ESTABLISHMENT OF A 1-KM PAN-EUROPEAN LAND COVER DATABASE FOR ENVIRONMENTAL MONITORING

Sander MÜCHER
Alterra, Green World Research, the Netherlands
s.mucher@alterra.wag-ur.nl

Klaus STEINNOCHER
Austrian Research Centre Seibersdorf, Austria
klaus.steinnocher@arcs.ac.at

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

Silvio GRIGUOLO
Instituto Universitario di Architettura, Italy
silvio@cidoc.iuav.it

Kjell WESTER
Satellus, Sweden
kwe@satellus.se

Camiel HEUNKS
National Institute for Public Health and the Environment (RIVM), the Netherlands
camiel.heunks@rivm.nl

Victor van KATWIJK
Geodan IT, the Netherlands
victor@geodan.nl

KEY WORDS: land cover, pan-Europe, NOAA-AVHRR, database, classification

ABSTRACT

In the current policy frameworks, such as the European Environmental Outlook and the Economic Assessment of Priorities for a European Environmental Policy Plan, the need for up-to-date and reliable information for the whole of Europe is well known. A step in this direction has been taken by the Pan-European Land Cover Monitoring (PELCOM) project. In this project a concise classification methodology was established for mapping major land cover types (e.g. forest, arable land and grassland) by using a regional and integrated approach of NOAA-AVHRR satellite data and ancillary information. One of the main products of this project was the 1-km pan-European land cover database. An extensive accuracy assessment has been implemented on basis of 40 interpreted Landsat-TM satellite images distributed over pan-Europe. The accuracy of the land cover class ' arable land' exceeded the 75%, while the accuracy of highly fragmented classes such as ' grassland' did not exceed the 50%. The total average accuracy was 69.2%, which was a good result considering the mixed pixel and geo-referencing problems of AVHRR data. Integration with thematic ancillary data sources covering Europe was hampered by their large variation in geographic accuracy, reliability and acquisition date. The PELCOM land cover database was applied successfully in various environmental and climate studies. However, the use of AVHRR satellite data has its limitations for monitoring purposes due to the fine scale at which thematic land cover changes take place. Medium resolution satellite imagery (eg. MODIS and MERIS) will play a crucial role in the near future to monitor the European landscape.

1 INTRODUCTION

For many environmental and agricultural studies up-to-date and reliable information on land use and land cover is urgently needed. At the moment there are, however, no geo-referenced land use or land cover databases with a high spatial accuracy that cover entire Europe and that can be updated frequently.

During the last decades, land use has suffered dramatic changes in time and space in Europe. For example agriculture in the European Union (EU) is going through a phase of accelerating changes (WRR, 1992). The continued increase in production per unit of land area and per unit of livestock, due to improved production circumstances, better cultivation methods and external inputs, has led to significant increases in agricultural productivity. Land use intensification in one area might cause marginalisation in other areas. A net gain of forest with 144980 km^2 (+10%), a net loss of arable land with 157320 km^2 (-11%), and a net loss of permanent pasture with 102280 km^2 (-11%) are some examples of land use changes in Europe over the 3 last decades (FAO statistics - http://apps.fao.org) - Russian federation was not included)/

In the mid-nineties, the ten-minutes pan-European land use database of the Dutch National Institute for Public Health and the Environment was a first step towards meeting the demands of environmental models on a European scale (Van de Velde *et al*. 1994, Veldkamp *et al*. 1995). A major drawback of this database is that it is based on statistical and spatial data from different sources that differ in spatial accuracy, reliability, age and nomenclature. Use of remotely sensed data decimates this problem, as actual land cover data can be derived in a consistent manner with a high spatial accuracy.

Relevant activities on the derivation of pan-European land cover databases from remotely sensed data including all major land cover classes are: the CORINE (Coordination of Information on the Environment) land cover project (CEC 1993), now under supervision of the European Environment Agency (EEA) and the recent developed 1-km global land cover product, ‘ DISCover’ (Loveland and Belward 1997), under the co-ordination of the International Geosphere and Biosphere Programme (IGBP-DIS). These are described briefly below.

