Available online at www.sciencedirect.com
ScienceDirect
Procedia Environmental Sciences 24 (2015) 215 – 221
Procedia Environmental Sciences

The 1st International Symposium on LAPAN-IPB Satellite for Food Security and Environmental Monitoring

Random forest classification for mangrove land cover mapping using Landsat 5 TM and ALOS PALSAR imageries

Romie Jhonnerie[a,b], Vincentius P. Siregar[c,]*, Bisman Nababan[c], Lilik Budi Prasetyo[d], and Sam Wouthuyzen[e]

[a]*Marine Technology Graduate Program, Faculty of Fisheries and Marine Science, Bogor Agricultural University, Bogor, Indonesia*
[b]*Departmen of Aquatic Resources Utilization, Faculty of Fisheries and Marine Science, Riau University. Indonesia*
[c]*Department of Marine Science and Technology, Faculty of Fisheries and Marine Science, Bogor Agricultural University, Bogor, Indonesia*
[d]*Departmen of Forest Resources, Conservation and Ecotourism, Faculty of Forestry, Bogor Agricultural University, Bogor, Indonesia*
[e]*Research Center for Oceanography, Indonesian Institute of Science, Jl. Pasir Putih I Ancol Timur, Jakarta, Indonesia*

Abstract

The objective of this research was to evaluate the accuracy of random forest classification rule using object based image analysis (OBIA) application (eCognition Developer) and the results were compared with common pixel-based classification algorithm (maximum likelihood/ML) for mangrove land cover mapping in Kembung River, Bengkalis Island, Indonesia. Seven data input model derived from Landsat 5TM bands, ALOS PALSAR FBD, and spectral transformations (NDVI, NDWI, NDBI) were examined by both classifiers. Feature objects statistical parameters were selected and implemented on random forest classifier. Overall accuracy (OA) as well as user and producer accuracies and Kappa statistic were used to compare classification results. Our results showed that the more data model used produced higher overall accuracy and kappa statistics for RF classifier. For each data input model, random forest classifier has higher overall accuracy than maximum likelihood. The best mangrove discrimination in RF classifier was achieved when the combination of Landsat 5 TM, SAR, and spectral transformation were used, while in ML classifier, the best mangrove discrimination was achieved when the combination of Landsat 5 TM and ALOS PALSAR was used. The overall accuracy achieved by RF classifier was 81.1% and 0.76 for Kappa statistic. Meanwhile, for ML classifier, the overall accuracy achieved was 77.7% and 0.71 for Kappa statistic.

Selection and peer-review under responsibility of the LISAT-FSEM Symposium Committee

Keywords: Random forest classification; object based image analysis; pixel based classification; Landsat TM; ALOS PALSAR; mangrove; Kembung River; Bengkalis island

* Corresponding author. Tel.: +62-251-862-2909; fax: +62-251-862-2907.
E-mail address: vincents@biotrop.org

Selection and peer-review under responsibility of the LISAT-FSEM Symposium Committee
doi:10.1016/j.proenv.2015.03.028

1. Introduction

Mangrove is a fragile and vulnerable ecosystem from natural disturbances such as hurricanes, tropical storms, and tsunamis that could change its shape and size. People could also change its function and role to agriculture, farms, settlements, and other forms. Currently, the mangrove ecosystem in Kembung River as a buffer region and habitat for many types of biota has been degraded by both human and nature [1]. In order to manage and maintain the mangrove ecosystem, it needs to take several actions such as to determine the coverage area of mangrove ecosystem. Accurate and up to date maps of mangrove are needed to observe and evaluate the existence of mangrove ecosystems by stakeholders such as local communities and the government. By having an accurate and up to date mangrove map, the stakeholders can use it to monitor and develop policies that favor the preservation and sustainability of the mangrove ecosystem.

