

DATA INTEGRATION FOR PAN EUROPEAN LAND COVER MONITORING

Victor van KATWIJK

Geodan IT
victor@geodan.nl

Wideke BOERSMA

CSO, The Netherlands
w.boersma@bunnik.cso.nl

Sander MÜCHER

Alterra, the Netherlands
s.mucher@alterra.wag-ur.nl

Klaus STEINNOCHER

Austrian Research Centre Seibersdorf
klaus.steinnocher@arcs.ac.at

Jean-louis CHAMPEAUX

Centre National de Recherches Météorologiques, France
jean-louis.champeaux@meteo.fr

Silvio GRIGUOLO

Istituto Universitario di Architettura, Italy
silvio@cidoc.iuav.unive.it

Kjell WESTER

Satellus, Sweden
kwe@satellus.se

Camiel HEUNKS

National Institute for Public Health and the Environment (RIVM), the Netherlands
s.mucher@alterra.wag-ur.nl

Working Group TC IV-1

KEY WORDS: Landcover, Remote Sensing, Fusion, Integration, Land Use

ABSTRACT

The need for reliable and up-to-date Pan European Land Use (LU) and Land Cover (LC) information is well known. Some steps have already been taken; one of them in the Pan-European Land Cover Monitoring project (PELCOM) in the EC 4th Framework Program. A concise classification methodology was established for mapping major land cover types (e.g. forest, grassland, and arable land) using a regional and integrated approach of NOAA-AVHRR satellite data and ancillary information.

It is the sensible integration of the numerously available sources of LU/LC related information that is expected to greatly enhance the value of the information derived from satellites. The available ancillary LU/LC information ranges from continuous topographic information (altitude, distance from the sea, latitude), to climatic maps, to existing thematic land cover maps (SEI map, CORINE, wetlands map), all providing complementary information on the actual and potential land cover.

Combining the ancillary sources of information with the processed land cover signatures derived from, for instance the NOAA/AVHRR satellites, it becomes feasible to produce more reliable land cover information.

In the PELCOM project, a technique has been developed and tested for integrating various sources of LU/LC information. The most critical part of this technique is to capture the qualitative knowledge that is typically available; valuable knowledge of both the reliability of the information and of the content of the available sources. What does the presence of a particular source class infer about the likelihood of the target classes? The challenge is to convert this

qualitative information into quantitative rules. In the proposed technique, matrices will be established expressing the value of the content and the reliability of the sources relative to the desired land use classes. These matrices will be used to convert the source information into LU/LC probabilities, resulting in probability maps for each desired land cover class. As a post-processing step, the probabilities can be converted to a land cover map, using a maximum likelihood classification.

In preliminary tests, the prototype shows promising results. An important potential improvement is back modelling functionality, thus providing a mechanism for knowledge extraction in addition to knowledge capturing. A main advantage of the prototype is that information can be combined in a flexible way meeting specific user needs in a better way.

1 INTRODUCTION

The need for reliable and up-to-date Pan European Land Use (LU) and Land Cover (LC) information is well known (Mücher et al., 2000). Some steps have already been taken; one of them in the Pan-European Land Cover Monitoring project (PELCOM, http://www.geodan.nl/ec_lu) in the EC 4th Framework Program. A concise classification methodology was established for mapping major land cover types (e.g. forest, grassland, and arable land) using a regional and integrated approach of NOAA-AVHRR satellite data and ancillary information (Mücher et al., 2000).

Dealing with the integration of multiple sources of geographic information is an essential part of a methodology for establishing a land use database. The classification of satellite data alone is often not sufficient for mapping the environment. This certainly applies to the classification of NOAA-AVHRR data. The classification of remotely sensed satellite imagery for thematic mapping applications can be improved by the use of ancillary geographic information. In the EC funded PELCOM project (see 5), a methodology is being established for creating a European Land Use database for environmental modelling. An essential part of this methodology is a tool/method for integrating multiple sources of land use information with the information resulting from the classification of remote sensing imagery. Spatial data in a GIS have been shown to improve classification accuracy and aid in the extraction of information from remotely sensed imagery (Jansen 1994, Bolstad and Lillesand, 1992). In the context of the PELCOM project a preliminary literature study of methods for integrating satellite imagery and ancillary data is being done. Methods include incorporation of ancillary data before, during, or after a classification.