The CORINE land cover database is produced by visual interpretation of high-resolution satellite images, e.g. Landsat-TM and SPOT-XS, at a scale of 1 : 100 000, with simultaneous consultation of ancillary data (CEC 1993). The CORINE legend distinguishes 44 classes grouped in a hierarchical nomenclature and is landscape and ecology oriented. The CORINE land cover database is the most detailed database at the moment that covers a large part of Europe. Despite the uniqueness of the CORINE land cover database its use is still limited for pan-Europe because a complete coverage is still not accomplished.

IGBP-DIS began a project in 1992 to produce a global land cover data set at a spatial resolution of 1-km (scale ~ 1:5M), derived from the Advanced Very High Resolution Radiometer (AVHRR) onboard the US National Oceanic and Atmospheric Administration’ s (NOAA) polar-orbiting satellite series (Loveland and Belward 1997). The methodology is based on unsupervised clustering of monthly NDVI maximum value composites on a continental basis. The clusters are labeled by expert knowledge. Major limitation of the approach is that it is implemented on a continental basis without any stratification. Therefore, the result may be more closely related to agro-ecological zones, i.e. zones of similar phenology, than to the different land cover types existing in each agro-climatic zone. The European landscape is heterogeneous and fragmented and requires a stratified approach. Moreover, experiences indicate that the clustering technique does not identify forests satisfactorily (Champeaux et al. 1998 a). An additional limitation is that the 1-km database according to the DISCover legend contains complex classes, e.g. cropland/natural vegetation mosaics (about 27 per cent of the pan-European land surface), which are difficult to apply in environmental studies. However, it must be stressed that the project is unique and enormous effort had to be invested in order to establish an up-to-date global land cover database at a 1-km resolution in a consistent manner. Application of the database in environmental and climate studies for pan-Europe may be limited.

Besides the detailed CORINE land cover database and the global DISCover database with their own advantages and disadvantages, there is a need for additional land cover data sets derived from remotely sensed data that fulfils the needs for pan-European environmental and climate modeling. These needs comprise a consistent land cover database that covers pan-Europe, that can be easily updated, and that contains main land cover classes such as arable land, grassland, urban areas, water bodies, wetlands, barren land and the various forest types with sufficient regional detail (Van de Velde et al. 1994). Therefore a study was initiated to investigate the applicability of AVHRR satellite data for Pan-European Land Cover Monitoring (PELCOM). PELCOM was a three years' EU project under the Environment and Climate section of the European Union 4th Framework RTD programme (http://cgi.girs.wageningen-ur.nl/cgi/projects/eu/pelcom/). End users were actively involved in the PELCOM project: both dealing with the development of the methodology and the applications.

2 DATA

2.1 Satellite data

The AVHRR (Advanced Very High Radiometric Resolution) sensor on board of the meteorological polar orbiting satellite NOAA (National Oceanic and Atmospheric Administration) can cover Europe by two or three full scenes on a daily basis. In comparison with the METEOSAT satellite sensor, the AVHRR sensor has a high spatial resolution. In comparison with the high resolution satellite data, such as SPOT and Landsat TM, AVHRR satellite data has a high temporal resolution, which makes it more likely to find a cloud-free image for a specific area at a certain time period (see Table 1). In finding an equilibrium in a high spatial resolution, a high temporal resolution and covering large areas, the AVHRR sensor seems to be a very suitable sensor for mapping and monitoring the European continent.

Table 1 Characteristics of various satellite sensors

	METEOSAT	NOAA/AVHRR	LANDSAT-TM	SPOT-XS
Spatial resolution	2.5 km (VIS) 5 km (TIR)	1.1 km (nadir)	30 meter	20 meter
Temporal resolution	½ hr	1/2 day	16 days	26 days
Swath width	-	2400 km	185 km	120 km
Altitude	35,900 km	805 - 870 km	715 km	832 km

One of the AVHRR data sources was the MARS (Monitoring Agriculture by Remote Sensing) archive which is maintained by the Space Applications Institute (SAI) of the Joint Research Centre (JRC). This archive contains unique pre-processed daily multi-spectral mosaics of AVHRR images covering entire Europe. NDVI maximum value composites were selected from the DLR (Deutches Zentrum für Luft und Raumfahrt) archive for the year 1997. This data set was used as the main data source within the classification. In addition, other AVHRR data sources such as the IGBP-DIS (EROS Data Center) and Lannion archive were used in specific cases.

