New Strategies for European Remote Sensing, Olujić (ed.)

Procedure for classification of forests for CORINE land cover in Sweden

Olle Hagner, Mats Nilsson, Heather Reese, Mikael Egberth & Håkan Olsson
Remote Sensing Laboratory, Swedish University of Agricultural Sciences, Umeå, Sweden

Keywords: land cover, CORINE, production line, forest classification, Landsat

ABSTRACT: The EU-CORINE Land Cover (CLC 2000) project for Sweden was carried out by the Swedish National Land Survey. The classification of forested land cover types, covering about 65% percent of the land area in Sweden, was contracted out to the Swedish University of Agricultural Sciences (SLU). The method of forest classification used detailed plot information from the National Forest Inventory (NFI), map masks and Landsat data, and was developed to be automated and robust. Within each Landsat scene an average of 2,000 NFI plots were available to use as field data. One use of the NFI plots was to use them as spectral reference targets for correction of within-scene haze differences and correction of slope effects. The forest classification was made with a self-calibrating maximum likelihood algorithm, using the NFI plots as ground truth. In the classification, prior probabilities were iterated until the frequency of each class in the classification corresponded to the frequency of that class according to the field plots. All pre-processing of the image data and the NFI data, in addition to the classification, was done using the automated production line written in-house called MUNIN.
In total three products were produced: the CLC 2000 product with 4 forest classes and 25 ha minimum mapping unit (MMU); and two national products with 7 forest classes and a MMU of 1 ha and 25 m (pixel level), respectively. The seven forest classes in the national product were: clear felled areas; young forests; deciduous forests; mixed forests; coniferous forests 5 to 15 m in height; coniferous forests greater than 15 m in height; and, lichen dominated coniferous forests. Accuracy assessment was carried out using both photo interpretation and cross-validation. The accuracies using cross-validation at the pixel level were on the order of 50-70 %. When aggregated to the four CORINE forest classes, the accuracy increased to about 80 % at the pixel level.

1 INTRODUCTION

The EU-CORINE (Coordination of Information on the Environment) Land Cover project, begun in 1985, had the objective of mapping land cover using satellite imagery and a common classification scheme which would allow comparisons between EU countries. A land cover database for 1990 was created by several, but not all, EU countries with an update planned for 2000.

While Sweden did not participate in the 1990 CORINE Land Cover project, it has taken part for the year 2000. The responsibility for the Swedish CORINE project resided with the Swedish National Land Survey, however the classification of forest was contracted out to the Swedish

University of Agricultural Sciences (SLU). This paper will focus on the methods used for the classification of forest in the CORINE project.

Of importance to the CORINE project was to produce a land cover map using a repeatable and time-efficient method. Approximately 50 Landsat scenes were to be classified in two years' time. Therefore, for the classification of forested land cover types including forested mires and bedrock outcrops, a robust method and a production line were developed at the Swedish University of Agricultural Sciences. The production line was automated and carried out all steps of the method, including retrieval and updating of forest inventory plots, illumination and haze correction of satellite imagery, and classification of the satellite data into forest classes using discriminant analysis with prior probabilities. The production line included tools for metadata management and provided a cross-validation matrix for each classification.

The production line, called MUNIN, processed a Landsat scene from input to output in six hours time. MUNIN's functionality was critical to the smooth flow of work, leading to completion of the forest classification on time and within budget. In addition, access to the forest inventory field plots was a crucial component in the forest classification method. However, copious pre-processing of the inventory data was needed, as was pre-processing of the satellite data. Again, the production line was written to provide the tools necessary to achieve an accurate forest classification in an automated and objective way.

In total three products were produced: the European CLC 2000 product with a 25 ha minimum mapping unit (MMU) and four forest classes, and two national products referred to as Swedish Landcover Data (Svenska Marktäcke Data or SMD) having seven forest classes with 25m (single pixel) resolution and a 1 ha generalized product.

