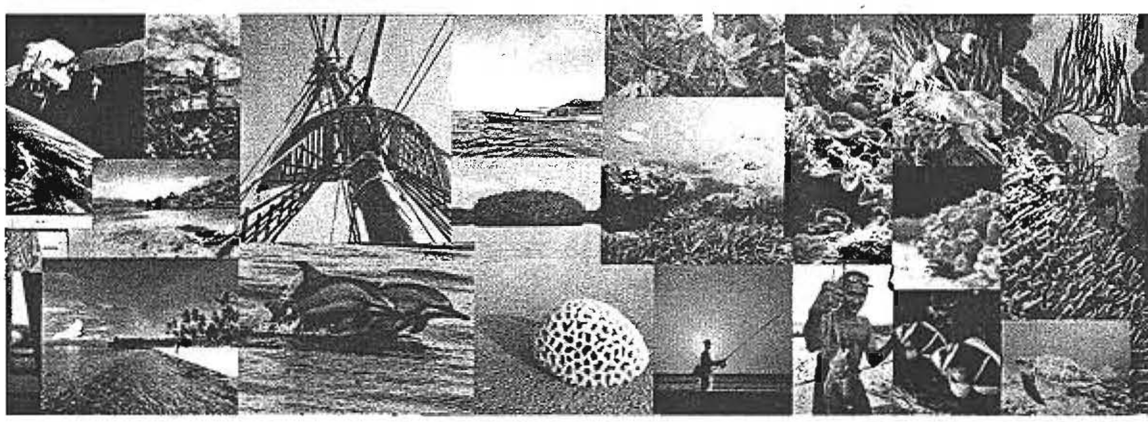
ISBN : 978-979-1266-18-5

MPULAN RISET KELAUTAN

MENUJU SUMBERDAYA ALAM YANG LESTARI



GEO-MARINE RESEARCH FORUM (FORUM RISET GEO MARINE)

Katalog Dalam Terbitan (KDT)
Kumpulan Riset Kelautan: Menuju Sumber Daya Alam Yang Lestari / Dewayany Sutrisno, Ati Rahadiati, dan Niendyawati – Cibinong: Pusat Survei Sumber Daya Alam Laut - BAKOSURTANAL, 2007

xi, 374 hal; 17 X 24 cm
ISBN : 978-979-1266-18-5

1. Sumber Daya Alam – Riset Kelautan

Diterbitkan oleh:

Pusat Survei Sumber Daya Alam Laut
BAKOSURTANAL
Jl. Raya Jakarta - Bogor KM 46
Cibinong 16911

Editor:
Dewayany Sutrisno
Ati Rahadiati
Niendyawati

Desain Sampul :
Rahmad Nugraha

ON THE DECISION TREE ANALYSIS FOR COASTAL AGRICULTURE MONITORING

By: Dyah R. Panuju, Ernan Rustiadi, Ita Carolita, Bambang H. Trisasongko, and Susanto

Abstract

Coastal region in Indonesia is highly dynamic with various land uses including agriculture. Uncontrolled expansion of agriculture may create a conflict with conservation programs. Java, as the largest populated island, experiences this problem. In order to minimize disputes, agricultural intensification techniques have been introduced, including seed technology and remote estimation. For the latter, remotely-sensed data plays an important role which provide updated information for food security. In general, remotely sensed data provide two basic information, i.e. spatial extent of current agricultural observation and estimation of yields. This paper discusses first theme of the role using Landsat multispectral data on two major agricultural sites in Java. An approach of decision tree analysis so-called Quick, Unbiased and Efficient Statistical Tree (QUEST) is demonstrated to provide various growth stages of paddy. The algorithm was designed to improve widely used Classification and Regression Trees (CART) model, improving variable selection, handling missing values and ability to incorporate categorical dataset. Two test sites were selected covering different land tenure. Although the rate of classification accuracy was similar, we found that the decision tree approach was consistently superior to maximum likelihood algorithm. We obtained around 99% on East Java site, in comparison with 98% using maximum likelihood. On West Java, we achieved about 99% for both algorithms.

