristics of case study

The monitoring research as applied in this report can be best described as 'GIS-Monitoring Land Use'. In this research the three most prevailing national databases have been used: a database based on the national topographical map, a database based on national land use statistics and a database containing the national land use data. At present, only these three national databases contain data on actual land use and/or land cover. All these databases are multi-temporal and are updated with a frequency of once per four years. Instead of performing a full database analysis, a representative study area was chosen to establish the suitability of different databases for monitoring. A reference data set was constructed for the representative study area to validate the monitoring results for the different databases.

3.1 The municipality of Soest as study area

The municipality of Soest was chosen as study area for several reasons. First of all, the municipality was covered multi-temporally in all databases. Secondly, Soest was considered representative for the Netherlands in rural, urban and natural areas. Finally, the willingness of the local authorities to co-operate in the study was important. In Map 1 the location of the municipality is shown.

Soest is a municipality with about 44,000 inhabitants, covering an area of 4,615 hectares. It contains the residential nucleus Soest, Soesterberg and Soestduinen.

Map 1. Location of Soest.

3.2 Databases

The three national databases considered in this study will be treated in the sections below. Figure 4 shows a detail of these spatial databases to illustrate the different types of information in the data sets.
The constructed reference data set is also treated in this section. Monitoring results obtained with the three multi-temporal national land use databases can be compared with the real 'world' changes determined with the reference data set. This comparison will give an idea of the usefulness of these multi-temporal databases in determining real land use changes.

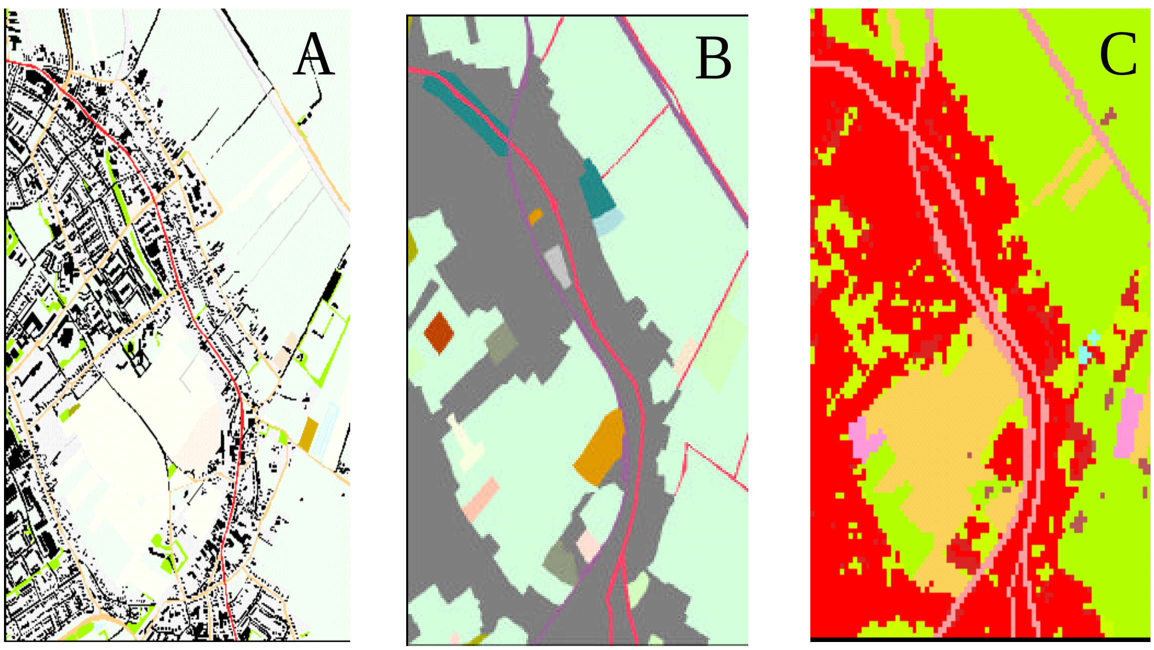


Figure 4. A map detail within the municipality of Soest of respectively the Top10vector (A), CBS-land use statistics (B) and LGN (C).

3.2.1 Land use statistics of Central Bureau of Statistics

The national Land Use Statistics (LUS) are collected by the Central Bureau for Statistics (CBS) and have been stored in ArcInfo format since 1989. The scale of the polygon file is approximately 1:50,000. The database discerns 35 classes of land use. The database contains information on land use, but is mainly focused on urban functionality and the legend is very limited in rural classes. For the first update in 1993 the 1989 version of the database was used as a starting point.

3.2.2 National topographic map: Top10vector

The topographic map in digital form is available at a scale of 1:10.000. The database is denominated Top10vector and is available in ArcInfo format. The database content is based on fieldwork and air photo interpretation (Topografische Dienst, 1995).
The semantic information in the database refers mainly to land cover classes. Functional characteristics are not attached as label to the entities. For the purpose of this study only the surface elements (polygon objects) are used, including the layer with information on buildings. Annotation, point elements and line elements are not considered.
The topographic map 32 West, which was first created with data of 1991, covers the study area. In 1996 the database was updated for the first time based on new ancillary data of 1995 rather than on the previous version of the database.

3.2.3 National Land Use database (LGN) of Alterra

The LGN database distinguishes five main classes and twenty-five sub-classes of land use. The main feature of the database is the differentiation between agricultural crops. It is a raster database that consists of elements of 25m x 25m, covering the whole of The Netherlands (Noordman et al., 1997).
Alterra actualised the National Land Use database (LGN) for the second time (de Wit et al., 1999). The database has a raster format with cells of 25m x 25m. The most important feature of this database is the possibility to discern agricultural crops. The geographical database is derived from remotely sensed data sources. Multi-temporal data sets from SPOT (20m x 20m resolution) and Landsat-TM (30m x 30m resolution) have been used. The applied classification procedures were partly automatic, partly manual. Classes that are difficult to distinguish from remote sensing data are incorporated with the aid GIS techniques from other data sources (like roads, houses and glasshouses). In 2000-2001LGN3 in combination with satellite images of 1999 and 2000 are being used as a basis for the development of LGN4.
The first release of the LGN database dates from 1986. This version has a different legend structure and accuracy level than the second (1992), third (1996) and fourth (2000) version of the database and is therefore not further considered in this study.

3.2.4 A reference data set

A reference data set for the Soest study area was constructed to have ‘real world’ land use data for a considerable period. The reference data set describes the yearly land use between 1989 and 1997. The description of land use takes places at sub-class level (Table 2) and at scale 1:10.000. For validation and comparison the reference data set needs to meet the requirements of a desired data set for monitoring purposes:

- At least recordings at two points in time;

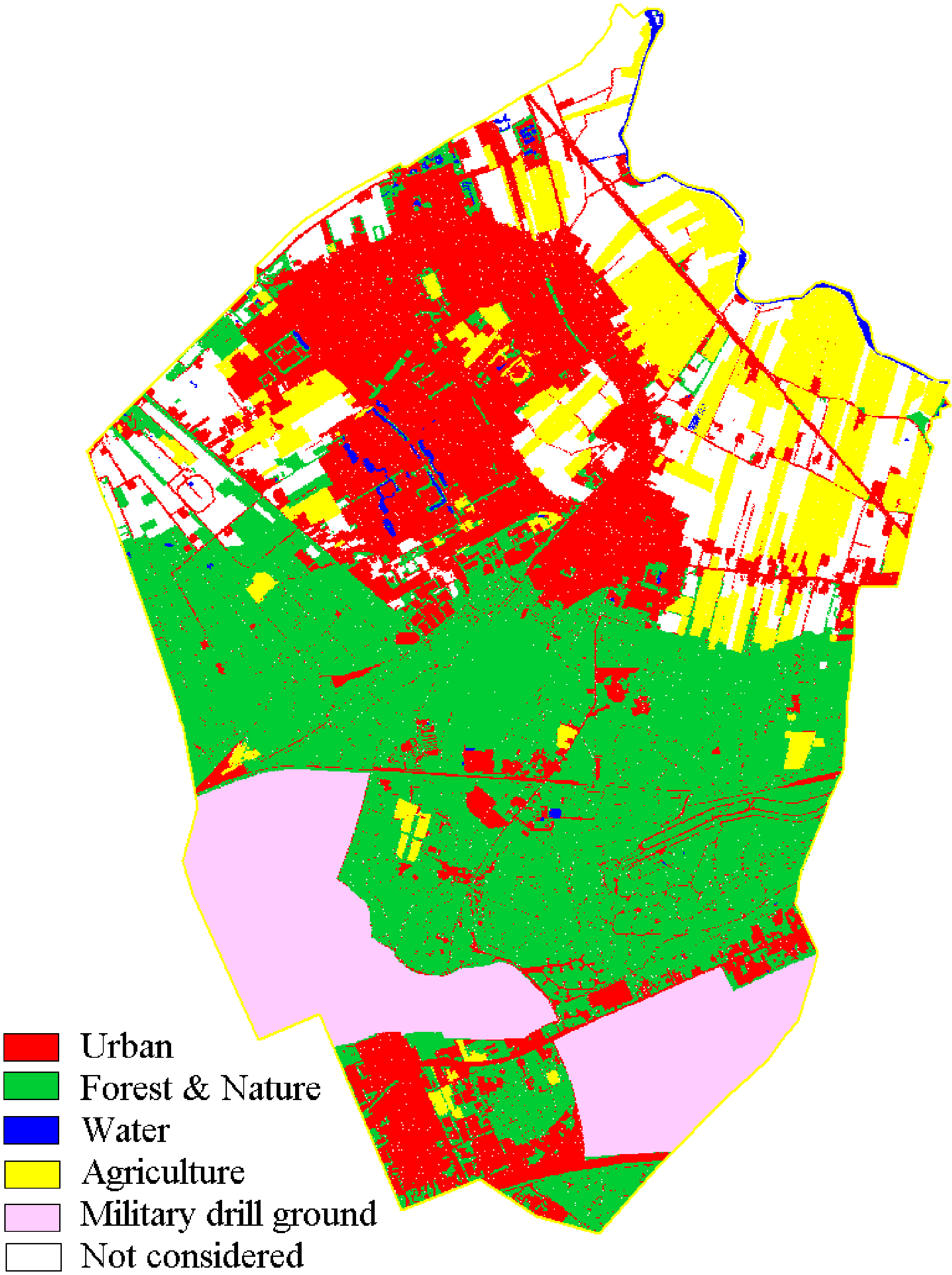
- The recordings should be performed in like-manner;
- The typology of the recordings should not change in time;
- The scale of the data set should be the same.

According to Dijkstra and Roos-Klein Lankhorst (1995) two types of monitoring exist: quality focussed monitoring and problem focussed monitoring. The reference data set as used in this study is focussed on quality. This means that the characteristics of the parameters are of importance instead of causes and consequences in a chain of processes.
For the period 1989 – 1997 a reference data set was constructed containing a yearly representation of summer crop situation. The reference data set was made with the topographic map of 1996 (Top10vector) as geometric base. Additional information was obtained through aerial photo interpretation (ortho-photos of 1989, 1991, 1994 and 1997 were available). Also information from existing detailed studies and maps was used and a three weeks fieldwork campaign was done. A representative sample was taken in the rural area covering 40% of this area. For the urban areas, natural areas and forests an inventory was done covering the study area fully. A legend with three hierarchic levels has been developed. In Table 1 the statistics on land use during the analysed period are given, at the level of superclasses. A cartographic representation is given in Map 2.

Table 1. Land use statistics (in hectares) of the reference data set, from 1989 till 1997 at superclass level.

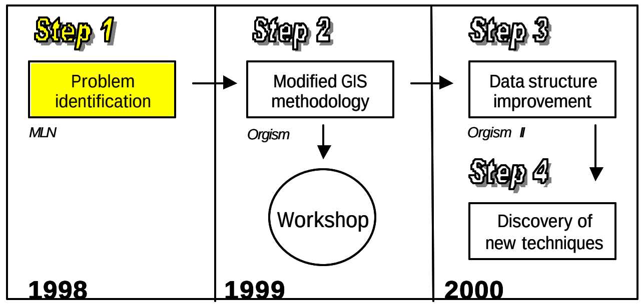
Year	Urban area & infrastructure	Forest & nature	Water	Agriculture
1989	1355	1512	29	1135
1990*	1359	1512	29	1131
1991	1363	1510	29	1128
1992	1367	1508	29	1126
1993	1366	1510	29	1125
1994	1371	1509	29	1121
1995	1391	1509	30	1101
1996	1405	1507	30	1089
1997	1412	1504	30	1084

* 1990 is not fully interpreted, because no data set was required for this year.



Map 2. Reference data set for 1995 at the superclasses level.

4 Multi-temporal Overlay Technique



4.1 Methods

Research approach

First of all, databases and a reference data set for the study area were created. Secondly, for every multi-temporal national database the overlay technique (simple GIS functionality) has been used to generate a database containing the changes in land use between two years. It was analysed what the different databases are giving as monitoring result using their own original legends. This was done both at regional and local scale (region and point wise approach). The results are compared with the reference data set. Thirdly, transition keys were developed to translate the different legends towards a common legend structure (ontology) at three different hierarchic levels. Using this common ontology the data sets were compared mutually and with the reference data set, both locally and regionally. In total three data sets at two different scale levels and three different levels in classification hierarchy were analysed, resulting in 18 sets of output results. Afterwards, the possibilities of a combined use of spatial databases were investigated to obtain the optimum use of databases for monitoring at a region wise approach. Based on the results of a thorough literature review, causes and types of noise were identified. This insight is required for further development of a more sophisticated methodology of GIS-monitoring land use. Figure 5 shows a schematic representation of the research steps.

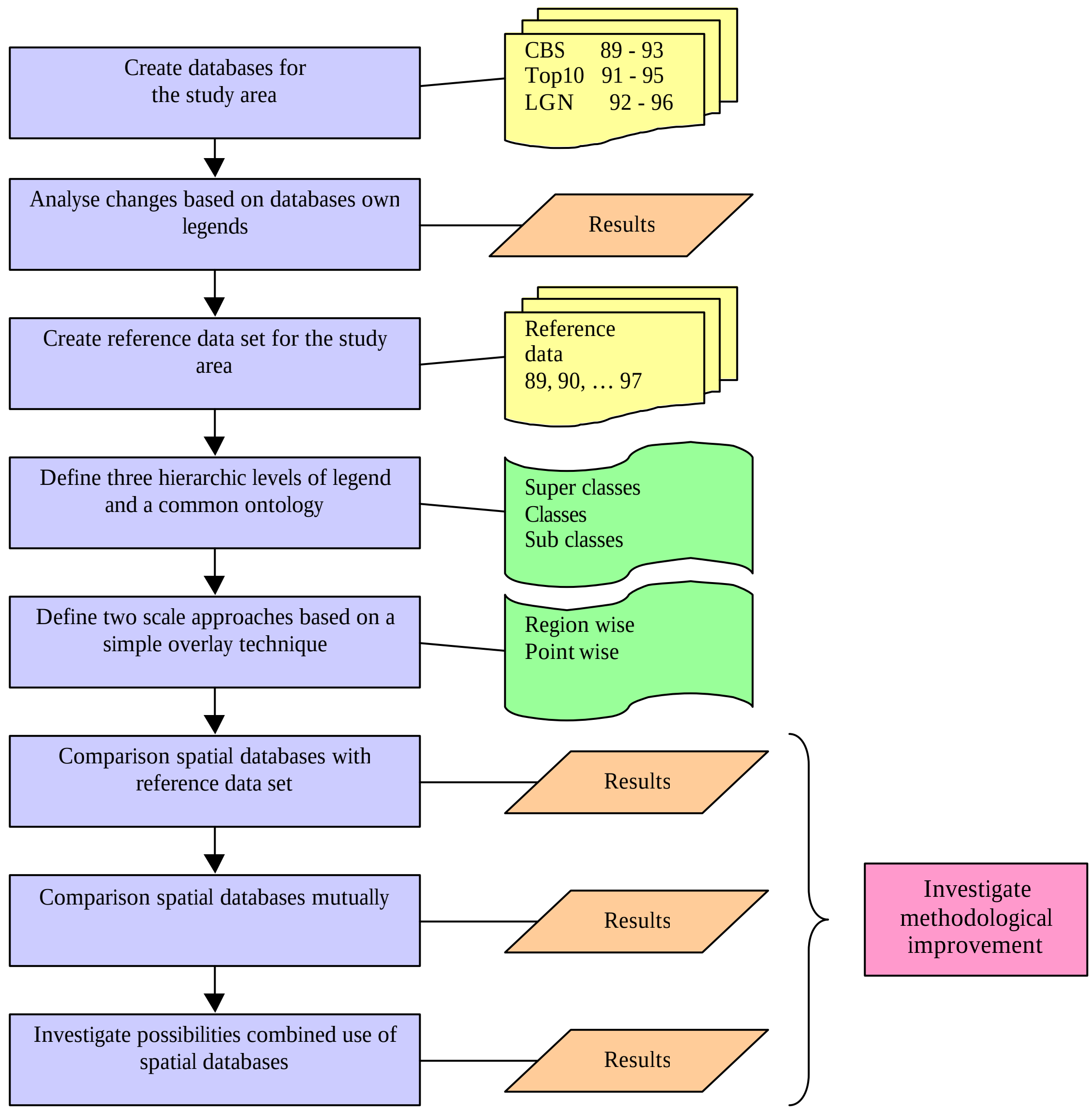


Figure 5. Schematic representation of the research approach (step 1).

