

## KLASIFIKASI HUTAN-NON HUTAN DATA ALOS PALSAR MENGUNAKAN METODE RANDOM FOREST

Katmoko Ari Sambodo<sup>\*)</sup>, Mulia Ina Rahayu<sup>\*)</sup>, Novie Indriasari<sup>\*)</sup>, M.Natsir<sup>\*)</sup>

<sup>\*)</sup>Pusat Teknologi dan Data Penginderaan Jauh, LAPAN

e-mail: katmoko\_ari@lapan.go.id

### Abstract

Extraction of forest and non-forest from remote sensing data can be used in monitoring carbon stocks and stock changes. SAR (Synthetic Aperture Radar) data can be an alternative data source as it is always cloud free. The presence of the cloud is often becomes an obstacle in monitoring using optical data, especially in parts of Indonesia. In this research, extraction of forest and non-forest information from SAR data was done experimentally in two main stages, i.e : 1) Performed land cover classification by grouping objects in the image into nine classes, namely forest, rubber, mangrove trees and shrubs mixed, oil palm and coconut, shrubs, agricultural, baresoil, settlement, and water; 2) Combined the 9 classes into two classes, namely forest (forest, rubber, mangrove trees and shrubs mixed) class and non-forest (palm and coconut, shrubs, agricultural, baresoil, settlement, and water). The first stage is necessary because the position of the class distribution of forests in the HH - HV feature space relatively surrounded by other classes and tend to overlap. Classification process carried out by non parametric Random Forest classifier, i.e, a combination of several tree classifier where each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The data used is ALOS PALSAR mosaic year 2010, ortho and slope corrected, 25 meters resolution dual polarization (HH and HV) around the province of South Sumatra and Jambi. From the experimental result, 95.46 % of accuracy was obtained.

**Key Words :** *Forest-Non Forest, classification, Random Forest (RF), ALOS-PALSAR.*

### Abstrak

Ekstraksi informasi hutan dan non hutan dari data penginderaan jauh dapat digunakan dalam memonitor stok karbon dan perubahannya. Data SAR (*Synthetic Aperture Radar*) dapat menjadi alternatif karena selalu bebas awan dimana keberadaan awan tersebut sering menjadi kendala pada pemantauan menggunakan data-data optis khususnya di wilayah Indonesia. Pada penelitian ini dilakukan eksperimen pengekstraksian informasi hutan dan non hutan dari data SAR dengan melalui dua tahap utama, yaitu : 1) Melakukan klasifikasi penutup lahan dengan mengelompokkan obyek pada citra menjadi 9 kelas, yaitu hutan, karet, *mangrove* & semak campur pohon, kelapa sawit dan kelapa, semak, pertanian, lahan terbuka, permukiman, dan air; 2) Menggabungkan 9 kelas tersebut menjadi 2 kelas yaitu kelas hutan (hutan, karet, *mangrove* & semak campur pohon) dan kelas non hutan (kelapa sawit & kelapa, semak, pertanian, lahan terbuka, permukiman, dan air). Tahap pertama perlu dilakukan karena posisi distribusi kelas hutan dalam ruang fitur HH-HV relatif dikelilingi oleh kelas-kelas lainnya dan cenderung saling overlap. Proses klasifikasi dilakukan dengan metode non parametrik *Random Forest classifier*, yakni kombinasi beberapa *tree classifier* dimana setiap *tree* bergantung pada random vektor yang diambil secara independen dan dengan distribusi yang sama pada setiap *tree* dalam *forest*. Data yang digunakan adalah data mosaik ALOS PALSAR tahun 2010 terkoreksi ortho dan *slope*, resolusi 25 meter *dualpolarization* (HH dan HV) di sekitar wilayah Provinsi Sumatera Selatan dan Provinsi Jambi. Dari hasil eksperimen diperoleh nilai akurasi adalah 95,46 %.

