Land cover classification in Riau, Indonesia using 50m ALOS PALSAR MOSAIC Based on Amplitude and Texture Analysis

P Rakwatin, N Longepe, O Isoguchi, M Shimada, and Y Uryu
Earth Observation Research Center (EORC)
Japan Aerospace Exploration Agency (JAXA)

16 June 2009

K&C meeting, RESTEC, Tokyo

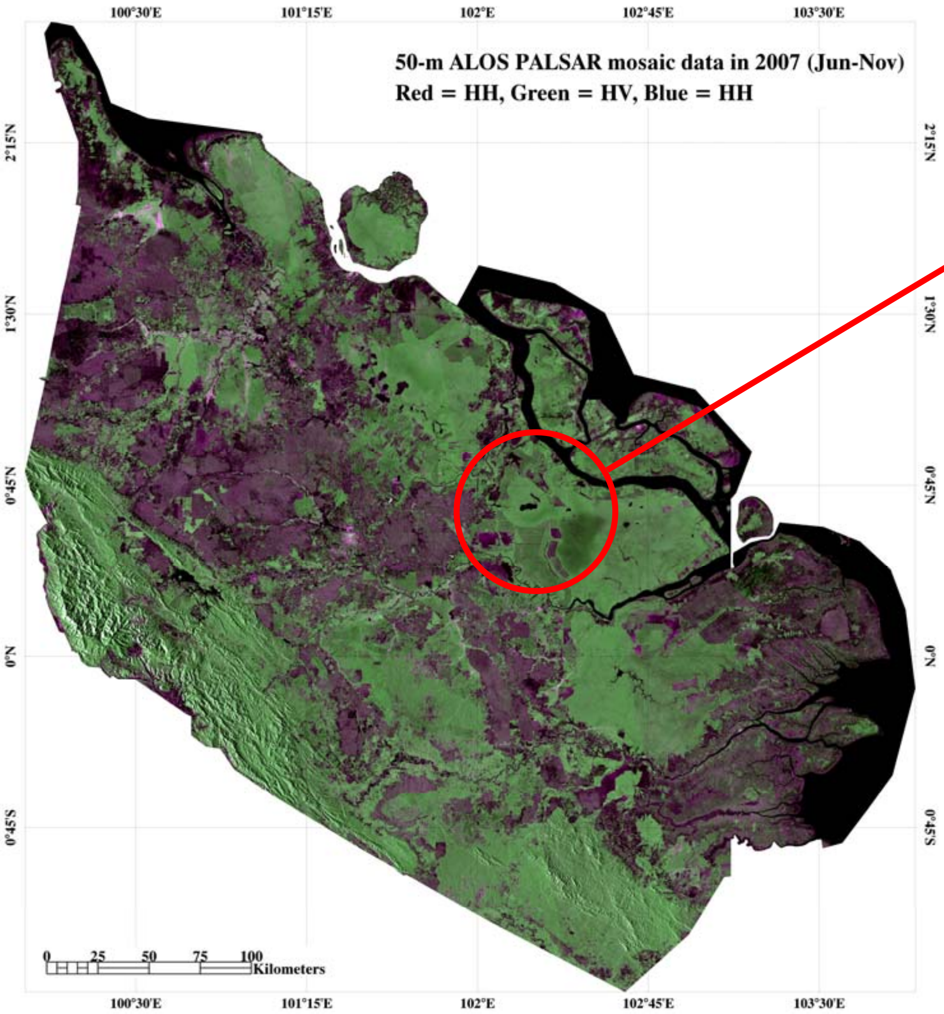
Research Objective

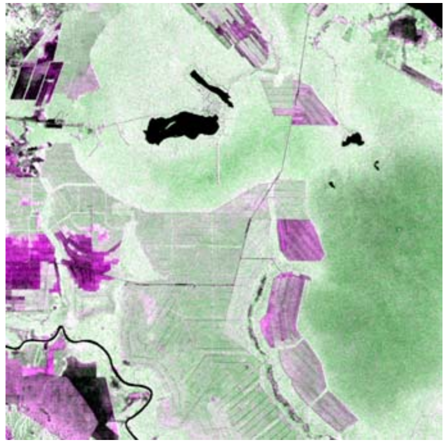
- Aim to assess the potential of ALOS PALSAR 50m mosaic for application to tropical rain forests
- Analysis of image texture based on grey level co-occurrence (GLC) approach involves choices concerning
  - GLC attribute, displacement length, quantization scale and window size.
- Explored a new method called Support Vector Machines (SVMs) for textural classification
  - Integrating spectral and textural information
  - Evaluating the effectiveness of proposed texture measures

Methodology and Data use

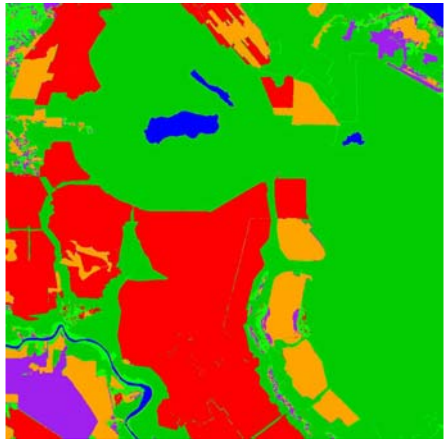
- **Methodology**
  - ➜ Gray Level Co-occurrence Matrix
  - ➜ Semivariogram
  - ➜ Separability of land use types
  - ➜ Support Vector Machines
- **Data used**
  - ➜ 50-m ALOS PALSAR mosaic data in 2007 (Jun-Nov)
  - ➜ WWF Land cover map in 2007

