EVALUATION OF CLASSIFICATION METHODS WITH POLARIMETRIC ALOS/PALSAR DATA

Anne LÖNNQVIST[a], Yrjö RAUSTE[a], Heikki AHOLA[a], Matthieu MOLINIER[a], and Tuomas HÄME[a]

[a] VTT Technical Research Centre of Finland, Digital Information Systems, Earth Observation Team, P.O. Box 1000, FI-02044 VTT, Finland, Email: anne.lonnqvist@vtt.fi

ABSTRACT

This article compares four land cover classification methods using ALOS Palsar fully polarimetric data. Two methods are based on supervised classification and two on unsupervised classification. Both full polarimetric and intensity data were used. Fully polarimetric data gave better results (85.5%–80.8%) than intensity data only (84%–76.1%) but the differences in the overall accuracies between the methods were not more than 6%. When using intensity data only, classification results with HH and HV features were as good as using in additionally also VV, HH+VV and HH-VV bands.

Keywords: Polarimetric synthetic aperture radar, land cover

1 INTRODUCTION

Three new or near-future space borne SAR sensors, ALOS PALSAR (Phased Array type L-band Synthetic Aperture Radar), TerraSAR-X and Radarsat 2 provide fully polarimetric images. Optical satellite imagery has already been used extensively to monitor land cover. Optical data can be of limited use in certain regions of the globe where cloud cover or darkness limits acquisitions (e.g. at high latitudes). Synthetic Aperture Radar (SAR), on the other hand, is capable of day-and-night, all weather observation.

ALOS PALSAR is already in use, TerraSAR-X has been launched and Radarsat 2 will be launched in the near future. The spatial resolution of these new sensors is high as TerraSAR-X promises a 1-m resolution and Radarsat a 3-m resolution. The sensors are expected to e.g. advance methods for measuring global forest distribution, monitoring the changes of agricultural production, measuring biomass density and its variation, and disaster monitoring. As a basis, all these applications need automated interpretation and classification of data to assist decision making.

In this paper, four land cover land classification methods are compared using polarimetric ALOS PALSAR data. Two supervised and two unsupervised methods are presented.

2 MATERIALS AND METHODS

2.1 TEST SITE

Three polarimetric quad-pol PALSAR scenes used in this experiment are from Kuortane, located in the central Finland (62°48'33"N, 23°30'50"E). Kuortane area is relatively flat (38–230 m) and dominated with small areas of different land cover types. There are a few built up areas but these mostly belong to discontinuous fabric of single-family houses. Built-up areas contain lots of mature trees. Forest areas are like a mosaic-pattern, small areas of different age, sparse and dense forest. Forest stands are often long and narrow, in some cases only 25 m wide. Coniferous forest on mineral soil is the dominating forest type. Second type is mixed forests on mineral soil. The dominant soil type is glacial drift, but sand areas exist also. The main tree species are Norway spruce (*Picea abies*), Scots pine (*Pinus sylvestris*), and birch (*Betula pendula* and *Betula pubescens*).

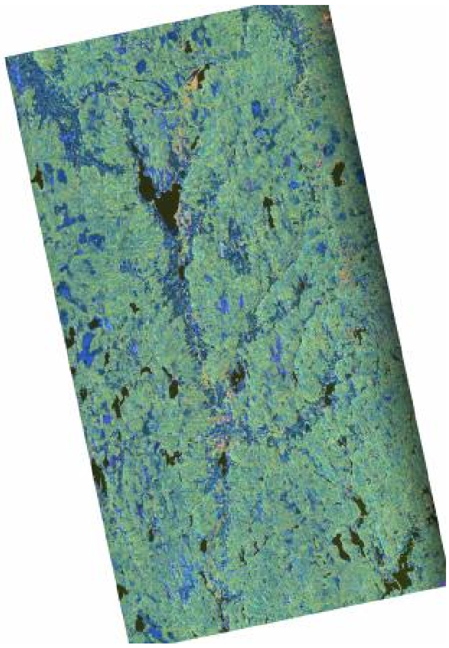

Figure 1. Pauli colour-coded presentation of the Kuortane test site. ©JAXA and METI 2006.

