Pan-European Forest/Non-Forest Mapping based on Landsat data

Anssi Pekkarinen, Lucia Reithmaier, Peter Strobl
European Commission – DG Joint Research Centre, Institute for Environment and Sustainability,
I–21020, Ispra (VA), Italy

INTRODUCTION

Information of the spatial distribution of the entire European forest is needed for several forest–related applications, e.g. forest protection, conservation planning, and forest resource analysis. Up to date there are several efforts of mapping European forests at different, from regional to continental scales; 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. In order to derive spatially detailed forest maps at continental scales, earth-observation data are currently the most obvious solution due to their consistency over large areas, growing computing power and relative low processing costs (Franklin and Wulder 2002; Wulder et al. 2003). For Europe, forest mapping has been tested with the low resolution (1 km) Advanced Very High Resolution Radiometer (AVHRR) data (Häme et al. 2001) and the 180 m Wide Field Sensor (WiFS) data (GAF 2001). However, in order to fulfil the requirements for studies where the level of spatial detail is important, such as forest spatial patterns, imagery with a spatial resolution of better than 50 m needs to be used. One of the most common sources for this kind of data is the Landsat Thematic Mapper series of instruments.

In this paper, we represent a harmonized pan-European forest/non-forest map based on Landsat ETM+ imagery and representing year 2000 forested area extent. The aim of the employed methodology is to guarantee the consistency of the product, its independency of national borders and to base it on one forest definition[1], while run in a fully operational mode.

MATERIAL AND METHODS

Area covered

The area of interest is currently covering the countries of the European Union and the following neighbouring countries: Albania, Bosnia-Herzegovina, Croatia and Macedonia. The main reason for this coverage is due to the availability of a harmonized land cover information source, i.e. the CORINE Land Cover 2000 (CLC2000). For processing Serbia and Switzerland the CLC2000 information of the adjacent countries was sufficient for training extraction since the satellite scenes were covering a substantial part of their area.

Satellite Imagery

For the pan-European forest monitoring project freely available Landsat data from the Global Land Cover Facility (GLCF) (http://glcf.umiacs.umd.edu/data/landsat/) was used for the target year 2000 (with a deviation of one to two years). The data was re-sampled (cubic convolution) from 28.5 m to 25 m and re-projected to the European ETRS89 Lambert Azimuthal Equal Area Coordinate Reference System (ETRS-LAEA). For complete coverage of the area of interest, 422 scenes were required.

Land cover Information

The Land cover information needed for the forest mapping is required to have consistent and comparable pan-European coverage with a harmonized nomenclature. Currently, only the CORINE Land Cover 2000 (CLC2000) (EC - DG JRC 2005) is available as land information source covering the countries of interest. The CLC2000 nomenclature includes 44 land cover classes covering the agricultural as well as the urban and natural sector. The smallest unit identified in CLC2000 is 25 hectares and the minimum width of linear feature is 100 m.

[1] Applied Forest Definition based on CORINE Land Cover (Bossard et al. 2000): "Areas occupied by forest and woodlands with a vegetation pattern compose 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"
.

Data Analysis

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 images and related problems such as equalising the radiometric content of all images, the processing is based on a scene by scene approach.

The processing consists of a preparation step and the actual supervised classification procedure. Within the preparation step the scenes were segmented applying an in-house developed edge-preserving segmentation algorithm and preliminary mapped into spectral categories related to main land cover classes (vegetation, non-vegetation, water, clouds, snow) using a spectral rule-based approach developed by Baraldi et al. (2006). A selection of these segments was then used as training set for the actual segment-level classification of the whole image.

The core of the mapping process is the training selection and extraction process, in which for each scene the spectral category map, segmentation output, k-means clustering, and the original imagery were used, combined with CLC2000 data. The main challenge within this study was related to the characteristics of CLC2000; namely, its minimum mapping unit of 25 hectares and the spectral heterogeneity of some of its land cover classes. In order to address these challenges, the extraction of the 'pure' training areas of each CLC class taking into account its spectral variability was done within an adaptive spectral representativity analysis. During the analysis the representative combination of a cluster and a CLC class is determined with help of their contingency table. Those combinations are then used for training extraction. The actual training set is first extracted from a dense equidistant grid (250 m) of sample plots, in order to assure a spatial uniform distribution and is then further sampled according to the results of the spectral representativity analysis. The segment average images are classified using the final training set applying a nearest neighbour classifier. The classified images were then mosaicked after an overlapping analysis was performed.