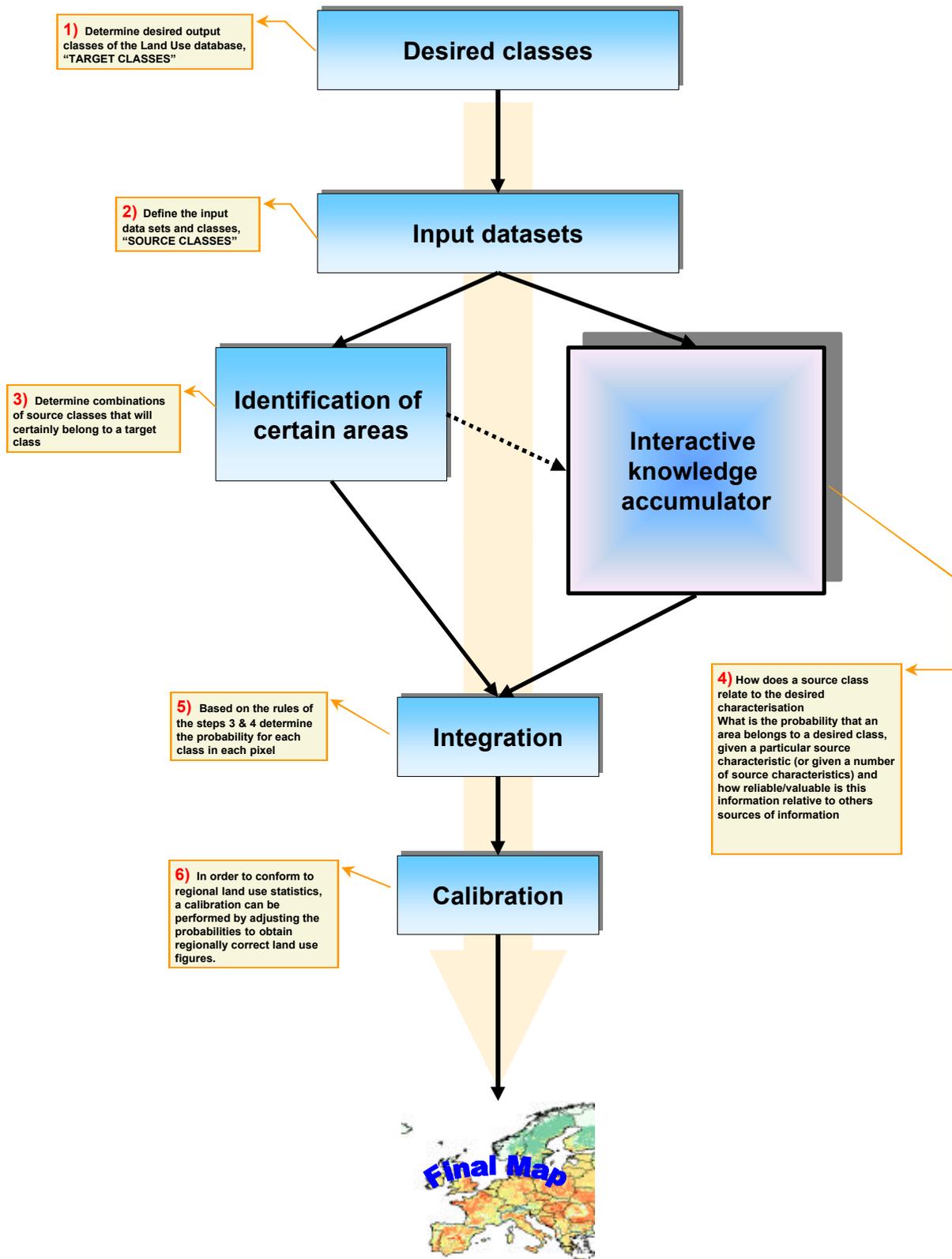
In the Land Use Integrator (LUI) the purpose of the methodology is to use and include the additional knowledge in the information extraction process. The aim of the method is to include knowledge of the thematic content of the information layers as opposed to the use of a priori knowledge on spatial characteristics as is done in other projects such as the CAMOTIUS project.

It has been found in the previous Land Use projects that there are numerous sources of ancillary land use information. These sources can have different thematic content with different relationships to the desired land use categories. In addition, the reliability of these sources will vary, and even within a data set the classes may have different reliabilities. It is also important to note that the analysis of remote sensing imagery can provide different types of information depending on the method used.

With a LUI (Land Use Integrator) the information from various analyses techniques can be combined with the ancillary geographic information to provide the best possible land use data set.

2 METHODOLOGY

The main objective of the Land Use Integrator is to combine data sets with geographic land use information as efficiently as possible, into a desired land use data set. In the methodology proposed the fuzzy knowledge about data sets and classes and the meaning of these data sets and classes to the actual land use at a location, is converted into relationships with probabilities and reliabilities. By bringing all the information and knowledge together simultaneously, the loss of information in the data is minimized and further enhanced by capturing the knowledge of classes etc.. The methodology to be used for integrating different data sets should be flexible, making it possible to add or remove data sets in the future.



The data integration is performed in six steps, as can be seen in the flowchart.

Step 1) Desired Classes

At first, the desired output classes of the Land Use database will be determined. These desired output classes are called "TARGET CLASSES".