The field areas are fragmented. There are marsh areas used as peat harvesting peat land, and also marsh and open bog areas in natural state.

In Fig. 1, Pauli colour-coded presentation of the test site can be seen. In this representation, power scattered by single or odd-bounce targets (HH+VV) is represented with blue colour, double-bounce (HH-VV) with red and volume scattering (HV) with green. Blue colour (surface scattering) can be found in areas containing for example field or water, red and white in built up areas and green in vegetated areas.

2.2 ALOS PALSAR DATA

ALOS PALSAR (Advanced Land Observing Satellite) is a spaceborne fully polarimetric SAR operating at L band (Jaxa, 2007). Fully polarimetric scene is 35 km wide and 70 km long. The resolution of the rectified 6-look data is 25 m.

The scenes used in this experiment were acquired in November 2006, March 2007, and May 2007. The off-nadir angle is 21.5 degrees, incidence angle 24 degrees, and resolution 25 m. During the November acquisition, some snow was on the ground but the water surfaces were mostly open. In the March scene, the lakes were covered with ice with some floating water on top.

2.3 CORINE LAND COVER 2000

Corine Land Cover 2000 (CLC2000) was used to verify classification results and label clusters of unsupervised classification. The land cover database of Finland was produced by Finnish Environment Institute (SYKE) (Härmä *et al.*, 2004). It is based on automated interpretation of optical satellite images and data integration with existing digital map. In the whole Finland, 44 classes were used. The resolution of the land cover data is 25 m. For the purpose of this study, Corine classes were merged to larger classes: water, field, sparse forest, dense forest, marsh and built-up areas.

2.4 PREPROCESSING AND CLASSIFICATION METHODS

In this paper, classification experiments were done with non-rectified PALSAR scenes. The classification results were then rectified (Rauste *et al.*, 2007). Since no digital elevation model (DEM) of the study site was available at the time of the experiments, a dummy constant-value DEM was used in ortho-rectification programs. A true ortho-rectification with accurate DEM will be used in future experiments. No radiometric correction was performed.

The classification was performed with four different methods. Autochange is an unsupervised change detection and recognition system for forestry, which was developed at VTT (Häme et al., 1998). It can also be used for general classification purposes. Commercial software ER Mapper was used for supervised classification of the data (ER Mapper, 2007). Both supervised and unsupervised classifications were done with PolSARpro (provided by ESA, Pottier *et al.*, 2007).

In order to use AutoChange and ER mapper for clustering polarimetric data, five intensity channels were selected, HH, HV, VV, HH+VV, and HH-VV. The sum represents single or odd-scattering part, difference indicates a scattering mechanism characterized by double-bounce and HV characterizes volume scattering. Forest canopy produces volume scattering so HV can be used as an indication of the biomass of a pixel.

3 RESULTS

3.1 CLASSIFICATION USING FULLY POLARIMETRIC DATA

PolSARpro was used to classify fully polarimetric data. The data was read using the knowledge that it is 6-look, and coherency matrices (T3) were formed. Before classification / clustering, Lee filter (Lee *et al.*, 1999) was used for speckle filtering with window size 3.

3.1.1 Unsupervised classification

The clustering was done using Wishart H/A/Alpha classification (Lee *et al.*, 2004). In this segmentation scheme different clusters are initialized using the results of H/A/Alpha decomposition. A maximum likelihood (ML) statistical clustering is done on the polarimetric data sets based on the complex