Legend hierarchy

A reference data set was constructed containing yearly information on land use in the period 1989 – 1997. Of the three national geographical databases at least two versions were available in this period. To be able to compare the suitability for monitoring of the different databases, a common legend for these databases was constructed. Three levels in classification hierarchy were specified on basis of the reference data set making monitoring at different thematic levels possible. The resulting 'common ontology' should be considered as the optimal compromise of three legends rather than a perfect fit. Semantic, syntactic and schematic heterogeneity are limiting a possible combined use. Semantic heterogeneity occurs due to differences in category definition, class definition and geometric descriptions, while syntactic heterogeneity depends on the thematic and geometric representation as well as the topological relation of spatial objects. The schematic heterogeneity is the difference in the class hierarchies and attribute structures (Bishr, 1997). Heterogeneity is increasing with legend detail. The derived hierarchic legend structure is shown in Table 2. Without pretending correct denomination, the three levels are denominated to superclasses, classes and sub-classes. The three different land use levels used in the legend are considered as generally accepted at the user level.

Two scale approaches

Two different approaches can be used when determining changes in land use:

- Region wise approach;
- Point wise approach.

At the regional level no importance is given to the position of different types of land use. Only the statistical appearance at regional level is considered. Changes in total area of land use are independent of location.
At the local scale the position is indeed of influence to the monitoring result. Land use change at every point is taken into account in the point wise approach. Figure 6 shows an example of the differences between these two approaches.

Table 2. The three levels of the legend for the reference data set.

Superclasses	Classes	Sub-classes
Continuos urban area	Built-up area	Buildings City-grounds Other urban use
	Infrastructure	Infrastructure
Forest & nature	Forest	Coniferous Deciduous Mixed Withy ground
	Nature	Heath Sand
Water	Water	Water
Agriculture	Horticulture	Orchard Tree nursery Glasshouses
	Grassland	Grassland
	Farming	Maize Grains Beets Other crops

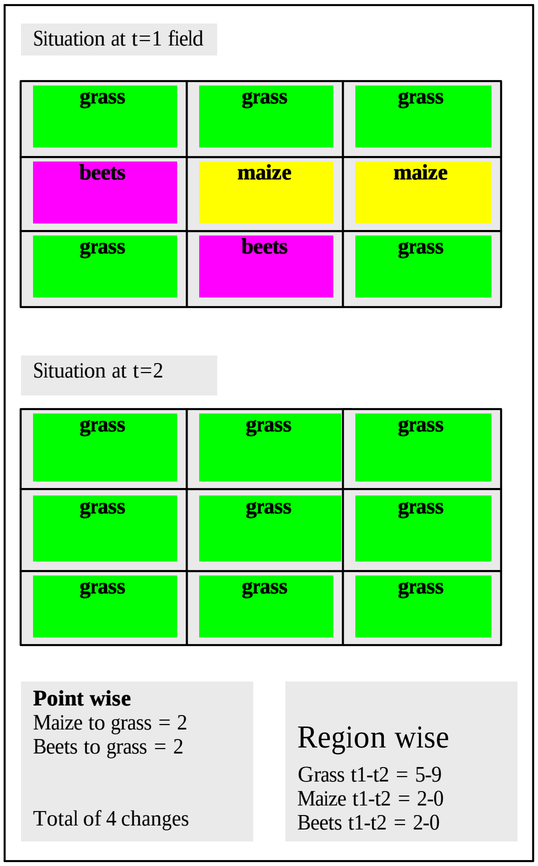


Figure 6.a. Fictive spatial situation where the total change calculated point wise is smaller than region wise.

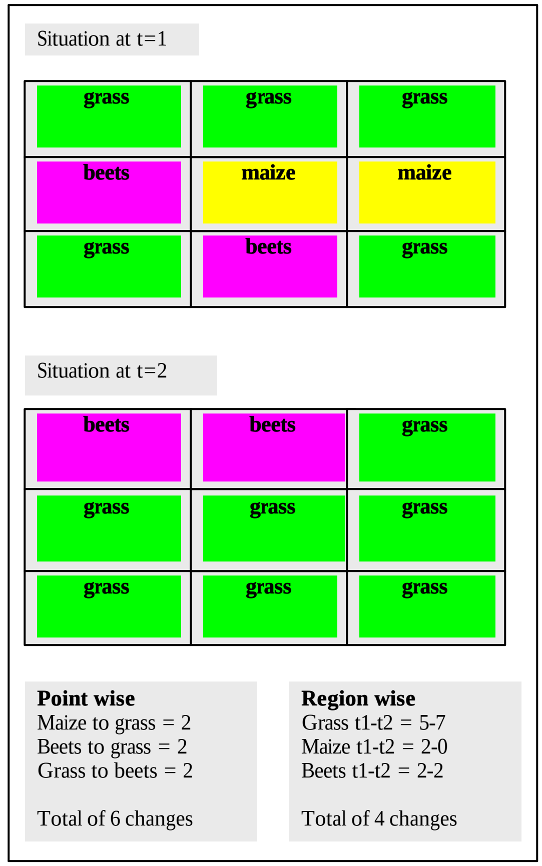


Figure6.b. Fictive spatial situation where the total change calculated point wise is larger than region wise.

As a consequence of this approach, at the regional scale it is not possible to obtain information on the type of changes that occurred during the period of monitoring. In other words, it cannot be observed what classes changed to what classes. At the local scale however, this can be analysed. The overlay operation between the two databases results in a new database (or map) containing so-called ‘classes of change’. Using these classes of change, conclusions can be drawn about the suitability of classes and groups of classes for monitoring purposes.
The variable of frequency (multi-temporality) was not investigated as the available databases had been updated only once. Therefore it is not possible to compare the monitoring over a four years cycle to an eight years cycle for example.

4.2 Results

In Table 3 the percentages of change are given in case the original legends of the databases are used. These figures refer to a four years period. The CBS-land use statistics shows minor changes compared to the other databases. Although a direct comparison between the databases is not possible as time spans and legend content vary, the magnitude of changes resulting from the CBS-land use statistics is the most realistic compared to the reference data set.

Table 3. Changes (% of total area) in the databases, considering their own legends.

		Region wise	Point wise
Top10vector	1991 – 1995	14,9	17,1
CBS – LUS	1989 – 1993	2,0	1,1
LGN	1992 – 1996	15,1	11,8

Table 4. Results for the region wise approach. The overall percentage of change in the database is given. The rural area is fully covered by means of a statistical representative sample. Between brackets the percentage is given representing changes in reality according to the reference data set.

	Top10vector 1991-1995	CBS-land use statistics 1989-1993	LGN 1992-1996
Superclasses	6,5 (1,4)	1,4 (0,6)	7,2 (2,0)
Classes	6,6 (2,0)	-	7,2 (3,3)
Sub-classes	-	-	-

Table 5. Results for the point wise approach. The overall percentage of change in the database is given. For the rural area only the census data are used (no statistical representative sample is applied). Between brackets the percentage is given representing changes in reality according to the reference data set.

	Top10vector 1991-1995	CBS-land use statistics 1989–1993	LGN 1992-1996
Superclasses	7,4 (1,4)	0,9 (1,0)	4,2 (1,7)
Classes	9,7 (2,2)	-	5,3 (2,4)
Sub-classes	-	-	-

In Tables 4 and 5 the results are given in case the legends of the spatial databases are converted to the legend structure of the reference data set. At a more detailed level not all classes could be derived from the national land use databases, explaining the empty fields in the tables (Meijners, 1998).

From these tables it can be concluded that in general the databases over-estimate the changes in land use at all legend levels and scale approaches. Only the CBS-land use statistics at a superclass level approximate the changes as recorded in the reference data. However, the CBS database does not have sufficient distinguishing capacity at class and sub-class level. The over-estimation of changes (noise) for the LGN database can be explained by pixel size, classification of mixed pixels and addition of classes in updated versions. Also the Top10Vector database overestimates the land use changes. Classification techniques for forest areas and farms and topography (small areas along roads and buildings) have a large contribution to the amount of noise measured. For a more elaborated presentation of the results we refer to Meijners (1998).

4.2.1 Regional approach

At the regional scale the overall statistics of the situation at t=2 are compared to the situation at t=1 (see Figure 6). In this case it is not possible to analyse changes between classes. It was only investigated which classes perform ‘acceptable’ at which level in the classification hierarchy (i.e. observed changes agree with changes in the reference data set). The performance is considered acceptable when the majority of the observed changes is explained by changes in the real world and not by ‘noise’ in the databases. The subjective criteria as derived from the data sets are the following (Meijners, 1998; de Zeeuw et al., 1999a).

- Per legend unit the suitability for monitoring is considered acceptable in case a deviation to the reference data set of maximum 1%, 0,5% and 0.25% is observed at respectively the level of superclasses, classes and sub-classes.
- To classify the complete legend as suitable for monitoring purposes deviations should not pass 2%, 4% and 8% at respectively the level of superclasses, classes and sub-classes.

Table 6. Results of the region wise scale approach. The first number (X/x/x) is the total number of classes. The second number (x/X/x) is the number of classes that could be analysed at that specific level in classification hierarchy. The third number (x/x/X) represents the number of classes for which the performance was considered acceptable.

	Top10vector	CBS-land use statistics	LGN
Superclasses	4/4/2	4/4/4	4/4/2
Classes	8/8/6	8/6/5	8/8/5
Sub-classes	19/15/8	19/7/4	19/16/11

Table 6 shows how many classes could be analysed and how many classes performed well at the different levels in classification hierarchy, in case the analyses were performed region wise (de Zeeuw, 1998). It is shown that in general the performance for the monitoring of land use is rather poor. Especially if a complete legend has to be obtained through one single data-source, even at the superclass level the results are not convincing, with the exception of CBS-land use statistics.

Theoretically it would be an improvement if the available databases could be used in a combined way. However, in practice this is not possible because the different databases represent the situation in different years (see also Table 3). Therefore, a synchronisation technique as proposed by Wijngaarden et al. (1998) is not valid for these data sets. This implies that the same classes derived from another database cannot simply replace classes performing poor in one database. As an example, in Table 7 the theoretic optimum combined use of the databases at the regional scale is shown by means of colour shadings. However, it should be realised that by applying this theoretic approach, unsatisfactory results still would be obtained.

Table 7. The hierarchical structured legend of the reference data set. The theoretic optimum combined use of the different databases at a region wise approach is indicated by colour shadings: light shading = Top10vector, moderate shading = CBS-land use statistics, dark shading = LGN.

Superclasses	Classes	Sub-classes
Continuous urban area & infrastructure	Built-up area	Buildings
		City-grounds
		Other urban use
	Infrastructure	Infrastructure
Forest & nature	Forest	Coniferous
		Deciduous
		Mixed
		Withy ground
	Nature	Heath
		Sand
Water	Water	Water
Agriculture	Horticulture	Orchard
		Tree nursery
		Glasshouses
	Grassland	Grassland
	Farming	Maize
		Grains
		Beets
		Other crops

4.2.2 Point wise approach

Using a point wise approach it can be analysed what changes took place. The so-called 'classes of change' are the result of an overlay operation with the aid of GIS software. Theoretically, at the sub-classes level, 19^2 classes of change could be generated (= 361 classes). In this approach no statistical representative sampling technique is allowed, as the geographical position of the classified entities is of importance for the result. Only recorded data truly covering the area (census data) can be used for validation. Therefore the sampling area has been reduced for this approach from 4030 hectares to 3391 hectares.

Almost without exception the databases overestimate changes in land use compared to the real world situation. Similar analyses have been done at the classes and sub-classes level. The deviations observed were larger than at the superclasses level. The quantity of output generated is very extensive and for that reason not fully presented in this paper. It was found that deviations to the reference data set are increasing if a more detailed level of hierarchy is chosen as legend. In Figure 7 an example is given of the classes of change obtained for the Top10vector database at the sub-classes level. Eye catching is the large deviation for the class of change "Grassland >> other urban use". This deviation can be mainly explained by the difficult translation of the land cover class 'grassland' to the land use classes 'urban use', 'grassland' and 'city grounds'.

In general it was found that monitoring results for all databases at all levels in classification hierarchy overestimate the changes considerably.

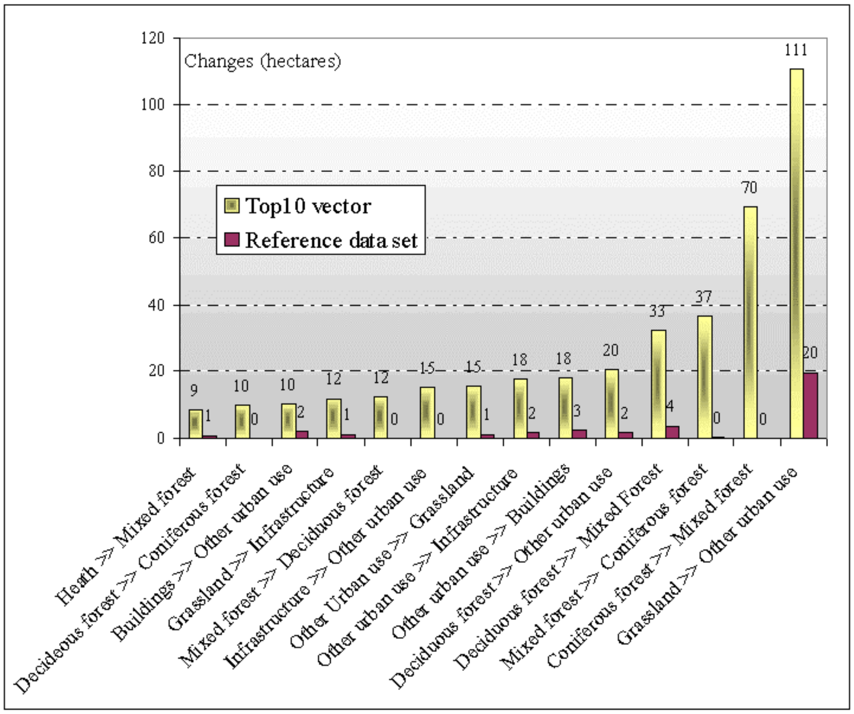


Figure 7. Example of 'classes of change' as derived for the Top10Vector database compared to the reference data set at the level of classes in the period 1991 – 1995.

4.3 Discussion and conclusion

The monitoring results for the three national land use databases have shown that the databases cannot be used for monitoring purposes by applying a simple overlay technique. The majority of changes observed were explained by the noise in the databases rather than by changes in the real world. It should be emphasised that this conclusion only relates to the suitability for monitoring land use and it is not a judgement on the reliability and accuracy of the various databases as such.

At the superclass level and based on its own legend structure, only the CBS land use statistics database had an acceptable performance. But if more detailed thematic classes are required, the Top10vector database presented better results in certain classes. If detailed information on agricultural use is required, the only database available is the LGN database. However, this database did not give a realistic impression of the changes in land use, especially in a point wise approach.

An optimum combination of national land use databases can be found. However, even this combination does not meet the requirements of a noise free monitoring result. Also a practical limitation was found in combining the different data sets, as the update cycles of the different data sets is not the same.

The poor suitability of the spatial databases for the monitoring of land use using a simple overlay technique is mainly caused by changes in the applied recording techniques. In general, the update of a database not only comprehends the processing of observed changes in the real world, but also implies an improvement of reliability and accuracy of the databases. In many cases, also the semantic and scale characteristics are changed, in an effort to meet the changing user requirements.

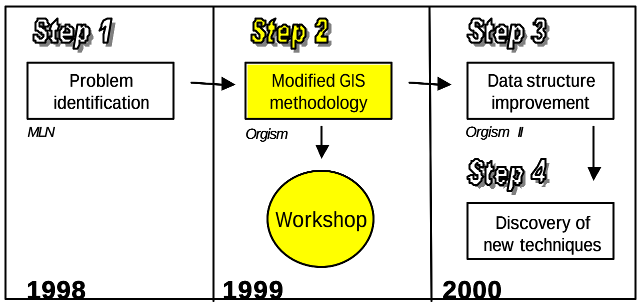
Another important factor influencing the suitability for the monitoring of land use are the difficulties encountered in defining a common ontology for the different data sets. Also the conversion of land cover classes to land use classes introduces an error.