Step 2) Input datasets

The second step is to define the input data sets and classes, the so called "SOURCE CLASSES". If a data set is added the LUI automatically searches for the existing source classes and the frequency of occurrence of these classes.

Step 3) Identification of certain areas

In this step the pixels for which the land use/ land cover class (target class) are almost 100% certain are identified. If a pixel has the same source class value, for example forest, in different data sets one can assume the target class in that pixel actually is forest. These pixels will not be changed in the next step of the data integration. The target classes in these pixels will be considered as locked. It could be considered, in a later stage, to change certain pixels into pixels with a high degree of probability that may be influenced by other information if the current method proves to be too rigid.

Step 4) Interactive knowledge accumulator

The fourth step is the most important step in the LUI. In this step the question will be answered: "How does a source class relate to the desired characterization".

What is the probability that an area belongs to a desired target class, given a particular source characteristic (or given a combination of source characteristics) and how reliable/valuable is this information relative to others sources of information. A matrix will be generated for each LU target class where for each combination of a source class and a data set the probability of this LU-class/data set/source-class combination must be indicated (probability matrix).

The probabilities indicated for the different sources will have a different reliability. For each source data set the relative reliability must be indicated. For example, one data set may be twice as reliable as another data set etc. In addition, it is likely that a certain source has classes that do not have the same reliability. When looking at the forest LU class, the forest class (e.g. in ESA) may, for instance, have a higher reliability than the urban class. For all combinations, this class reliability may be important, it can therefore be input in the so-called reliability matrix.

Step 5) Integration

Based on the rules of the steps 3 & 4 the probability for each target class in each pixel will be determined. In this step, other sources with probabilistic information can be included. From these, a resulting land use map can be made using a maximum likelihood classification.

Step 6) Calibration

A prerequisite of the resulting data set can be that the regional class areas comply with the existing statistical areas. For instance if the land use data set will be used for environmental models influencing national and European political decisions it is important that the regional figures match the statistical information. For this reason, the methodology will include a procedure for adjusting the land use classification to match the statistical values. By integrating several data sets into one land use/land cover database, the quality of the result will exceed the reliability of the separate data sets. The integrated land use/land cover database will be used for monitoring land use on European level.

2.1 The prototype

A prototype has been developed according to the Geodan software specification, using Delphi and Geolib. Geolib is the geographic software library/toolbox (C++) developed by Geodan.

3 DISCUSSION

It should be realized that the integration of information from the classification often needs further interpretation and information from other sources before land use can be reliably determined. It is to be expected that the procedures developed for integrating ancillary data will prove to be useful during the classification process also.

A major new project is proposed to the EU for realizing an Internet based environment for establishing LC/LU information. The project will be called STEMS but at the time of publication of this paper the project still had to be approved. The overall objective of the proposal is to build a Spatio-Temporal European Monitoring System for monitoring land cover changes in order to derive environmental indicators, which can then be applied in the fields of

water balance modelling, wetlands conservation, climate modelling, and biodiversity assessment. The project will explore new techniques such as data mining, geographic visualisation, and decision support systems to develop new tools, which can be used to integrate and explore geo-referenced data for monitoring land cover changes. The emphasis will be on detecting significant land cover changes and trends from regional to European scales using different Earth Observation systems. The results of this project will be the development of a concise methodology and the use of appropriate technology to implement a prototype of the Spatio-Temporal European Monitoring System, which will be used to improve change analysis, data management and information access to users from environmental institutes, regional governments, environmental policy making organisations and research centres.

REFERENCES

- Bolstad, P.V. and T.M. Lillesand, 1992. Rule-Based Classification Models: Flexible Integration of Satellite Imagery and Thematic Spatial Data. In: Photogrammetric Engineering and Remote Sensing, Vol. 58, No.7, pp. 965-971. 6 pp.
- Janssen, L.F., 1994 Methodology for updating terrain object data from remote sensing data: the application of Landsat TM data with respect to agricultural fields. Thesis, Wageningen, 173 pp.
- Mücher, S, Steinocher, K, Champeaux, JL., Griguolo, S, Wester, K, Heunks, C, Katwijk, V. van, Establishment of a 1-km pan-european land cover database for environmental monitoring, ISPRS-2000 proceedings, Amsterdam, 2000.
- Wel, F.J.M. van der and B.G.H. Gorte, 1995. CAMOTIUS Hoofdfase, 1 januari 1993 - 31 december 1994. BCRS report nr. 95-06, Delft, the Netherlands. 89 pp.
- Wel, F.J.M. van der and B.G.H. Gorte, 1996. CAMOTIUS Eindfase, 1 januari 1995 - 30 juni 1996. BCRS report nr. 96-31, Delft, the Netherlands. 25 pp.