2.2 Ancillary data

During the last two decades, significant advances have been made in developing techniques for the use of satellite data jointly with ancillary data to improve the results of satellite image analysis (Brown et al., 1993; Palacio-Prieto and Luna-Gonzalez ,1996). Ancillary data are now considered of crucial importance for methodology improvement of land cover mapping on a European scale. Ancillary data may exist as discrete, thematic or continuous data (Brown et al. 1993).

Topographic features from the Digital Chart of the World (DCW) and Bartholomew Euromaps, e.g. rivers, lakes, coastlines, played a role when additional geometric corrections were necessary. In combination with the CORINE land cover database some important masks for *water*, *urban areas* and in a later phase (after validation) *wetlands* were extracted from these data sources. One of the most important ancillary data sources in the PELCOM project was the CORINE land cover database that provides highly detailed land cover information for a large part of Europe. Homogeneous samples from the CORINE database were used for training the classifier.

Stratification was based on data from the FIRS (Forest Inventory by Remote Sensing) project. In the framework of the FIRS project of SAI/JRC a regionalization and stratification was made for European forest ecosystems (EC 1995, Kennedy et al. 1996). On the first level (regions) Europe has been divided into a small number of ecosystem regions based on *geofactors*, such as climate, soil and topography. On the second level (strata) the division has been guided by *biofactors* (e.g. potential forest species) to identify the various forest ecosystems. When required the strata (115 in total) were aggregated or divided into several strata depending on the expert knowledge of the interpreter.

In Veldkamp et al. (1998) the role of ancillary data for improving classification results at a European scale was investigated. The main conclusion was that the use of ancillary data for entire Europe is severely hampered due to large variation in thematic and spatial accuracy of the data sets.

3 METHODOLOGY

The classification methodology within the PELCOM project is based on a regional and stratified approach. Figure 1 shows a flowchart demonstrating the general classification methodology. The FIRS strata were used for the stratification with a 10-pixel boundary buffer. The classifications were implemented per stratum with thematic maps for urban areas, water bodies and forest as optional for masking.

Concerning forests, classification of NDVI maximum value composites often fails to correctly discriminate forests from grasslands (Champeaux *et al*. 1998a). In a supervised classification forests can be identified on individual multi-spectral scenes using AVHRR channels 1, 2 and 3 (ESA 1992, Mücher *et al*. 1994). However, such an approach is often hampered for pan-Europe due to frequent occurrence of clouds. Therefore Champeaux *et al*. (1998 a, b) proposed another method to improve forest mapping. Advantage was taken of the photosynthetic efficiency of the forest canopy that results in low visible reflectances. This is particularly true in summertime for northern Europe and in spring and autumn for southern Europe. The identification of forests was therefore implemented on basis of thresholding the synthesis of visible reflectance of AVHRR channel one. Because north-eastern Europe was not covered by the threshold forest map, the map was merged with the ESA forest map (ESA 1992) for this part.