It is assumed that the next updates of different databases will be more consistent in its methods and characteristics. However, as long as these databases are not made for monitoring purposes it is not expected to reach a satisfactory level on the short term. A combined use might improve the results with future updates, but in that case a closer co-ordination between the data providers is required in operating simultaneously concerning the update cycle and data characteristics.

Other possibilities to improve the monitoring results can be twofold. The development of a GIS methodology that improves the data processing side or adaptation of the way data are captured for a new database version.

5 Improved GIS methodology for monitoring

The development of an improved GIS methodology for monitoring is the second step in the monitoring research executed by Alterra. In the text we often refer to the MonGis approach when we are dealing with this improved GIS methodology. MonGis is the name of a developed ArcInfo tool.



5.1 Approach

Using multi-temporal overlay techniques for monitoring results in a lot of unreal changes (noise) caused by data processing and registering of multi-temporal databases. The methodology to improve monitoring results is based on the reduction of this noise. In the improved vector-based GIS methodology presented in this chapter, only the criteria for reliable change, significant overlap and shape factor are user inputs, which can be indirectly derived from the databases meta-data, or empirically from test sets. The resulting outcome is not a noise-free monitoring result but a noise-reduced monitoring result. The magnitude of noise filtered out can be quantified stepwise and is fully dependent on the chosen criteria. The advantage of the method is that no detailed knowledge on the data quality is required and that it is applicable to different databases at different hierarchic levels of legends.

Figure 8 presents the steps in the research to develop an improved GIS methodology for monitoring. It can be seen that the development of the object-structured approach is based on new research and on results of the map overlay approach. The results of the improved GIS methodology are compared with the results of the map overlay technique. The improved GIS methodology for monitoring that has been developed is called the MonGis procedure.

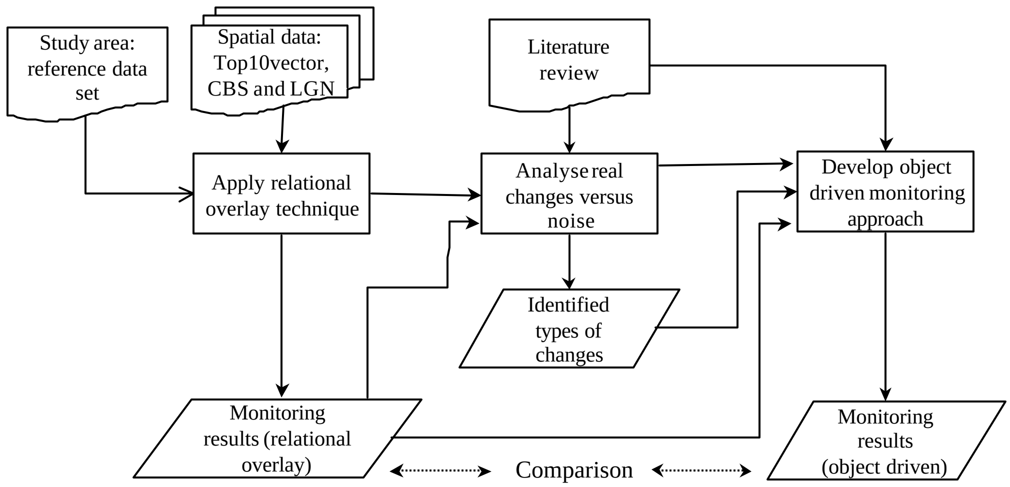


Figure 8. Schematic presentation of the research approach (step 1 and 2). The part to the right represents the steps in the development of an improved GIS methodology for monitoring (step 2).

In order to develop a GIS methodology that reduces the influence of noise, it is important to understand what different types of changes one encounters when multi-temporal geographic databases are compared. Also the decision rules to distinguish real changes from noise should be identified. To do so, an object-structured approach is required as both geometric and thematic characteristics are to be considered (Fritsch and Anders, 1996). Although topological aspects can help in analysing monitoring results (Hendriks and Ottens, 1997), in this study they are not considered. Likewise, only objects with a polygon attribute value on land use have been considered, excluding point and line objects.

The *MonGis* procedure was developed on the basis of the Top10Vector database and can be applied in an ArcInfo environment. This was done because Top10vector has been fully updated by means of new recordings, so that all types of errors may potentially occur in a Top10vector overlay-set.

5.2 The MonGis procedure

5.2.1 Identification and analysis of changes

The method developed not only considers individual objects, but also the relationships between them. An overlay of two databases has been made to recognise polygons at t=1 and t=2 as the same object. Many new polygons, i.e. so-called slivers, are created due to intersections of objects of different recordings (t=1 and t=2). These new polygons result from changes between the two databases. The schematic representation in Figure 9 shows that the following changes for polygon objects have been identified:

- Thematic changes (continuation of geometric appearance);
- Geometric changes (displacement, addition and disappearance);
- Combined thematic and geometric changes (displacement & thematic change, addition & thematic change and disappearance & thematic change).

Three different causes can be defined for geometric changes: displacement, addition and disappearance. Displacement occurs when an object is recognised equal at t=1 and t=2 but its geometry has changed. Addition refers to objects occurring at t=2 but not at t=1. Disappearance resembles the opposite situation where an object occurs at t=1 but not at t=2. In case an object does not change geometrically in time it is referred to as a continuation. The existence of non-changed objects is mainly dependent on the method applied for updating the geographic database (e.g. newly digitised databases versus modification of existing features).

A thematic change appears when the attribute value of an object changes. Depending on the criteria used for reliability and expectation, these changes are regarded as real changes or unreal changes (noise). In case an object changes both geometrically as thematically the observed change should be evaluated on both criteria.

For each of these possibilities, the question can be raised whether the intersection has resulted from real change or whether it can be attributed to noise. A so-called 'stepwise refinement approach for polygons' is developed to deal with this question. In this approach the schematic order of analysis is of importance. As starting point all polygons are considered noise. First, all polygon objects that are classified as thematic displacements are selected. Using predefined threshold values for criteria on the shape factor and overlap percentage it is decided if a change is considered real thematic displacement (results *a* and *b* in Figure 9) or noise. At this stage, a sliver polygon can be a real displacement with respect to multiple objects. This procedure is applied to all polygons that appear in the overlay coverage of the study area. All polygons remaining classified as noise, can still be considered real change in a later stage of the analysis. There-upon all polygons are analysed again in the same manner, but now for the combination of displacement and non-thematic change (results *c* and *d* in Figure 9). The next step is to establish if added polygons are considered added objects (results *e* and *f* in Figure 9) and if disappeared polygons are equal to disappeared objects (results *g* and *h* in Figure 9). Now all polygons that are considered as noise can be dissolved back to the object they belong to in the t=2 situation. All polygons in this resulting map can be considered objects now. The final check is on the polygons that are classified as a (geometric) continuation. If there is a thematic change that is considered reliable it concerns a really changed object (result *i* in Figure 9), otherwise it is a really unchanged object (result *j* in Figure 9).

The decision rules identified for the distinction between real changes and unreal changes are numbered 1 up to 5 in Figure 9. In Table 8 these numbers are explained. Based on the definition of the threshold values required in the decision rules, changes are classified as real or unreal.

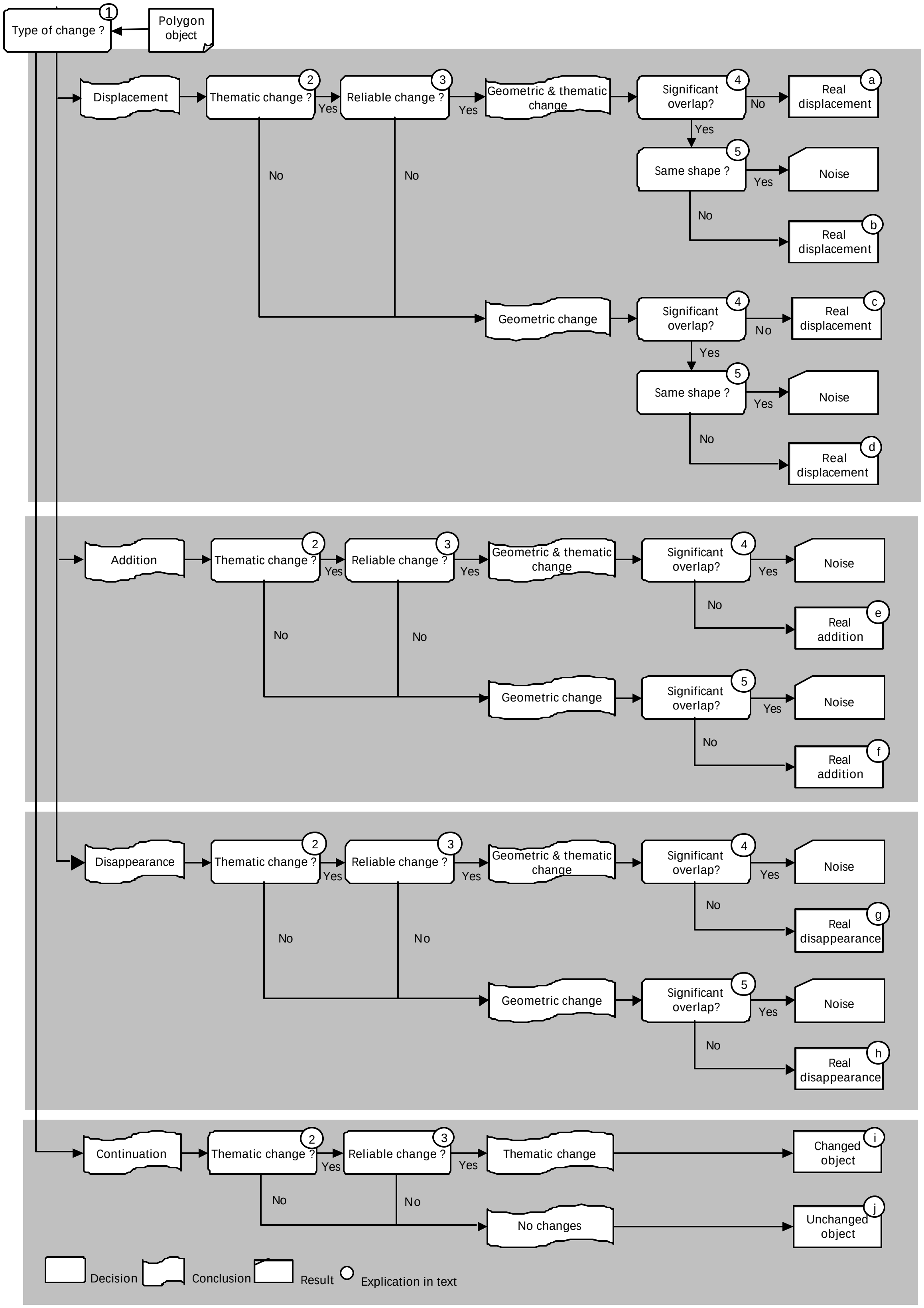


Figure 9. Schematic representation of the different types of changes for a polygon object and decision structure for the distinction between real changes and unreal changes.

Table 8. Explanation of the type of decision rules (see Figure 9) required to distinguish real changes from different types of noise that appear in the monitoring result using geographical databases.

N°	Decision on	Explication
①	Type of change	For every object in map t=2 it is checked if the object already existed in map t=1. If the same geometry is encountered the object is classified as “Continuation”. If an object is recognised but geometry changed it concerns “Displacement”. If it concerns a new object it is classified “Addition” while in the opposite case it is classified as “Disappearance”.
②	Thematic change	For all polygons created through a simple overlay of map t=1 and map t=2 it is checked if the attribute value on land use has changed or not.
③	Reliable change	Factors considered are the syntactic-, semantic- and schematic heterogeneity between the maps at t=1 and t=2. But also the logic of a change in practice is taken account for.
④	Significant overlap	With threshold values it is defined what overlap (%) is considered as noise or real change. Also minimum and maximum absolute threshold values (m^2) are used to avoid scale influence.
⑤	Change of shape	A shape factor is calculated by Area/(Perimeter)^2. Using threshold values for the shape factor it is established if the geometric shape of an object has changed or not.

5.2.2 Proposed algorithm

Based on the methodology discussed above and extensive literature review an algorithm has been developed. The algorithm is based on the assumption that equal identified objects overlap (see also Uitermark et al., 1998; in Kim, 1999). We used two geometric criteria: the percentage of overlap (P_o), which is the percentage by which objects overlap, and the shape-factor (S), which is defined as area/(perimeter)^2. Intersecting objects were considered to be identical if their degree of overlap was above a certain threshold and when the ratio of their shapes did not deviate by more than $1 \pm S_{dev}$, where S_{dev} is a threshold value.

The percentage of overlap (P_o) is the percentage of common area of an object at t=1 and at t=2 with the same attribute value. Using this percentage the overlap can be denominated significant or not. In Figure 10 this is illustrated by means of an example. Object A has a very large common area (80%) if an overlay of map t=1 and map t=2 is made. The deviation is more explained by inaccuracies in the data than by changes over time in the real world situation. Object B however has a common area (15%) that is too small to represent the same object. The changes in the geometric characteristics of the object are more bound to be caused by changes in the real world.

To evaluate for similarity in shape a shape factor (S) is defined (Stoorvogel, 1995). The shape factor is calculated as the area of an object divided by the square of its perimeter multiplied by four times pi ($S = A \ / \ 4 p \ P^2$). The shape factor of a circle is equal to 1. The ratio of the shape factors at t=1 and t=2 ($S_{ratio} = \ S_{t=1} \ / \ S_{t=2} \ |$) is the basis for the second geometric criterion, S_{dev}, the deviation of S_{ratio} from 1 (expressed in %). If S_{ratio} = 1 no change in shape took place and S_{dev} = 0%. If for example S_{ratio} =

0.75 or 1.25, then S_{dev} = 25%. In Figure 10 the shape factor of object C is almost exactly the same at t=1 and t=2 (resp. 0.80 and 0.81 resulting in a S_{dev} of 1%). In combination with the overlay percentage the changes can be considered noise. Object D has a high overlay percentage as well. However the shape factor deviates considerably (resp. 0.80 at t=1 and 0.50 at t=2, resulting in S_{dev} = 60%) which is interpreted as a real change rather than noise.

Thematic changes were evaluated using a characteristic class size and shape band. If an object from t=1 fell outside this band, its attribute value was considered to be unreliable and hence the thematic change was not considered to be a real change. To create this band, the average class size and shape were determined at t=1 and the class limits were calculated by respectively adding or subtracting twice the standard deviation. This criterion was not applied to the objects at t=2, since these were not the subject of discussion as they formed the data set of higher quality. The same thematic criterion was also applied to objects that had disappeared or had thematically changed. Both new and lost objects were evaluated using an overlap-percentage similar to that of intersecting objects but with a different threshold value. Only if 'within'-objects (i.e. those overlapped by objects at t=2) showed significant overlap with its surrounding (e.g. >95%) was the addition or disappearance considered to be noise. Continued objects (i.e. those showing an exact geometric match) were considered to have changed only if their attribute values had changed reliably. As a constraint in determining the threshold values, we stipulated that no real changes (according to the reference data set) should be identified as noise.

In theory it is possible that a polygon might be regarded as representing a change from the point of view of one object but as representing noise from that of another object. It was therefore stipulated that change should take over noise, resulting in the decision rule that a polygon which is evaluated as change can not be evaluated as noise later on.

Finally, it was decided that the output of the procedure should be logically consistent and should cover the study area fully.

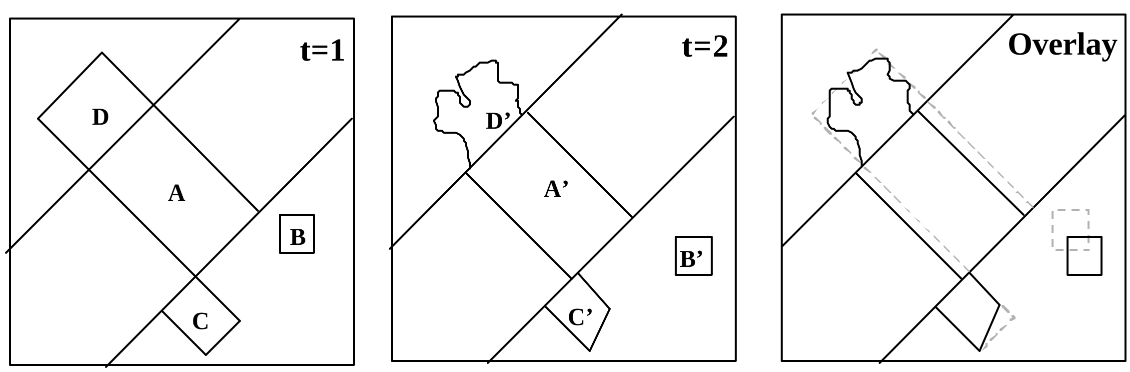


Figure 10. Example of geometric criteria for the distinction between noise and real changes. Object A is considered unchanged (high P_o, low S_{dev}), Object B is considered changed (low S_{dev} but low P_o), Object C is considered unchanged (high P_o, low S_{dev}) and Object D is considered changed (high P_o, but too high S_{dev}).