Study Area





PALSAR



WWF map

Texture Analysis

- Texture is a repeating pattern of local variations in image intensity
- There are many different method to extract textural information
  - **Statistical Approach** : a quantitative measure of the arrangement of intensities in a region (Moment of Intensity, Gray Level Co-occurrence Matrix)
  - **Modeling Approach** : texture modeling techniques involve constructing models to specify textures. (Markov random fields)
  - **Frequency Approach** : texture is a set of texture element in some regular or repeated relationship (Gabor filters)

Gray Level Co-occurrence

- Shows how frequent every particular pair of grey levels in the pixel pairs is separated by a certain distance (d) along direction (θ). (Haralick, 1979)

Example matrix:

0	0	1	2	1
0	2	2	0	2
1	1	1	2	1
0	2	0	1	0
0	1	2	2	0

C-O matrix (angle=0°, distance=1):

	0	1	2
0	1	$\boxed{3}$	3
1	1	2	3
2	3	2	2

Haralick Attribute

$$\text{Contrast} = \sum_{i=0}^{M}\sum_{j=0}^{N} C_{ij}(i-j)^2$$

$$\text{Correlation} = \sum_{i=0}^{M}\sum_{j=0}^{N} \frac{(i-\mu_r)(j-\mu_c)C_{ij}}{\sqrt{\sigma_r^2\sigma_r^c}}$$

$$\text{Entropy} = -\sum_{i=0}^{M}\sum_{j=0}^{N} C_{ij}\log(C_{ij})$$

$$\text{Energy} = \sum_{i=0}^{M}\sum_{j=0}^{N} C_{ij}^2$$

$$\text{Homogeneity} = \sum_{i=0}^{M}\sum_{j=0}^{N} \frac{C_{ij}}{1+|i-j|}$$

$$\text{3rd Order Moment} = \sum_{i=0}^{M}\sum_{j=0}^{N} C_{ij}(i-j)^3$$

$$\text{Inverse Variance} = \sum_{i=0}^{M}\sum_{j=0}^{N} \frac{C_{ij}}{(i-j)^2}$$

$$\text{Sum Average} = \frac{1}{2}\sum_{i=0}^{M}\sum_{j=0}^{N}(iC_{ij}+jC_{ij})$$

$$\text{Variance} = \frac{1}{2}\sum_{i=0}^{M}\sum_{j=0}^{N}\left[(i-\mu_r)^2C_{ij}+(j-\mu_c)^2C_{ij}\right]$$

$$\text{Cluster Tendency} = \sum_{i=0}^{M}\sum_{j=0}^{N}(i-\mu_r+j-\mu_c)C_{ij}$$

$$\text{Maximum Probability} = \text{Max}(C_{ij})$$

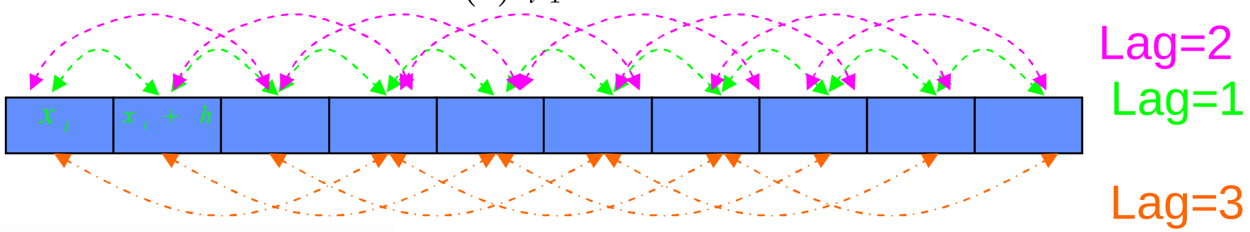
GLCM features

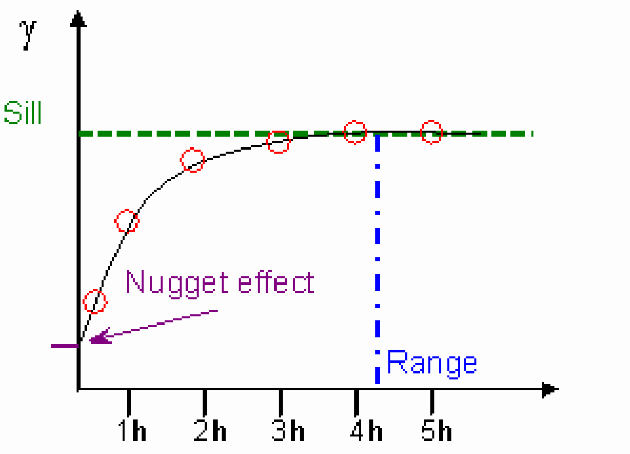
- **Displacement length** can be used, d=1,2,3, etc.
- **Quantization scale**
  - 256x256 matrices for 8 bit image.
  - 64x64 matrices for 6 bit image will result in
- **Attribute** can be calculated from each GLCM.
- **Window size**

Semivariogram

- **A plot of average variance between points vs. distance between those points**