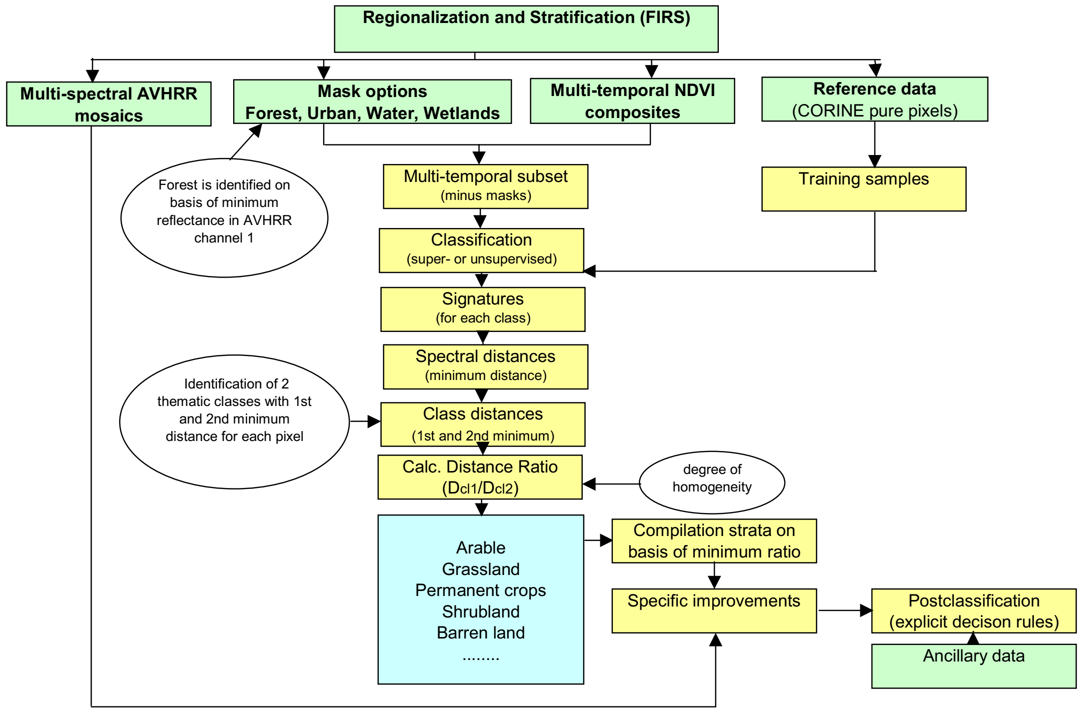


Figure 1 General framework of the PELCOM classification methodology

To classify the satellite data around 10 clusters/classes were defined for each stratum. Assessment of the adequate number of clusters/classes is based on analysis of AVHRR scenes, CORINE, and/or statistical data. Depending on the complexity of the region, a supervised classification or clustering is performed on the NDVI composites, not including the masked areas. For each pixel the spectral distance to each class/cluster is determined and it gets assigned the two classes with the smallest distance. In a third layer the ratio between the two distances is given. This ratio can be used as a measure for the homogeneity of the single pixel and is valuable information to the end-user. So, with each classification three layers are produced: a first-best classification (highest ‘ probability’), a second-best classification (second-highest ‘ probability’) and a distance ratio. These components offer also the possibility to calibrate the classification results with statistical information. If for example, according to the statistics insufficient grassland is found for a certain administrative region, the second-best classification can help to identify more grassland areas.

For each stratum the results from the classification of the NDVI composites was compared interactively with the visual information in the multi-spectral AVHRR scenes. The information in either the multi-temporal or multi-spectral AVHRR data will be strongly influenced by the quality of the concerned data and the specific land cover features present in the specific stratum. If specific features, e.g. linear features, and specific land cover classes were only visible in the multi-spectral AVHRR scenes they were derived from these scenes by a supervised classification.

4 RESULTS

For the compilation of the final database only the first-best classification results have been used due to time constarints. Because for each stratum a 10-pixel boundary zone was used, there was an overlap between all classifications. On basis on visual interpretations it was decided which regional classification should overrule which adjacent classification. Irregularities in terms of unclassified pixels were removed by a majority filter. Eastern-Europe (east of Poland) was not considered within the defined regional classification experiments. This problem was solved with another project, called INDAVOR - a Dutch collaboration of Resource Analysis, TNO-MEP and SC-DLO (Boer et al., 2000). Within this project there was also need for a 1km pan-European land cover database for application within a tropospheric ozone model (LOTOS). It was decided to follow the classification methodology of the PELCOM project, using the same data. The only difference now was that the classification was not performed per stratum but per region – largely reducing the number of classifications. The 1-km pan-European land cover database established within the INDAVOR project has been called PELINDA database (referring to PELCOM and INDAVOR). In the PELCOM project the part that has not been covered within the regional classification experiments has been retrieved from the PELINDA database. The final result is demonstrated in Figure 2.