5.2.3 Application of the improved GIS methodology

All the possibilities for real change and noise, i.e. thematic, geometric and a combination of those changes, that could potentially occur were explored, and constituted the basis of an implementation of the concept in Arc/Info's Arc Macro Language. The requirements for processing were mainly related to the overlay data set that was given as input. It had to contain information, for each polygon, on the identification, the class, the area and the shape of the objects to which it belonged at t=1 as well as at t=2.

In applying *MonGis* to the data set for Soest at class level, the Top10vector class 'Other Urban Use' was excluded because of the fuzzy definition of this class with respect to other classes. It was found from intersect-relations between objects from this class and objects from other classes that 'Other Urban Use' should be given a lower object ranking priority than other objects. The geometric parameters were calibrated by studying the effect of varying these parameters on the identified change Soest sub-set. These parameters then had to be tuned in such a way that no real change according to the reference data set would be evaluated as being noise in the Top10vector overlay data set. To evaluate the effect of the proposed thematic criteria, a default statistics file for classes was created containing the minimal and maximal values of shape and area occurring for a particular class. Every object will by definition be situated on or within these boundaries and an indication of the effect of applying these criteria can be obtained by comparing a calculation using this default file with one using the proposed statistics file.

MonGis was applied to the original Top10vector classification scheme for Soest as well as to the common ontology legend. Using the common ontology legend, the procedure was applied at superclass A and B (i.e. superclass and class) levels.

5.3 Results

5.3.1 The MonGis module

The *MonGis* application can be used in an Arc/Info environment and uses some 200 Kb of disk space. Within this application, the user has the option of altering several criteria through a menu driven environment. The application is relatively simple to use since the only input is an overlay data set, containing all the original items, and a minimum of four user-defined geometric and thematic criteria. In addition to these input variables, there is an option for using a relatively simple object ranking system to study the effects of individual classes.

A noise-reduced multi-temporal data set is given as output containing the objects of the original data set for t=2 and indicating the thematically changed areas. For these areas, the information on their former as well as their more recent land use is provided. Statistics on specific changes in land use can then easily be obtained.

5.3.2 Noise reduction

Figure 11 shows the relationship between the various geometric parameters and the changes found for the Soest sub-set. The overall behaviour shows that extending the limits for isomorphism by either reducing the percentage of overlap or increasing the shape deviation factor, results in a reduction of the total amount of change measured.

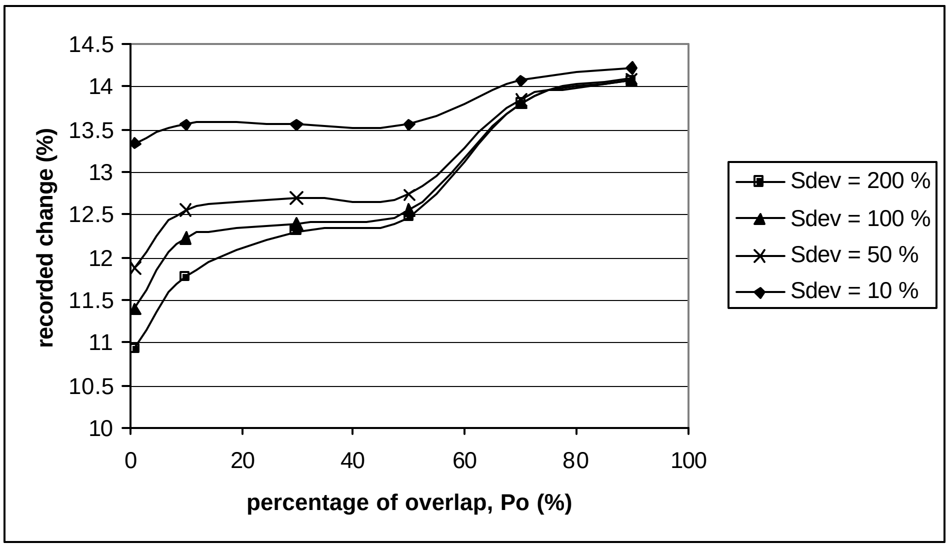


Figure 11: Relation between geometric parameters and recorded change. Note the appearance of a stepwise morphology. Changed objects fulfilling the requirements for noise seem to group around specific geometric parameter combinations. In the present study, no further attention was paid to this.

It was found that no real change in the reference data set corresponded with intersecting interactions of objects in the Top10vector overlay data set, which made it difficult to actually tune the parameters. We decided to use those parameter values for which a significant amount of noise could be removed and for which visual checking ensured that no real changes would be identified as noise. The percentage of geometric overlap, P_o, was set at 10%, while S_{dev} was set at 200% and the overlap percentage for 'within' objects was set at 95%. The results of applying this object-structured procedure to the Top10vector data set for Soest are summarised in Table 9.

Table 9: Values obtained for changes in land use (%) for the area of Soest between 1991 and 1995. The first column indicates the change in land use according to the reference data set, while the last column indicates the relative noise reduction compared to the calculated change using ordinary overlay.

	Change using reference data set (%)	Change using ordinary overlay (%)	Change using MonGis (%)	Relative reduction in change
Top10vector: Class level	2.0	14.3	10.4	27.3%
Common ontology: Sub-class level	2.0	11.5	8.1	29.6%
Common ontology: Class level	1.4	7.1	6.4	9.9%
Common ontology: Superclass level	1.4	5.4	5.0	7.4%

The table shows that not only the amount of change but also the relative amount of noise identified decreases when going from sub-class level to superclass level. This decrease in efficiency is explained by the fact that the total length of the borders between various classes is larger than the total length of the borders between the various superclasses, resulting in fewer thematically changed polygons. Another factor influencing the results is that definitions of classes are less precise than definitions of superclasses.

Figure 12-I indicates how the various superclasses contributed to the total noise reduction for superclass level.

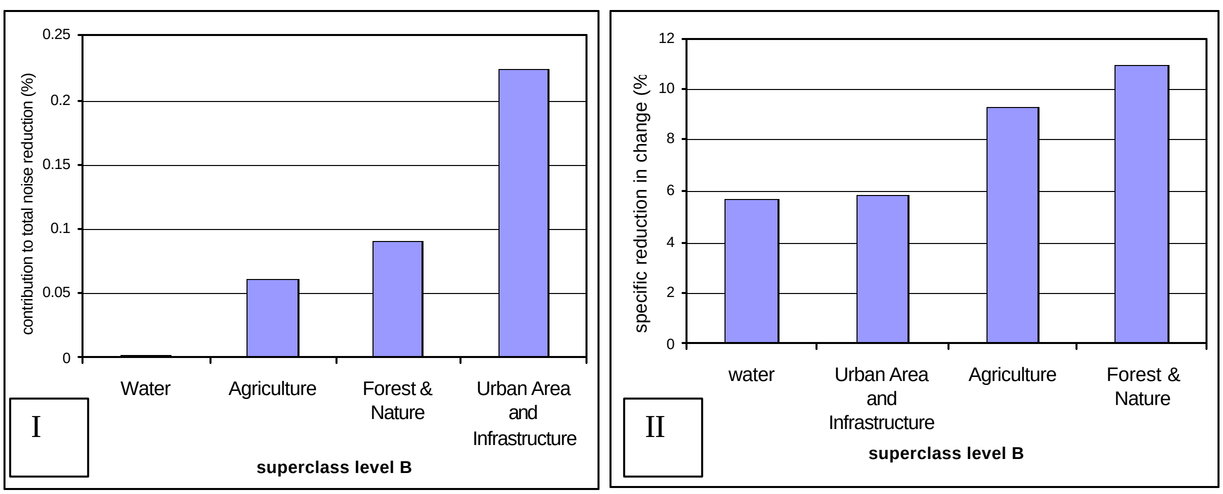


Figure 12. Contribution to the total noise reduction achieved at superclass level (I) and specific reduction in change per superclass (II).

The superclass distribution in Figure 12-I is roughly preserved when going to class level and to sub-class level. Two factors influence the interpretation of these figures. Firstly, and most obviously, the contributions are related to the amount of noise that is removed when a particular sub-class, class or superclass is involved. Secondly, these contributions are also related to the relative abundance of a particular sub-class,

class or superclass within an area, as a low relative abundance will never result in a high contribution. This effect can also be deduced from Figure 12-II. Although the relative reduction in change for 'Water' was almost identical to that of 'Urban area and Infrastructure', their contributions to the total noise reduction differed considerably. Although one might expect a higher efficiency of the procedure for sub-classes, classes or superclasses with a more crisp character, Figure 12-II does not support this assumption. It indicates that there was no significant difference in the efficiency of the procedure for the various superclasses.

A further relevant aspect of the present study is the exclusion of the class 'Other Urban Use' from the procedure, resulting in a somewhat higher contribution of the superclass 'Urban Area and Infrastructure' to the total noise reduction. It can however be observed that there was a considerable difference in contribution between the various superclasses, ranging from 1% for Water to 59% for Urban Area and Infrastructure.

Evaluation of the reduction in noise using thematic criteria revealed that only a very small fraction (< 0.1 %) of the values given in Table 9 at sub-class level, was due to the application of this criterion. The main cause of this poor performance is the poor construction of the data set from 1991 from which these statistics were derived, resulting in irregular characteristic size and shape distributions. Even in an optimised situation, however, the question remains how much noise would really be removed by using this criterion. If there were a number of classes with smooth size and shape distributions but with a considerable overlap between these distributions, then hardly any additional noise would be identified.

Table 9 shows that there were still changes in the data set which were in fact noise but which had not been identified as such in the procedure presented here. A comparison with the reference data set revealed that the main causes of unidentified noise were:

- The complexity of aggregated objects. The aggregation of several individual objects at T=1 or T=2 sometimes resulted in large deviations in shape, that exceeding the criteria for the changes to be classified as noise.
- The fuzzy nature of classes. This was especially true for the original classification scheme used in the Top10vector. In the present study, the class 'Other Urban Use' was excluded from the intersection algorithm because of its conceptual overlap with other classes. More classes may be excluded.
- Thematically changed 'within' polygons. This is related to the fuzzy nature of objects discussed above. An island polygon might have changed from one fuzzy class to another, making it impossible to decide whether this was noise in the overlay. For example, the class 'Other Urban Use' cannot be excluded from the 'within' polygon – algorithm because real changes (such as new buildings) may also have occurred in an area belonging to this class.

The introduction of a common ontology can be seen as an example of how to deal with the latter two problems. In our case study, a common ontology was applied to match the reference data set with the Top10vector data set and this already resulted in a significant reduction of noise (from 14.3% to 10.4%). However, any other fully covering classification scheme may also be applied to aggregate objects into one class.

It was also observed that certain thematic changes in the reference data set were not included in the Top10vector multi-temporal data set. Such changes are then impossible to identify in the data set without applying knowledge from the nominal ground to these data sets, something which in most cases will not be possible.

The method described here uses considerable processing time on contemporary PC's, which is basically a function of the number of area objects that occur in a data set. A more efficient implementation of the approach described here may in future improve the flexibility of the procedure.

5.4 Discussion and conclusion

An improved procedure, *MonGis*, for monitoring land use with vector based data sets was introduced to meet the need for more accurate data on land use changes. A data set for Soest, a small municipality in the Netherlands, was used as a case study for the application of *MonGis*. A reference data set was available for this area, containing the land use objects as they should have been recorded (nominal ground). It was found that almost a third of the noise could be identified and removed from multi-temporal data sets by applying the *MonGis* object-structured approach at class level, using an adjusted thematic legend structure. The study further showed that relative geometric criteria were most successful in identifying noise, whereas the thematic criteria used only made a minor contribution to this noise reduction.

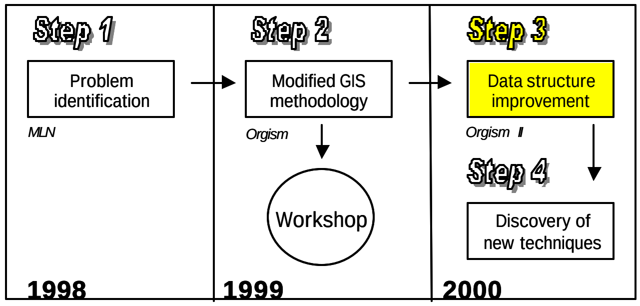
The results indicate a significant reduction in change when a query for land use changes was performed at superclass level. Relative contributions of the various superclasses differed considerably in their degree of total noise reduction, mainly due to the relative abundance of the various superclasses but also to the relative efficiency of this procedure for particular superclasses.

MonGis is attractive in the sense that it requires no extra information on the quality of databases for monitoring purposes. It is clear that the present study benefited greatly from the availability of a multi-temporal reference data set but the absence of such a data set should not be a reason for not using the type of approach presented here. It is suggested that parameter settings first have to be tested for a particular database and this can, in the absence of a complete reference data set, be performed, for example, by specific field checks.

The concept of nominal ground in the context of the performance of *MonGis* indicates the need for specification of monitoring requirements. It may then be more beneficial to transform the classification scheme of a particular data set to a possibly adjusted legend structure.

Finally, the results of *MonGis* indicate the possibility of considerably reducing noise within a multi-temporal data set by approaching it at the level of noise expression. The question whether this reduction is relevant to a real application can only be answered by taking the monitoring requirements into account. Total elimination of the noise within a vector-based multi-temporal data set is not considered possible without additional information on data quality and meta-data.

6 Alternative data capturing technique for monitoring



Monitoring of land use with geographical databases is a technique that is widely used. However, it was not investigated if the quality of monitoring results was sufficient. The simple overlay technique (step 1) resulted in monitoring results of mainly unreal changes (noise). For all national databases at different scale and legend levels the monitoring results were not reliable. The combination of the used geographical databases and GIS functionality did not give reliable monitoring results (chapter 4).
In the second step the GIS technique of overlaying was tried to be optimised. Although the improvements in GIS methodology have reduced noise considerably, a great part of the monitoring results is still noise. Therefore, it is necessary to look for another approach in which the central question will be how to create suitable databases for monitoring land use (chapter 5).

In the third step the data capturing side will be important in improving the quality of databases used in monitoring land use. Intervention in the data capturing process will be necessary. Alterra has made the National Land Use (LGN) database and therefore it is easy to intervene in the data capturing process. The adaptation of data collection, structure and/or processing will be the way to improve the quality of the LGN database for monitoring. Other databases, like Top10Vector and Land Use Statistics of CBS are not considered in this chapter.

6.1 Method

6.1.1 General

Objectives

The main objective of the approach is adapting the data capturing of multi-temporal geographical data in such a way that multi-temporal analysis of geo-information will result in more reliable results. An integrated approach of improving data collection, structure and processing should result in a suitable database for multi-temporal spatial analysis.

Another objective is to establish the hierarchical level at which monitoring gives convincing results. We are especially interested in changes within urban, forest and agricultural areas for the LGN database.

The third objective is to compare land use changes for more than one cycle of 4 years. With the completion of LGN4 we have three databases to monitor land use changes for the period of 1992 till 2000.

Approach

The approach of making multi-temporal data suitable for monitoring purposes can be subdivided in the following three steps:

- Establish for updated versions identical semantics (legend hierarchy) and common object definition (combination of topographic map and satellite images);
- Adapt data processing technique (digitising real (world) changes);
- Create a separate layer that distinguishes real from unreal changes (containing real world changes).

These three steps to adapt data need to be incorporated in the data collection, structuring and processing for new versions of land use databases. The way LGN4 is constructed implies identical semantics and common object definition.

The establishment of identical legend hierarchies for the different versions of the LGN database is important to make data sets suitable for monitoring. Common object definition is a useful tool to compare objects in time. For example, the use of field boundaries makes it easier to monitor agricultural crops. The objects or fields are classified as one entity and therefore can be compared in time. Other possibilities to compare data sets in time are the establishment of consistent metadata (versioning) and the registration of mutations at object level. However, these possibilities have not been covered by this research.

A new data processing technique is used to create a layer containing real changes. Real changes are digitised comparing satellite images (t and t+1). Only satellite images are used which were used in the classification procedure to generate the different LGN versions (t and t+1). The satellite images represent the land cover of the study area. The detection of real changes comparing two satellite images is subjective, it depends on expert knowledge and scale.

The layer containing real changes is a combination of three different sub-layers. These three sub-layers are created in different ways and contain different information:

1. A visual comparison of satellite images on which LGN versions (t and t+1) are based;
2. A comparison of satellite images on basis of NDVI (Normalised Difference Vegetation Index) indices followed by a visual comparison;

3. A comparison of agricultural areas (objects) of the LGN databases followed by a visual comparison of those areas for the concerning satellite images.