**Kata Kunci :** *Forest-Non Forest, klasifikasi, Random Forest (RF), ALOS-PALSAR.*

## 1. Pendahuluan

Hutan adalah suatu kesatuan ekosistem berupa hamparan lahan berisi sumber daya alam hayati yang didominasi pepohonan dalam persekutuan alam lingkungannya, yang satu dengan lainnya tidak dapat dipisahkan (UU No.41 Tahun. 1999). Menurut UNFCCC, definisi hutan adalah suatu area dengan luas 0.05 – 1 hektar dengan tutupan kanopi minimum 10%-30%, dan tinggi minimum 2-5 meter. Sedangkan pengertian hutan menurut FAO adalah area seluas minimum 0,5 ha, dengan tutupan kanopi minimum 10% (kepadatan kanopi ditentukan dengan mengestimasi bidang tanah yang dinaungi oleh mahkota pohon) dan tinggi pohon minimum 5 meter.

Dengan kemampuan hutan dalam menyerap dan menyimpan karbon maka hutan memiliki peran penting dalam hal perubahan iklim. Pemantauan terhadap hutan perlu dilakukan untuk memelihara hutan

dari ancaman dalam bentuk degradasi dan deforestasi. Kegiatan pemantauan hutan ini dapat dilakukan dengan menggunakan data penginderaan jauh, sehingga dapat dilakukan dengan waktu yang relatif cepat dan efisien. Ekstraksi informasi hutan dan non hutan dari data penginderaan jauh dapat digunakan dalam memonitor stok karbon dan perubahannya.

Definisi hutan sebenarnya masih diperdebatkan dalam beberapa penelitian, terutama untuk jenis tanaman kelapa sawit. Ada yang berpendapat bahwa tanaman kelapa sawit dimasukkan ke dalam kelas hutan, dan ada yang berpendapat sebaliknya. Dalam penelitian ini, tanaman kelapa sawit dimasukkan ke dalam kelas non hutan karena meskipun kebun sawit menyerupai tanaman hutan dalam kaitannya dengan fisiognomi (tinggi dan penutupan tajuk), sawit sesungguhnya adalah tanaman pertanian, khususnya ditinjau dari siklus waktunya, jenis dan intensitas pengelolaannya (Dijk dan Savenije, 2011).

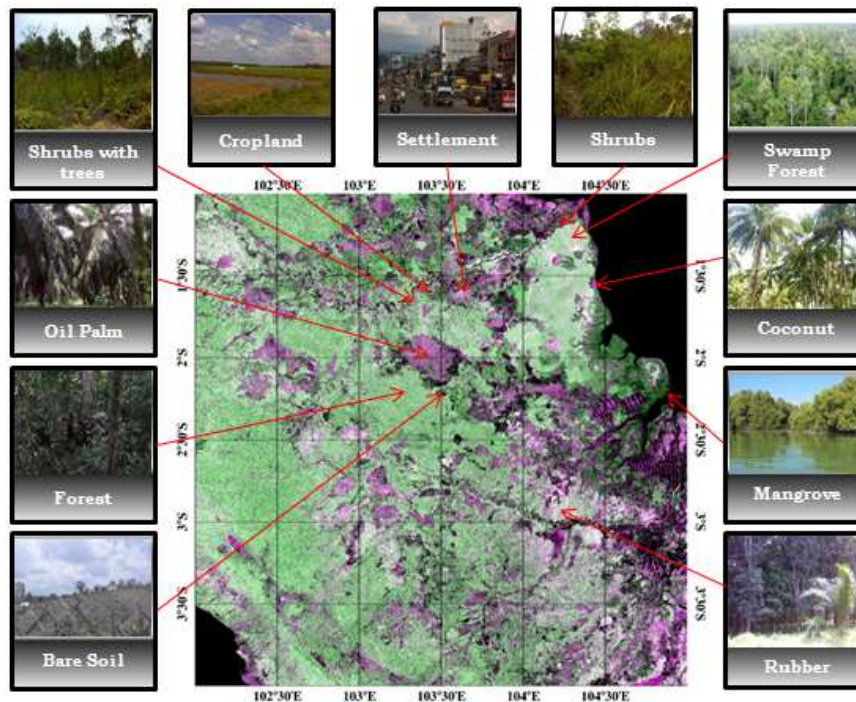
Citra penginderaan jauh dengan sensor gelombang mikro aktif memiliki sumber energi sendiri. Rentang panjang gelombang yang lebih lebar memungkinkan penetrasi kondisi atmosfer seperti hujan es, kabut, kabut, asap, curah hujan, dan awan (Haack *et al.*, 2000). Data SAR (*Synthetic Aperture Radar*) dapat menjadi alternatif karena selalu bebas awan dimana keberadaan awan tersebut sering menjadi kendala pada pemantauan menggunakan data-data optis khususnya di wilayah Indonesia.