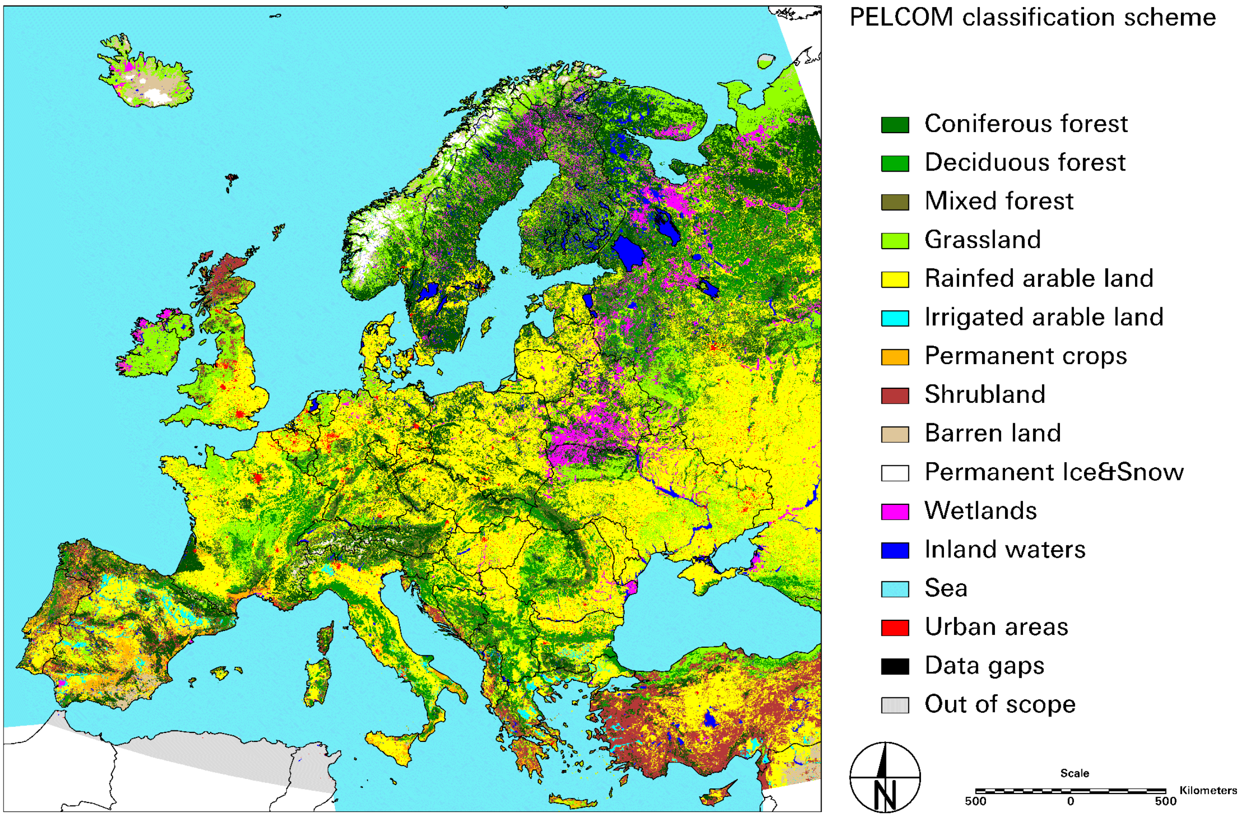


Figure 2 The PELCOM database: a 1km pan-European land cover database

The PELCOM database contains 16 classes with a total surface of 17,603,669 km^2 (Table 2). Note that the land cover classes urban areas, wetlands and water bodies have been derived from ancillary data sources.

Table 2 Land cover statistics of the 1km PELCOM land cover database

Nr	Class name	Area (ha)	Nr	Class name	Area (ha)
1	Coniferous forest	148545000	9	Barren land	16259300
2	Deciduous forest	82033900	10	Ice&Snow	8783270
3	Mixed forest	51345900	11	Wetlands	34204800
4	Grassland	127978000	12	Inland waters	23010700
5	Rainfed arable land	309130000	13	Sea	7081737
6	Irrigated arable land	7067000	14	Urban areas	104335
7	Permanent crops	11695500	15	Data gaps	2550
8	Shrubland	48608100	16	Out of scope	180852