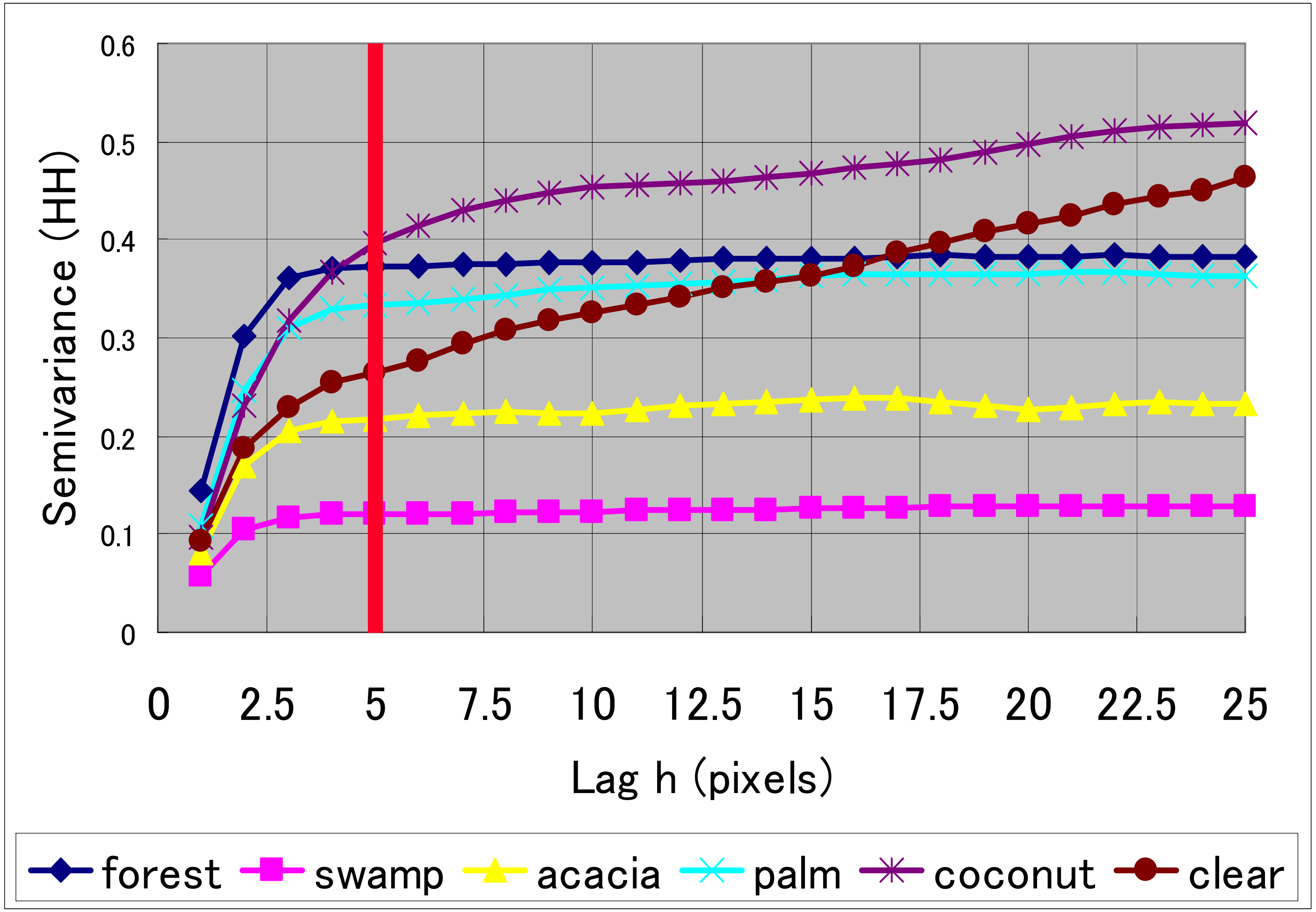
$$\gamma(h)=\frac{1}{2N(h)}\sum_{i=1}^{N(h)}\left[\mathrm{DN}(x_i)-\mathrm{DN}(x_i+h)\right]^2$$





- Range – **Distance at which max. variance is reached (data considered decorrelated)**
- Nugget – variability at zero distance, represents analytical, or theoretical errors
- Sill – **Level of max. variability**