Untuk membedakan area hutan dan non hutan dari data penginderaan jauh secara efisien dapat dilakukan dengan klasifikasi secara digital. Secara garis besar ada 2 jenis metode klasifikasi, yaitu metode parametrik seperti Maksimum Likelihood, dan metode non-parametrik seperti *neural network*, *support vector machine*, dan *decision tree*. *Non-parametrik classifier* telah menjadi pendekatan yang penting untuk mengklasifikasi data-data yang berasal dari berbagai sumber (Lu and Weng, 2007). Salah satunya adalah metode klasifikasi *Random Forest* (RF) yang merupakan pengembangan dari metode *single Decision Tree*. Metode RF terdiri dari beberapa *tree* dimana di setiap *tree* dilakukan training terhadap sampel data. Penentuan kelas dilakukan berdasarkan jumlah vote dari setiap *tree*. Metode *ensemble learning* memiliki tingkat ketelitian yang lebih tinggi dibandingkan dengan algoritma *machine learning* lainnya karena kumpulan beberapa *classifier* memberikan hasil yang lebih akurat daripada *single classifier* (Ghimire *et al.*, 2010; Akar and GÜngör, 2012). Tujuan dari penelitian ini adalah melakukan ekstraksi informasi hutan dan non hutan terhadap data ALOS PALSAR tahun 2010, resolusi 25 meter, polarisasi HH+HV, dengan menggunakan algoritma non-parametrik *Random Forest*.

## 2. Data dan Metode

### 2.1 Data

Data SAR yang digunakan dalam penelitian ini adalah mosaik citra ALOS-PALSAR, L-band, resolusi 25 meter, dual polarisasi (HH+HV). Wilayah penelitian di sekitar Provinsi Sumatera Selatan dan Provinsi Jambi. Data ALOS-PALSAR yang digunakan adalah hasil perekaman tahun 2010 dan telah melalui proses orthorektifikasi dan koreksi slope yang dilakukan oleh JAXA-EORC (*Japan Aerospace Exploration Agency – Earth Observation Research Center*).



Gambar 2-1. Mosaik citra ALOS PALSAR dan informasi survey lapangan

Gambar 2-1. menampilkan cakupan data dan informasi hasil survei lapangan. Dari hasil survei lapangan yang dilakukan pada bulan Maret 2013, terdapat 12 jenis tutupan lahan yang dapat diidentifikasi dari data ALOS-PALSAR, yaitu hutan, hutan rawa, akasia, karet, *mangrove*, belukar, kelapa sawit, kelapa, lahan pertanian, lahan terbuka, permukiman, dan air.