The final change layer represents the visible changes in land use between two versions of the LGN database. This final change layer is a combination of all observed changes in the three steps. It is stressed that the final change layer contains only real changes, but not all of them. Depending on expert knowledge and scale only a part of the real world changes are covered in the final change layer.

Sub-layer 1 is made on the basis of a visual comparison of satellite images (t and t+1). Large areas that have been changed in land cover are digitised in this layer. Only changes in land cover of a certain size can be detected. As stated above, this detection of changes is dependent on knowledge of the expert interpreting the satellite images. To detect land use changes, which are not directly observed, two approaches are proposed:

- Establishment of sub-layer 2 to monitor changes within forest and urban areas;
- Establishment of sub-layer 3 to monitor changes within agricultural land.

Monitoring within these areas seems interesting for policy decisions. Monitoring makes it possible to see if a certain policy has the expected effects. E.g. has the total area of grassland decreased for the period of interest?

Sub-layer 2 is made on the basis of NDVI indices. For urban and forest areas NDVI indices are calculated for different satellite images. Only satellite images are used representing a common month/season at t and t+1. It is tried to differentiate areas with deciduous and coniferous forest and areas with buildings and city grounds on basis of these NDVI indices. A change layer is created representing land use changes within urban and forest areas between t and t+1. It is verified if the size and form of the changes are plausible. For the remainder of changes, it is visually verified by comparing satellite images of t and t+1 if these changes are real.

With help of digitised field boundaries sub-layer 3 is created, representing visible real changes in the agricultural areas. The field boundaries make it possible to create objects with one agricultural use for the LGN version. The zonal majority tool is used to generate unambiguous agricultural crop entities. These objects are compared for different LGN versions. A change layer is created with the overlay technique. The changes are verified if size and form are plausible. Visual verification of these changes with the satellite images resulted in sub-layer 3. It is assumed that changes are real if you can see differences in reflections between satellite images at t and t+1. Differences in reflection between satellite images (t and t+1) at one moment (one season/month) and/or for crop cycles (more than one season/month) are indications for changes in agricultural land use.

Out of the three sub-layers a total change layer is constructed. Of all three sub-layers the land use changes (areas) are copied into a new total change layer. The created total change layer is used as an intermediate layer between the different LGN versions. Total change layers are established between LGN2 and LGN3 and between

LGN3 and LGN4. With help of these intermediate layers updated versions for LGN2 and LGN4 are made. It means that if an area is not considered as being changed the data of LGN3 are taken as true characteristics for that area. LGN3 is used as reference database, because it was the newest LGN version at the start of the project. It is better (technical aspects and no mixture classes) than LGN2 and it is used as basis for the creation of LGN 4 (geometric basis). The adapted LGN2 and LGN4 versions are compared with LGN3 to obtain monitoring results.

In Figure 13 a schematic representation is presented of this process. With help of the change layer, non-changed and changed LGN data of t and t+1 are combined to create an adapted LGN data set that will be used in the monitoring process. For example in Figure 13, only area B is taken from LGN t+1, because it was detected as a real change. Furthermore, it is important to notice that the change layer does not contain any information about what land use changed to what, but only indicates which area have changed.

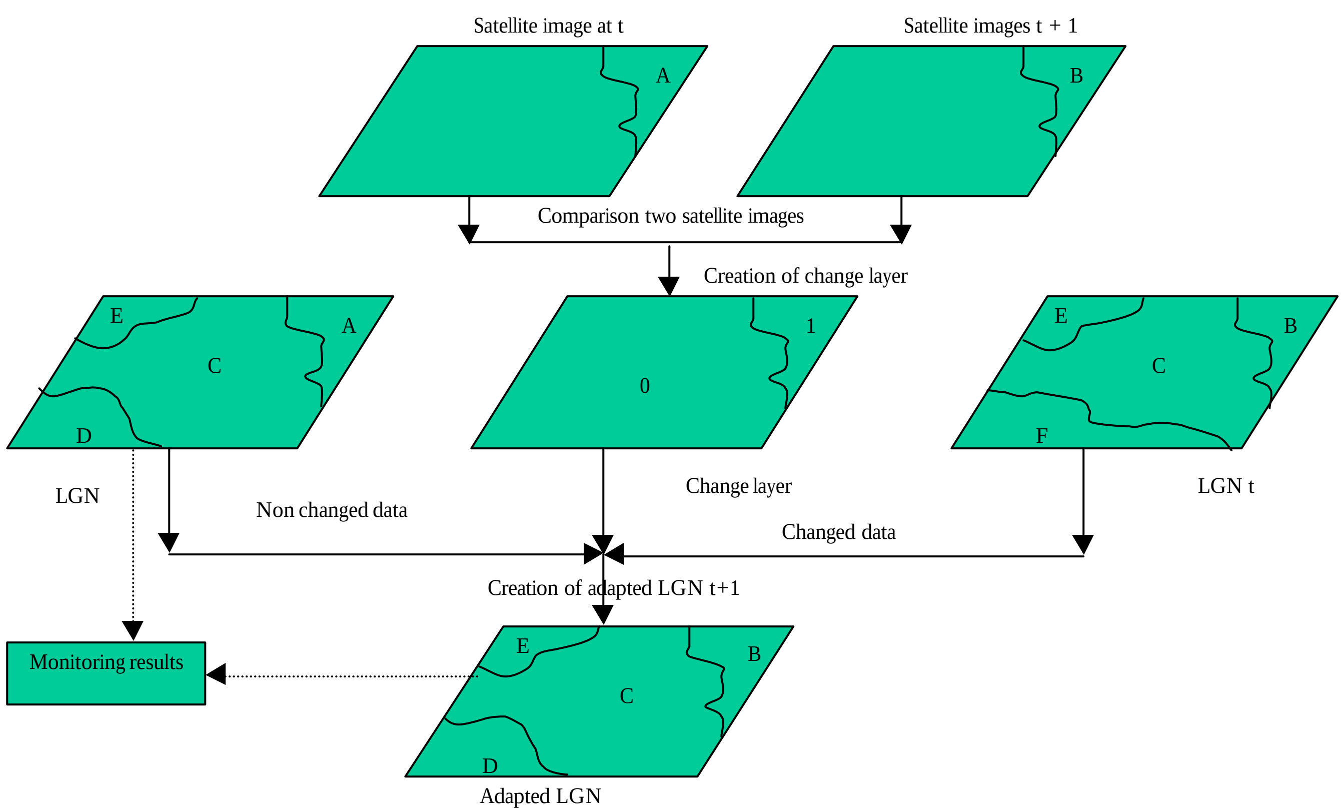


Figure 13. Schematic presentation of research approach (step 3).

For the study area Soest the results of the different approaches will be verified with help of the reference database. The monitoring results of LGN3 with adapted LGN versions are compared with the overlay results of LGN3 with LGN2 and LGN4. However, the results of step 2 (improving GIS methodology) reflect land use changes between 1991 and 1995 for the Top10Vector database, which are different compared to the LGN database (time period, vector database). Also, the monitoring results of step 3 are compared with results obtained with the overlay technique for an entire topographic map sheet (32West). The study area Soest is only a part of this topographic map sheet. Monitoring land use changes for an entire sheet gives a slightly broader scope of the approach used in step 3. Monitoring can take place for more classes, because some land uses do not exist in the study area Soest. In this case, only monitoring results between the LGN2, LGN3 and LGN4 versions will be presented.

6.1.2 Data capturing for different hierarchical levels

Legend

Monitoring of land use is only possible on the basis of a common legend structure. To compare land use at t with land use at t+1 you need to have the same land use classes defined in your legend for both versions. However, a common legend structure is also needed for the different approaches and the reference data set if you want to compare their monitoring results. Therefore, common monitoring classes need to be identified out of the different LGN legends, the reference data set legend and the legends used in chapter 4 and 5. The same semantic levels are used as defined for the reference data set legend: superclass, class and sub-class level. These levels were also used in the approaches described in chapter 4 and 5. The definition of superclasses and classes is the same in all cases. However, the number and definition of sub-classes is different between LGN and the reference data set. Therefore, only sub-classes are mentioned which can be derived from the LGN database. The sub-classes of the reference data set are grouped to compare the monitoring results with the results for LGN.

The LGN legends have changed with time. Differences in technical aspects (classification of satellite images) and in user demands have caused differences in land use classes and legend structure for the different LGN database versions. For example, LGN2 has agricultural mixture classes which LGN3 and LGN4 do not have. Therefore it is impossible to monitor between LGN2 and LGN3/4 on agricultural crops. Another example is the absence of class 26 ‘urbanisation in agricultural areas’ in LGN2. The applied legend structure is presented in Table 10.

For the three semantic levels as presented in Table 10 the procedure to intervene in the data capturing process will be described. Aggregation of legend classes will result in superclasses, classes and sub-classes. Only relevant monitoring classes will be taken into consideration. Monitoring on the level of the original legend means that no aggregation of classes took place (exception: nature classes LGN4).

Table 10. The three semantic levels of the legend used in this approach of adapting data capturing techniques for monitoring (shade – monitoring investigated in step 3; ***bold*** *– monitoring possible in step 3).*

Superclasses	Classes	Sub-classes
Continuous urban area	**Built-up area**	Buildings
		City-grounds and other urban uses
	Infrastructure	**Infrastructure**
Forest & nature	**Forest**	Coniferous
		Deciduous
	Nature	**Nature**
Water	**Water**	**Water**
Agriculture	**Horticulture**	**Orchard**
		Bulb fields
		Greenhouses
	Grassland	**Grassland**
	Farming	Maize
		Grains
		Potatoes
		Beets
		Other crops

Superclass

Monitoring results at superclass level are fully defined by the observed changes of sub-layer 1. The change layer does not contain any information about what land use changed to what, but only indicates which areas have changed land use at superclass level. Only comparison of LGN classes at t and t+1 can reveal the type of change for the area of interest. Overlaying the adapted LGN version with LGN3 at superclass level determines the land cover changes for the monitoring superclasses urban, forest & nature, water and agricultural areas.

Class

At class level, the definition of monitoring results is also based on the visual interpretation of satellite images (sub-layer 1). Changes between the classes water, forest, nature, built-up areas and infrastructure are monitored in this way.

For the classes horticulture, grassland and farming areas another approach has been used. A vector database is generated containing parcels, which are based on field boundaries distinguished by crop. For each parcel, the dominating agricultural land use is defined for every LGN version. At this object level the overlay technique is used to generate a change layer between LGN(t) and LGN(t+1). Subsequently, a visual interpretation of the satellite images t and t+1 for the object oriented change layer is done, subjectively. The unreliable changes are eliminated and a real change layer is generated for agricultural land at class level. Land use changes for above-mentioned classes are changes defined in sub-layer 3.

Sub-class
At sub-class level, the monitoring result is defined by changes in the total change layer (a combination of sub-layers 1 – 3). Apart from the visual interpretation, two other approaches are used to define changes. For the monitoring sub-classes orchards, bulb fields and greenhouses the object-oriented approach described at class level is used (sub-layer 3).

In case of monitoring changes between deciduous and coniferous forest and between city-grounds and buildings another approach has been used. Sub-layer 2 encompasses this type of changes. The approach comprises the generation of NDVI images (for a certain month, e.g. May) on the basis of the satellite images in which different forest types and/or different built-up areas possibly can be recognised. The NDVI images are normalised to create a common basis. One NDVI image is used as reference. For each (normalised) NDVI image mean and standard deviations are calculated for the different monitoring classes. These values are analysed to look for a possibility to split up forest into coniferous and deciduous forest and built-up areas into buildings and city grounds and other urban uses on the basis of NDVI values. Subsequently, a decision rule is established to define if a change has occurred between image (t) and image (t+1) or not. Clusters are established to cope with scale. Finally, the result is visually interpreted to create real change layers for coniferous and deciduous forest and buildings and city grounds and other urban uses.

6.2 Results

The approach of adapting data capturing techniques has been performed for the municipality Soest. The results of this approach can be compared with the results of the simple overlay technique and the reference data set. For the Soest area it was possible to monitor for two periods of four year. Also results are presented for the topographic map sheet 32 West. These results situate the monitoring research of step 3 in a broader perspective. However, for map 32West monitoring results are only available for the period 1992 – 1995. LGN4 is not yet available for the entire map sheet.

For the municipality of Soest reference data were available. This allowed for comparison with former results in the study. Map 32 West was used for monitoring of a wider range of sub-classes. Therefore, this study area is more representative for the land use in the Netherlands.

6.2.1 Study area Soest

The results for the Soest study area are summarised in Table 10 and 11. The classes on which monitoring is possible with the data capturing approach of step 3 are presented in Table 10 (in **bold**). At sub-class level, monitoring on some sub-classes is not possible, because the spectral reflectance is not different enough to separate the sub-classes. In Table 11 the percentages of change for all semantic levels and for the original LGN legends are presented comparing LGN3 with LGN2 and LGN3 with LGN4 using the simple overlay technique and the approach as defined in this chapter (overlaying LGN3 with adapted LGN2 and LGN4). Also change percentages are presented for the reference data set. Hereafter, the results per semantic level are discussed and presented. The main reasons for the differences are discussed in section 6.3.

Superclass

Land use changes between continuous urban areas, water, nature and forest and agriculture areas are monitored (Table 10 - superclasses in **bold**). Overlaying LGN3 and adapted LGN versions resulted in change areas at superclass level (Table 11). The results obtained are an under-estimation of the real world changes.

The major changes are between built-up and agricultural areas. The spectral reflectance for the agricultural land uses and built-up areas is different. Therefore, changes can be detected easily when comparing two satellite images. However, scale is a limiting factor in detecting all changes. The only difficulty is bare ground. From the reflectance of bare ground you cannot detect the function of it: land preparation for agricultural use or for extension of urban areas.

Monitoring land use changes for forest and nature areas are not a problem if they correspond with land cover changes. Spectral reflections are different if a forest or nature area is converted to an urban area. However, changes from other classes to the forest and nature class are difficult to monitor. A different spectral reflectance is not established in a 4 years period. Therefore these changes are not visible when you compare satellite images representing t and t+1.

Areas dedicated to water can be easily recognised at satellite images. Therefore, real changes in surface area dedicated to water can be monitored without problems.

Table 11. Changes in land use (% of total area) for the area of Soest between LGN2, LGN3 and LGN4.

	LGN2 – LGN3						LGN3 – LGN4			
	Region wise approach			Point wise approach			Region wise approach		Point wise approach	
	ref. data	Overlay	Adapted data capturing	ref. Data	Overlay	Adapted data capturing	Overlay	Adapted data capturing	Overlay	Adapted data capturing
Superclass	1.34	4.59	0.74	1.14	2.43	0.38	1.43	1.31	0.72	0.66
Class	1.59	4.59	1.01	1.6	3.53	0.58	2.75	1.31	2.22	0.92
Sub-class	1.59	4.59	1.01	1.6	3.53	0.58	2.75	1.31	2.22	0.92
Original legend*	2.44	13.88	1.1	1.97	9.13	0.63	9.96	1.33	5.86	0.93

* class 26 is not longer joined with class 1 (grassland); nature classes for 1995-1999 are grouped according to LGN3 legend

According to Table 11, the amount of changes in land use that are monitored with the change layer of LGN are considerably smaller than the percentages obtained with the ordinary overlay technique. However, this difference in amount of changes is much smaller for LGN3 - LGN4 (0.1%) than for LGN3 – LGN2 (2 – 4%). In particular the change percentage obtained by the overlay technique has reduced considerable for LGN3 - 4.

Comparing the region and point wise approach, the point wise approach, in which the location of land use is of importance, always has lower change percentages. The difference is between 0.2 and 2% for all three techniques. The differences are relatively small for the reference data set. Only the differences for the overlay technique between LGN2 and LGN3 are around 2%.

The changes in land use as determined with the alternative data capturing approach are an under-estimation of the real world changes. For both the region and point wise approaches the detected changes are between 0.6 and 0.8% lower as those of the reference data set (see Table 11). However, the results are more realistic than those obtained with the overlay technique. The changes are real changes and they approach the percentages of change for the reference data set.

Class

The legends of the layers LGN2, LGN3, LGN4, adapted LGN2 and adapted LGN4, are aggregated according to the classes of Table 10. Overlaying these data sets resulted in change percentages at class level (Table 11). This procedure was possible for the classes built-up areas, water, horticulture, grassland and farming areas.

The land use class infrastructure is based largely on actual information derived from the Top10Vector database. Therefore, adapting data capturing for the infrastructure class is not considered in this approach. No changes in infrastructure between LGN versions are observed. Land use changes between built-up areas and the agricultural classes are the main contributor to the overall land use change percentage at class level.