Semivariogram for major land use in Riau province

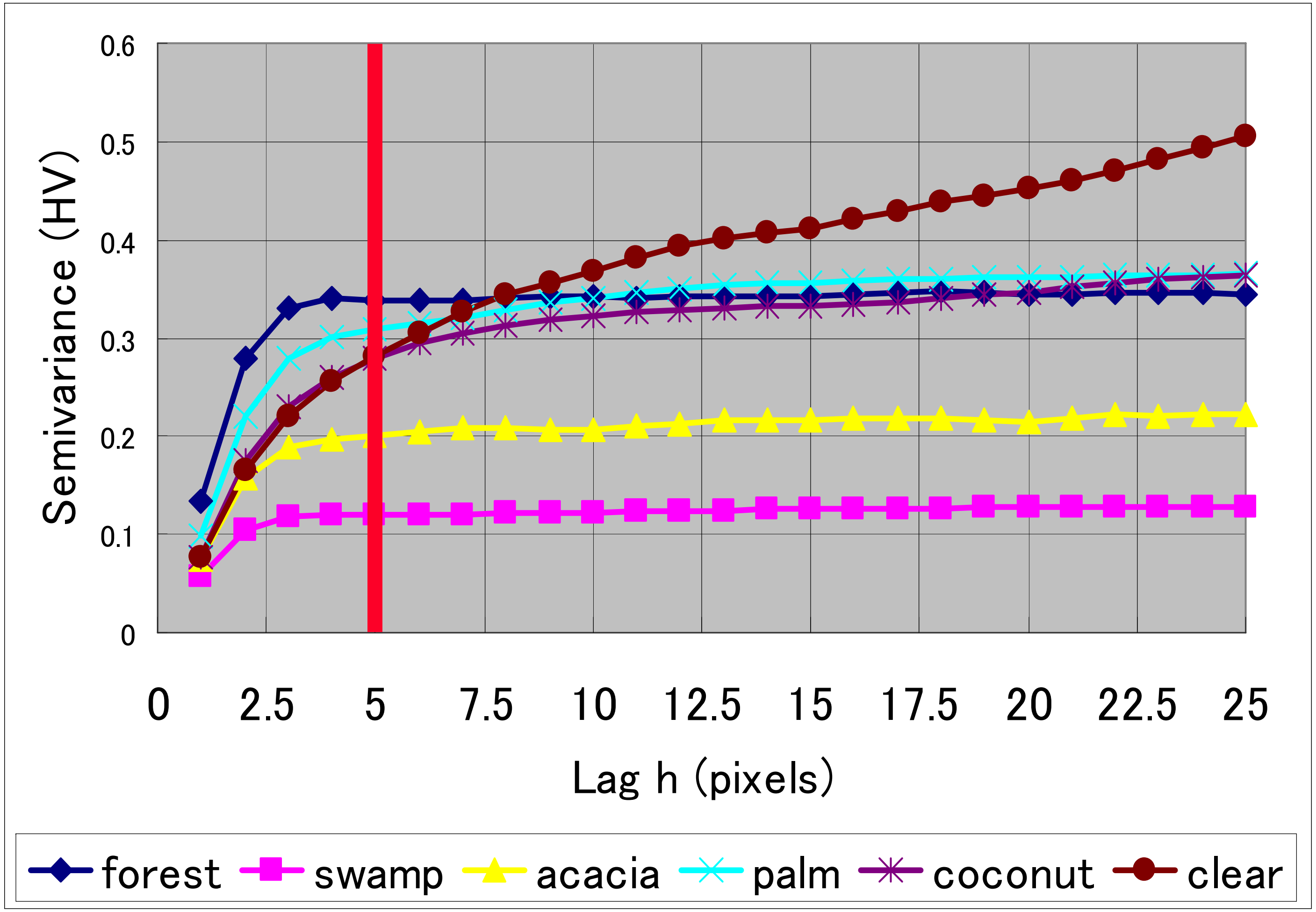


ALOS

K&C Initiative

An international science collaboration led by JAXA

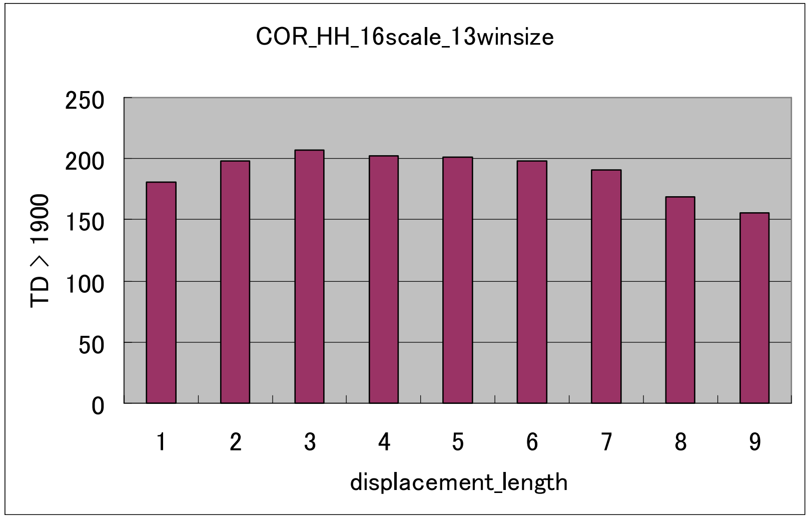
Semivariogram for major land use in Riau province

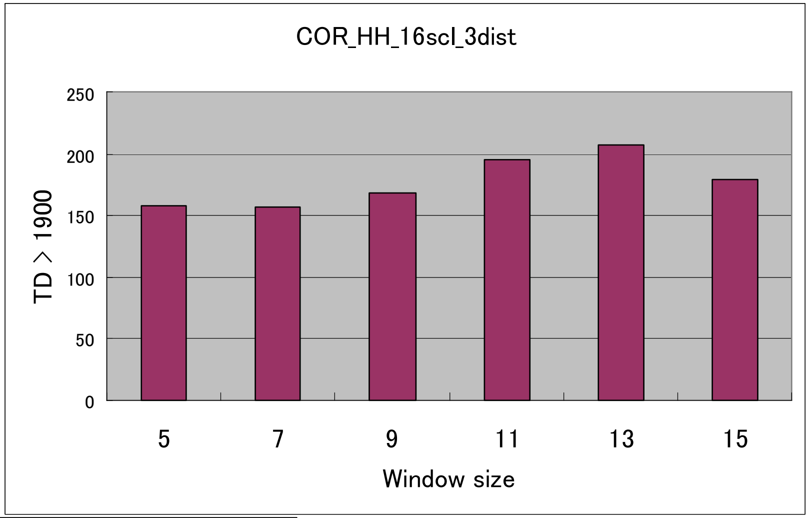


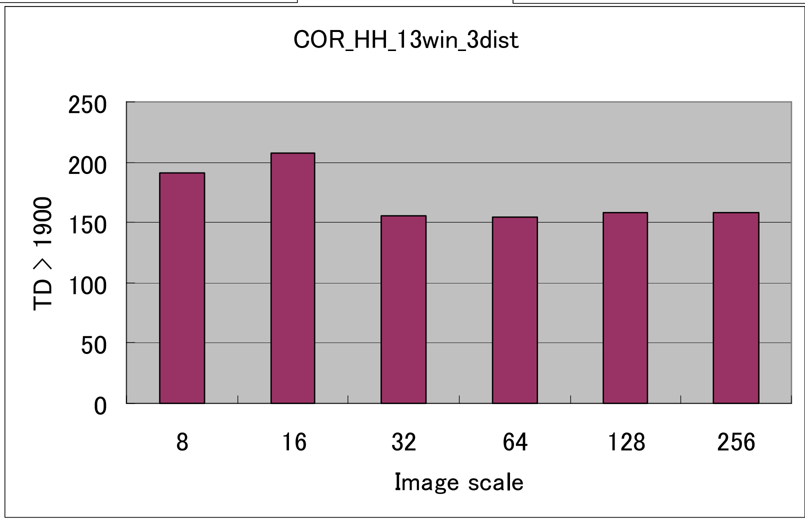
Separatability

- **Transformed Divergence (TD)**
  - **Statistical distance between class pairs**
  - **an indirect and a priori estimate of the probability of correct classification.**
  - **Divergence values scaled to 0 – 2000.**
  - **Between-class separation**
    - **2000 ~ excellent**
    - **1900-2000 ~ good**
    - **1700-1900 ~ moderate**
    - **below 1700 ~ poor**