2 METHODS AND MATERIALS

The objective of the forest classification was to derive from satellite data an accurate classification of seven forest classes using a robust and repeatable method that would be automated using an in-house written production line. Input data were National Forest Inventory (NFI) plots, Landsat ETM+ or TM satellite data, and map data. In brief, the method involved pre-processing and updating of the NFI data, processing of map data, illumination and haze correction of the satellite imagery, radiometric modeling, matching of plot and image data, assembly of the training dataset, an iterative maximum likelihood classification of the image data into forest classes, and a cross-validation of the classification results.

The forest classes to be derived were clear felled areas; young forests; deciduous forests; mixed forests; coniferous forests less than 15 m high; coniferous forests higher than 15 m; and, lichen dominated coniferous forests. This was determined by previous pilot studies (Swedish Space Corporation, 1999). These classes were to be aggregated into the four EU-CORINE classes of young forest, deciduous forest, mixed forest, and coniferous forest. The classes are defined by crown closure, tree species proportion, and height. In the forest classification process, sub-classes were also identified. Both the main classes and the subclasses are given in Table 1.

The project area covers the country of Sweden, which is approximately 44 million ha in size (National Atlas of Sweden, 1990). Forested land cover types comprise 65% of the land area, and the mix of deciduous and coniferous species varies widely from the northern to southern reaches of the country. In the northern two-thirds of Sweden, the coniferous species of Norway spruce and Scots pine dominate with birch species also present. In the southern third of Sweden, more deciduous species can be found, such as Oak, Beech and Aspen, along with Norway spruce, Scots pine and birch. The mountain areas along the border with Norway also have forest that consists of Mountain birch. For the Swedish forest classification only forestland as defined by the National Land Survey's 1:100 000 scale map was classified.

Table 1. The seven main classes of the Swedish Land Cover data and their sub-classes used for classification.

Main class	**Definition Height (m)**	**Definition Canopy closure**	**Definition Species proportions**	**Sub classes used in the classification**
Clear-felled	0-2	< 20 %	none	Bedrock New clear-cut Open mire
Young forest	2-5m	> 20%	none	< 70% of each species
Coniferous 5-15m	5-15m	> 20%	>70% Coniferous	Coniferous on mire Coniferous on bedrock >70% Pine >70% Spruce >70% Lodgepole
Deciduous	> 5m	> 20%	>70% Deciduous	Deciduous on mire Deciduous on bedrock >70% Birch >70% Oak >70% Beech
Mixed Forest	> 5m	> 20%	30-70% Deciduous	Mixed on mire Mixed on bedrock > 50% Pine > 50% Spruce > 50% Deciduous
Coniferous > 15m	> 15m	>20%	>70% Coniferous	Coniferous on mire Coniferous on bedrock >70% Pine >70% Spruce >70% Lodgepole
Coniferous, lichen dominated	> 5m	> 20%	Lichen>12% Pine>50% Decid<30%	Volume > median for lichen dominated forest Pine forest on bedrock

While 28 Landsat TM satellite scenes were theoretically required to cover all of the country once, more scenes were needed to fill in cloud-covered areas. In all, 50 Landsat scenes were processed to obtain a cloud-free classification of forest. The Landsat data was obtained from the Image 2000 dataset, which was primarily Landsat 7 ETM+ data from 1999 to 2002. Occasionally Landsat 5 TM data were necessary, and a few scenes from 1994 and 1997 were used. Data were geometrically corrected with an RMSE of ± 1 pixel, with local errors of up to three pixels allowed (Bossard et al., 2000). For the classification, TM bands 3, 4, 5 and 7 were used.

MUNIN, the production line programmed at SLU, was written in Visual Basic, and involved calling various different programs including Microsoft Access for inventory data processing, ERMapper for image processing, and Minitab for statistical calculations. From input to output, the operations required very little manual intervention. Registration and retrieval of all input data occurred in the first 15 minutes using three operations, and the pre-processing and classification was begun with one operation, running until the output was created.

