GeoInformation in Europe, M.A. Gomarsca (ed.)

High resolution pan-European forest/non-forest map based on Landsat data and CORINE Land Cover 2000

A. Pekkarinen, L. Reithmaier & P. Strobl
European Commission – DG Joint Research Centre, Institute for Environment and Sustainability, I – 21020 Ispra (VA), Italy

Keywords: large area mapping, pan-European, forestry, Landsat, CORINE, INSPIRE

ABSTRACT: The paper presents a harmonised pan-European forest/non-forest map based on Landsat imagery. This high resolution forest map is aimed to be consistent over Europe, independent of national borders, and it can be used for European level forest resource and spatial pattern analysis. The applied fully automatic forest mapping methodology employs publicly available Landsat satellite images and Corine Land Cover 2000 data. The classification process was carried out on a scene-by-scene basis for the target year 2000. The image analysis consisted of several steps such as image segmentation, cloud masking, clustering, adaptive spectral representative analysis, and supervised classification.

The methodology proved to be applicable for different bio-geographical regions and ecological conditions. Preliminary validation results state that the classification accuracy is at the level of 90% in central European conditions.

1 INTRODUCTION

Forests cover approximately 30% of Europe's land area (EC-EUROSTAT 2007), and therefore play an important role in general welfare, preservation of natural biodiversity and in climate change mitigation. In order to plan and analyse the use of forests efficiently, information on their spatial distribution is needed. For many applications, e.g. forest protection, conservation planning, and forest resource analysis, relatively coarse spatial information about forest distribution may be sufficient but for other purposes, e.g. analysis of the spatial pattern of forests (i.e. information on the size, shape of forests in a landscape), detailed information about their spatial distribution in the form of maps, is required.

To date there have been several efforts of mapping European forests at different scales ranging from regional to continental. While the regional products may be accurate and of high resolution, the remaining problem is that they vary in level of detail, use diverse sources of information, and are based on different forest definitions. Therefore, their use for international comparison for various scientific, policy and reporting purposes is complex.

Earth-observation data are currently the most obvious solution for deriving spatially detailed forest maps at continental scales due to their consistency over large areas and their relative low processing costs (Franklin & Wulder 2002, Wulder *et al.* 2003). For Europe, forest mapping has been tested with the Advanced Very High Resolution

Radiometer (AVHRR) data (Häme *et al.* 2001) and the Wide Field Sensor (WiFS) data (GAF 2001). However, in order to fulfil the requirements for studies where the level of spatial detail is important, imagery with a spatial resolution of better than 50 m needs to be used.

In this paper, we present a harmonised pan-European forest/non-forest map based on Landsat ETM+ imagery and representing year 2000 forested area conditions.

2 STUDY MATERIALS

2.1 *Study area*

The area of interest covers the countries of the European Union and the following neighbouring countries: Albania, Bosnia-Herzegovina, Croatia, Macedonia, Serbia and Switzerland. The main reason for this coverage is due to the availability of a harmonised land cover information source, i.e. the CORINE Land Cover 2000 (CLC2000). For Serbia and Switzerland the CLC2000 information of the adjacent countries was employed.

2.2 *Satellite imagery*

For the pan-European forest monitoring project, freely available Landsat data from two different data sets were used, one of which is part of the NASA's Global Orthorectified Landsat Data Set (Tucker & Grant *et al.* 2004) freely available at the Global Land Cover Facility (GLCF) http://www.landcover.org, while the other data set was the IMAGE2000 (European Commission–DG Joint Research Centre 2005) available at http://image2000.jrc.it/. Both data sets comprise Landsat 7 ETM+ scenes for the target year 2000 and provide pan-European coverage.

The final input image data set consists of 415 Landsat ETM+ scenes, 285 of which were downloaded from the GLCF and 130 of which were found in the IMAGE2000 set. The scenes employed were chosen on the basis of consistency in positional accuracy and suitable acquisition date for vegetation classification. All the scenes were re-projected to ETRS-LAEA projection and resampled to 25 metre pixel size in order to be compliant to INSPIRE standards (Strobl *et al.* 2007).