## 2.2 Metode

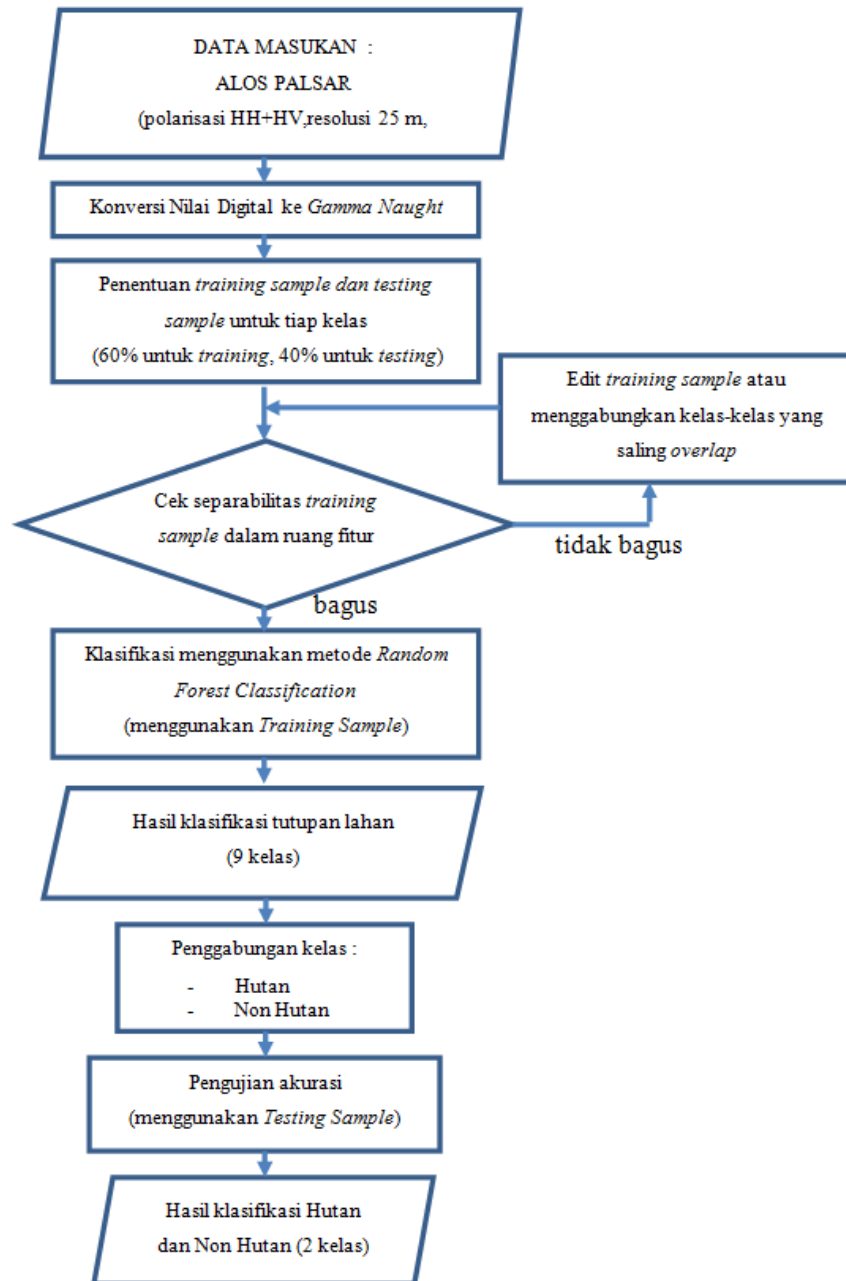
Diagram alir penelitian ditampilkan dalam Gambar 2-2. Pengolahan tahap awal dilakukan dengan mengkonversi nilai digital data ALOS PALSAR ke nilai *Gamma Naught* dalam satuan desibel.

Sebelum melakukan klasifikasi hutan dan non hutan, terlebih dahulu dilakukan analisa nilai digital terhadap training sampel untuk setiap kelas yang diperoleh dari hasil survei lapangan. Berdasarkan analisa tersebut, terutama untuk kelas hutan, distribusi datanya relatif dikelilingi oleh kelas-kelas lainnya dan bahkan saling *overlap* satu sama lain. Hal ini pada umumnya akan menjadi kendala bagi *classifier* untuk menentukan batas kelas yang optimal bagi kelas tersebut dengan kelas di sekelilingnya. Oleh karena itu untuk membantu *classifier* dalam menentukan batas kelas yang optimal dan meningkatkan akurasi klasifikasi, perlu dilakukan klasifikasi dengan jumlah kelas tutupan lahan yang memungkinkan berdasarkan hasil survei terlebih dahulu semaksimal mungkin. Untuk tujuan tersebut, dilakukan pemilihan training sampel masing-masing kelas tersebut yang dilanjutkan dengan pengecekan separabilitas antar kelas dalam ruang fitur HH-HV. Proses ini dilakukan secara iterasi. Jika terdapat dua kelas atau lebih yang saling *overlap*, maka kelas-kelas tersebut akan dikelompokkan menjadi satu kelas. Setelah pengecekan separabilitas tersebut selesai, barulah kemudian dilanjutkan pada proses klasifikasi dengan menggunakan metode *Random Forest* yang akan dijelaskan pada sub-bab berikut.

Setelah hasil klasifikasi sejumlah kelas tutupan lahan selesai, selanjutnya dilakukan proses penggabungan kelas. Pada penggabungan kelas tersebut digunakan 2 aturan, yaitu :

- Kelas hutan yang terdiri dari kelas hutan, karet, *mangrove* & belukar campur pohon
- Kelas non hutan yang terdiri dari kelas kelapa sawit & kelapa, belukar, pertanian, lahan terbuka, permukiman, dan air

Pada bagian akhir dilakukan pengujian akurasi hasil klasifikasi dengan menggunakan *Confusion Matrix*.