Currently non-parametric classification algorithm has been developed rapidly and has been widely used in a variety of fields such as in medical and ecological fields [2]. One of the classifications is the Random Forest classification (RF) [3]. The use of RF has become popular in the field of remote sensing for land cover mapping [4, 5]. Rodriguez-Galiano *et al*. [6] stated that RF has several advantages including: the nature of non-parametric algorithm, a high classification accuracy, and has ability to determine important variables and able to predict the missing values. However, the use of the RF classification to map the mangrove coverage is still scarce.

In this study, we used seven data input models derived from Landsat 5 Thematic Mapper (TM), Advanced Land Observation Satellite Phase Array L-band Synthetic Aperture Radar Fine Beam Double Polarization (ALOS PALSAR FBD), and spectral transformations to map mangrove land cover in Kembung river, Bengkalis Island using object based image analysis (OBIA) application. Furthermore, we then test and compare the classification accuracy with commonly used classification algorithm (maximum likelihood).

2. Methodology

2.1. Data

The field study was conducted in June-December 2012 in mangrove ecosystems of Kembung river, Bengkalis Island, Riau Province, Indonesia (Fig. 1). Mangrove communities of Sungai Kembung were composed by 68 species of mangrove vegetation which was consisted of 22 species of true mangrove and 46 species of associate mangrove.

Satellite imageries used in this research consisted of: (1) Landsat 5 TM Level 1T recorded on February 2nd, 2010 with path/row 126/59. The data was obtained from USGS GLOVIS; and (2) L-Band SAR ALOS PALSAR Fine Beam Double Polarizations (FBD) Level 1.1 acquisition on September 19th, 2010. The data were obtained from JAXA.

Due to limited access, field observations were only conducted around 500 m from mangrove edges. We took 447 observation points made randomly using ArcGIS Desktop in which156 of them were used in training area and 291 were used for accuracy test of mapping result for each land cover. Land cover classification scheme referred to [7] and consisted of 8 land cover such as: coconut tree, rubber tree, bush, mangrove, mangrove transition vegetation, bare land, built-up area, and water bodies.

The method proposed in the study consisted of four steps: (i) Data preparation, including pre-processing of Landsat 5 TM, such as atmospheric correction [8], ALOS PALSAR backscattering calibration [9], noise reduction [10], spectral transformation of Landsat 5 TM data, NDVI [11], NDWI [12], and NDBI [13]. We composed the source data into seven input data model for RF classification, *i.e.*: (I1) SAR data (HH+HV); (I2) Spectral transformation (NDVI+NDWI+NDBI); (I3) SAR+spectral transformation; (I4) Landsat 5 TM (Blue+Green+Red+NIR+MIR+FIR); (I5) Landsat 5 TM+SAR; [I6] Landsat 5 TM+Spectral transformation and (I7) Landsat 5 TM+SAR+Spectral transformation; (ii) Segmentation optimization. There was no such fix threshold value due to segmentation process. To determine optimal segmentation parameter value in both segmentation algorithm (multiresolusion segmentation (MRS) and spectral different algorithm (SDA)) was crucial. In this research examined object scale value as optimum parameters is described in Table 1. (iii) RF parameters optimization and classification. RF is assembler machine learning and it is very efficient in handling big data and qualified to outlier and over fitting. The use of RF on remote sensing application has been found in many studies.

RF is a combination of a number of non-parametric classification and decision tree/CART (classification and regression trees). Decision tree is similar with hierarchy, composed of root node, including all samples, node separator which has decision rules, and the end of the leaf node, which represents desired classes. RF has several parameters that can be examined. To obtain the optimum parameter values, we examined several RF parameters, such as: maximum total of tree depth (depth), minimum number of sample per node (sample), and maximum tree number (tree number) (Table 2). The assessment was tested to I7 data input model. The optimum values determined based on accuracy assessment (overall accuracy) then the optimum parameters values implemented to other data input models. (iv) classification validation using accuracy assessment [14]. As a comparison, we classified seven input model by using ML classifier. For a fairly comparison, ML was used the same training area as well as RF. Accuracy assessment used matrix error (contingency matrix) by measuring, producer, user, overall accuracy and Kappa statistic [15]. Accuracy validation was also supported by SPOT 6 multispectral and panchromatic imageries.