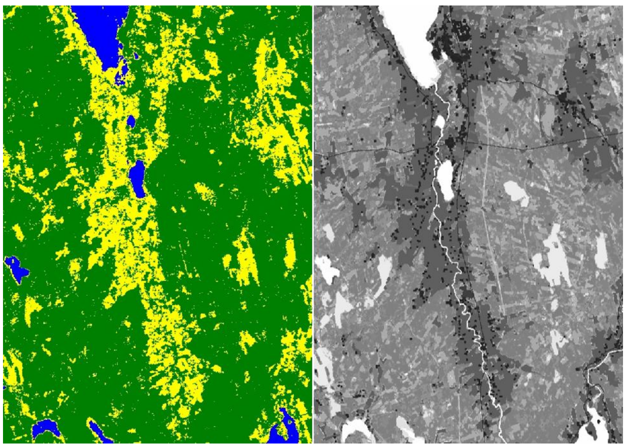

Figure 2. Classified map (lhs) and Corine 2000 land cover map of the same area (rhs). CLC2000 land cover (25m): ©SYKE (partly © MMM, MML, VRK).

Table 1. Confusion matrix of the classification with PolsarPro unsupervised Wishart classification. November dataset.

		Corine ground data				User's
		Water	Open	Forest	Total	accuracy
Palsar class	Water	2.9	0.1	0.1	3.1	91.7
	Open	0.7	14.9	4.2	19.9	75.0
	Forest	0.3	8.9	67.7	76.9	88.0
	Total	3.9	24.0	72.1	100.0	
	Procucer's accuracy	73.7	62.2	94.0		85.5

Table 2. Confusion matrix of the classification with PolsarPro supervised Wishart classification. November dataset.

		Corine ground data				User's
		Water	Open	Forest	Total	accuracy
Palsar class	Water	2.6	0.0	0.1	2.7	94.3
	Open	1.0	15.3	5.4	21.7	70.5
	Forest	0.4	8.6	66.5	75.5	88.1
	Total	3.9	24.0	72.1	100.0	
	Procucer's accuracy	65.2	63.9	92.3		84.4

Wishart probability density function. This procedure produces 16 segments. A confusion matrix of these results against the Corine 2000 data was calculated. The classes were merged on the basis of proportion of hits in classes water, field, sparse forest, dense forest, marsh and built-up areas. Since with all methods it was possible to separate classes water, open (includes field and open marsh areas) and forest, only those were used to evaluate performance of methods. In Fig. 2, a map of classified SAR data and Corine 2000 land cover of the same area can be seen. An example of the confusion matrix can be seen in table 1.

3.1.2 Supervised classification

In supervised classification, the user first defines the training areas of each class. Then the classifier calculates the Wishart statistics of the training areas and then assigns each pixel to the closest class using maximum likelihood decision rule. The accuracy of the classification was evaluated against the Corine data. As an example, confusion matrix of the November dataset is shown in table 2.

3.2 CLASSIFICATION USING INTENSITY DATA

In order to use AutoChange (CHDED) for clustering polarimetric data, the data was first transformed into intensity data. During the clustering, within image scaling was used, meaning weight of each channel is similar. Output data is given in the original intensity values. The same channels were used also in ER Mapper for the supervised classification.