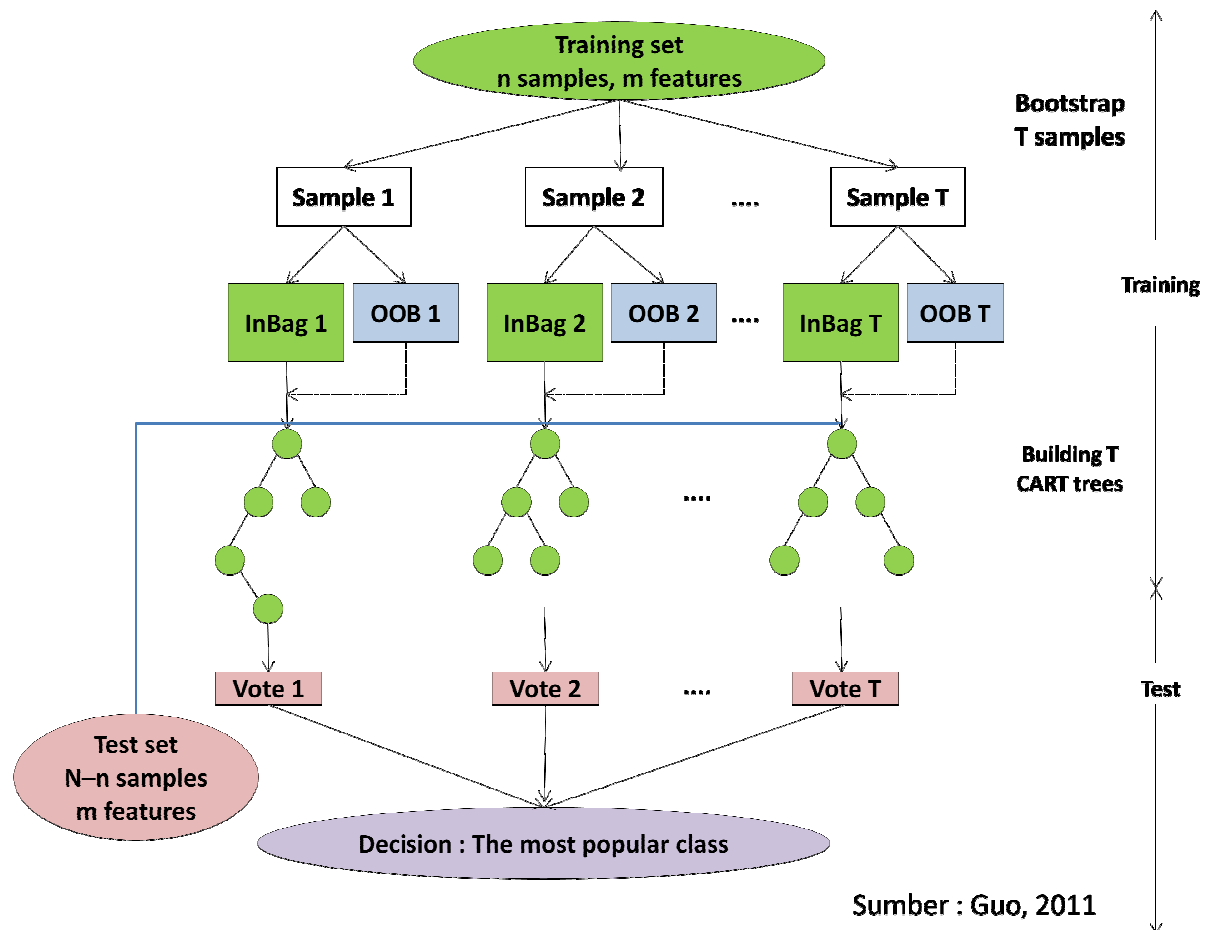


Gambar 2-2. Diagram alir klasifikasi hutan dan non hutan

### 2.3 Klasifikasi Random Forest

Proses klasifikasi dilakukan dengan metode non-parametrik *Random Forest classifier*. *Random Forest* merupakan suatu kumpulan dari beberapa *tree*, dimana masing-masing *tree* bergantung pada nilai piksel pada tiap vektor yang diambil secara acak dan independen (Breiman, 2001). *Random forest* tidak berkecenderungan untuk *overfit* dan dapat melakukan proses dengan cepat, sehingga memungkinkan untuk memproses *tree* sebanyak yang diinginkan oleh pengguna (Breiman and Cutler, 2005).