Land use changes for the forest and nature classes are difficult to monitor by comparing satellite images. The extension of forest areas is difficult to detect at short and intermediate temporal scale. A solution can be to define reflection class for different forest development stages, but this is not considered in this approach. Therefore, land use changes will not be detected for the period considered between versions of the LGN database. However, decline in forest area (cutting trees) is a sudden/abrupt activity. These changes in surface area can be monitored visually.
The approach of adapting data capturing for monitoring changes from nature to other classes and vice versa will be useful at a long temporal scale (+/- 10 years). In the case of shorter periods these changes are invisible; the land cover does not change immediate and changes in land use functions can not be seen. By comparing satellite images only land cover changes (and not land use functions) can be detected. Therefore, monitoring at short and intermediate temporal scale (< 10 years) is only possible on basis of databases delivered by external institutes (like Natuurmonumenten and Staatsbosbeheer).

The class water is not subdivided. Monitoring changes in surface area between salt and fresh water is not useful, neither possible. Salt water is not present in the study area. Difference in reflectance between salt and fresh water is not observable.

Horticulture, grassland and farming areas can be monitored with help of digitised field/crop boundaries (see 6.1.). This subdivision is adapted to the legends used in other monitoring approaches. The increase in land use changes from superclass to class level for the step 3 approach is largely caused by land use changes between the horticulture, grassland and farming area classes. The increase in areas changed for the step 3 approach from superclass level to class level is largely based on changes between those classes.

According to Table 11, the amount of changes in land use that are determined with the approach of step 3 are considerably smaller than the percentages obtained with the overlay technique (1.3 – 3.5%). However, the increase of area changed determined with the overlay technique is obvious when compared with the amount of changes at superclass level. The increase for LGN3 – LGN4 is especially great (+/- 1.5%).

Comparing the region and point wise approach, the point wise approach always has lower change percentages. However, the differences are much smaller than at superclass level (0 – 1%). Also, most changes in land use are detected by the approach in which location is not of importance.

The monitoring results for this approach at class level are also an under-estimation of the real world changes. For both the region and point wise approach the land use changes in the reference data set are 1.6% at class level, whereas only 1.0% respectively 0.6% of area changes in land use in case of the step 3 approach (see Table 11). However, the results at this level are more realistic than those obtained with the overlay technique. The changes are real changes and they approach the percentages of change for the reference data set. It can also be noticed that the results for the region wise approach (1.0%) correspond better with the changes in land use for the reference data set (1.6%). The differences in land use change percentages between the reference data set and this approach are approximately the same at superclass and class level.

Sub-class

Monitoring land use changes at sub-class level has been investigated with a combination of visual comparison of satellite images, overlaying agricultural parcels and NDVI images. Only for a very limited number of sub-classes monitoring is possible (Table 10 – sub-classes in **bold**).

The buildings and city grounds and other urban uses monitoring sub-classes are complex LGN classes. The building monitoring class comprises the LGN classes 18 and 19. The city grounds and other urban uses sub-class is composed of LGN classes 20 – 24. The normalised mean pixel values +/- their standard deviation for buildings and city grounds and other urban uses for the May 1992, 1995 and 1999 NDVI images are shown in Table 12. For the 1992 and 1995 images it seems possible to separate the monitoring classes. Subsequently, a decision rule is established to define if a change has occurred between image (t) and image (t+1) or not. The generated change layer is clumped (4 pixels contiguous) and sieved (25 pixels) to establish clusters. These values give useful results. However, the NDVI map of May 1999 shows a great overlap. The overlap is too much to separate these classes on basis of NDVI values. Separation in this case is only possible at very low confidence levels. Figure 14 is a histogram showing the overlap between the different monitoring classes within built-up areas for the NDVI image of 1992 and 1999.

Table 12. Normalised mean values and standard deviations for buildings and city-grounds and other urban uses at three moments (May 1992, 1995 and 1999) for the topographic map sheet 32West.

	1992		1995		1999	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Buildings	52.8	26.4	55.2	30.4	49.6	25.6
City-grounds	98.3	25.7	94.2	26.1	78.2	27.8

Monitoring between 1992 and 1995 for buildings and city-grounds and other urban uses is possible. However, monitoring between 1995 and 1999 is not possible due to the overlap in histograms. At a high confidence level it is not possible to separate both monitoring classes. The main problem in separating these monitoring classes is the difference in scale. The spatial scale of the changes is smaller than the resolution of the pixels (25*25 meter). It is decided that monitoring within built-up areas is not possible with NDVI values. There is not sufficient confidence to separate both monitoring classes in all cases.

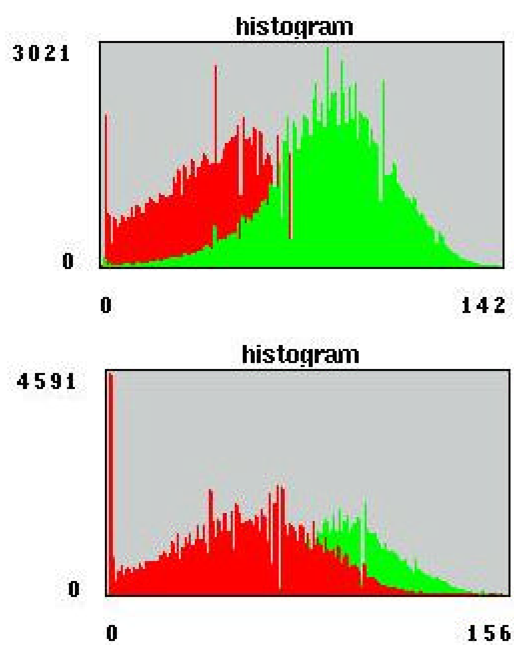


Figure 14. Histogram of for buildings (dark) and city-grounds and other urban uses (light) for the 1992 (above) and 1999 May images of topographic map sheet 32W (X-axis: DN value, Y-axis: frequency).

At sub-class level it is not possible to monitor changes between coniferous and deciduous forest. The mean values +/- their standard deviation for coniferous forest and deciduous forest for the May 1992 and 1995 NDVI images show great overlaps (Table 13 and Figure 15). For both NDVI images it is not possible to split up the forest area into coniferous and deciduous forest. Confidence levels will be very low if you split up the NDVI images. The reflection of coniferous and deciduous forest in May images is not different enough to separate these forest types. Possibly, winter images are more suitable to separate these forest types. However, these images are not available. Also it is difficult to imagine that in a 4 year cycle a forest changed from one type to the other. Without separating the two forest types it is impossible to monitor them.

Table 13. Mean values and standard deviations for coniferous and deciduous forest for the May 1992 and 1995 NDVI images.

	1992		1995	
	Mean	Standard Deviation	Mean	Standard Deviation
Coniferous forest	78.4	12.1	92.4	15.6
Deciduous forest	95.7	14.7	112.3	17.2

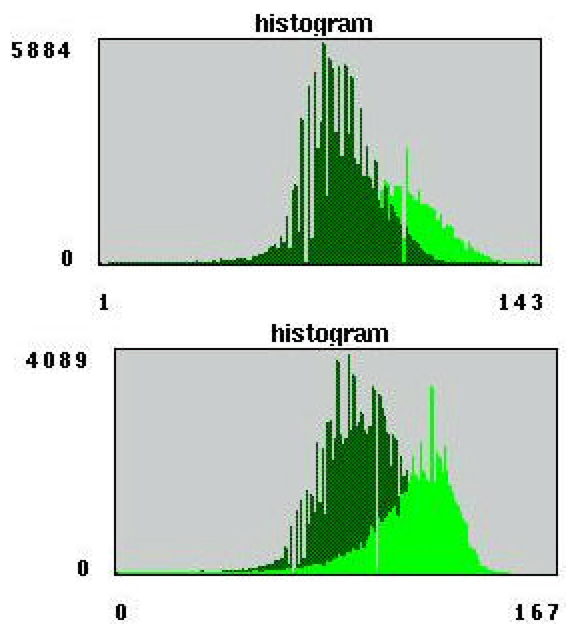


Figure 15. A histogram that shows the overlap between coniferous (dark) and deciduous (light) forest for NDVI image of May 1992 (above) and 1995 (X-axis: DN value, Y-axis: frequency).

Monitoring land use changes for the sub-classes orchards, bulb fields and greenhouses seems possible. Overlaying parcels (objects with one agricultural use) with orchards, bulb fields and greenhouses as land use will result in a change layer. Dependent on recognition of specific reflection patterns for these land uses, visual verification will be possible to detect real changes. However, surface areas with orchards, bulb fields and greenhouses in the Soest study area are very small.

Therefore, it seems too early for a reliable conclusion on monitoring these classes with the alternative data capturing approach.

Monitoring crops with the adapted data capturing technique is not taken into account. Due to crop rotations land use is very dynamic and monitoring with the point wish approach does not seem useful. Land use changes within years and between years. However, monitoring results with the region approach might be interesting for policy makers. Unfortunately, due to the use of agricultural mix classes in LGN2 it is impossible to monitor specific crops.

Comparing the monitoring results at sub-class level with the results at class level it is obvious that there is no difference. The monitoring results for all data sets, approaches and techniques are equal to the results at class level (see Table 11). No additional land use changes are monitored at sub-class level.

Original legend

Monitoring land use with the original legend classes means that the original class hierarchy as dominated by the data provider is used. However, the LGN4 nature classes are an exception. These are aggregated in the classes as defined in LGN3. In Table 11 it is obvious that the land use change percentages for the original legend are the highest for all techniques, approaches and data sets. The percentages are higher than those obtained at sub-class level.

6.2.2 Map sheet 32West

In this section the monitoring results for the entire topographic map sheet 32West are presented. These results situate the monitoring research of step 3 in a broader perspective. Only results are presented for the 1992 – 1995 period as LGN4 is not yet ready for the entire map sheet. Although results for the Soest study area indicate that separation of the classes buildings and city-grounds on basis of NDVI indices is impossible at high confidence levels for the total 1992 – 1999 period, for 32West results at sub-class level are presented with these changes (Table 14: data between brackets). The results for the topographic map sheet 32West cannot be compared with real world data as no reference data set for the entire topographic map sheet 32West was made.

Table 14. Changes in land use (%) for map sheet 32W between LGN2 and LGN3.

	LGN2 – LGN3			
	Region wise approach		Point wise approach	
	Overlay	Adapted Data capturing	Overlay	Adapted data capturing
Superclass	2.6	1.7	1.63	0.25
Class	4.8	3.0	5.80	1.92
Sub-class*	4.8 (11.8)	3.0 (3.1)	5.80 (9.66)	1.92 (2.01)
Original legend	15.8	3.3	11.59	1.97

* change percentages including monitoring results for buildings and city grounds (between brackets)

The difference between the change percentages for both overlay and alternative data capturing technique are smaller compared with those of the study area Soest. In general, the change percentages for the overlay technique are lower and those for the data capturing technique are higher. As observed for the Soest study area, the change percentages for the point wise approach are smaller then those of the region wise approach. At sub-class level the difference in percentages between the overlay and the data capturing technique increases strongly when you incorporate the changes between buildings and city-grounds. The technique to monitor changes in building and city-ground areas strongly underestimates the changes. Although we do not have reference data, it seems reasonable that all percentages for the adapted data capturing technique in Table 14 are under-estimations of real world changes.

6.3 Discussion and conclusion

After analysing the results presented in Table 11 it can be concluded that the technique to adapt data capturing produced monitoring results that are an under-estimation of real world changes. To explain this under-estimation various factors can be mentioned:

- Subjectivity and level of expert knowledge;
- Spatial and temporal scale;
- Relation between land cover and land use.

The process of visual comparison of two satellite images is open to subjectivity. The experience and knowledge of the person who is interpreting satellite images to create change layers determine the quality of the results. Size and form of changes, differences in reflections (at one moment and within a year) and field knowledge are factors to be taken into account. The person determines if a change is observed and if it is a real change. Normally, a person will not detect all changes and only define it as a real change if he is sure about it. Therefore, an under-estimation of changes can be foreseen.

A number of real changes will not be detected due to scale problems. In the reference data set changes are observed at a more detailed level than can be observed by comparing satellite images. The used satellite images have a resolution of 25*25 meters resulting in LGN database versions that are suitable for applications at scale 1:50.000 or more global. Whereas, the reference data set is made at scale 1:10.000. The number of changes observed with the adapted data capturing technique will be lower than for the real world due to spatial scale differences. Also temporal scale has its effects on the number of changes detected by the adapting data capturing technique. For example, a change in land cover from agricultural to forest land is not detected in a relatively short period of four years. However, such abrupt land use change is registered in the reference data set.

The third factor that causes an under-estimation of real changes with the adapted data capturing technique is based on the difference between land cover and land use.

The technique of adapting data capturing can only detect land cover changes. Differences in reflections between satellite images can be observed. Land use changes, which are related to land use functions, are not observed. However, some of these changes are registered in the reference data set. For example, an area covered with grass that has changed its function from agricultural use to a recreation area (football field) will not be observed as a change when applying this method. However, this information is integrated in the reference data set and will be monitored as a change.

Another important result is the limited possibility to monitor at sub-class level. Due to the scale/resolution of satellite images, changes are too detailed to observe them by comparing satellite images. Additional techniques were used to detect these changes. Separation of deciduous and coniferous forest and built-up areas and city grounds on basis of NDVI indices was not possible due to small differences in reflections. Therefore, monitoring changes between these sub-classes could not be considered as an option. As a result of object definition it seems possible to monitor changes between orchards, bulb fields and greenhouses. However, no changes occurred between these sub-classes that have a very limited surface area. Monitoring other sub-classes was not possible due to the difficulty to create identical legend sub-classes.
Hence, it can be stated that monitoring changes at sub-class level by adapting data capturing does not present useful results for the case study area.

Monitoring changes at crop level was not taken into account due to the use of agricultural mix classes in LGN2. However, monitoring agricultural crops from LGN3 onwards will be possible if the same geometric base will be used (Top10Vector). Another condition will be that for every version additional field crop limits are digitised. On basis of this object definition, monitoring at crop level looks promising. However, due to crop rotations land use is very dynamic and therefore attention should be focussed on the regional approach.

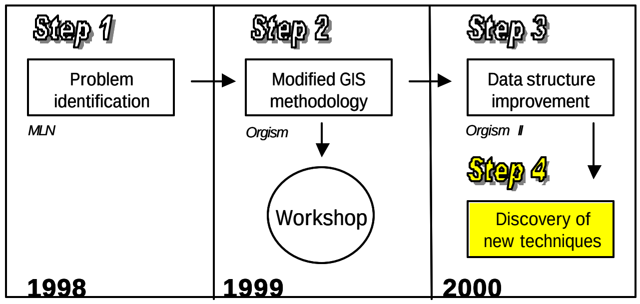
Comparing the monitoring results between LGN3 – LGN2 with LGN3 – LGN4 it can be concluded that the way LGN4 is created results in a considerable reduction of unreal change (noise) for the simple overlay technique. The difference in monitoring results between the two techniques is much smaller in the LGN3 – LGN4 case. Unfortunately, the results cannot be compared with a reference data set due to the absence of these data. The use of common legends and geometric basis is evident in the results. To create new database versions with the idea to use them for monitoring will improve the monitoring result.

Looking at the monitoring results it is obvious that at superclass and class level the under-estimation of changes is relatively large. However, the under-estimation is much smaller than the over-estimation by the simple overlay technique. Another advantage of the adapted data capturing technique is that the changes are real changes. The technique of adapting data capturing is a useful tool in reducing noise, but not all changes are monitored. Therefore, it needs to be investigated how this technique can be combined with the other techniques (step 2 and 4) to optimise the monitoring results.

The overall conclusion of the adapted data capturing technique is as follows:

- Monitoring at superclass and class level is possible;
- A reduction of noise can be obtained;
- Under-estimation of real changes;
- Monitoring at sub-class level is restricted due to scale;
- New database versions made with help of a change layer result in a sharp reduction of noise.

7 Knowledge Discovery in Databases for Monitoring



7.1 Methods

Knowledge Discovery in Databases (KDD) has been emerging as a new research field and a new technology for the discovery of interesting, implicit, and previously unknown patterns, trends, and relationships from large databases. A KDD process usually involves experimentation, iteration, user interaction, and many design decisions and customisations. Different delineations have been proposed for a KDD process, including the nine-stage process described as following (Fayyad, 1996, p. 23):

1. Developing an understanding of the knowledge domain, the relevant prior knowledge, and the goals of the user.
2. Creating a target data set, selecting a data set, or focusing on a sub-set of variables or data samples, on which discovery is to be performed.
3. Data cleaning and pre-processing.
4. Data reduction and transformation.
5. Choosing the data mining task.
6. Choosing the data mining algorithm(s).
7. Data Mining for a particular form of representation such as classification rules or trees, regression, clustering, etc.
8. Evaluating the results.
9. Consolidating discovered knowledge: incorporating this knowledge into the performance of the system, or simply documenting it and reporting it to users.