1. Introduction

Food security has been one of major issues in developing countries such as Indonesia. Expansion of agricultural fields were adopted, however land disputes inhibit the program, mostly related to environmental problems. Intensifying agricultural fields become an alternative to maintain availability of food, relying on technological inventions to improve the productivity. Many efforts have been sought involving recent integrated technologies, span from seed technology to the use of satellite data. However no significant breakthrough was observed since the Indonesian rice sufficiency in 1984.

The use of remotely sensed imageries has been used for decades, mostly using spaceborne sensors which are suitable for monitoring purposes. The know-how is considerably maturing for optical dataset with various spatial resolutions, despite the lack of reports on tropical agriculture fields. Maxwell et al (2004) employed Landsat Thematic Mapper data coupled with Mahalanobis distance to map corn extent autonomously. Xiao et al (2005) exploited vegetation indices on MODIS data in attempt to study flooding and transplanting period. Synthetic Aperture Radar (SAR) data have been used for some extent, for instance see a paper by Wang et al (2005) or, in case of Indonesia, a recent publication of Raimadoya et al. (2007).

To date, limited reports, if any, were published on comparing various algorithms to extract information on the extent and growing stages of paddy field, in particular on tropical sites. In Indonesia, the fields are characterized by small land parcel and heterogeneous management system. The nature of land ownership implies difficulty in data processing and urges to evaluate various presently-available methods. This paper discusses a variant of decision tree analysis in comparison to "standard" maximum likelihood.

2. Test Sites

Two test sites, which are widely recognized as the main production centers, were used to demonstrate the problem. Both sites represented different land tenure, hence reflected different land parcel. The first site was in West Java province, located on north coastal region. The site was previously studied on its importance on national food supply (Winoto et al., 1996). Despite on its significance, vast conversion has been witnessed, in particular on its alteration to industrial uses (Firman, 1997). The second was located on the Brantas River delta which widely known as one of prime lands in East Java province. The site was under major threads, not only due to land use change (again to industrial expansion) (Damayanti, 2003) but also vulnerable to recent mud flooding after a failure in energy exploration (for details, see www.eastjavamud.net). Figure 1 shows both site locations.

Both sites were studied by means of Landsat data. We selected fairly clear scenes from our database to avoid atmospheric effects on the images. The data were acquired on 27 August 1999 (path/row 122/064) and 10 May 1996 (path/row 118/065) for West and East Java respectively. No atmospheric correction prior the analysis to prevent bias on the comparison. All bands were used, except thermal bands. Four classes of rice production stages were identified for West Java site: fallow, fallow in wet condition, vegetative and generative. In East Java site, we were able to detect fully wet condition which associated with water-logged field indicating the start of a growing season.

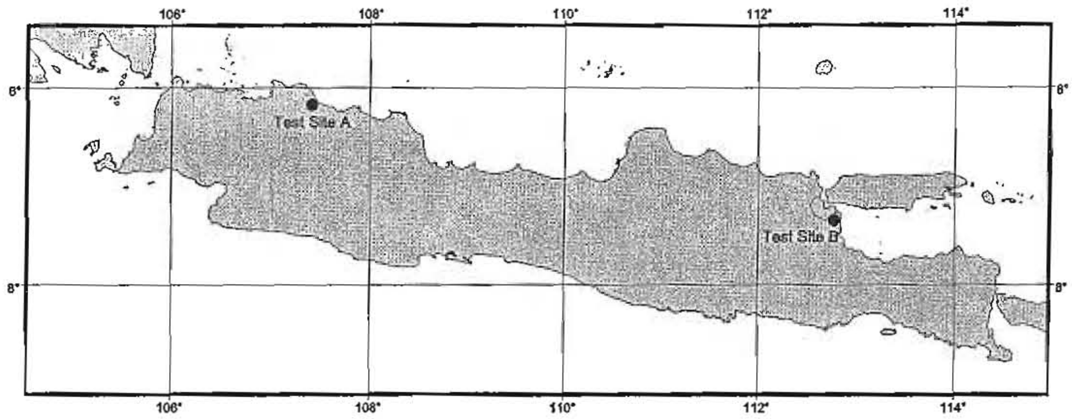


Figure 1. Test sites

3. Data Processing

Decision tree (also known as classification or statistical tree) has been a popular approach for data mining or segmentation for remotely sensed imagery. Classification and Regression Tree (CART) presented by Breiman et al. (1984) was used considerably for various problems, including agriculture (Waheed et al., 2006) and forestry (Herold et al., 2003).