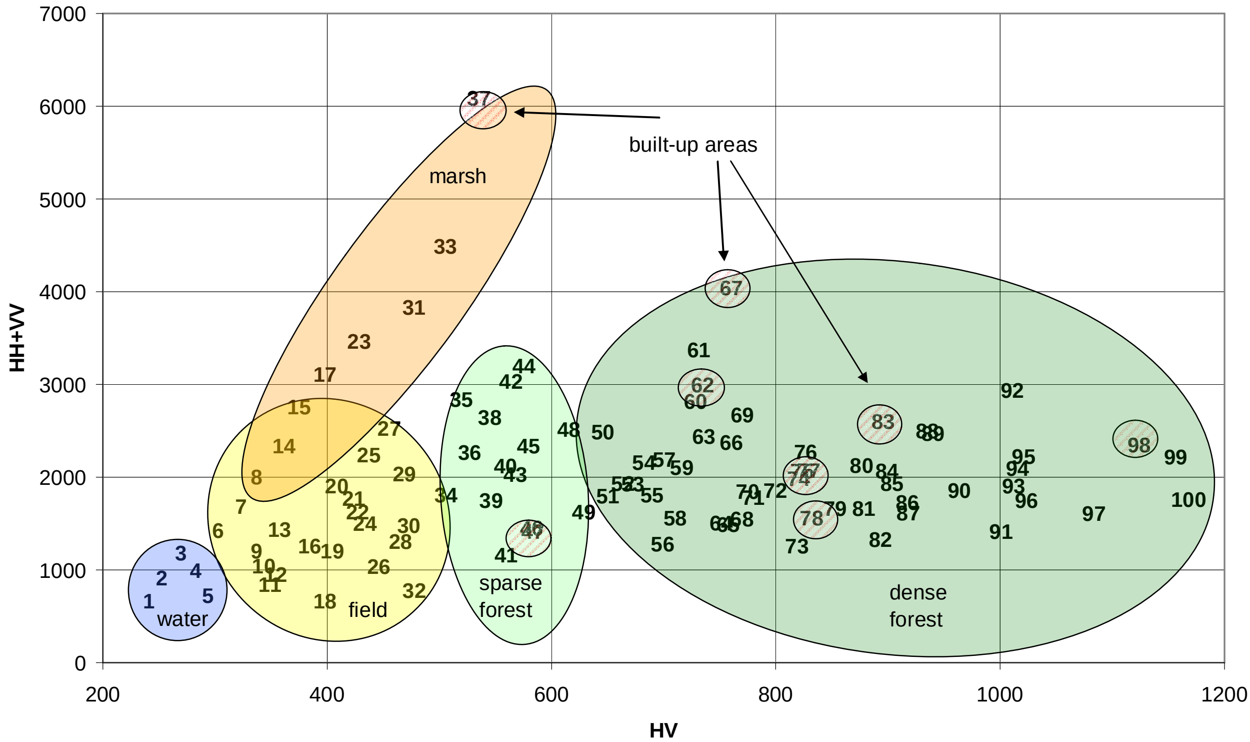

Figure 3. Clustering result of November dataset with 100 classes. Different type of land covers are divided into different classes. Blue = water, yellow = field, orange = contains marsh, light green = sparse forest, green = dense forest, red stripes = contains built-up areas.

3.2.1 Unsupervised classification

To be sure that also small classes are represented by clusters, it was decided to use 100 clusters for the clustering. They were assigned to classes the same way as with unsupervised PolSARpro results. The location of the clusters in the HV vs. HH+VV plane can be seen in Figure 3. Clearly, HV axis is proportional to the biomass content. The sum channel can be used to differentiate for example marsh. If HH-VV was used as the *y*-axis in this figure, it would also be possible to distinguish the clusters containing built-up areas (high values of difference). Confusion matrix of the November dataset can be seen in table 4.

3.2.2 Supervised classification

Training areas for classes were defined from Pauli colour-coded presentation of the data. According to their statistics, maximum likelihood enhanced classifier classifies them to the selected classes. The confusion matrix of the November dataset can be seen in table 3.

3.2.3 Comparison with dual polarisation results

To compare classification of dual polarized and fully polarized data, unsupervised classification with CHDED was done on the same data twice, first using 5 intensity channels, HH, HV, VV, HH+VV, and HH-VV, and then using only two intensity channels, HH and HV. The results can be seen in tables 4 and 5. In this evaluation, there seems to be no significant difference between the results. This might be due to the fact that when only classes water, open and forest are used, it is a biomass related classification and HV contains most of that information. Also the difference is quite small when compared to using fully polarimetric data, less than 6 % in the datasets studied here.

No data of Kuortane in the dual-polarisation mode of PALSAR was available at the time of these experiments, but a similar evaluation can be done on that data.