Although this list might suggest a linear process, a KDD process is an *interactive circular process* in which the stages are often repeated and do not follow a linear sequential order.

7.1.1 What does data mining bring to land use monitoring systems?

Data mining techniques are useful when pursuing application objectives of classification and prediction, optimisation, summarisation, or exploration (Wachowicz, 2000a). They usually improve the quality and effectiveness of decision making processes. They complement, and can often replace, other decision assistance techniques, such as statistical analysis, and data reporting and querying. Data mining is a necessity when exploring how different stakeholders interact to the land, especially when the intentions are to provide a knowledge base for land use management. Scientists and policy makers are expected to understand and respond to a variety of complex issues and their inter-relations concerning processes such as those related to global concerns, regional disparities, and local implications. Therefore, new methods and associated tools are needed to facilitate science and policy decision (Wachowicz, 2000b). Figure 16 illustrates our effort towards the development of the next generation of land use monitoring systems based on the current technology and a sustainable land use development.

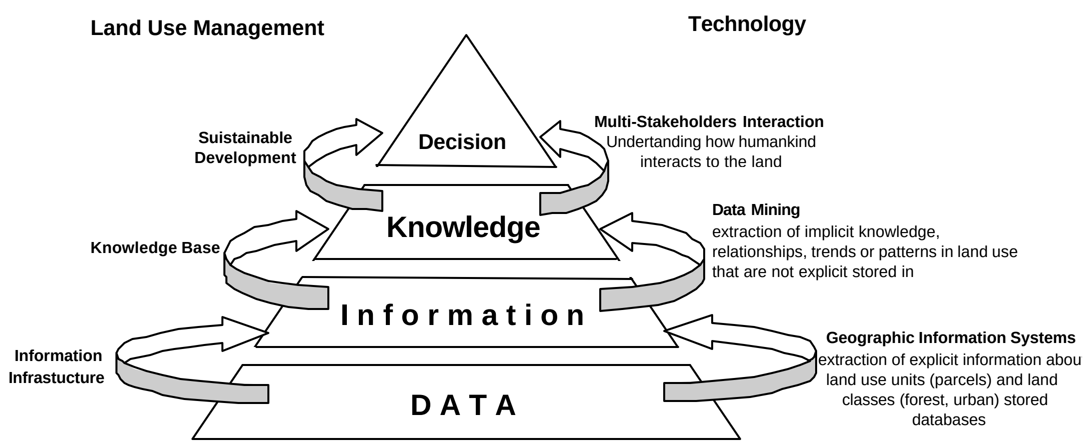


Figure 16. Overview of the next generation of land use monitoring systems

Having the right information about land use changes is not only important for helping to frame and monitor policies required for improving the state of the environment. A knowledge base is also important for changing societal behaviour and influencing in a positive manner the impact that society as a whole has on land use. The data-information-knowledge-decision strategy illustrated in the Figure 16 works on the assumption that policy, however devised, works better when policy makers and stakeholders are better informed. The knowledge base will support this assumption by generating land use monitoring information from ever increasing volumes of data from different time scales, multiple sensors, and higher spatial, temporal and spectral resolutions.

7.1.2 What are the challenges to which data mining could contribute?

The main challenge is focussed on the importance of understanding how changes in land use (thematic, spatial, and temporal changes) that have been stored in a GIS relate to global drivers indicators in sustainable development, economic reforms, globalisation, urbanisation, and technology. Therefore, data mining contributes to the discovery of local patterns for the identification of interesting regions of a high dimensional input space, or models (in form of consolidated best system of patterns). Several data mining techniques have been developed and they include all types of rules, decision trees, change and trend patterns, and analyses of contingency tables. These techniques will be important for the land use applications. The following questions represent areas in which data mining techniques may be applicable:

- Is land use changing?
- What are the signatures of land use change? (The land use fingerprint).
- If it is changing, why is it? (Anthropogenic versus natural processes).
- How we can use this to predict future land use, given different policy choices?
- How can we relate land use changes with global driver indicators?

7.2 Mining association rules in land use changes

Land use change is one of the most critical dynamic elements of a system. A suite of complex factors, including policy, population change, culture, economics, and environment characteristics, drive land use change. Human-induced changes to the land often result in changes to patterns and processes in systems such as alterations to vegetation cover, species diversity, and changes to the economies of a community. The LGN data sets are a sequence of snapshots or discrete displays at sequential moments in time. Researchers interested in capturing the complexities of underlying processes, however, are often dissatisfied with the snapshot approach, because this method overlooks the events, each of which occurred separately, which is the cause of such changes. Therefore, data mining was used to discover relationships and patterns in land use classes in the case study area - a classic data-driven knowledge discovery process in which correlations that were not immediately apparent become clear, and can be displayed in visual form. In our case, association rules were particularly useful in discovering patterns in LGN1, LGN2, and LGN3 data sets, with rules of association indicating the frequency of items occurring in the target data set (i.e. the case study area).

In data mining, an association rule states that given that X is true, there is a certain probability that Y is also true. X is referred to as the left-hand side (LHS) of the rule and Y as the right-hand side of the rule (RHS). For example, an association rule can state that 80% of urban areas that are close to recreation areas are also close to nature conservation areas. The components of an association rule quantify the strength of this association. They are:

1. Support: The support of a rule quantifies how prevalent the rule is throughout the data set, or how often X and Y occur together in the data set as a fraction of the total number of records. For example, if the support is 1%, X and Y occur together in 1% of the total number of records. The lower the minimum support, the more rules are generated, and the slower the performance of the tool might be.
2. Confidence: The confidence of a rule quantifies the number of records in which both sides of the rule appear, divided by the number of records in which the LHS rule appears. For example, if the confidence is 50%, Y occurs in 50% of the records in which X occurs. Thus, knowing that X occurs in a record, the probability that Y also occurs in that record is 50%.
3. Expected Confidence: The expected confidence measures the confidence of a rule as if there were no relationship between the left and right sides of the rule. It is computed based on the number of records in which the RHS item appears in the data set. So the difference between expected confidence and confidence is a measure of the change in predictive power to the presence of the LHS item.
4. Lift: It is the ratio of confidence to expected confidence. The greater the number, the more unexpected the rule. It tells you how much additional information the LHS

The MineSet System was used for mining the association rules. In MineSet, a rule can be described as following:

- *Association rule: all pixels which were class 1 in the LGN1 data set are also class 1 in the LGN2 data set.*
  col3 in (LGN 1, class 1] && col4 in (LGN2, class 1]
- *Association rule: all pixels which were class 1 in LGN 1 data set are also class 1 the the LGN 3 data set.*
  col3 in (LGN 1, class 1] && col5 in (LGN3, class 1]

MineSet is a scalable client/server tool set for extracting information from databases, mining the data with analytical algorithms, and revealing newly discovered patterns and connections through intuitive, interactive, visual displays. A complete set of tools is provided for data mining, and it supports the generation and the analysis of association rules and classification models, used for prediction, scoring, segmentation, and profiling. Combining these models with animated, interactive visualisation, the system delivers a large set of utility functions to enhance and accelerate data analysis, including automatic data binding and column selection, critical to analytical accuracy and relevance. The results of the data mining are discussed in Section 7.2..

7.3 Results

Several target data sets have been created for the case study area according to the set of variables being used to discover association rules. The algorithm used was an unsupervised learning algorithm because it does not focus on any particular attribute.

Instead it treats all attributes equivalently, and does a global search for interesting rules. Table 15 illustrates the set of parameters used in the KDD process.

Table 15 - An example of the variables used within the target data set.

LGN1 – land use class	LGN2 - land use class	LGN3 - land use class
Landsat TM – band 1 Landsat TM – band 2 Landsat TM – band 3 NDVI	Landsat TM - band 1 Landsat TM - band 2 Landsat TM - band 3 NDVI	Landsat TM - band 1 Landsat TM - band 2 Landsat TM - band 3 NDVI
Classification Accuracy	Classification Accuracy	Classification Accuracy
Pixel Coordinates (X and Y)	Pixel Coordinates (X and Y)	Pixel Coordinates (X and Y)
Texture	Texture	Texture

The development of Association Rules has proceeded in two phases: the target data file is first processed by the Association Rule Generator, which creates a file usable by the visualiser. Then the visualiser displays the file, which in this case is a Scatter Visualiser. Four numbers quantify the strength of the association. They are:

1. *Support*: frequency of LHS and RHS occurring together, that is P(LHS $\cap$ RHS);
2. *Confidence*: of all occurrences of LHS, the fractions where RHS is also seen, or the support divided by the frequency of occurrence of LHS items, that is P(RHS $|$ LHS);
3. *Expected Confidence*: frequency occurrence of RHS items, that is P(RHS);
4. *Lift*: ratio of confidence to expected confidence, that is P(RHS $|$ LHS) : P(RHS).

7.4 Rules Visualisation

Figure 17 illustrates the association rules in a graphic form. Rule Visualiser tool was used for graphically displaying the results from the association rule algorithm. Analysing rules that have been discovered during data mining process gives us greater insight into the nature of the data sets. In the Figure 17, the left-hand side (LHS) items are on the X-axis and right-hand side (RHS) items are on The Y-axis. Attributes of a rule are displayed at the junction of its LHS and RHS axes. The attributes of a rule are the variables of the data sets used in the data mining (See Table 15), and they are shown as col1 to col18 in the figure. Moreover, the squares represent the existence of a rule, and the different colours represent the lift values ranging from 1 until 29.29.

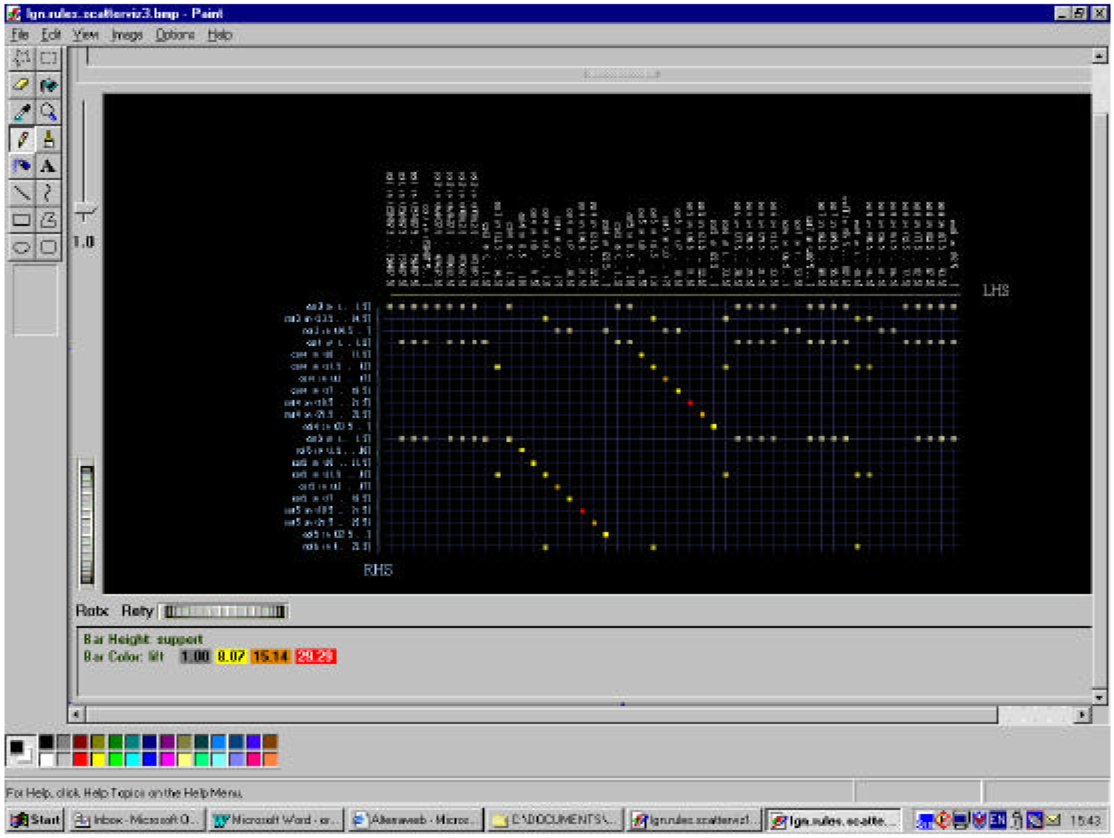


Figure 17. Overview of the Association Rules Visualiser's Main Window

As shown in Figure 18, a legend indicating the mapping between displayed attributes (such as bar heights and colours) and values associated with underlying rules (such as support and lift) is displayed at the bottom of the main window. For example, bar heights correspond to the support and bar colours correspond to the lift. Placing the mouse pointer over the Association Rules object displays that object's information.

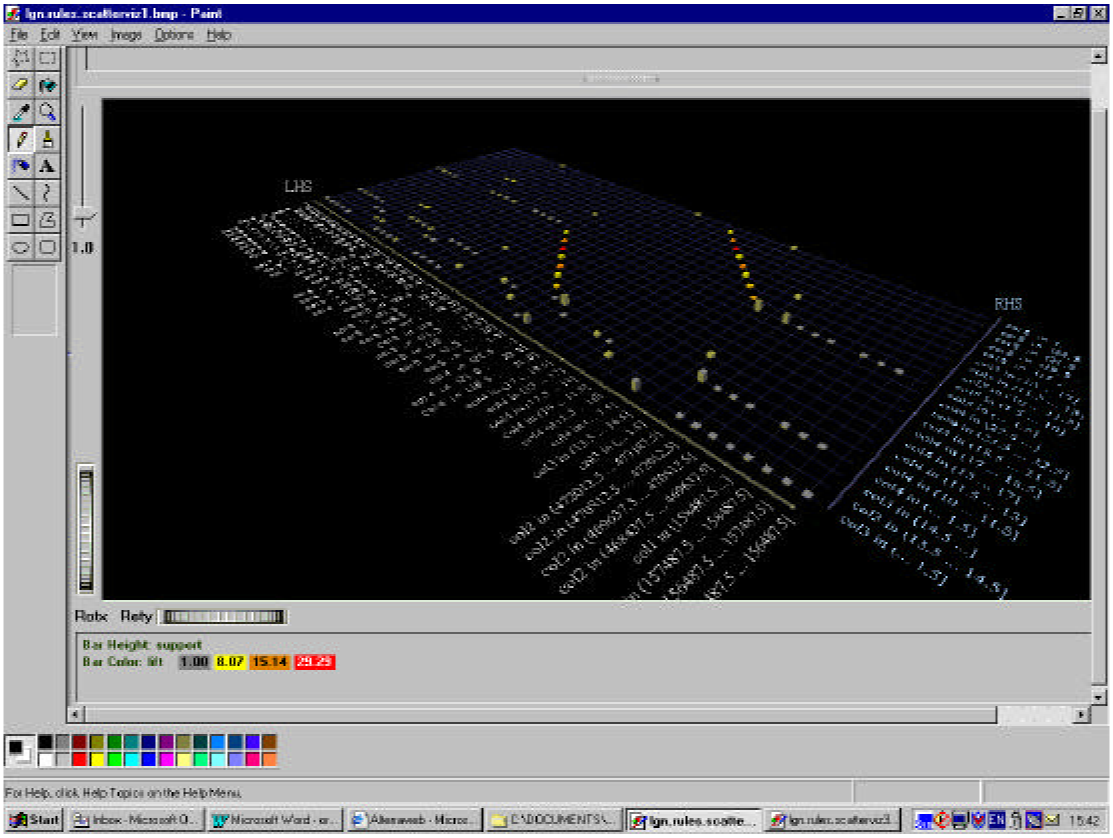


Figure 18. Detailed view of the Association Rules

7.5 Interpreting Association Rules in the Scatter Visualiser

The outcomes from the KDD process can be described using three sets of relevant relations:

- Space Dominant Association Rules that associate one or more pixels with other pixels independently of the land use class they belong to,
- Thematic Dominant Association Rules that associate pixels according to their land use classes,

Time Dominant Association Rules that associate pixels with different land use classes over different periods of time.

7.5.1 Space Dominant Rules

Two main space dominant rules were found in the target data sets. First, linear patterns were found in pixels located between the X co-ordinates of 155,487m and 158,487m. Second, linear patterns were also found in pixels located between the Y co-ordinates of 468,437m and 473,187m. The pixels showing these patterns belong to the grassland class for all LGN1, LGN2, and LGN3 classifications. Therefore, no spatial change has occurred in these pixels over a period of 10 years time.

7.5.2 Thematic Dominant Rules

The association rules have discovered the most stable land use classes in all three classifications (LGN 1, LGN2, and LGN3). Table 16 illustrates these land use classes with the data mining results. For example, knowing that a pixel was classified as deciduous forest in LGN 2, the probability that this pixel will be the same in LGN3 is that 99.91%. This rule is 15% prevalent throughout the target data set. However, having 6.9% as its lift value, this low value implies that this rule was well expected in the target data set. The most unexpected rules were found in the classes 22 and 14, 15, and 16 with lift values of 13.9% and 15.0% respectively.