Gambar 2-3 menampilkan alur kerja algoritma *Random Forest*. Dalam pembentukan *tree*, algoritma *Random Forest* akan melakukan training terhadap sampel data. Pengambilan sampel dilakukan dengan cara *sampling with replacement*. Sebanyak sepertiga dari sampel akan digunakan untuk menentukan data *out of bag* (OOB). Penentuan data *out of bag* dilakukan untuk mengestimasi *error* dan menentukan *variable importance*. Variabel yang akan digunakan untuk menentukan pemisahan (*split*) terbaik ditentukan secara acak. Setelah seluruh *tree* terbentuk, maka proses klasifikasi akan berjalan. Penentuan kelas dilakukan dengan cara *voting* dari masing-masing *tree*, kelas dengan jumlah *vote* terbanyak akan menjadi pemenangnya.



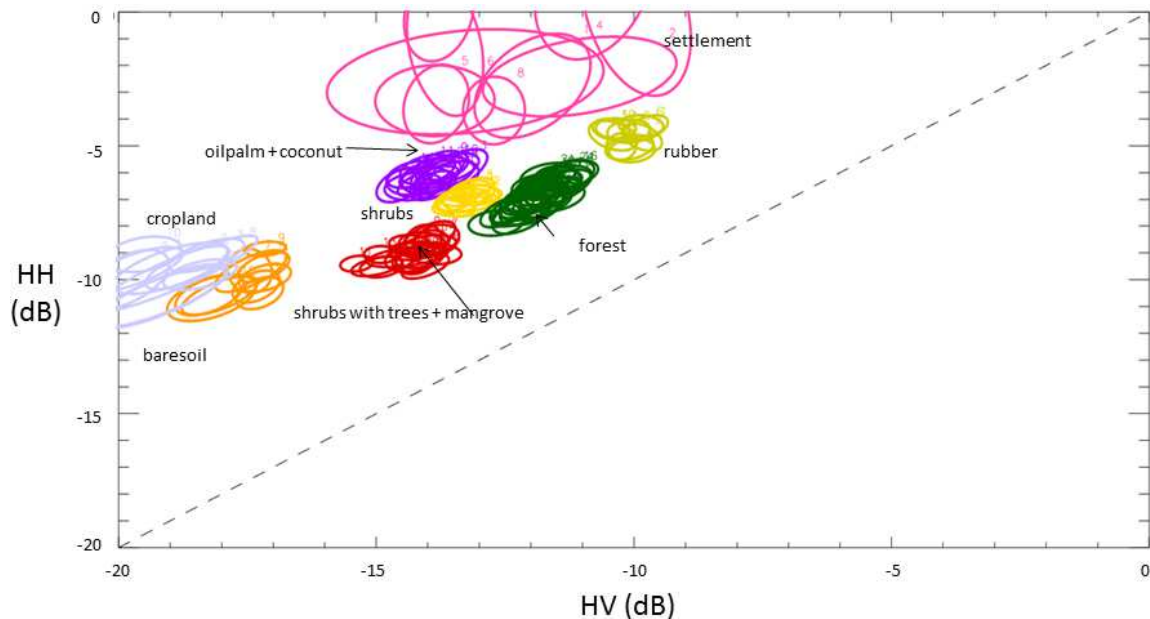
Gambar 2-3. Proses Klasifikasi *Random Forest* (Sumber: Guo, 2011)

### 3. Hasil dan Pembahasan

Berdasarkan hasil dari survei lapangan terdapat 12 jenis tutupan lahan yang dapat diidentifikasi pada citra ALOS-PALSAR, yaitu hutan, hutan rawa, akasia, karet, *mangrove*, belukar, kelapa sawit, kelapa, lahan pertanian, lahan terbuka, permukiman, dan air. Akan tetapi dari hasil pengecekan separabilitas *training sample* dalam ruang fitur HH-HV, ada beberapa kelas yang saling tumpang tindih dan sulit untuk dipisahkan. Oleh karena itu dilakukan penggabungan kelas dari 12 kelas menjadi 9 kelas.