The resulting forest/non-forest map was validated using an independent systematic reference data set of a regular point grid of 10 by 10 kilometres. The grid points were visually interpreted where very high resolution imagery (e.g. Google™Earth, freely available orthophotos) was available. Accuracy assessment measures were calculated according to Stehman (2000) and Congalton and Green (1999). Additionally, comparison studies were carried out using available European land cover data sets (e.g. LUCAS 2001 inventory data (EC - DG JRC 2002)). In case of LUCAS 2001 data, the comparison was done in two different ways in order to analyse the effect of locational errors. First, all survey points were considered in the evaluation and second, only those points, later referred to as homogeneous points, were considered, which fell into a 3x3 pixel block consisting of either only forest or non-forest pixels, respectively.

RESULTS AND DISCUSSION

The final pan-European forest/non-forest map is illustrated in Figure 1. The methodology applied proved to run in a fully automatic mode and to be applicable to all bio-geographical regions. Hence, no further adjustments had to be implemented for different ecological conditions or forest formations. However, some CLC classes, i.e. wooded parks or fruit tree plantations, were still difficult to separate due to the similarity of spectral signatures and may in certain cases also 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.

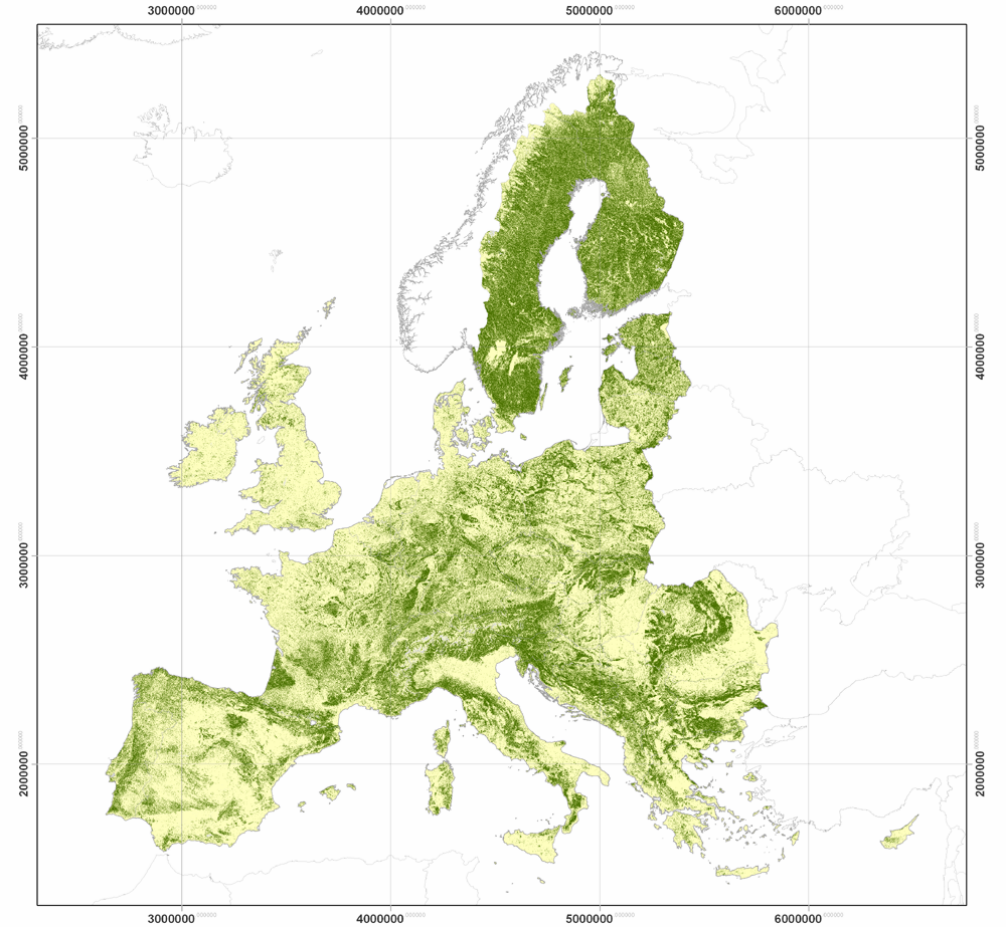

Figure 1: Pan-European forest/non-forest map based on Landsat data representing year 2000 forested area conditions, coordinate reference system applied is ETRS-LAEA, with a spatial resolution of 25 m.