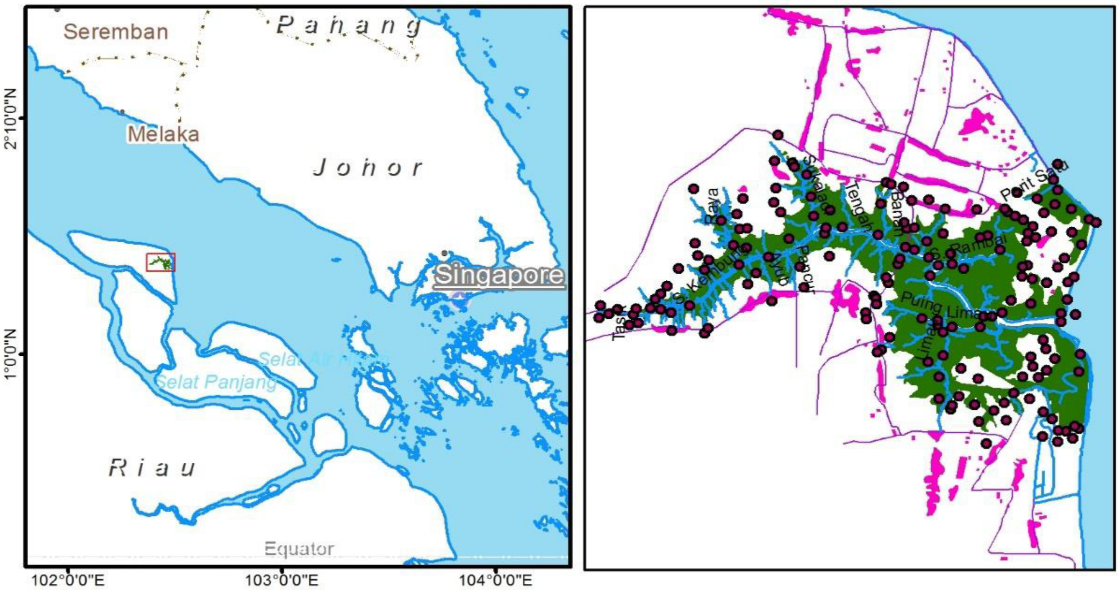

Fig. 1. Sampling sites in Kembung river, Bengkalis Island, Indonesia.

Table 1. Scale parameter values which tested by segmentation algorithm

No	MRS scale	SDA scale
1	1	0.1
2	5	0.5
3	10	1
4	20	5

Table 2. RF parameters and values used to obtain the optimum value of the classification result

No	Parameters	Values
1	Depth	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 1000
2	Minimum number of sample per node	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 1000
3	Maximum number of trees in the forest	50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000

For optic data preprocessing and ML classification, we used ERDAS IMAGINE application. Meanwhile, SAR data was preprocessed by ASF MapReady. Segmentations and RF parameters optimization were examined by using eCognition Developer. Finally, data visualization was performed using ArcGIS suite application.

3. Results and Discussion

Two segmentation algorithms were used in producing objects and the objects size defined by the scale used. The bigger scale used by the algorithms, the larger object produced (Fig. 2). Object scale value obviously contributed to the accuracy of classification result. Accuracy produce by using MRS object scale was about 59.0% - 77.7%. Best accuracy was produced by scale of 1. Additional algorithm SDA did not increase the accuracy result but it was succeeded in decreasing the total object. Best accuracy (77.7%) was produced by scale of 0.1 with total numbers of object 71,626. The result of segmentation algorithm testing MRS and SDA is described in Table 3.