After registering the satellite data into the production line, the preparation of the forest inventory plots was the first task. The NFI annually collects information on 13,000 plots over the entire country. The plot inventory is taken on over 200 variables per plot. Most variables are collected within a 7 or 10 m radius from the plot center and some within up to 20m radius (Anon 2002). For the classification, all NFI plots from a minimum of 5 years to a maximum 10-year time period (depending on the number of plots) were extracted for a scene. This resulted in an average of approximately 2,000 plots per scene, with an absolute minimum of 600 plots up to a maximum of 7,000 were used. There were approximately 84,000 NFI plots used in total for the classifications. Due to use of plots from a time period other than the image date, it was necessary to update forest growth reflecting the state of the forest at the time of image acquisition. Identification and re-labeling of plots that had been clear-felled between date of inventory and image acquisition was also required. Calculation of variables not directly measured in the field, such as crown closure and percent of lichen cover, required calculation from other measured variables.

Land cover categories from available map data were associated with the NFI plots. The map data used were 1:50 000 scale available only for parts of Sweden and 1:100 000 maps available for all of Sweden. The map classes of forest, wetland, bedrock, clear-cut, coniferous forest or deciduous forest, when available, were assigned to each plot as the map land cover class.

Pre-processing of the satellite data included reduction of within scene haze differences and illumination correction. By dividing the NFI plots into the categories of dense forest, water and open land and relating the satellite data's digital numbers (DNs) to these land cover types, a haze index was calculated using regression (Hagner and Olsson, 2004). The index was applied to the satellite imagery to make the haze level within the image constant within the scene, rather than completely removing haze effects. The haze index was then used in the radiometric modeling, which was the next step. A regression model was used which described the spectral signature associated with each plot as a function of the following variables: land use category, timber volume, tree species proportion, age, ground moisture, stand index and vegetation type (herbaceous or lichen). From this, the c-coefficient needed in the C-correction (Teillet et al. 1982) for illumination correction could be derived. The illumination correction reduced shadow effects due to topography and sun angle. A 50 m digital elevation model was available for the correction.

Each plot had geographic coordinate information, which has been collected using differentially corrected GPS since 1996. However, due to local geometric errors in the image data and the lower accuracy of the pre-GPS NFI plot coordinates, registration errors between plot and image data could have meaningful consequences. Therefore, a routine for matching plot and image data was implemented. An estimate of the most likely placement of a plot, given the errors of both the plot and image, was made for a 3x3 pixel area, and the "best match" between the image spectral signature for that class and the estimated locational error determined the placement of the plot.

When all the NFI plots were updated and selected for use, and the satellite data were corrected, the NFI plots to be used for classification were assigned one of the seven SMD forest classes. The assignment was based on tree height, tree species, percent crown closure, and the map's land cover type. In a further step, they were assigned a sub-class as well (see Table 1).

The classification used a maximum likelihood algorithm, in which each pixel within the satellite scene was assigned a sub-class according to its similarity with the training data's spectral signature as well as employing a prior probability of class occurrence within the satellite scene. Without the use of prior probability, there was a risk that dominant classes would be over-represented at the cost of more unusual classes. A value describing spectral similarity typical of a sub-class, referred to as the typicality value (from ERMapper software), was calculated for each pixel of the likelihood of belonging to the sub-class assigned by the maximum likelihood classification. If the typicality value was >30%, the sub-class assignment was accepted. If the typicality value was between 10-30%, the pixel was assigned the majority class from a 3x3 window. If the typicality value was <10%, it was assigned the class given by a minimum-distance classification. After classification into sub-classes, the result was recoded hierarchically into the seven SMD forest classes. A cross validation of the seven forest classes was calculated using the NFI plots, as well as a fuzzy accuracy assessment (Gopal and Woodcock, 1994).

3 RESULTS

For the forest classification, 50 Landsat scenes were processed at a rate of approximately one scene per week or every other week. Without the automated production line, MUNIN, such a production rate would not have been likely. The classification results from the images were of fairly consistent quality, although classifications from the southern part of the country had generally higher accuracy, perhaps because of a denser sampling of NFI plots. Also, the forests of the north are at an average height of around 15m, which is the height definition distinguishing between two of the coniferous forest classes (Hagner 2003).