In this paper, we evaluate a classification tree analysis introduced by Loh and Shih (1997), then improved by Kim and Loh (2001) with unbiased multiway splits. More enhanced version of this type by incorporating bivariate linear discriminant node models was presented by Kim and Loh (2003), however the last approach is not being discussed in this paper. A variant of decision tree analysis knows as QUEST (Quick, Unbiased, Efficient Statistical Trees) was evaluated in this research, in comparison with the maximum likelihood algorithm (MLA). The model allows to cut down tree size, develops class prediction and builds up data visualization. Reduction of tree size can be accomplished by allowing the discriminant models share parts the data complexities. In addition, the models enhance estimation accurateness. The analysis allows obtaining better decision on classification problems.

4. Results and Discussion

Employing selected training data, the QUEST algorithm automatically built a decision tree model for West Java site presented in Figure 2. As shown, the tree is fairly complex and makes use almost all available bands. Notation B represents corresponding Landsat band number, e.q. B2 denotes band 2 and so on, with exception on B6 which indicates band 7.

The model exploited band 7 of Landsat as the starting point to separate bare soil (or dry fallow in this case) from vegetated and/or wet surface. At band 7, low digital numbers (DNs) cover areas which were ready for planting (waterlogged) and already in growing sequence. After the separation, visible bands play role fallow and vegetative cover. At the end of branches, band 1 was used to discriminate vegetative and generative stages. Detail of alternative combination characteristics for each class was presented in Table 1.

Table 1 shows all alternatives of classification results using DTA. There were two alternatives of combination for fallow class, one for wet fallow class, and nine alternatives for vegetative and generative classes as well. Apparently, only 3 bands were involved in generating alternatives for fallow, i.e. band 3, band 4 and band 7; while alternative combination of wet fallow was only involving 2 band, i.e. band 4 and band 7. Segmentation of generative and vegetative stages mostly required band 7 and band 4, however some alternatives involved band 3, band 5 or band 1.

The DTA result showed that the most important band to differentiate the classes was band 7, which contributed in creation of all classes. The next important bands were band 4, then band 3. Band 1 was involved only on four alternatives.

From the table, we were able to differentiate dry and wet fallow using the characteristics of band 7. A pixel can be categorized as fallow if it has DN of band 7 more than or equal with 51, while wet fallow wet has it less than 51. Vegetative and generative stages were very complicated to be discriminated, but we found almost all combination involving band 7. The simplest approach to separate was using a threshold of 60 on band 7. Vegetative stage appeared to have DN less than 60 or 35, otherwise it could be classified as generative.