GLCM Correlation as a function of displacement length, window size, and Image scale

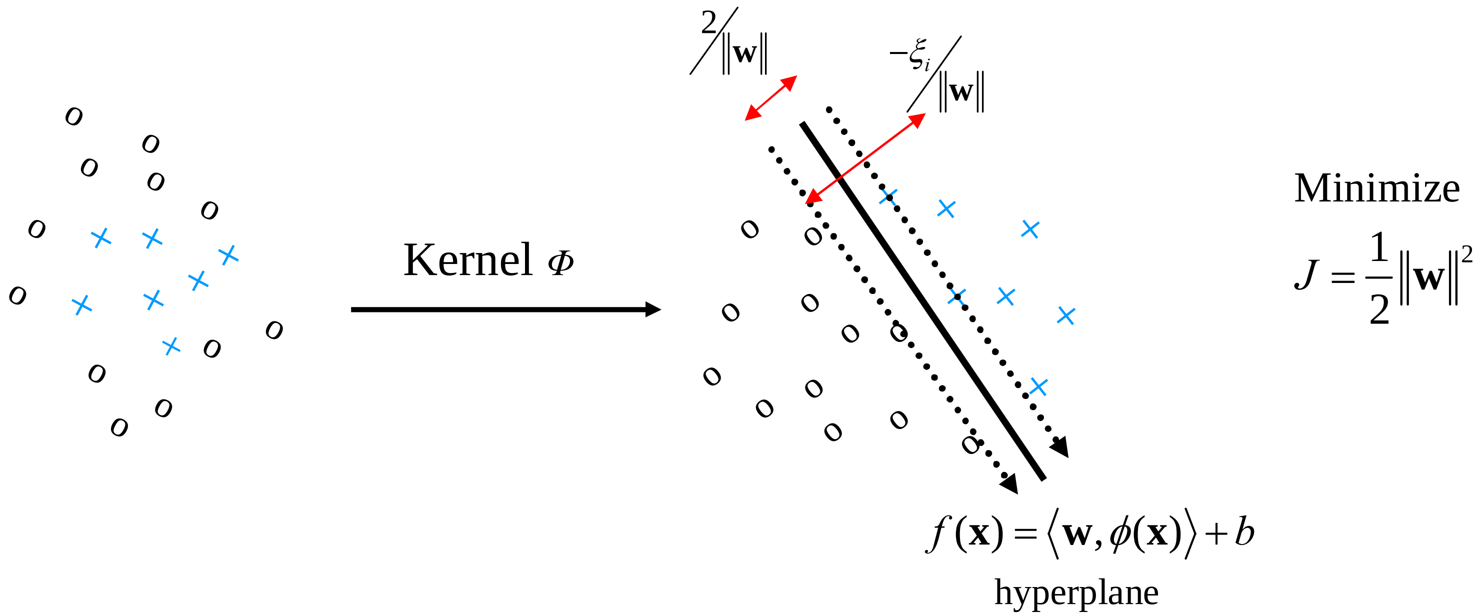


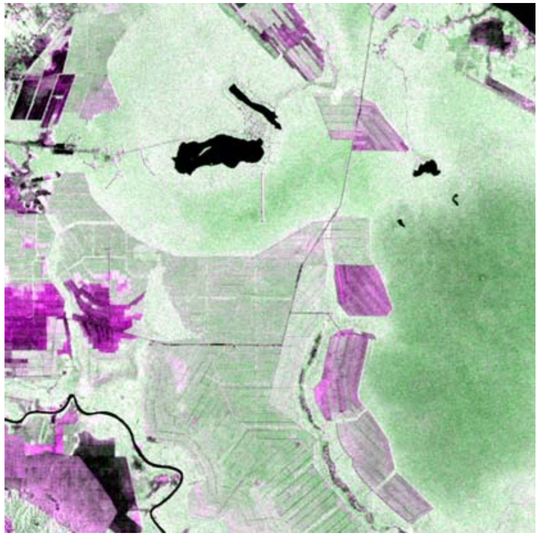




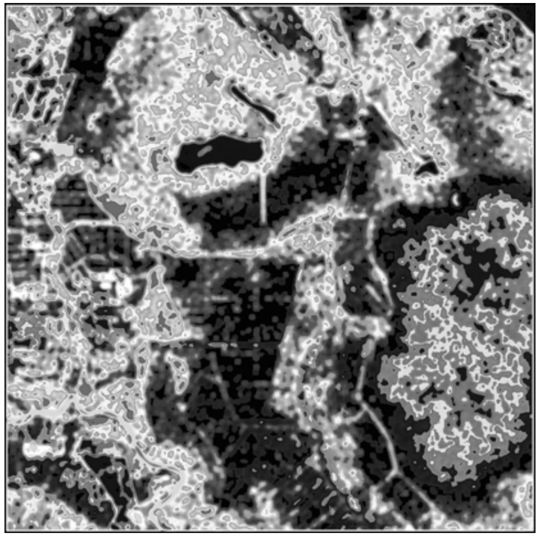
Support Vector Machine (SVM) Classifiers

- First SVM was designed for binary classification
- Possible to separate non linear case by using higher dimension through a Kernel function
- Deal with classification errors with a penalty parameter