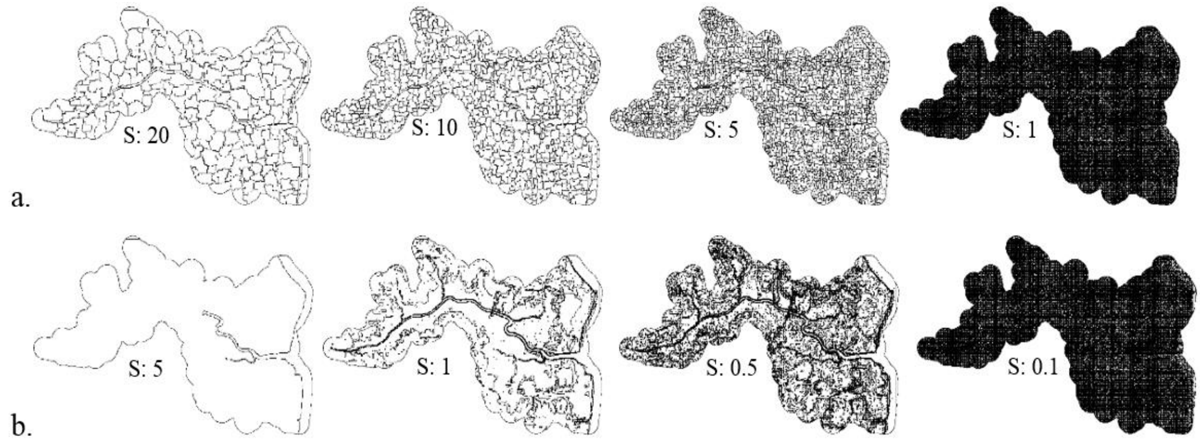

Fig. 2. Segmentation result by using MRS (a) and SDA (b). S indicated scale parameter value used

Table 3. Objects number and properties produced by MRS and SDA segmentation algorithms based on scale parameters tested. The bold values indicated the best result for each algorithm. OA = overall accuracy, K = Kappa statistic

Segmentation	Scale	Number of objects	Area (m^2)			OA	K
			Mean	Minimum	Maximum		
MRS	**1**	**73981**	**965.97**	**0.001**	**6,875.00**	**77.7%**	**0.71**
	5	3765	18,981.00	0.014	161,796.88	76.6%	0.70
	10	1069	66,850.78	0.140	429,218.75	74.0%	0.66
	20	308	232,024.28	0.140	1,219,687.50	59.0%	0.49
SDA	**0.1**	**71626**	**997.73**	**0.001**	**118,804.90**	**77.7%**	**0.71**
	0.5	13023	5,487.48	0.001	3,402,187.50	75.8%	0.69
	1	2795	25,568.33	0.001	12,304,921.88	67.4%	0.57
	5	3	23,821,159.76	118,503.01	65,962,167.52	15.8%	0.04

Despite of using similar segmentation parameter, the total object and size produced by segmentation of each image layer input combinations were different. They produced 6,064 to 125.230 objects. The maximum number object created was produced by input I1 and the minimum was produced by input I2. Tabulation of information number of object generated from segmentation process is described in Table 4.

Table 4. Objects area attribute, numbers, minimum, maximum and mean, produced by each input model

Input	Numbers of object	Area (m^2)		
		Minimum	Maximum	Mean
I1	125,230	0	27,734.40	570.7
I2	6,064	11.561	830,724.90	11,784.90
I3	51,357	0.001	128,686.20	1,391.50
I4	32,433	0.083	465,342.90	2,203.40
I5	98,119	0.001	30,680.20	728.3
I6	17,640	0.083	841,500.00	4,051.20
I7	71,626	0.001	118,804.90	997.7

The use of default depth parameter values (0), the RF produced 77.7% overall accuracy. Meanwhile, changing depth parameter to be 1, the RF turned down the overall accuracy value into 17.9% (Fig. 3). The increase of accuracy appeared along with the increase of depth parameter value used, and tended to be stable (78.0%) when the depth parameter value reached 9 to 1000. Parameter of total sample used in each node gain average accuracy for 77.3% (the sample parameter value used 1 – 10). Higher accuracy obtained when sample value was 7 (80.6%) and the lowest on was 10 (71.1%). Accuracy tended to be lower when parameter sample used of 100 to 1000. Meanwhile, the increase of total trees parameter used tended to produce a constant accuracy. Average accuracy and best accuracy of 80.1% and 80.6%, respectively were obtained by using 300 trees. (Fig. 3).

