ON THE USE OF PALSAR DATA AND SUPPORT VECTOR MACHINES FOR LAND COVER ANALYSIS AND CLASSIFICATION

Nicolas Longépé[1], *Preesan Rakwatin*[1], *Osamu Isoguchi*[1], *Masanobu Shimada*[1], *Yumiko Uryu*[2]

[1]Earth Observation Research Center, Japan Aerospace Exploration Agency
Tsukuba Space Center, 2-1-1 Sengen 305-8505 Tsukuba, Japan
e-mail: nicolas.longepe@jaxa.jp

[2]World Wildlife Fund
Kantor Taman A9 Unit A-1, Kawasan Mega Kuningan, Jakarta 12950, Indonesia

1. INTRODUCTION

This study introduces the status of one methodology for land cover classification in tropical rainforest as part of the ALOS K&C initiative project. Incorporated with World Wildlife Fund (WWF), Riau province in central Sumatra (Indonesia) was selected as test site. Riau hosts some of the most biodiversity ecosystems and unique species in the world. It is covered by vast peat lands estimated to hold Indonesian largest stock of carbon. However, Riau has been under serious threat because of rapid large-scale deforestation. In this study, 50m ortho-rectified dual-polarized PALSAR mosaic products are solely used to make an attempt to monitor these changes. The following points will be detailed through the paper.

2. QUALITATIVE ANALYSIS OF SCATTER PLOT

WWF listed 27 different land covers within this area. In the framework of this study, a set of 14 classes is analyzed for the sake of clarity and of negligible land cover types. Using dual-polarized SAR data only (HH and HV channels), the capability of PALSAR data for discriminating these different classes is first analyzed. Some preliminary scatter plots (HH vs. HV) are studied showing some limited possibilities with radiometric information only. Oil palm and acacia plantations widely spread over this area. Their dual-polarized signatures are slightly different from the natural forest ones providing some information in a classification context. Clear cut areas clearly induce a strong modification of the dual-polarized signature. However, in a general manner, the variance of these histograms is very large and it is obvious that a simply threshold-based method can not work properly over this dataset.

3. QUALITATIVE ANALYSIS OF TEXTURAL INFORMATION

By using PALSAR mosaic products, no time-dependent information is available. In that framework, spatial statistics may be useful in classifying these natural media. Many methodologies exist in computer vision in order to analyze textural statistics. Variogram, gray-level co-occurrence matrix (GLCM), Markov-Random fields (MRF) and wavelet transforms are some of the most common methods. Variogram is simply defined by the squared difference between paired pixels separated by a varying

The authors wish to thank WWF Indonesia and all the participants to the flyover mission held in Riau in August-September 2009.

distance lag. The “sill” and “range” values might be correlated with biomass and species as referred to experimental studies [1]. At least, it depends in a theoretical point of view on the density of scatterers within the resolution cell and the correlation between them [2]. For this study, variograms with data in logarithmic scale are plotted for each class and are summarized in the following.

For one specific land cover, the sill value varies in a large manner over clear-forested and dry forested areas. For these two land covers, the variance strongly depends on the selection of the Region of Interest (ROI). The dependence is smaller for the others, especially for peat swamp or swamp forest, oil palm and clear-cut acacia where these variations are of the order of 0.5 (log. scale). Since ROIs have been selected over the entire Riau province, the more the natural media is homogeneous at regional scale, the less the variations of variance are important. Acacia and oil palm plantations are each exploited through similar processes by several companies in Riau, resulting in a well-organized network of plants whose biomass is about 100 t/ha [3].

Finally, some polarization particularities can be observed over some land cover types. Coconut, dry forest or mangrove clearly induces higher sill values for the co-polarized HH channel as opposed to shrubs, acacia or rubber plantations whose variance is superior for the cross-polarized HV channel. Assuming that HH and HV channels are linked at L-band to the single/double bounce and volume contribution respectively, one can state that the canopys spatial distribution is relatively more heterogeneous than the trunk ones over shrubs or rubber than over coconut plantation or mangrove. The leaves and branches structure for these two first species might be more chaotic and diversified at decametric distance.