Gambar 3-1 menunjukkan hasil pengecekan separabilitas *training sample* dalam ruang fitur HH-HV. Dapat dilihat bahwa setelah penggabungan terdapat 9 kelas tutupan lahan, yaitu hutan, karet, *mangrove*+belukar campur pohon, kelapa sawit+kelapa, belukar, lahan terbuka, lahan pertanian, permukiman, dan air. Selanjutnya dilakukan proses klasifikasi dengan 9 kelas tutupan lahan. Pembahasan lebih lanjut mengenai klasifikasi dengan 9 kelas tutupan lahan ini dapat dilihat dalam Rahayu dan Sambodo, 2013.

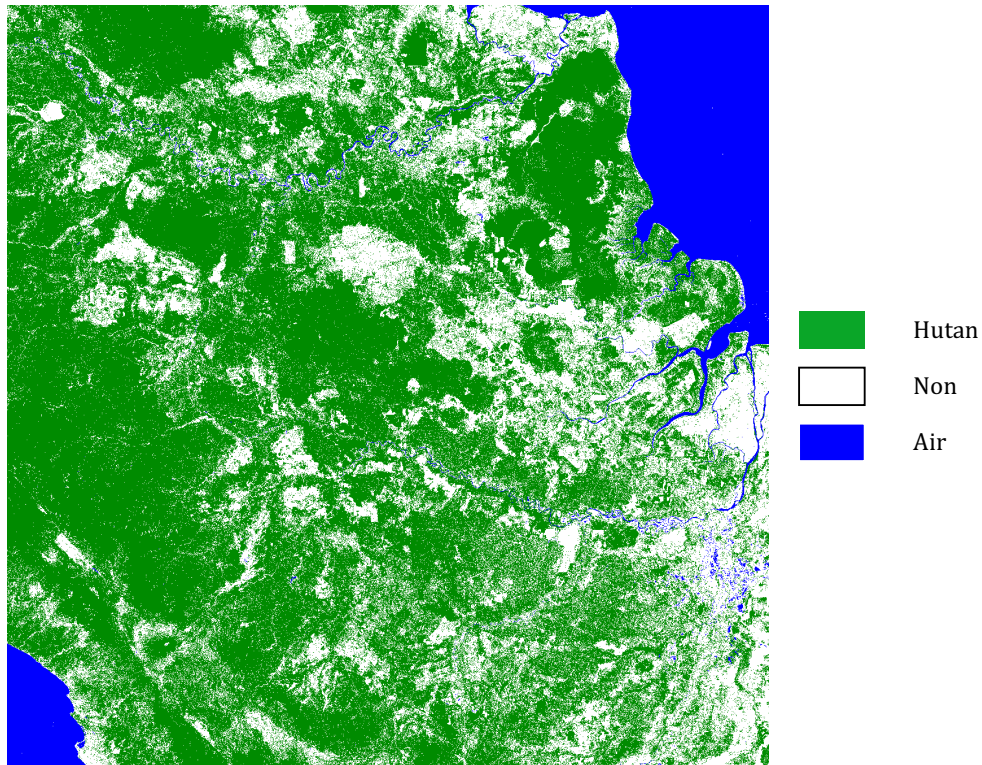
Dari 9 kelas tutupan lahan tersebut akan dikelompokkan lagi menjadi 2 kelas yaitu kelas hutan dan non hutan. Kelas karet, hutan, dan *mangrove*+belukar campur pohon dikelompokkan menjadi kelas hutan, sedangkan kelas kelapa sawit+kelapa, belukar, permukiman, lahan terbuka, lahan pertanian, dan air dikelompokkan menjadi kelas non hutan. Setelah itu dilakukan pengujian hasil klasifikasi.