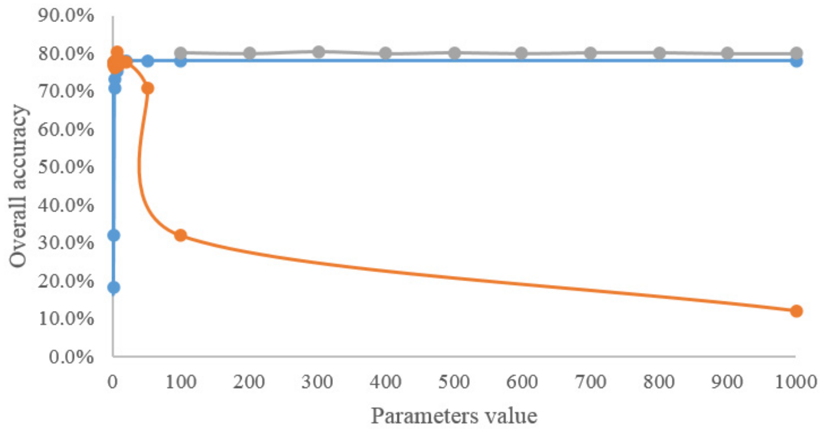

Fig. 3. The influences of RF parameters to the accuracy assessment

Three parameters such as depth, sample, and tree produced optimum RF overall accuracy classification for the values of 10, 7, and 300, respectively. Classification result by RF and ML algorithms showed on Fig 4. The overall accuracy produced by the above parameters was 81.1% and Kappa statistic was 0.71. The best input model in land cover mapping of RF algorithm was achieved by using all input data (I7) within overall accuracy of 81.7% and Kappa statistics of 0.76. Meanwhile, the best result in employing ML algorithm was achieved by using Landsat 5TM and SAR data (I5) input models within overall accuracy of 76.8% and Kappa statistics of 0.71 (Table 5).

Although mangrove land cover mapping can be performed better through RF algorithm, some problems are still found such as object misclassification. Object misclassification was normally found in the form of transition, water body, and bare land, with commission interval 0.03%-25.3% and ommision interval 9.6%-37.7%. Fatoyinbo *et al.* [16] found the same problem in mapping mangrove land cover. Heumann [17] which employed hybrid classification technique found misclassification object in transition class incorporation (association) with mangrove class and possible for other classes. Other problem such as time difference of field observation with imagery satellite recording can produce object interpretation [18]. These problems always become obstacles in remote sensing research, moreover, using optic data also affected by atmospheric condition. Several times, land cover classes experienced changes to another land cover classes. This problems could be resolved by using several series of older and newer imagery satellites, to determine the consistency of observed land cover change for observation of valid field data.