3.3 COMPARISON OF CLASSIFICATION RESULTS IN DIFFERENT SEASONS

The overall accuracy of different methods is compared in Fig. 4. Methods using the whole polarimetric data performed better than the ones using only intensity data. It seems that the winter pictures (containing some snow) are more difficult to classify than the spring picture.

In addition to the three classes used in this study, it was also possible to separate other classes in some cases. The number of the classes which can be separated can be seen in table 6. Due to limitations of PolSARpro, it was not possible to use the same training areas as in ER Mapper if classes other than water, field, marsh and forest were used. Here,

Table 3. Confusion matrix of the classification with ER Mapper supervised classification. November dataset.

		Corine ground data				User's
		Water	Open	Forest	Total	accuracy
Palsar class	Water	2.5	0.0	0.1	2.7	93.6
	Open	0.9	12.9	3.3	17.1	75.2
	Forest	0.4	11.1	68.7	80.2	85.6
	Total	3.9	24.0	72.1	100.0	
	Procucer's accuracy	64.3	53.6	95.3		84.0

Table 4. Classification accuracy of November data using intensity channels HH, HV, VV, HH+VV, HH-VV. November dataset.

		Corine ground data				User's
		Water	Open	Forest	Total	accuracy
Palsar class	Water	2.9	0.3	0.2	3.4	85.2
	Open	0.5	9.8	4.5	14.8	66.4
	Forest	0.6	14.0	67.4	81.9	82.3
	Total	3.9	24.1	72.0	100.0	
	Procucer's accuracy	73.3	40.7	93.5		80.0

Table 5. Classification accuracy of November data using intensity channels HH and HV. November dataset.

		Corine ground data				User's
		Water	Open	Forest	Total	accuracy
Palsar class	Water	2.7	0.2	0.1	3.0	88.8
	Open	0.6	8.0	3.2	11.9	67.5
	Forest	0.6	15.8	68.6	85.1	80.7
	Total	3.9	24.1	72.0	100.0	
	Procucer's accuracy	68.8	33.4	95.3		79.4

Table 6. Comparison of the number of classes which can be separated.

	November	March	May
CHDED 5 Channels	3	4	4
CHDED HH, HV	3	4	4
PolSARpro unsupervised	4	6	6
PolSARpro supervised	4	4	4
ER Mapper supervised	6	-	-

Table 7. Confusion matrix of the classification with PolsarPro unsupervised Wishart classification. May dataset.

		Corine ground data							User's
		Water	Field	Dense forest	Sparse forest	Marsh	Urban	Total	accuracy
Palsar class	Water	2.4	3.1	0.1	0.2	0.2	0.1	6.2	38.7
	Field	0.2	6.0	0.6	0.7	2.4	0.4	10.4	57.8
	Dense forest	0.2	3.8	45.0	12.1	1.5	1.1	63.7	70.7
	Sparse forest	0.2	3.1	4.9	4.3	1.2	0.7	14.4	30.1
	Marsh	0.1	3.2	0.5	0.5	0.6	0.3	5.3	12.0
	Urban	0.00	0.01	0.00	0.00	0.01	0.02	0.05	35.1
	Total	3.1	19.2	51.1	17.8	6.0	2.8	100.0	
	Procucer's accuracy	78.4	31.2	88.1	24.3	10.6	0.6		58.4

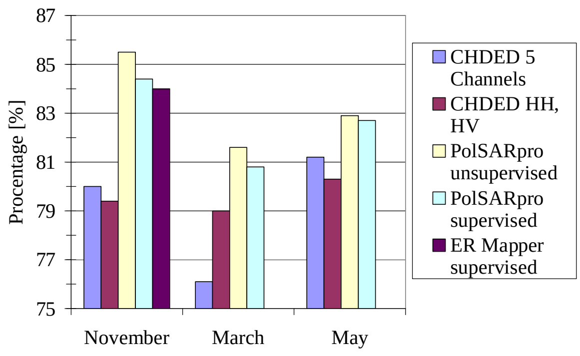

Figure 4. Comparison of overall accuracy of different classification methods.

unsupervised PolSARpro had the best performance (table 7., 6 classes with overall accuracy of 58.4%, three classes 82.5%), but supervised might be even better if it would work properly. Next step is to identify the cause of the problem in the training process.