5 VALIDATION

For validating the PELCOM land cover database high-resolution satellite images have been used. For validation ideas have been followed from the IGBP-DIS global land cover database validation (IGBP-DIS, 1996). It was decided to limit the validation to the confidence site mapping due to the amount of work. IGBP-DIS provided 30 Landsat-TM images for Europe. In addition, 10 high resolution satellite images were provided by PELCOM partners. This resulted in 40 high-resolution satellite images distributed over pan-Europe that had to be interpreted. Visual interpretations of the high-resolution satellite images were done independently of the PELCOM land cover database, but use of ancillary data was allowed (topographic maps, national land cover databases). In the ideal situation one identifies on each TM image polygons with a homogenous land cover for each PELCOM class. The polygons should be large enough meaning polygons of at least 4 km^2. Each polygon got a label (code) according to the PELCOM classification scheme. Before using the confidence sites for validating the PELCOM database, the coverages had to be converted from polygon to raster format. Next, there was an aggregation process (from the 25 meter spatial resolution of the Landsat-TM data to a spatial resolution of 1100 meters of the PELCOM land cover database). For this purpose a majority rule was used with a 80% threshold. So, only those 1100 meter pixels (PELCOM pixels) were accepted for validation that were covered for more than 80% with interpreted 25 meter pixels (Landsat-TM) of one specific class. This resulted in a total area of (interpreted) confidence sites of 7700 km^2. As can be seen in Figure 3 not all PELCOM classes are equally represented within the confidence sites. Moreover, for permanent crops no confidence sites were identified within the available images. Land cover classes such as arable land and coniferous forest are more common in Europe, less fragmented and easier to interpret than classes such as permanent crops and wetlands and this is reflected in the interpreted confidence sites. Of course, this distribution affects the validation results. From the contingency table Figure 4 has been derived showing the accuracies and reliabilities of the various land cover classes.

The total average accuracy was 69.2%, which can be considered as a good result considering the mixed pixel and geo-referencing problems of AVHRR data. The mixed pixel problem is especially present in heterogeneous areas such as the Mediterranean region. Accurate geo-referencing is a general problem of AVHRR data that reduces the apparent spatial resolution when multi-temporal series of AVHRR data are being used. Major classes such as arable land and coniferous forest still have a high reliability and accuracy of around 80%, while small and fragmented classes such as shrubland and wetlands have a very low accuracy and reliability. Exceptions are barren land and permanent ice and snow. These validation results show in general the same trends as in those cases that the PELCOM database was validated using CORINE land cover database and statistical data.

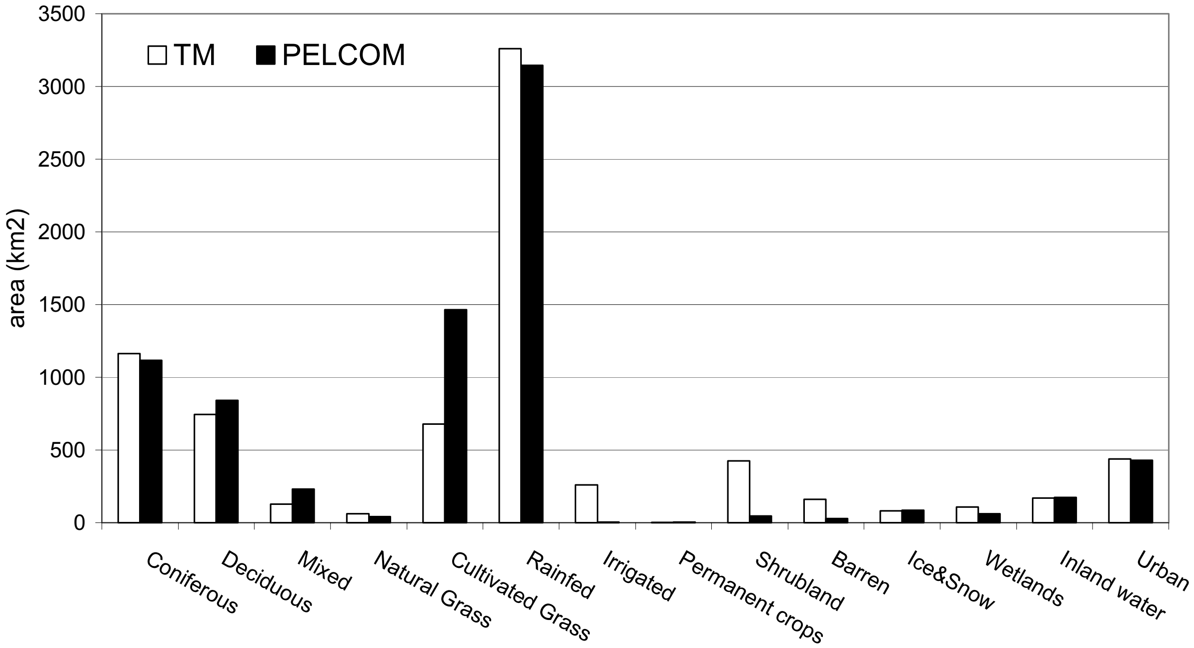
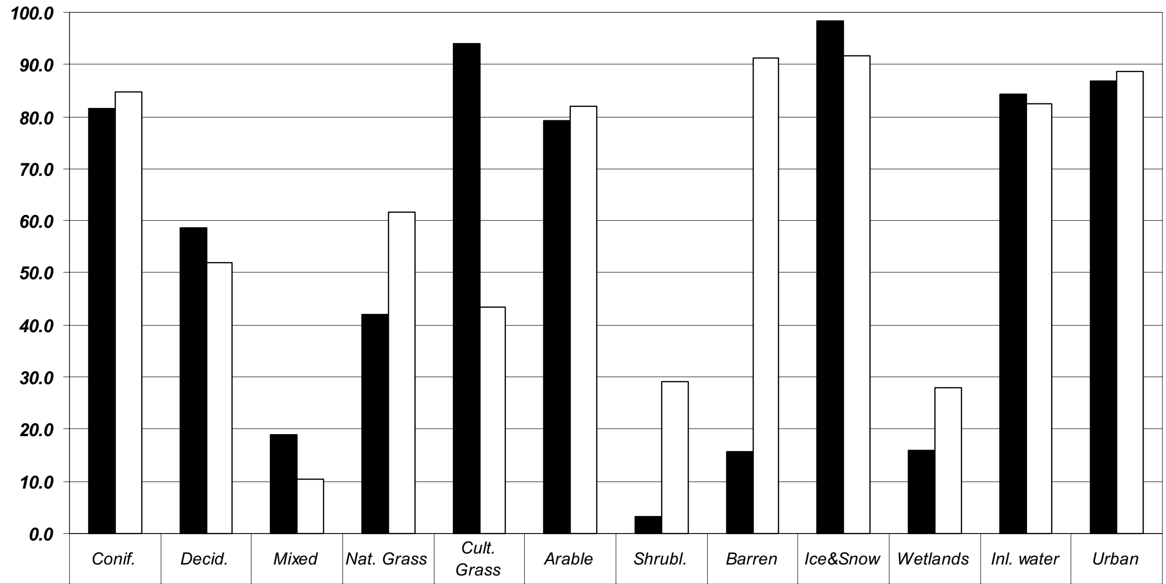