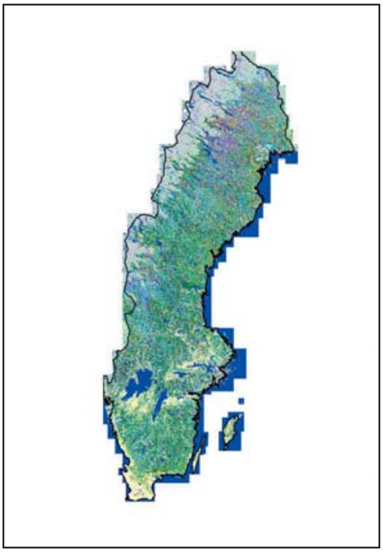

Figure 1. The pixel-wise classification of forest with the seven SMD classes.

The output from MUNIN was a pixel-wise classification of the seven SMD forest classes (Figure 1). This data layer was sent to the National Land Survey who then combined this product with map data indicating whether classes were on wetland or bedrock, producing a total of 13 forest classes, such as coniferous forest on wetland and coniferous forest on bedrock. They combined this with classifications made primarily at the National Land Survey of non-forested wetland, mountain vegetation, urban, water, and other open land, to achieve the final pixel-wise resolution national product. They used a generalization routine (Wester and Olsson, 1997) to derive a 1 to 2 ha MMU generalized national product. A further generalization was used for the EU-CORINE product, with a 25 ha MMU and fewer classes than the national SMD product.

For the forest classification, accuracy assessment occurred at two levels: photo-interpretation and cross-validation. An accuracy assessment using photo-interpretation was limited to three areas in the country because doing so for every scene was too expensive for the project. Aerial photos were randomly chosen and interpreted, and checked against the 1 ha generalization of the national product. This resulted in a traditional accuracy assessment matrix for all land cover types. The result compiled over all three areas showed a 74% area-weighted accuracy for all 58 classes (Röst and Ahlcrona, 2004). A compilation of the accuracy over the three areas for the forest classes alone gives a result of approximately 65% accuracy for forest types.

Because an accuracy assessment from photo-interpretation could not be done for each scene, cross-validation was used for monitoring the quality of the pixel product for each scene. The individual cross-validation results ranged from a minimum of 46% up to 70%. The mean overall accuracy for all scenes was 59%. The class with the highest accuracy tended to be coniferous forest over 15m and clear-cut, perhaps because there classes were some of the most prevalent. There are some explainable and notable differences between the cross-validation and photo-interpretation results. It should be noted that the cross-validation result was in general lower than the result given by photo-interpretation, due to the nature of using all NFI plots to make the cross-validation. In addition, the cross-validation covers the entire country, while the photo-interpretation result is based on the more southern part of the country (with more deciduous forest). Another difference is that the cross-validation was carried out at the pixel-level, while the photo-interpretation was carried out on the 1 ha generalized product. A compilation of the cross-validation results is given in Table 2.

Problems with misclassification were encountered when exceptional forest types were present, such as sparse mature coniferous forest on sandy soils or forests infected with the *Gremeniella Abietina* fungus. A few satellite images that had a small percentage of forest in the scene did not have enough inventory plots for classification, and in most of these cases a minimum distance classification was run.

Table 2. A summary of the cross-validation results at pixel-level for forest classification on all scenes.