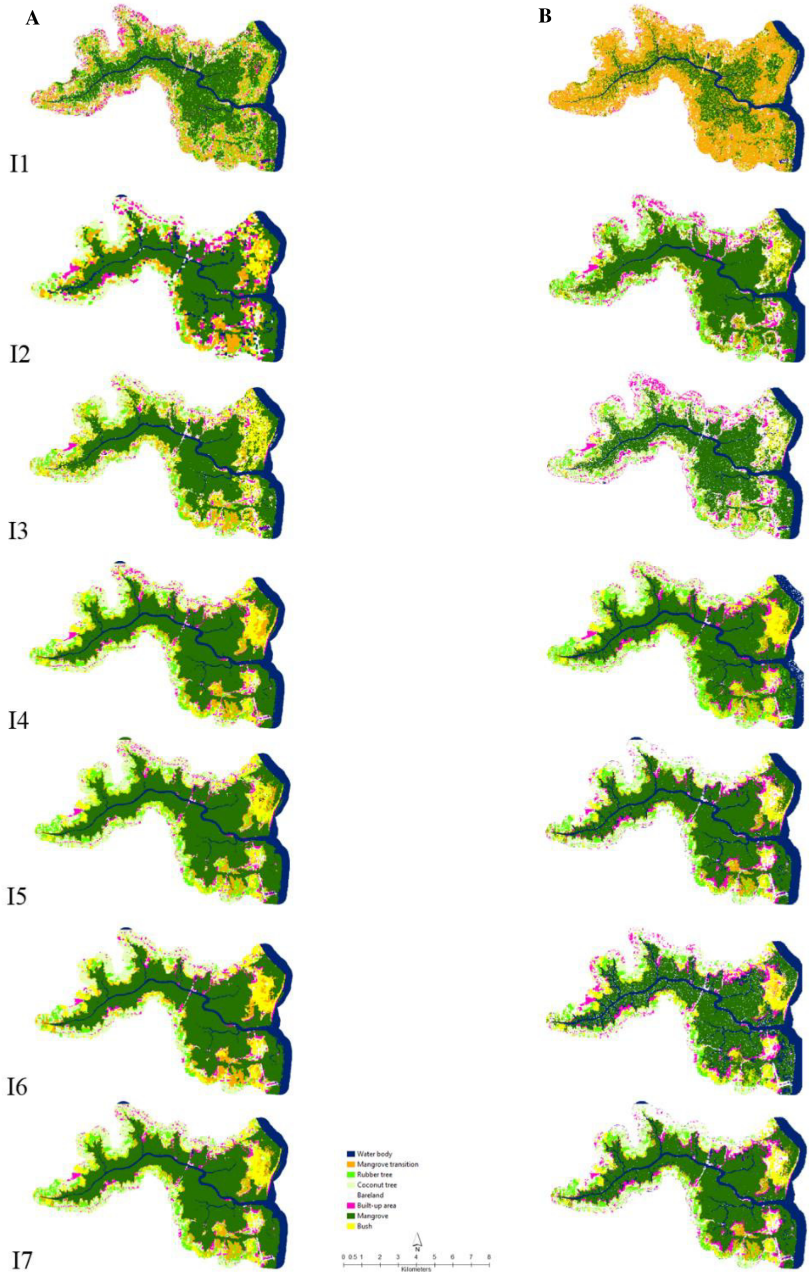

Fig. 4. Classification result (A) Random forest algorithm, (B) Maximum likelihood

Table 5. Mangrove class accuracy (user accuracy (UA) and producer accuracy (PA) and overall accuracy (OA) and Kappa Statistic (K) calculated from RF and ML algorithms (values in bold indicated the best results for both algorithms)

Input Model	Random Forest				Maximum Likelihood			
	UA	PA	OA	K	UA	PA	OA	K
I1 (SAR Data)	74.7%	62.3%	45.4%	0.321	85.0%	44.7%	35.4%	0.240
I2 (Spectral Transformation)	92.4%	74.6%	61.8%	0.527	80.2%	81.6%	58.2%	0.464
I3 (SAR + Spectral Transformation)	89.1%	71.9%	63.6%	0.549	92.4%	74.6%	62.9%	0.536
I4 (L-5 TM all bands)	97.0%	84.2%	76.8%	0.710	93.5%	87.7%	76.8%	0.707
I5 (L-5 TM + SAR)	92.8%	90.4%	76.1%	0.696	**98.0%**	**84.2%**	**76.8%**	**0.710**
I6 (L-5 TM + Spectral transformation)	94.4%	88.6%	76.1%	0.711	90.9%	78.9%	76.8%	0.642
I7 (L-5 TM + SAR + Spectral transformation)	**95.4%**	**90.4%**	**81.1%**	**0.760**	95.9%	82.5%	73.9%	0.673