4 DISCUSSION AND SUMMARY

Different classification methods which can be used for land cover classification were compared in this article. Fully polarimetric data from a new instrument, ALOS PALSAR, was used. Both fully polarimetric and intensity data drawn from the polarimetric data were used. As expected, fully polarimetric data gave better results than intensity data only. Supervised classification worked better with intensity data, with fully polarimetric data there was no significant difference. This might also be due to the fact that the interface for selecting the training areas was better in ER Mapper than in PolSARpro.

According to the results described in section 3.2.3, dual polarisation data is an attractive alternative to fully polarimetric data. The difference of overall accuracy was less than 6 % in each dataset. It can be expected that dual-pol classification performs better with finer-resolution data. The refined resolution is the advantage of the dual polarization mode: ground range resolution 14.9 m (pixel area 236.2 m^2) in dual-pol vs. 22.9 m (pixel area 487.8 m^2) in quad-pol.

Forest inventory data has been acquired to be able to separate labelling of the classes from validating the classes. The inventory data is also more precise than Corine 2000 data. With that data, for instance proportion of different tree species could be evaluated. Also DEM for the test site has been acquired. In the next step of this study, both DEM and forest inventory data will be taken into account.

5 ACKNOWLEDGMENTS

This work was supported by the national Technology Development Agency of Finland (TEKES) in the context of project NewSAR. Alos/Palsar data were provided by the European Space Agency in the context of ALOS/Aden AO project AOALO.3713.

Data from following resources have been used in the production of CLC2000 material: SYKE, MML, MMM (fields 1999), VRK (built-up areas 2001). Material from Metsähallitus and UPM Kymmene has been utilized in the interpretation of the satellite images.

6 REFERENCES

ER Mapper, 2007, http://www.ermapper.com/.

Häme, T; Heiler, I; San Miguel-Ayanz, J. 1998. An unsupervised change detection and recognition system for forestry. *International Journal of Remote Sensing*, vol. 19 no. 6, 1998, 1079 – 1099.

Härmä, P., Teiniranta, R., Törmä, M., Repo, R., Järvenpää, E., Kallio, M. 2004. Production of CORINE2000 land cover data using calibrated LANDSAT 7 ETM satellite image mosaics and digital maps in Finland. *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, 20-24 Anchorage US. 20-24 Sept. 2004, vol. 4, pp.2703 - 2706.

Jaxa, 2007, http://www.eorc.jaxa.jp/ALOS/about/palsar.htm

Lee, J.-S., Grunes, M. R., and de Grandi, G. 1999. Polarimetric SAR Speckle Filtering and Its Implication for Classification. *IEEE Transactions on Geoscience and Remote Sensing*, vol 37, no. 5, part 2, pp.:2363 – 2373, Sept. 1999.

Lee, J.S., Grunes, M.R., Pottier, E., Ferro-Famil, L. 2004. Unsupervised terrain classification preserving polarimetric scattering characteristics. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 4, pp. 722-731, April 2004.

Pottier, E., Ferro-Famil, L., Allain, S., Cloude, S., Hajnsek, I., Papathanassiou, K., Moreira, A., Williams, M., Pearson, T., and Desnos, Y.-L. 2007. An overview of the PolSARpro v2.0 software. the educational toolbox for polarimetric and interferometric polarimetric SAR data processing. *POLinSAR07 ESA Workshop*, Frascati, Italy, January 2007.

Rauste, Y., Lönnqvist, A., Molinier, M., Henry, J.-P., and Häme, T., 2007. Ortho-Rectification and Terrain Correction of Polarimetric SAR Data Applied in the ALOS/Palsar Context. *Proceedings of IEEE International Geoscience and Remote Sensing Symposium IGARSS 2007*. In press.