All the texture analysis (GLCM, Wavelet...) require a method to define the size of the convolution filter window. The size of the filter window is important as it defines the area used for texture calculations. More recent research has found that the range values are useful for determining the maximum size of processing windows and for limiting spatial resolution [4, 5]. Through this qualitative analysis of semivariograms, it is shown that textural parameters over natural media in Riau are very complex. The characterization of an optimal processing window is not straightforward for a given land cover. It appears that an enhanced approach accounting for the high complexity of textural signature is necessary.

4. SUPPORT VECTOR MACHINES FOR LAND COVER ANALYSIS

Spatial statistics in SAR backscatter data acquired over forests are of high interests for land cover characterization and classification [6]. However, it is still very difficult to assess the complex nature of textures. On the other hand, classification algorithms based on statistical learning methods such as the supervised Support Vector Machines (SVM) approach are used in a wide range of data mining applications [7]. During the last decade, SVM has been successfully introduced as a classifier in remote sensing [8], with recent studies dealing with SAR data and forestry [9].

An original methodology is introduced in order to analyze textural information. In order to compute an optimized set of textural parameters at regional scale, SVM is also used as a technique for feature selection. In a similar vein as in [10, 11] which is based on wavelet frames, a new tool for analyzing multi-scale information from SAR data is built. Based on the Support Vector Machine (SVM) technique, the Recursive Feature Elimination algorithm, namely the SVM-RFE [12], is a simple but efficient algorithm which was first implemented in the context of cancer gene selection. As inputs of this algorithm, the Haralick’s parameters [13] based on the GLCM are computed from the 50m dual-polarized PALSAR data for a large range of lag distances, windows sizes and quantization levels. All of these parameters are then ranked by the SVM-RFE algorithm based on their respective weight in the separating hyperplane. By analyzing the ranked GLCM parameters, the textural information can be assessed for each land cover type.

This methodology represents a real asset for a future operational method of land cover mapping. In this study, the area of interest is very wide (about 14 000 by 24 000 pixels at 50m resolution) and the use of a limited number of textural parameters is crucial for any future operational applicability. By knowing the most relevant textural parameters for a given land cover once and for all, a limited number of parameters has to be computed while including an optimized range of spatial information.

5. CLASSIFICATION RESULTS

Finally, the radiometric information (HH and HV channels) and the best Haralick's textural information are used as input of the SVM classifier so as to discriminate the different land covers. Accuracy assessment is carried out at local scale, showing the good capabilities of 50m dual-polarized PALSAR data for land cover classification. The results over the entire Riau province are shown in Fig. 1. Some outlooks such as the implementation of pre/post processing methods (clustering, pattern recognition) are discussed. The forest/non forest map over the whole Sumatra island will be also processed using the same technique.