4. Conclusion

This study explained the use of some data input combination to map mangrove land cover, derived from Landsat 5 TM, ALOS PALSAR, and spectral information implemented by RF algorithm and then compared to ML algorithm. Various accuracy results were obtained from each layer input combination. The best result of RF algorithm was achieved by all layer input combination. Meanwhile, the best result for ML algorithm was obtained by combination of Landsat 5 TM and ALOS PALSAR. The RF Algorithm could map mangrove land cover better and could also reduce noise existence in classification result compared to ML algorithm. Adding total layer input in RF algorithm tended to increase the classification result, but not in ML algorithm.

References

1. Jhonnerie R, Siregar VP, Nababan B, Praseto BP, Wouthuizen S. Deteksi perubahan tutupan mangrove menggunakan citra Landsat berdasarkan klasifikasi hibrida di Sungai Kembung, Pulau Bengkalis, Provinsi Riau. *J ITKT* 2014; **6**: 491-506.
2. Pantaleoni E, Wynne RH, Galbraith JM, Campbell JB. Mapping wetlands using ASTER data: a comparison between classification trees and logistic regression. *Int J Remote Sens* 2009;**30**: 3423-40.
3. Breiman, L. Random forests. *Mach Learn* 2001;**45**: 5-32.
4. Rodriguez-Galiano VF, Chica-Olmo M, Abarca-Hernandez F, Atkinson PM, Jeganathan C. Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sens Environ* 2012;**121**: 93-107.
5. Corcoran J, Knight J, Gallant A. Influence of multi-source and multi-temporal remotely sensed and ancillary data on the accuracy of random forest classification of wetlands in Northern Minnesota. *Remote Sens* 2013;**5**: 3212-38.
6. Rodriguez-Galiano, VF, Ghimire B, Rogan J, Chica-Olmo, M, Rigol-Sanchez JP. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J Photogram Remote Sens* 2012;**67**: 93-104.
7. Badan Standar Nasional. Standar Nasional Indonesia. Klasifikasi Penutup Lahan. Jakarta; 2010.
8. Richter RA. Spatially adaptive fast atmospheric correction algorithm. *Int J Remote Sens* 1996;**17**: 1201-14.
9. Shimada M, Isoguchi O, Tadono T, Isono K. PALSAR Radiometric and Geometric Calibration. *IEEE Geosci Remote S* 2009;**47**: 3915-32.
10. Baraldi, A. Parmiggiani F. An alternative form of the lee filter for speckle suppression in SAR images. *Graph Model Im Proc* 1995;**57**: 75-8.
11. Rouse JW, Haas RH, Schell JA, Deering DW. Monitoring Vegetation Systems in the Great Plains with ERTS. in Third Earth Resources Technology Satellite-1 Symposium. Greenbelt: NASA; 1974.
12. Xu H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Inter J Remote Sens* 2006; **27**: 3025-33.
13. Zha, Y, Gao J, and Ni S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int J Remote Sens* 2003;**24**:583-594.
14. Congalton, RG. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens Environ* 1991;**49**: 1671-8.
15. Congalton RG, Green K. Assessing the Accuracy of Remotely Sensed Data: Principle and Practices. Second Edition, Boca Raton: CRC Press; 2009.
16. Fatoyinbo TE, Fatoyinbo, Simard M, Robert A, Washington A, Shugart HH. Landscape-scale extent, height, biomass, and carbon estimation of Mozambique's mangrove, forests with Landsat ETM+ and Shuttle Radar Topography Mission elevation data. *J Geophys Res Biogeo* 2008;**113**: G02S06.
17. Heumann BW. An object-based classification of mangroves using a hybrid decision tree-support vector machine approach. *Remote Sens* 2011;**3**: 2440-60.
18. van Beijma S, Comber A, Lamb A. Random forest classification of salt marsh vegetation habitats using quad-polarimetric airborne SAR, elevation and optical RS data. *Remote Sens Environ* 2014;**149**: 118-29.