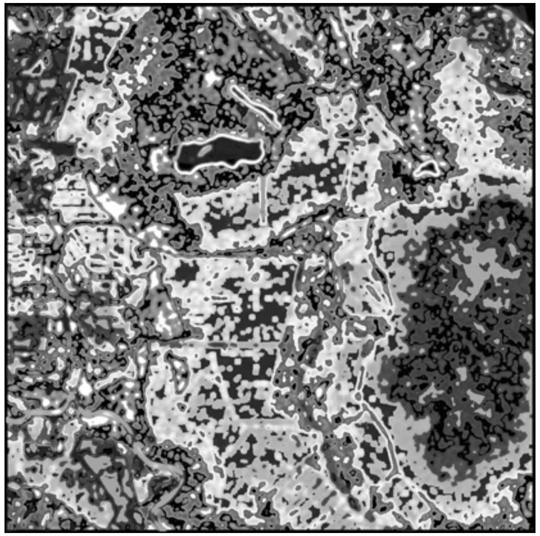




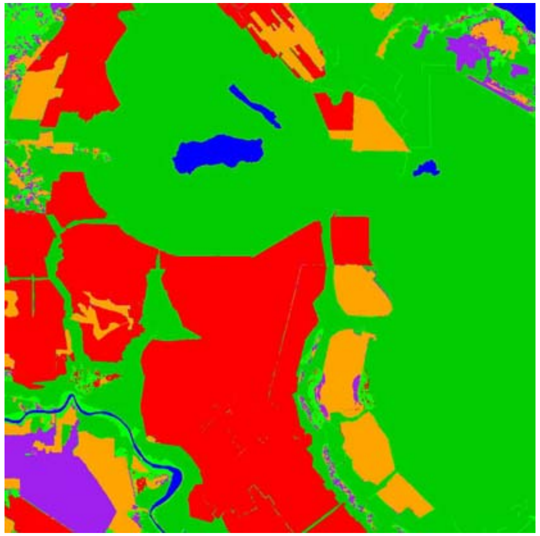
R = HH, G = HV ,B = HH



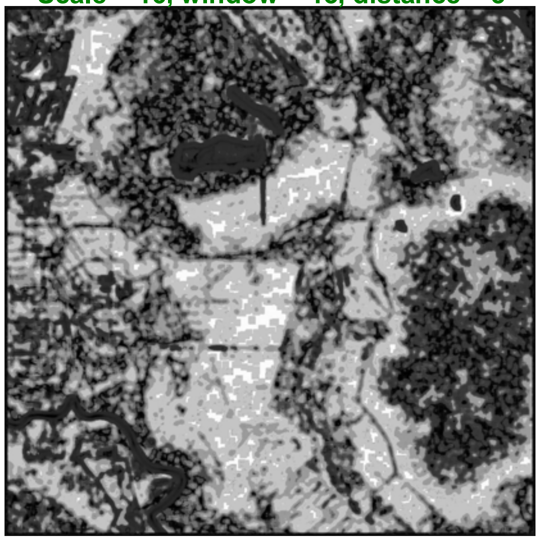
GLCM Mean
Scale = 16, window = 13, distance = 3



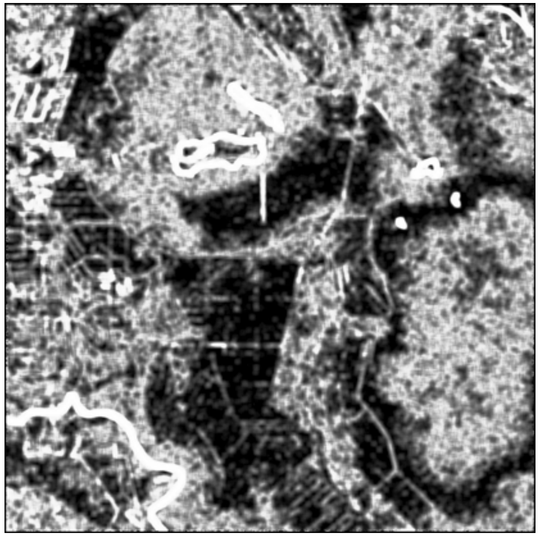
GLCM Variance



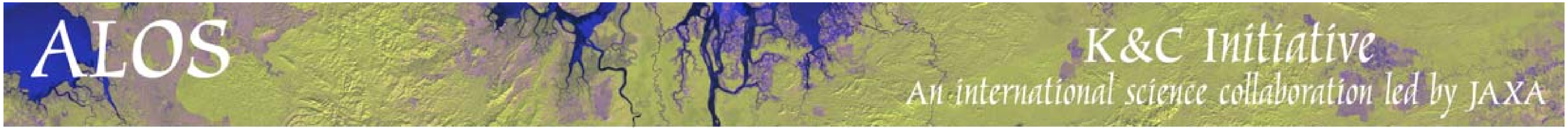
green = swamp, red = acacia , orange = clear, purple = palm, blue = water