2.3 *Land cover information*

The land cover information for the supervised classifier was extracted from the CORINE Land Cover 2000 (EC–DG JRC 2005) data as it provides consistent and comparable land cover information for our area of interest. The CLC2000 nomenclature includes 44 land cover classes covering the agricultural as well as the urban and natural sector. The forest in the nomenclature of CLC2000 is defined as "areas occupied by forest and woodlands with a vegetation pattern composed of native or exotic coniferous and/or deciduous trees and which can be used for the production of timber or other forest products. The forest trees are under normal climatic conditions higher than 5 m with a canopy closure of 30%

at least" (Bossard *et al*. 2000). The smallest unit identified in CLC2000 is 25 ha and the minimum width of linear feature is 100 m. Therefore, all land cover patches smaller than 25 ha are merged with the dominant land cover patches during the image interpretation.

2.4 *Reference data set*

The forest/non-forest mosaic of classified images was validated using an independent set of sample points. For that purpose a grid covering the whole Europe with a cell size of 10×10 km was established and a random evaluation point was generated for each of the cells. Those points that fell on an area covered by very high resolution scenes were visually interpreted using GoogleEarth® software (http://earth.google.com) and assigned forest or non-forest label.

Additionally, the results were evaluated using the Land Use/Cover Area frame statistical Survey (LUCAS) data set (EC–DG JRC 2002) that was provided by Eurostat for 15 European Union countries for the year 2001. LUCAS is a multi-purpose land cover/use survey that has been designed for supporting policy making by providing harmonised information on agri-environment for Europe. The definition for forest is given as "areas of >0.5 ha covered by tree-crown area density of more than 10% capable of achieving >5 m in height (accounting for the trees' genetic characteristics and the local agro-meteorological conditions)" (EC–EUROSTAT 2005). Further on, only the land cover information of the points was taken into account for the validation.

3 METHODOLOGY

For producing the forest/non-forest map we developed and applied a fully automated image processing methodology. In order to avoid problems linked to phenological differences between image acquisition dates and related problems such as need for radiometric calibration, the processing was based on a scene by scene approach.

The processing consisted of preparatory steps and the actual mapping. Within the preparation step the scenes were segmented applying an in-house developed edge-preserving segmentation algorithm, and a cloud mask was computed for each input scene using the Automatic Cloud Cover Assessment (ACCA) – algorithm (Irish & Barker 2006).

The core of the mapping process was the training selection and extraction process, in which for each scene the segmentation output, *k-means* clustering, and the original imagery were used in combination with the CLC2000 data.

The main challenges within this study were related to the characteristics of CLC2000; namely, the spectral heterogeneity of some of some its land cover classes and its minimum mapping unit (MMU) of 25 ha. The complex classes typically include trees or clumps of trees, hence also forest in addition to other land cover elements (e.g. agricultural fields, pastures, etc.). These classes were excluded from the training extraction since their use may result in ambiguous training observations carrying a spectral signature of forest while labelled as non-forest. The MMU of 25 hectares applied in the production of CLC2000 may result in some cases in merging of two or more land cover polygons that