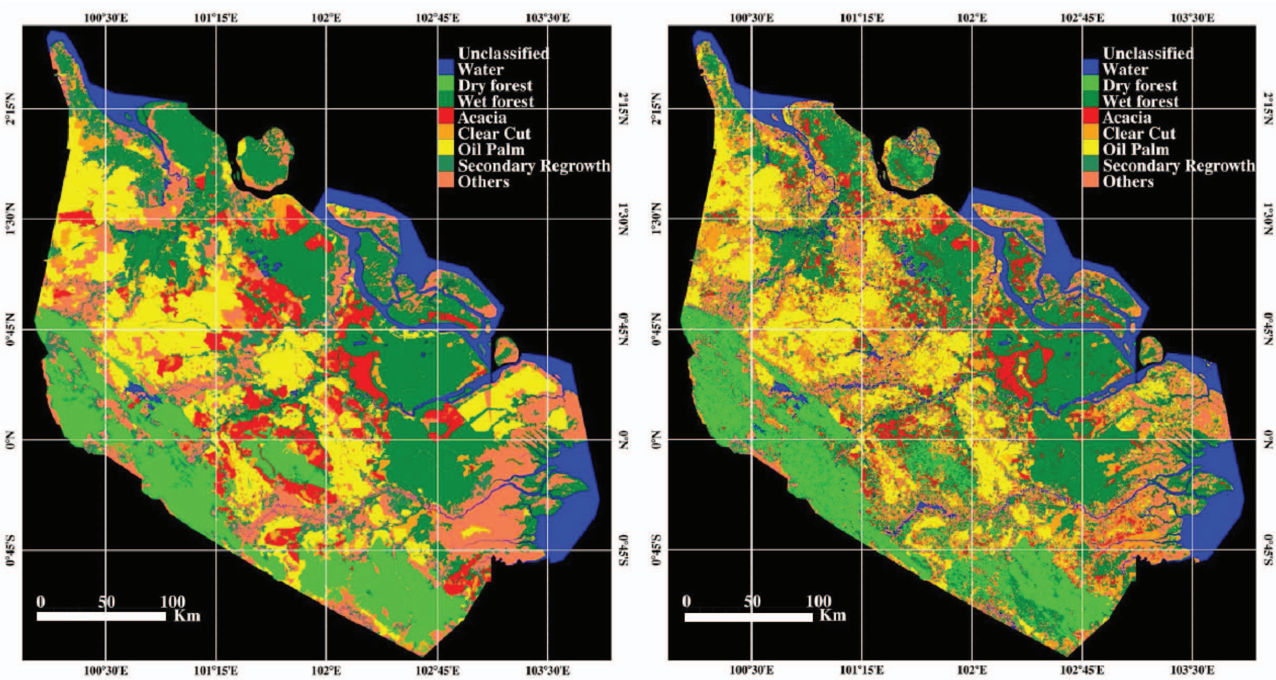

Fig. 1. WWF ground truth generated with Landsat images (left) Land cover map estimated by 50m dual-polarized orthorectified PALSAR data (right)

6. REFERENCES

[1] T.M. Kuplich and P.J. Curran, "Relating SAR image texture to the biomass of regenerating tropical forest," *Int. J. Remote Sens.*, vol. 26, pp. 4829–4854, 2005.

[2] C. Olivier, "Correlated K-distributed clutter models," *Opt. Acta.*, vol. 32, no. 12, pp. 1515–1547, 1985.

[3] Y. Uryu et al., "Deforestation, forest degradation, biodiversity loss and CO_2 emission in Riau, Sumatra, Indonesia," Tech. Rep., WWF Indonesia, 2008.

[4] W.B. Cohen and T.A. Spies, "Semivariograms of digital imagery for analysis of conifer canopy structure," *Remote Sens. Env.*, vol. 34, pp. 167–178, 1990.

[5] C.A. Coburn and A.C.B. Roberts, "A multiscale texture analysis procedure for improved forest stand classification," *Int. J. Remote Sens.*, vol. 25, pp. 4287–4308, 2004.

[6] F.T. Ulaby, F. Kouyate, B. Brisco, and T.H. Lee Williams, “Textural information in SAR images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 24, no. 2, pp. 235–245, 1986.

[7] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines (and other kernel-based learning methods)*, Cambridge University Press, 2000.

[8] M. Brown, H.G. Lewis, and S.R. Gunn, “Linear spectral mixture models and support vector machines for remote sensing,” *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2346–2360, Sep 2000.

[9] C. Lardeux, P.-L. Frison, C. Tison, J.-C. Souyris, B. Stoll, B. Fruneau, and J.-P. Rudant, “Support vector machine for multifrequency SAR polarimetric data classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 12, pp. 4143–4152, Dec. 2009.

[10] M. Simard, S. Saatchi, and G.F. De Grandi, “The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest,” *IEEE Trans. Geosci. Remote Sens.*, vol. 38, pp. 2310–2321, 2000.

[11] G.D. De Grandi, R.M. Lucas, and J. Kropacek, “Analysis by wavelet frames of spatial statistics in SAR data for characterizing structural properties of forests,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 2, pp. 494–507, 2009.

[12] I. Guyon, J. Weston, S. Barnhill, and V.N. Vapnik, “Gene selection for cancer classification using Support Vector Machines,” *Machine Learning*, vol. 46, pp. 389–422, 2002.

[13] R.M. Haralick and K. Shaunmugam, “Textural features for image classification,” *IEEE Trans. Syst. Man Cybernetics*, vol. 3, pp. 610–621, 1973.