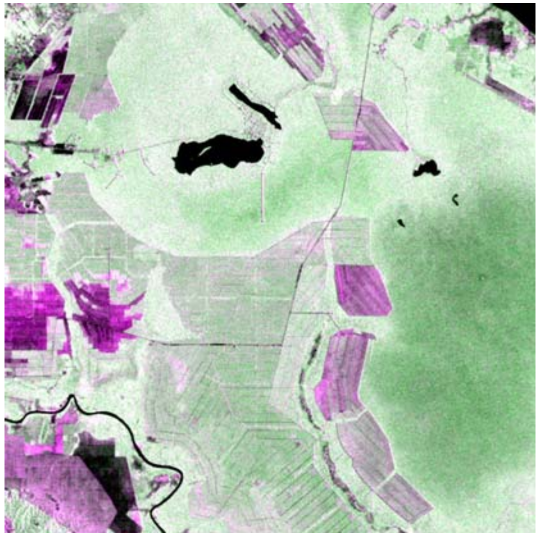


GLCM Correlation

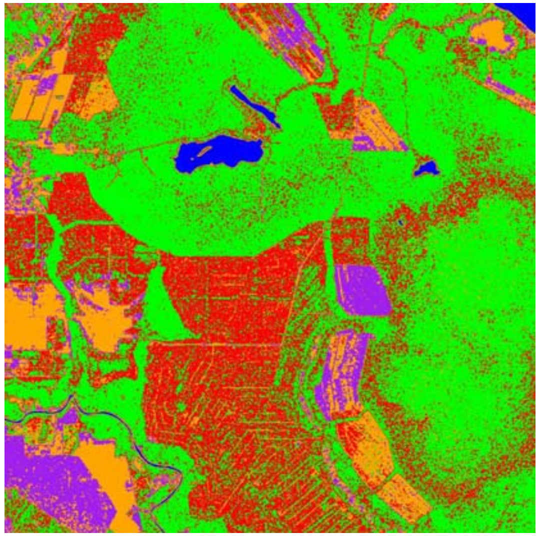


GLCM Dissimilarity

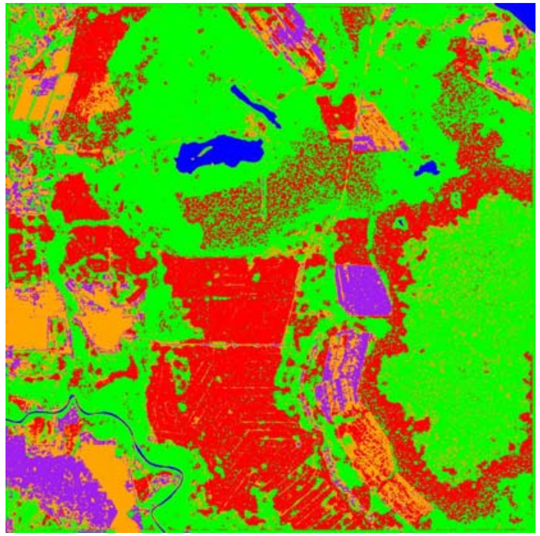




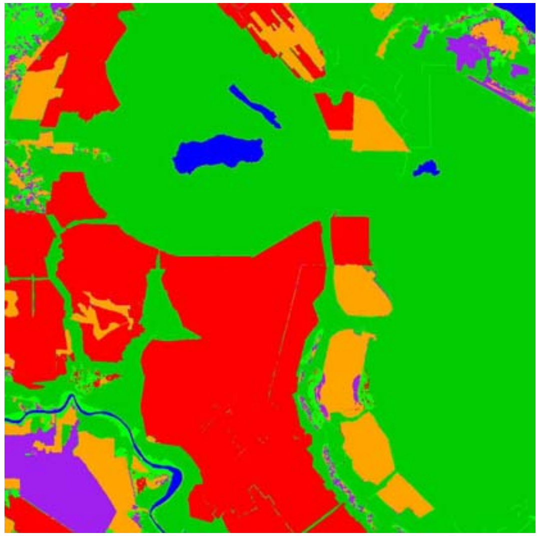
R = HH, G = HV ,B = HH



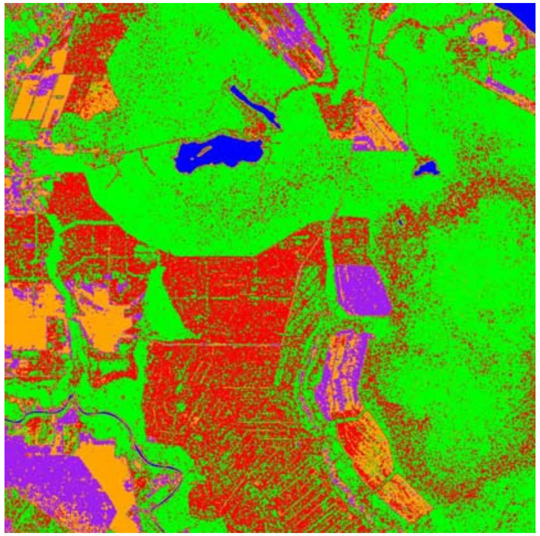
Maximum likelihood (71.35%)



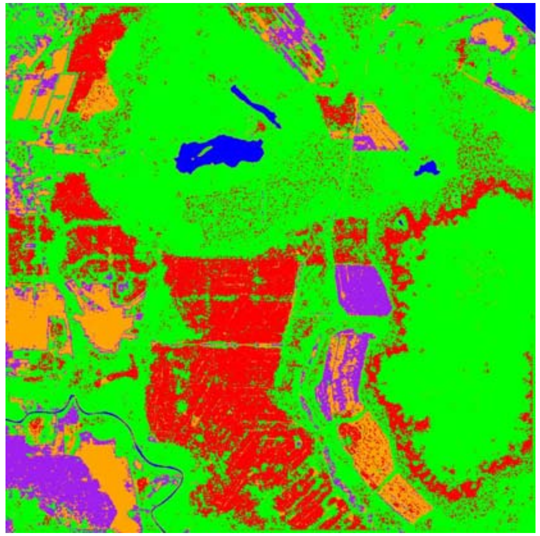
Maximum likelihood with texture (76.54%)



green = swamp, red = acacia , orange = clear,
purple = palm, blue = water



svm (72.56%)



svm with texture (81.12%)

Feature Selection Method (Nicola`s initiation)