Figure 3 Overall statistics (acreages) of the confidence sites and validated PELCOM areas



	Conif.	Decid.	Mixed	Nat. Grass	Cult. Grass	Arable	Shrubl.	Barren	Ice&Snow	Wetlands	Inl. water	Urban
■ Accuracy	81.5	58.7	19.0	41.9	94.0	79.2	3.2	15.8	98.5	16.0	84.3	86.8
□ Reliability	84.8	51.8	10.4	61.6	43.5	82.1	29.0	91.1	91.8	28.1	82.5	88.7

Figure 4 Validation results for the 1km PELCOM land cover database using the confidence sites of 40 high resolution satellite images distributed over Europe.

6 CONCLUSIONS

This paper presented an improved stratified and integrated classification methodology to map major land cover types of pan-Europe using NOAA-AVHRR satellite and additional geographic data in a consistent manner. Both multi-temporal NDVI profiles and multi-spectral AVHRR scenes were used as input in the classification procedure exploiting the advantages of both data types. Due to the limited accuracy of identifying forests on basis of unsupervised clustering of NDVI timeseries, the identification of forests was implemented on basis of thresholding the synthesis of the visible reflectance of AVHRR channel one. This resulted in a pan-European forest map. Additionally, urban areas and inland water, which were derived from ancillary data, were masked optionally.

The validation results showed reasonable results for major classes. In general land cover classes that were very fragmented and had complex spectral signatures were more difficult to identify using AVHRR satellite data. But an overall classification accuracy 69.2% could be seen as a good result considering the mixed pixel and geo-referencing problems of AVHRR satellite data. The spatial resolution limits also the possibilities to identify land cover changes in the European landscape. Comparison of two AVHRR derived land cover maps will result in identifying many spurious changes.

Medium resolution satellite sensors with a spatial resolution between 150-250 meters (eg. MODIS and MERIS) will play a major role in the near future to identify land cover changes on a European scale. Nevertheless, the 1-km PELCOM land cover database can be considered as a unique product, next to the CORINE land cover database, that fills a gap in environmental and climate studies for pan-Europe.