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 90% in parts of central and southern Europe with a 95% - confidence interval (δ) of 0.3% (Table 1, Figure 2).

Table 1: Classification accuracy statistics showing overall accuracy (OA), producer's accuracy (PA) and user's accuracy (UA) with their relative 95% - confidence limits (δ) for selected areas are shown based on independent systematic reference data set.

		Forest Map				
		forest	non-forest	total	PA %	δ (%)
reference points	forest	1278	106	1384	**92.3**	**1.4%**
	non-forest	360	3264	3624	**90.1**	**1.0%**
	total	1638	3370	5008		
	UA %	**78.0**	**96.9**		**OA**	**90.7%**
	δ (%)	**2.0%**	**0.6%**		**δ (%)**	**0.3%**

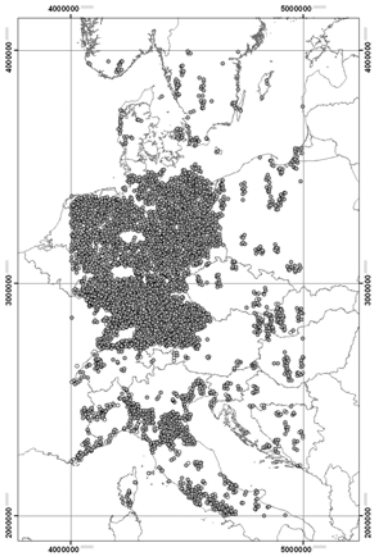

Figure 2: Distribution of available reference points covered by very high resolution imagery using Google™Earth.

The comparison of the LUCAS 2001 inventory data with the map gives insight to the performance of the classification in the EU15 countries (Table 2). For all survey points for the evaluation, the averaged overall agreement (OA) between the two data sets is 83% considering all EU15 countries, while for the homogeneous points the agreement is enhanced to 91%. Between countries the agreement results vary from 66.9% 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 positional errors are more likely to occur.

Table 2: Agreement of the forest/non-forest map with LUCAS 2001 inventory data.

COUNTRY	TOTAL Points	ALL		HOMOGENEOUS		
		OA %	δ %	OA%	δ %	used points %
AT	2528	84.8	0.4	94.2	0.5	69.9
BE	989	86.5	0.6	95.7	0.7	73.7
DE	10981	89.4	0.2	96.5	0.2	77.7
DK	1373	89.2	0.5	95.2	0.6	81.0
ES	12670	80.0	0.2	88.4	0.2	74.0
FI	10410	78.4	0.2	87.3	0.2	65.1
FR	16916	86.2	0.1	94.2	0.2	74.5
GR	4068	82.4	0.3	89.4	0.3	76.7
IE	2163	91.0	0.4	94.1	0.4	91.1
IT	9275	81.1	0.2	88.7	0.2	75.1
NL	1154	87.9	0.5	95.5	0.6	78.4
PT	2731	66.9	0.3	75.2	0.4	59.9
SE	13808	78.7	0.1	88.1	0.2	66.9
UK	7499	91.4	0.2	96.1	0.2	85.5

Note, that the disagreements can be partly explained by the differences in forest definitions: in the LUCAS 2001 inventory the canopy closure limit is set to 10%[2], while in the definition applied in this work

[2] Forest Definition applied in the LUCAS 2001 study (EC - Eurostat 2005):"Areas of >0.5 ha covered by tree-crown area density of more than 10% capable of achieving >5m in height and composed of more than 75% of broadleaved species (accounting for the trees genetic characteristics and the local agro meteorological conditions)"

the forest is defined with a canopy closure of at least 30%. Especially in Mediterranean countries, where open forests with a canopy closure of less than 30% are common, an underestimation of forest by the classification is likely.

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 years 1990s, again based on GLCF Landsat data.

BIBLIOGRAPHY

Baraldi, A., Puzzolo, V. Blonda, P., Buzzone, L., Tarantino, C.: Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images. IEEE transactions on geoscience and remote sensing 44(9): 2563-2586, 2006.

Bossard, M., Feranec, J., Otahel, J.: CORINE land cover technical guide - Addendum 2000. Technical Report No. 40, EEA, 2000.

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

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

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

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

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

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

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

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

Wulder, M. A., Dechka, J. A., Gillis, M. A., Luther, J. E., Hall, R. J., Beaudoin, A., Franklin, S. E.: 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, 2003.