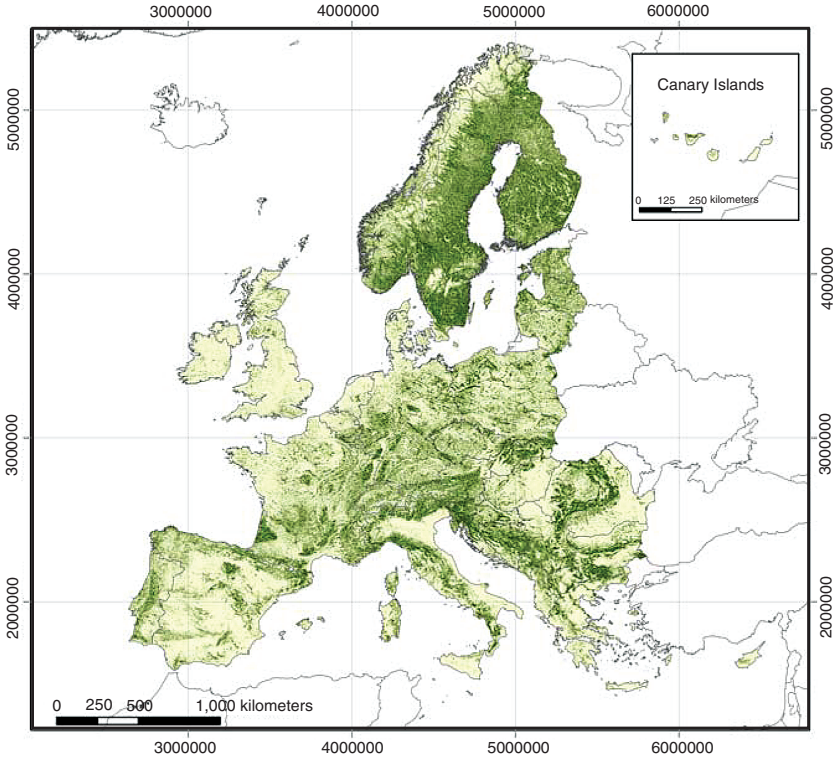

Figure 1. Pan-European forest/non-forest map based on Landsat data representing year 2000 forested area conditions, coordinate reference system applied is ETRS-LAEA, with a spatial resolution of 25 m.

have a very different spectral response. Because of this, the CLC2000 patches cannot be considered as being spectrally 'pure' and therefore, they cannot be used straightforwardly as training areas; their spectral characteristics need to be analysed in order to be able to select the relevant areas for the training. In addition, co-registration errors between the CLC layer and the imagery need to be taken into account.

In order to address these challenges, the extraction of the training areas for each CLC2000 class was done within an adaptive representativity analysis (ARA). During the ARA the representative combination of a cluster labels and CLC2000 classes was determined with the help of their contingency table. After that, the actual training set was extracted from an equidistant grid (250 m) of sample plots and then further sampled according to the results of the spectral representativity analysis. The actual classification of the input imagery was conducted at the segment level using the extracted training data and a nearest neighbour classifier. Finally, the classified images were merged into a pan-European mosaic.

Accuracy assessment was carried out using the independent reference data set based on GoogleEarth®. For the comparison with the LUCAS 2001 data, two different

methods were applied in order to analyse the effect of locational errors. First, all survey points were considered in the evaluation and second, only those points falling into a 3×3 pixel block consisting of either only forest or non-forest pixels (later referred to as homogeneous points), were considered. This was done in order to account for errors which occur at forest/non-forest boundaries due to misregistration of the data sets and mixed pixels.

4 RESULTS AND DISCUSSION

The final pan-European forest/non-forest map is illustrated in Figure 1. The methodology applied proved able to run in a fully automatic mode and to be applicable to European biogeographical regions. Hence, no further adjustments had to be implemented for different ecological conditions or forest formations. However, some CLC2000 classes, i.e. wooded parks or fruit tree plantations, cannot be distinguished from forest due to the similarity of spectral signatures and may in certain cases be classified as forest dependent on the tree density and background reflectance since no additional auxiliary data, e.g. cadastral data, could be used for all countries. The forest/non-forest map can therefore be regarded as a forest cover map rather than a forest land use map.

Concerning validation of the map, preliminary results based on the independent reference data set show that the overall accuracy (OA) is at the level of 88% in parts of central and southern Europe with a 95%–confidence interval (δ) of 0.25% (Table 1).