Gambar 3-1. Separabilitas training sample dalam ruang fitur HH-HV

Hasil klasifikasi dengan metode *Random Forest* dapat dilihat dalam Gambar 3-2, sedangkan *confusion matrix* ditampilkan dalam Tabel 3-1. Untuk kelas hutan diperoleh nilai *producer accuracy* sebesar 96,26 % dan kelas non hutan diperoleh nilai *producer accuracy* sebesar 93,28%. Sedangkan *overall accuracy* yang didapatkan adalah 95.46% (*Kappa Coefficient* 0.9207). Kesalahan (*error*) hasil klasifikasi masih banyak terjadi terutama yang bersifat bintang-bintang (*salt and pepper*) meskipun pada

area yang homogen. Hal ini disebabkan karakteristik dasar data SAR yang mengandung *noise speckle* yang kemudian akan berakibat hasil klasifikasinya juga cenderung berbintik-bintik apabila dilakukan dengan *classifier* berbasis piksel seperti halnya *Random Forest* ini. Oleh karenanya untuk mereduksi efek tersebut, perlu dilakukan tambahan informasi spasial kontekstual dari piksel-piksel sekelilingnya.



Gambar 3-2. Hasil klasifikasi Hutan dan Non Hutan dengan metode *Random Forest*

Tabel 3-1. *Confusion Matrix* hasil klasifikasi Hutan dan Non Hutan

<b>Data Referensi</b>	<b>Hasil klasifikasi</b>		
	Hutan	Non Hutan	<i>User's accuracy</i>
Hutan	8292	431	95.06
Non Hutan	322	5980	94.89
<i>Producer's accuracy</i>	96.26	93.28	

*Overall accuracy*: 95.46% *Kappa Coefficient*: 0.9207

#### 4. Kesimpulan

Dalam penelitian ini, klasifikasi hutan dan non hutan data ALOS PALSAR resolusi 25 meter dengan menggunakan metode *Random Forest* memberikan hasil yang cukup memuaskan dengan nilai *producer accuracy* untuk kelas hutan sebesar 96,26% dan untuk kelas non hutan sebesar 93,28%. Nilai *overall accuracy* yang diperoleh adalah sebesar 95,46% (nilai *Kappa* = 0.9207). Untuk lebih meningkatkan kualitas hasil klasifikasi, maka dalam penelitian selanjutnya akan dilakukan klasifikasi dengan menambahkan informasi spasial kontekstual seperti dari *image texture* atau metode lainnya.

#### 5. Ucapan Terima Kasih

Ucapan terimakasih disampaikan kepada JAXA yang telah menyediakan mosaik data ALOS PALSAR resolusi 25 meter dalam kerangka kerjasama JAXA Kyoto & Carbon Initiative.

#### 6. Daftar Rujukan

- Akar and Güngör, 2012. Classification of multispectral images using Random Forest algorithm. *Journal of Geodesy and Geoinformation* 1 (2), 105 – 112.
- Breiman, 2001. Random Forests. *Machine Learning* 45, 5 – 32.
- Breiman L., Cutler A., 2005. Random Forests. [http://www.stst.berkeley.edu/~breiman/RandomForests/cc\\_home.htm](http://www.stst.berkeley.edu/~breiman/RandomForests/cc_home.htm), diakses tanggal 11 Juli 2013.
- Departemen Kehutanan dan Perkebunan. 1999. Undang-undang Republik Indonesia Nomor 41 Tahun 1999 tentang Kehutanan. Dephutbun RI. Jakarta.
- Dijk, Kees van dan Savenije, Herman., 2011. *Kelapa sawit atau hutan? Lebih dari sekedar definisi*. Policy Brief. Tropenbos International Indonesia Programme. iv + 20.
- FAO. 2007. Definitional issues related to reducing emissions from deforestation in developing countries. *Forests and Climate Change Working Paper 5*. <ftp://ftp.fao.org/docrep/fao/009/j9345e/j9345e00.pdf>. Diakses tanggal 20 Maret 2014
- Ghimire et al., 2010. Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. *Remote Sensing Letters* 1 (1), 45 – 54.
- Haack et al., 2000. Radar and optical data integration for land-use/land-cover mapping. *Photogrammetric Engineering & Remote Sensing* 66 (6), 709 – 716.
- Lu and Weng, 2007. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing* 28 (5), 823 – 870.
- Rahayu dan Sambodo. 2013. Random Forest Classification of Jambi and South Sumatera using ALOS PALSAR Data. *34th Asian Conference on Remote Sensing*, Bali.
- UNFCCC (2001), Report of the Conference of the Parties on its seventh session. [http://unfccc.int/cop7/documents/accords\\_draft.pdf](http://unfccc.int/cop7/documents/accords_draft.pdf). Diakses tanggal 20 Maret 2014.