Table 16. Most stable land use classes with their data mining results.

Land Use Class	Support	Lift	Confidence
Deciduous forest (class 12)	14.5%	6.9%	99.9%
Forest in urban areas (class 22)	4.6%	13.9%	99.8%
Urban city areas (class 18)	11.3%	3.3%	99.7%
Coniferous forest (class 11)	8.6%	11.3%	99.4%
Rest open nature areas with vegetation (class14) Bare ground in nature areas (class 15) Fresh water (class 16)	6.9%	15.0%	98.9%
Grassland (class 1)	31.8%	2.6%	96.2%

7.5.3 Time Dominant Rules

These rules have discovered conflicting situations that have occurred when a certain pixel takes part in several land use classes. Such circumstance was greatly influenced by the classification accuracy obtained by the different methods and remote-sensing images used for the classification of land use in LGN1, LGN2, and LGN3 data sets. The rules indicate that the classification accuracy is not enough for the detection of land use change in the target data set. There is also the possibility that the target data set was inadequate for the proposed discovery of time dominant rules. However, the measure of change was used to calculate the difference between expected confidence and confidence. The results of the data mining have shown that the classes that had the most unreliable values of measure of change were:

- Maize (class2), potatoes (class 3), grains (class 5), other crops (class 6), bare ground (class 7), greenhouses (class 8), orchard (class 9);
- Rest open nature areas with vegetation (class 14), bare ground in nature areas (class 15), and fresh water (class 16);
- Rural built-up areas (class 19), coniferous forest in urban areas (class 20), deciduous forest in urban areas (class 21);

- Grassland in urban areas (class 23), bare ground in rural built-up areas (class 24), and infrastructure (class 25).

7.6 Discussion and conclusion

Knowledge discovery is a process that reflects not the end result a user would want, or what the system should produce, but rather the operations a user will perform using the tools of a knowledge construction environment. This is an important distinction because it places the user first in the knowledge discovery process, but it also highlights the semantic 'gap' between desired outcomes and the available tools. Existing knowledge environments such as MineSet focus on identifying what kind of knowledge the system needs to discover successfully. However, it is very important to take the design initiative of enabling the users' expertise within a knowledge discovery process, rather than attempting to supplant it.

The strength of the development of KDD methods lie in the powerful integrated system that they can provide for spatio-temporal database modelling. A system from which to store, explore, and evaluate data, and subsequently understand and communicate this understanding. The forthcoming tools will facilitate pattern noticing, whether this pattern is used for steering data mining, identifying, comparing, and analysing objects, or trying to link patterns to processes. The key goal is to find relationships among objects in thematic, temporal, or space within a single transparent environment that is intuitive and supportive of the heuristics of the domain expert while defining a flexible, adaptive control structure for algorithmic process and graphic user interface - the ideal paradigm for supporting spatio-temporal database modelling.

The results of the KDD process have demonstrated the necessity and the potential of data mining techniques in land use monitoring systems. Association rules are the first step for capturing the complexities of underlying processes without using the GIS overlay approach. The main goal is to look for relationships, each of which occurred separately, which is the cause of land use changes. In our case, association rules were particularly useful in discovering patterns in LGN1, LGN2, and LGN3 data sets, with rules of association indicating the frequency of items occurring in the target data set (i.e. the case study area).

Although this method has not produced a geometric description of the changed pixels, it has provided indicators for the land use classes. Two land use classes may vary in terms of their composition, spatial arrangement of units and variability through time. This is often insufficient to classify a given land use class with a single statistics. In fact, the identification and definition of components composing a land use class in itself will determine what form the land use change process.

KDD methods have shown that different land use changes and their corresponding processes MUST be established, and we must relate statistical significance to the change occurrence. In other words, for actually monitoring land use changes, we MUST look at the process (e.g. urbanisation, crop monitoring, etc) rather than the land use classes. This will require a good understanding of how certain processes are related to an observed land use class. Current approaches in GIS give us a better understanding of the range of types of classes (e.g. urban, grass) rather than the process responsible for their generation (e.g. fire, urbanisation, climate change).

A suite of complex factors can drive land use changes, including policy, population change, culture, economics, and environmental characteristics. Human-induced changes to the land often result in changes to processes such as alteration to vegetation cover, species diversity, and changes to the economies of a community. Therefore, we propose as future research goals directed towards discovering patterns in spatio-temporal data and the processes responsible for their generation. There are many research issues that need extensive studies to make mining knowledge in land use management a reality, including:

- Additional data mining tasks and methods such as pattern-based or similarity-based mining, and meta-rule guided data mining, should be further studied for mining land use data;
- Improvement of visualisation of data mining results, using interactive techniques, visual feed backs, and dynamic visualisation;
- The measurement of interestingness of discovered patterns and the handling of uncertainty and incomplete information in data mining is also an important research issue;
- Use of other databases that are relevant for the understanding of land use changes such as statistical data sets, economical data sets, biodiveristy data, etc…

8 Conclusions and recommendations

8.1 Conclusions

In the presented research project the role of geo-information in the process of monitoring land use was investigated. The conclusions are mainly based on a case study done in the municipality of Soest, but are believed to be representative for the Netherlands and of use to similar research activities abroad.

Three national spatial databases were selected (LGN, CBS-land use statistics and Top10vector) that could be used for the monitoring of land use changes. The municipality of Soest was multi-temporal available in all three databases and considered representative in its' land use types for the Netherlands.

It was shown that simple overlay techniques at different thematic and spatial aggregation levels do not provide reliable user information on the real world land use changes. Especially Top10vector and LGN performed poor. This is mainly caused by changes in the applied data recording, data structuring and semantic content over time; new versions of spatial databases are in general adapted to actual user requirements and available techniques, making multi-temporal analysis very difficult.

An improved GIS procedure (MonGis) was developed for monitoring land use with vector based data sets. MonGis was tested with the Top10vector database and achieved in certain cases a reduction of 30% in the quantity of unreal changes. However, this is still insufficient for reliable monitoring results. It is believed that further refinement and development of the method could improve the results further, but not till a reliable extent.

For the LGN database it was investigated if reliable monitoring results can be obtained by adapting the data capturing and structuring process. At a superclass level a 'layer of real world changes' was added tot the database, containing the real world changes as they could be established during the process of database actualisation. This layer of changes is mainly based on visual techniques and therefore operator dependent. The changes in land use captured with this technique are within an acceptable range compared to the reference data used, but are mainly under-estimated. This is caused by subjectivity, the spatial and temporal scale of the database and the sometimes difficult distinction between land use and land cover.

Knowledge discovery in databases (KDD) is a process that reflects not the end result a user would want, or what the system should produce, but rather the operations a user will perform using the tools of a knowledge construction environment. The strength of the development of KDD methods lie in the powerful integrated system that they can provide for spatio-temporal database modelling.

The results of the KDD process have demonstrated the necessity and the potential of data mining techniques in land use monitoring systems. Association rules are the first step for capturing the complexities of underlying processes without using the GIS overlay approach. However, in this study no geometric description of the changed pixels have been produced.

8.2 Recommendations

Users of geo-information doing multi-temporal analysis should be aware of the restrictions of the data. It is easy to derive trends and achievements from geo-information not reflecting the real-world situation at all.

Providers of geo-information should be aware that the demand for data, suitable for multi-temporal analysis, is growing. Inconsistency of the structure, scale and semantic content should be clearly documented in the accompanying meta-data of a data set. At the same time, this report might provide them a reference on how spatial data sets can be improved for monitoring practices. A closer co-ordination between different data providers in the future on up-date frequencies, semantic content and spatial resolution might allow for better possibilities for monitoring land use by combining data sets.

Although the development of MonGis has improved the monitoring results for existing data sets till some extent it is not recommended to extend this procedure without making adaptations to the data sets it selves. However, in combination with modified data capturing and structuring, a MonGis like GIS module might be able to improve monitoring results considerably.

With respect tot knowledge discovery in databases future research goals should be directed towards discovering patterns in spatio-temporal data and the processes responsible for their generation. To distinguish between changed objects representing real world changes and noise in a spatial database more profound research is required than performed in this research project.

References

Aalders, H.J.G.L. (1998). *The nominal ground in GIS*, CD-ROM proceedings, Gisplanet '98, Lisbon, Portugal.

Aspinall, R.J., Miller, D.R. and Birnie, R.V. (1993). *Geographical information systems for rural land use planning.* Applied Geography, 13, 54-66.

Bernhardsen, T. (1999). *Geographic information systems: an introduction*, (New York: John Wiley & Sons).

Bishr, Y. (1997). *Semantic aspects of interoperable GIS*. ITC publication series, no. 56, Enschede, the Netherlands.

Bruin, S de, 2000. *Geographical Information Modelling for Land Resource Survey*. Ph.D – thesis, Wageningen University, Wageningen, the Netherlands

Burrough, P. (1986). *Principles of Geographical Information Systems* (Oxford: Oxford University Press).

Dijkstra H. and Roos-Klein Lankhorst, J. (1995). *Haalbaarheidsstudie meetnet landschap.* Onderzoeksreeks nota landschap nr.4, SC-DLO rapport 390 (Dutch).

Egenhoffer, M.J. and Al-Taha, K.K. (1992). *Reasoning about gradual changes of topological relationships.* In: Theories and Methods of Spatio-Temporal Reasoning in Geographical Space, Frank, A. et al. (eds), Springer Verlag, 196-219.

Fayyad, U.M. (1996). *Data Mining and Knowledge Discovery. Making sense out of data*. IEEE Expert 11: (5) 20-25. IEEE, Los Alamitos, California

Frank, A. (1998). *Metamodels for Data Quality Description.* In: Data Quality in Geographic Information; from error to uncertainty (Paris: Hermes), pp. 15-30.

Frank, A.U. (1994). *Qualitative temporal reasoning in GIS – ordered time scales.* Proceedings of the SDH'94 Conference, **1**, 410-30.

Fritsch, D. and Anders, K.H. (1996). *Objektorientierte konzepte in Geo-Informationssystemen.* Geo-Informations-Systeme, Zeitschrift fur raumbezogene information und enscheidungen, 9, no. 2, pp. 2-14.

Harvey, D. (1969). *Explanation in Geography*, St Martin's Press.

Hendriks, P. and Ottens, H. (Eds.) (1997). *Geografische Informatie Systemen in ruimtelijk onderzoek*. Van Gorcum, Assen, The Netherlands (Dutch).

Holt, A., MacDonell, S.G., and Benwell, G.L. (1998). *Assessing the degree of spatial isomorphism for exploratory spatial analysis.* The Third International Conference on GeoComputation, University of Bristol, Bristol, England, CD-ROM Paper No. 17, 28 pages.

Jammer, M. (1969). *Concepts of space*. Cambridge, Harward University Press.

Kent, W. (1991). *The Breakdown of the Information Model in Multi-Database Systems.* Sigmond Record, 20, no. 4, 10-15.

Kent, M., Jones, A. and Weaver, R. (1993). *Geographical Information systems and remote sensing in land use planning: an introduction*. Applied Geography, 13, 5-8.

Kim, C-J (1999). *Implementation of Semantic Translation for Finding the Corresponding Geometric Objects between Topographic Databases.* MSc. thesis, International Institute of Aerospace Survey and Earth Science, Enschede, The Netherlands.

Kucera, G.L. (1996). *Temporal extensions to spatial data models: Final report.* US Army Construction Engineering Research Laboratory.

Langran, G. (1989). *A review of temporal database research and its use in GIS applications.* International Journal of GIS, **3**(3), 215-32.

Lindgren, D.T. (1985). *Land use planning and remote sensing* (Dordrecht, The Netherlands: Martinus Nijhoff publishers).

Meijners, R.H. (1998). *Gis Monitoring Landgebruik. Een casestudy in de gemeente Soest naar bruikbaarheid van de CBS-Bodemstatistiek, het LGN-bestand en de TOP10vector voor GIS monitoring landgebruik.* Interne mededeling 529, Wageningen (Dutch).

Michalak, W.Z. (1993). *GIS in land use change analysis: integration of remotely sensed data into GIS.* Applied Geography, 13, 28-44.

Molenaar, M. (1996). *The role of topologic and hierarchical spatial object models in database generalization*. In: Methods for the generalization of geo-databases, Delft: Netherlands Geodetic Commission, New Series, no. 43, pp. 13-36.

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

Noordman, E., Thunnissen, H.A.M. and Kramer, H. (1997). *Vervaardiging en nauwkeurigheid van het LGN2-grondgebruiksbestand. Achtergrondinformatie bij gebruik van het bestand.* SC-DLO Rapport 515 (Dutch).

Openshaw, S. (1998*). Building automated geographical analysis and explanation machines.* In: Geocomputation; A primer (Chicester, UK: John Wiley & sons), pp. 95-115.

Ornstein, R.E. (1969). *On the experience of time*. London, Penguin

Peuquet, D. (1994). *It's about time: a conceptual framework for the representation of temporal dynamics in geographic information systems.* Annals of the Association of American Geographers, **84**(3), 441-61.

Sonneveld, M.P.W., Brink, J.A.T. van den, Zeeuw, C.J. de and Bregt, A.K. (2000). *Noise reduction in GIS-based land use monitoring.* Proceedings 9th internationala symposium on spatial data handling, Beijing, China. Section 5a, pp 44-56.

Stoorvogel, J.J. (1995). *Geographical information systems as a tool to explore land characteristics and land use with reference to Costa Rica.* PhD Thesis, Wageningen, The Netherlands.

Topografische Dienst (1995). *Productbeschrijving Top10vector. Versie 1.0.* Emmen, The Netherlands (Dutch).

Uitermark, H., Oosterom, P., Mars, N. van., and Molenaar, M. (1998). *Propagating updates: Finding Corresponding Objects in a Multi-Source environment,* International Symposium for Spatial Data Handling SDH'98, Vancouver, Canada.

Vaughlin, F. (1998). *Statistical representation of relative positional uncertainty for geographical linear features.* In: Data Quality in Geographic Information; from error to uncertainty (Paris: Hermes), pp. 87-96.

Wachowicz , M. (1999).*Object-Oriented Design for Temporal GIS.* Taylor&Francis.

Wachowicz, M. (2000a). *How can knowledge discovery methods uncover spatio-temporal patterns in environmental data?* In: Data Mining and Knowledge Discovery: Theory, Tools, and Technology II, B.V. Dasarath (ed), Proceedings of SPIE, **4057**(2000), pp. 221- 229.

Wachowicz, M. (2000b). *GeoInsight: an approach for developing a knowledge construction process based on the integration of GVis and KDD methods.* To appear in: Geographic knowledge discovery and spatial data mining (Eds. Miller, H. J. and Han, J.), London: Taylor & Francis.

Wachowicz, M. (2000c). *The role of geographic visualisation and knowledge disocvery in spatio-temporal data modelling.* In: Time in GIS: Issues in spatio-temporal modelling. Heres, H. (ed). Publications in Geodesy 47, pp. 13-26.

Walsh, S.J., Lightfoot, D.R. and Ruler, D.R. (1987). *Assessment of inherent and operational errors in geographic information systems.* In: ASPRS-ACSM annual convention; Technical Papers, Baltimore, vol. 5.

Wit, A.J.W. de, Heijden, Th.G.C. van der and Thunnissen, H.A.M. (1999). *Vervaardiging en nauwkeurigheid van het LGN3-grondgebruiksbestand.* SC-DLO Rapport, in prep (Dutch).

Wijngaarden, F.A., Putten, J.D. van, Oosterom, P.J.M. van and Uitermark, H.T. (1998). *De kaartintegrator. Mutatiepropagatie tussen onafhankelijke topografische databases.* In: Geodesia, 98-3, pp. 109-116 (Dutch).

Zeeuw, C.J. de (1998). *Monitoring land use in The Netherlands, using national geographical databases.* CD-ROM proceedings, Gisplanet '98, Lisbon, Portugal.

Zeeuw, C.J. de, Meijners, R.H. and Bregt, A.K. (1999a). *Geo-informatie voor monitoring landgebruik. Veranderingen om ons heen of ruis in de databestanden?* In: Geodesia, 99-3, pp 133-139(Dutch).

Zeeuw, C.J. de, Bregt, A.K., Sonneveld, M.P.W. and Brink, J.A.T. van den (1999b). *Geo-information for monitoring land use; From map overlay to object-structured noise reduction.* Tenth international workshop on database and expert systems applications, 1-3 september 1999, Florence, Italy. IEEE, Los Alamitos, California. Pp 419-425.

Zeeuw, C.J. de, Bregt, A.K., Sonneveld, M.P.W. and Brink, J.A.T. van den (2000). *Geo-information for monitoring land use.* GIS2000 conference proceedings, March 13-16, 2000, Toronto, Canada.