The comparison of the LUCAS 2001 inventory data with the map gives insight into the performance of the classification in the EU15 countries. For all survey points during the evaluation, the averaged overall agreement (OA) between the two data sets is 85% considering all EU15 countries, while for the homogeneous points the agreement is enhanced to 91%. Between countries the agreement results vary from 73.5% in Portugal to 91.4% in United Kingdom. Countries with a lower agreement tend to be either in regions of open forest areas as in the Mediterranean or in areas with high percentage of peat bogs which could contribute to a more fragmented forest/non-forest map (reflected in a low percentage of used homogeneous points) where classification mismatch due to misregistration is more likely to occur.

Table 1. Classification accuracy statistics showing overall accuracy (OA), producer's accuracy (PA) and user's accuracy (UA) with their relative standard error (SE) and 95%-confidence limits (δ) for central and southern Europe based on the independent systematic reference data set.

	Fmap	forest	nforest	total	PA %	SE	δ (%)
googleE	forest	1321	224	1545	85.50	0.90	1.80
	nforest	380	3268	3648	89.58	0.50	1.00
	total	1701	3492	5193			
	UA %	77.66	93.59	OA	88.37%		
	SE	1.01	0.41	se	0.13		
	δ (%)	1.98	0.81	δ (%)	0.25		

Further work is geared towards ongoing accomplishment and improvement of the accuracy assessment procedure of the current map. Additionally, the methodology will be used for the production of a pan-European forest/non-forest map for the 1990s, again based on GLCF Landsat data.

REFERENCES

Bossard, M., Feranec, J. & Otahel, J. 2000. CORINE land cover technical guide – Addendum 2000. *Technical Report No. 40, EEA*.

Congalton, R.G., Green, K., 1999. Assessing the accuracy of remotely sensed data: principles and practices. Lewis Publishers, Boca Raton: pp 131.

EC EUROSTAT 2005. LUCAS 2005 (Land use/Cover Area Frame Survey)–Technical reference document C-3: Nomenclature. *D. AGRI*: pp. 82, 2005.

EC EUROSTAT 2007. Forestry statistics. Luxembourg: *Office for Official Publications of the European Communities*. ISBN 92-79-025956-8: 97 pp.

EC– DG JRC 2005. IMAGE2000 and CLC2000–Products and Methods. M. V. Nunes de Lima. Ispra, *European Commission, DG JRC, IES, LMU, EUR 21757 EN*: pp. 150.

EC–DG JRC 2002. Building Agro Environmental Indicators–Focussing on the European area frame survey LUCAS. *EUR Report 20521 EN*. J. Gallego. Ispra: pp. 176.

Franklin, S.E., Wulder, M.A. 2002. Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. *Progress in Physical Geography* 26(2): 173–205.

GAF 2001. IRS WIFS image mosaic classification of forest classes for the European Union. Final report. Contract Nr. 17240-2000-12 F1ED ISP DE. Company for Applied Remote Sensing: Munich, pp. 26.

Häme, T., Stenberg, P., Anderson, K., Rauste, Y., Kennedy, P., Folving, S. & Sarkeala, J. 2001. AVHRR-based forest proportion map of the Pan-European area. *Remote Sensing of Environment* 77: 76–91.

Irish, R.R., Barker, J.L., Goward, S.N. & Arvidson, T.J. 2006. Characterization of the Landsat-7 ETM Automated Cloud-Cover Assessment (ACCA) Algorithm. *Photogrammetric Engineering & Remote Sensing* 72(10): 1179–1188.

Stehman, S.V. 2000. Practical implication of design-based sampling inference for thematic map accuracy assessment. *Remote Sensing of Environment* 72: 35–45.

Strobl, P., Reithmaier, L., Mehl, W., Soille, P. & Bielski, C. 2007. Assembly of a remotely sensed reference image of Europe at 25m resolution. *Proc. 27th EARSeL*, 4–7 June 2007, Bozen/Bolzano, Italy, this issue.

Wulder, M.A., Dechka, J.A., Gillis, M.A., Luther, J.E., Hall, R.J., Beaudoin, A. & Franklin, S.E. 2003. Operational mapping of the land cover of the forested area of Canada with Landsat data: EOSD land cover program. *The forestry chronicle* 79(6): 1075–1083.