- **SVM-based test: SVM-RFE (Guyon et al. 2002)**
  - **SVM-based classifier minimizes the following cost function:**

$$J = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i$$

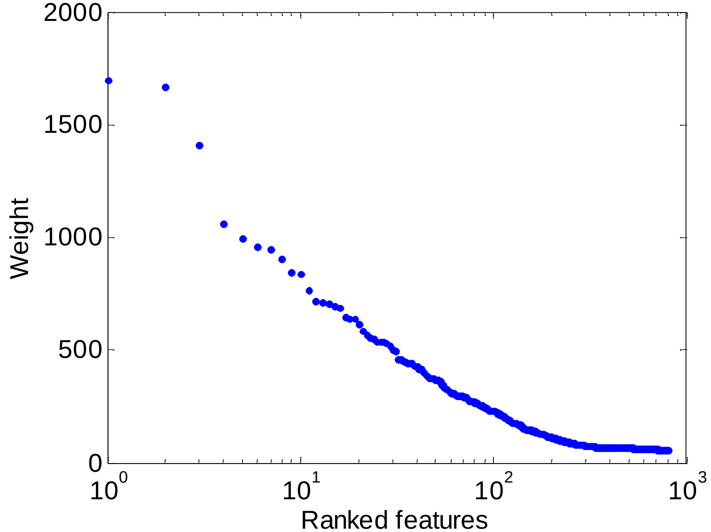
  - **When the j^th feature is removed:** $J(j) \approx J + w_j^{\ 2}$
  - **Removing the feature the lowest** w_j^2
    - **Remove the less relevant parameter**
    - **Increase the generalization performance**
- **Multiclass SVM**
  - **One versus All method: for k-class problem, k hyper planes** $f^k(\mathbf{x}) = \langle \mathbf{w}^k, \phi(\mathbf{x}) \rangle + b^k$
  - **When the j-th feature is removed (Zhou and Tuck 2007):**

$$J(j) \approx J + \frac{1}{k}\sum_{r=1}^{k}\left(w_j^r\right)^2$$

Guyon et al, “Gene selection for cancer classification using support vector machine”, Mach. Learning, 2002
Zhou and Tuck, “MSVM-RFE: extensions of SVM-RFE for multiclass gene selection on DNA microarray data”,Bioinformatics, 2007

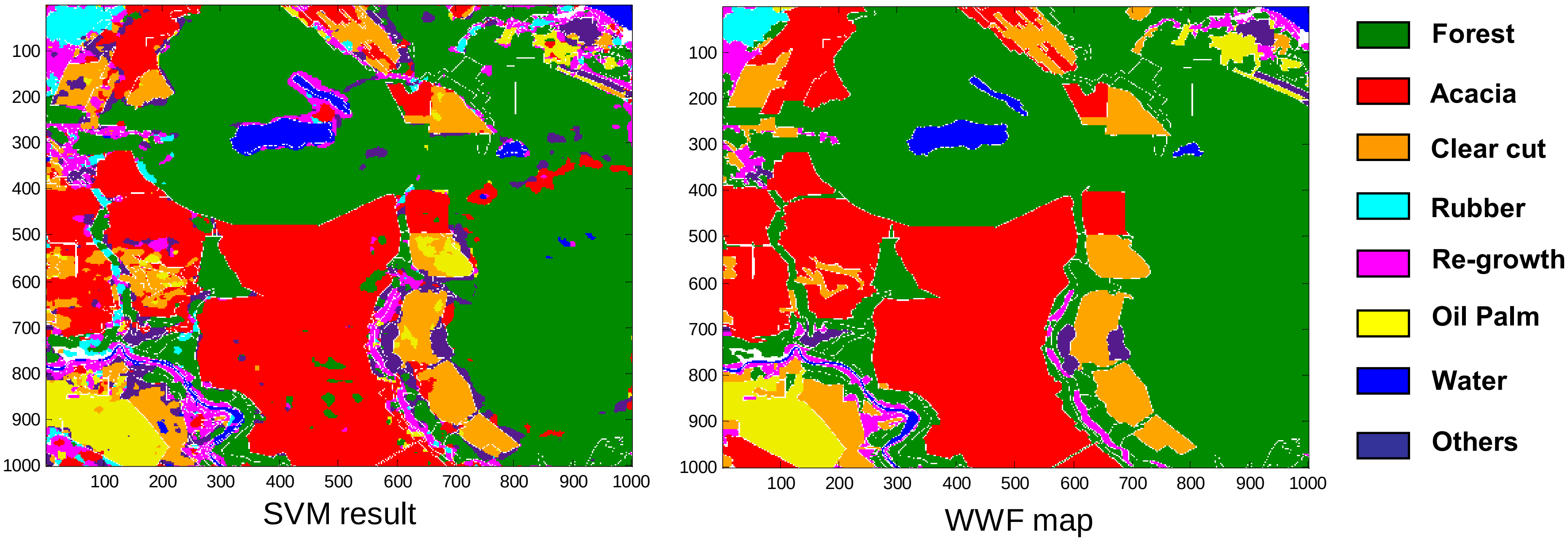
SVM-based Feature Selection

- Initial features
  - Channels (2) * Quantization Method (2) * Quantization levels (4) * Distance depending on windows size (41) * Haralick's parameters (9)
  - 5904 parameters
- MSVM-RFE with 8 classes:
  - Peat / non Peat Swamp Forest, Acacia, Oil Palm, Forest Regrowth, Clear cut, Rubber, Water, Others…
  - 1000 pixels are randomly selected (among 1000*1000 pixels)



SVM Classification Result

- **Results with 8 classes with 30 parameters and 300 training pixels per class (randomly selected)**
  - **Mean value for each classifier over a sliding box with delimitation based on WWF map so as to avoid the averaging of a priori mixed class**
- **Overall Accuracy = 84.2%**



Summary

- From the results, it can be found that SVM algorithm with GLCM texture gave better results compared with the other algorithm.
  - Acacia is very difficult to differentiate from peat swamp (depended on growth stage)
- Proposed for next step
- First Level Optical Classifier (WWF)
  - MODIS, Landsat, ASTER, AVNIR-2
- Multi temporal Dual-Pol SAR
  - HV sensitive to standing biomass for different forest types.
  - HH sensitive to both flooding (soil moisture) and land cover type.
  - Dry period will help in differentiation of dry agriculture class