				NFI Plot	**Data**				
	Clear-cut	**Young forest**	**Conif 5-15m**	**Mixed Forest**	**Decid-uous**	**Conif >15m**	**Conif on lichen**	**Total**	**% Correct**
Clear-cut	**11021**	1983	704	260	246	121	347	14682	75.06
Young forest	2491	**4148**	858	637	720	195	193	9242	44.88
Conif 5-15m	971	703	**8272**	2104	230	4062	782	17124	48.31
Mixed Forest	436	496	1789	**3854**	633	1027	98	8333	46.25
Deciduous	326	570	225	764	**2408**	36	25	4354	55.31
Conif > 15m	260	162	3911	1102	48	**15354**	862	21699	70.76
Conif on lichen	580	186	708	89	30	661	**1424**	3678	38.72
Total	16085	8248	16467	8810	4315	21456	3731	**79112**	
% Correct	68.52	50.29	50.23	43.75	55.81	71.56	38.17		
Total Correct /Total N =	**46481/**	**79112 =**	**58.75%**						

4 CONCLUSIONS

The classification of forest for Sweden's part of the EU-CORINE Land Cover (CLC2000) project was contracted out to the Swedish University of Agricultural Sciences (SLU). The classification of forest was done in an automated and repeatable way using the in-house written production line called MUNIN. The production line made it possible to classify the 50 Landsat scenes required within the short time frame available and with objective methods. The method of classification, which included use of the National Forest Inventory plots, haze and illumination correction of satellite data, modeling of signatures, and a maximum likelihood classification using prior probabilities, worked well. Cross-validation results at the pixel-level gave an overall result of 59% correct for the seven forest classes alone, while photo-interpretation of three areas showed a 65% correct result for the seven forest classes alone.

The production line MUNIN is currently being used at SLU for other similar projects requiring the use of satellite data and NFI plots. In addition, the method developed for this project will be a basis for future forest classification methods at SLU.

ACKNOWLEDGEMENTS

The Swedish research program Remote Sensing for the Environment (RESE), financed by the Swedish Foundation for Strategic Environmental Research (MISTRA), and the Swedish National Space Board financed the method development work. The National Land Survey (Lantmäteriet) provided the funding for the production work. Financial support for attending the 24th annual EARSeL symposium and workshop was obtained through the Swedish National Space Board.

REFERENCES

Anonymous. 2002. *Instruktion för Fältarbetet vid Risksskogstaxeringen*. Swedish University of Agricultural Sciences, Department of forest resource management and geomatics, Umeå, Sweden (In English: Instructions for Fieldwork of the National Forest Inventory).

Bossard, M., Feranec, J. and J. Otahel. 2000. *CORINE land cover technical guide – Addendum 2000*. Technical report No. 40. European Environment Agency. 105 p.

Gopal, S. and C. Woodcock. 1994. Theory and Methods for Accuracy Assessment of Thematic Maps using Fuzzy-sets. *Photogrammetric Engineering and Remote Sensing*, Vol. 60(2), pp 181-188.

Hagner, O., 2003. *Utvärderingen av SMD-skogsklassning I Värmland*. (In English: The evaluation of the SMD forest classes in Värmland). Manuscript.

Hagner, O. and H. Olsson. 2004. Normalisation of within-scene optical depth levels in multispectral satellite imagery using National Forest Inventory plot data. The 24th EARSeL Symposium, *New Strategies for European Remote Sensing*, 25-27 May 2004, Dubrovnik, Croatia.

National Atlas of Sweden. 1990. The Forests. Ed. N.E. Nilsson. SNA Publishing, Italy. 144 p.

Röst, T. and E. Ahlcrona. 2004. Tematisk noggrannhet I Svenska Marktäckedata. Lantmäteriet. Manuscript (In English: Thematic accuracy of the Swedish Land Cover Data)

Swedish Space Corporation, 1999. Pilotproduktion av CORINE Landtäckedata – beskrivning av produktionsprocess. Rapport för Lantmäteriverket, Miljödatacentrum, Naturvårdsverket, and Försvarsmakten. Rymdbolaget. (In English: The Pilot Production of CORINE Land Cover Data – A description of the production process.)

Teillet, P.M., Guindon B., and D.G Goodenough. 1982. On the slope-aspect correction of multispectral scanner data, *Canadian Journal of Remote Sensing* 8:84-106.

Wester, K., and B. Olsson. 1997. Generalisering för CORINE Landtäckning. Rapport för Rymdstyrelsen och Miljödatacentrum i Kiruna. Rymdbolaget (In English: Generalization of CORINE Landcover).