REFERENCES

De Boer M.E., De Vente J., Mücher C.A., Nijenhuis W.A.S., Thunnissen H.A.M, 2000. INDAVOR, Instrument for detecting land cover change for Europe. NPRS-2 project no. 4.2/DE-03, 98pp. BCRS report, Delft, the Netherlands. In prep.

Brown, J.F., T.R. Loveland, J.W. Merchant, B.C. Reed and D.O. Ohlen (1993). Using multisource data in global land cover characterization: requirements and methods. Photogrammetric Engineering and Remote Sensing, 59(6): 977-987.

CEC, 1993. CORINE Land Cover technical guide. European Union. Directorate-Generale Environment, Nuclear Safety and Civil Protection. Luxembourg.

EC, 1995, Regionalization and Startification of European Forest Ecosystems, Internal Special Publication of the Joint Research Centre of the European Commission. S.I.P.95.44, European Commission Joint Research Centre, Institute for Remote Sensing Apllications, Environmental Mapping and Modelling Unit, Italy, p.69.

Champeaux J.L, Arcos, D., Bazile, E., Giard, D. and Noilhan, J., 1998 a, Vegetation parameters maps over Europe using NOAA/AVHRR in meteorological models. IEEE International Geoscience and Remote Sensing Symposium held in Seattle, USA, on 6-10 July 1998, Vol III, pp 1726-1728.

Champeaux, J.L., Arcos, D., Bazile, E., Giard, D., Goutorbe, J.P., Habets, F., Noilhan, J. and Roujean, J.L., 1998 b, AVHRR-derived vegetation mapping over western Europe for use in numerical weather prediction models. International Journal of Remote Sensing, special issue on Global and Regional Land Cover Characterization from Remotely sensed Data. Submitted

ESA, 1992, Remote Sensing Forest Map of Europe. ESA/ESTEC, ISY Office, Noordwijk, the Netherlands, map and report, 18 pp.

IGBP-DIS, 1996. The IGBP-DIS global 1km land cover data set, DISCOVER, Proposal and implementation plans. Report of the Land Cover Working Group IGBP-DIS. Working paper 13, Belward, A.S. (Ed.), 61 pp.

Kennedy, P., Roy, D.P., Folving, S. and Mégier, J., 1996, The application of the FIRS Project' s Foundation Action 1, the regionalization and stratification of European forest ecosystems, for providing a European NOAA-AVHRR-based forest map. Proceedings of the 15th EARSeL Symposium on ' Progress in Environmental Research and Applications' held in Basel, Switzerland, on 4-6 September 1995. Published by A. Balkema, Rotterdam, 1996 (editor E. Parlow), pp 43-53.

Loveland and Belward 1997, The IGBP-DIS global 1 km land cover data set, DISCover: first results. International Journal of Remote Sensing, 18, 3289-3295.

Mücher, C.A., van de Velde, R.J., and Nieuwenhuis, G.J.A., 1994, Mapping Land Cover for Environmental Monitoring on a European Scale. Pilot project for the applicability of NOAA/AVHRR HRPT data. Report 93, SC-DLO, Wageningen, the Netherlands. RIVM report no. 402001004.

Palacio-Prieto, J.L. and L. Luna-González (1996). Improving spectral results in a GIS context. International journal of Remote Sensing, vloume 17 (no. 11): 2201-2209.

van de Velde R.J., W. Faber, V. van Katwijk, H.J. ScholtenT. Thewesen, M. Verspuy, M. Zevenbergen (1994). The preparation of a European land use database. Rep. 71240100, RIVM, Bilthoven.

Veldkamp, J.G., W.S. Faber, V.F. Van Katwijk, R.J. Van de Velde (1995). Enhancements on the European land use database. RIVM report no. 724001001, Bilthoven (the Netherlands).

Veldkamp,J.C., C.Heunks, V.F.Katwijk, M.Barkhof,C.A.Mücher,B.J.E.ten Brink, 1998. Towards an improved methodology for pan_European land cover classification. Applicability of NOAA-AVHRR and ancillary data for land cover based environmental monitoring on a European scale. Final report part2. BCRS. NRSP-2 98-06.

WRR, 1992. Ground for choices: four perspectives for the rural areas in the European Comminity. The Hague: SDU, 1992 – 144 p. ISBN 90-399-0367-0.