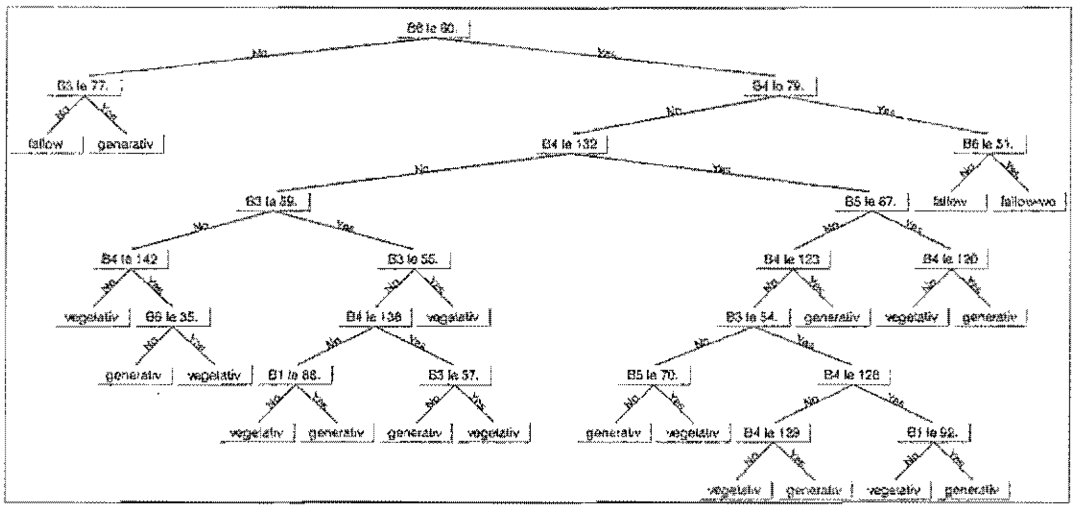


Figure 2. Decision tree of West Java dataset. See text for notations.

Table 1. Alternative combination DN characteristics of DTA result of West Java Dataset

Class	Alternative	DN characteristics				
		Band1	**Band3**	**Band4**	**Band5**	**Band7**
Fallow	1	-	DN≥77	-	-	DN≥60
	2	-	-	DN<79	-	51≤DN<60
Fallow wet	1	-	-	DN<79	-	DN<51
Vegetative	1	-	DN≥59	DN≥79	-	DN<60
	2	-	DN≥59	DN<42	-	DN<35
	3	DN≥88	55≤DN<59	DN≥79	-	DN<60
	4	-	55≤DN<57	DN<79	-	DN<60
	5	-	DN<55	DN≥79	-	DN<60
	6	-	DN≥54	79≤DN<132	67≤DN<70	DN<60
	7	-	DN<54	79≤DN<132	67≤DN	DN<60
	8	DN≥92	-	79≤DN<128	-	DN<60
	9	-	-	79≤DN<132	DN<67	DN<60
Generative	1	-	DN<77	-	-	DN≥60
	2	-	DN≥59	79≤DN<142	-	DN≥60
	3	DN<88	55≤DN<59	DN≥136	-	DN≥60
	4	-	57≤DN<59	DN<136	-	DN≥60
	5	-	DN≥54	79≤DN<132	DN≥67	DN≥60
	6	-	DN<54	79≤DN<129	DN≥67	DN≥60
	7	DN<92	DN<54	79≤DN<132	DN≥67	DN≥60
	8	-	-	79≤DN<132	DN≥67	DN≥60
	9	-	-	79≤DN<132	DN<67	DN≥60

In classification procedure, we noticed similarity in accuracy between decision tree analysis using QUEST algorithm (Table 2) and the Maximum Likelihood algorithm (Table 3). Using the same training set on Subang district, West Java, performance of DTA and MLA was reasonably accurate, about 99.7%. No distinct performance on both algorithms was able to be detected. However, apparently the DTA could differentiate slightly better on generative and vegetative class mixture than the MLA.

Table 2. Accuracy assessment on West Java dataset using QUEST. The overall accuracy is 99.70%.

Class	Fallow	Fallow-Wet	Vegetative	Generative
Fallow	99.91	0.00	0.00	0.10
Fallow-Wet	0.00	100.00	0.00	0.00
Vegetative	0.00	0.00	99.67	0.67
Generative	0.09	0.00	0.33	99.24

Table 3. Accuracy assessment on West Java dataset using Maximum Likelihood. The overall accuracy is 99.70%.

Class	Fallow	Fallow-Wet	Vegetative	Generative
Fallow	100.00	0.00	0.00	0.00
Fallow-Wet	0.00	100.00	0.00	0.00
Vegetative	0.00	0.00	99.45	0.86
Generative	0.09	0.00	0.55	99.14

On East Java site, we found a simpler decision tree. Several factors could be considered as the main reasons; however radiometric quality might be the main cause as we noticed differences on brightness between East and West Java scene. QUEST modeled a decision tree process of East Java dataset as shown in Figure 3. Table 4 was presented to assist in reading Figure 3.

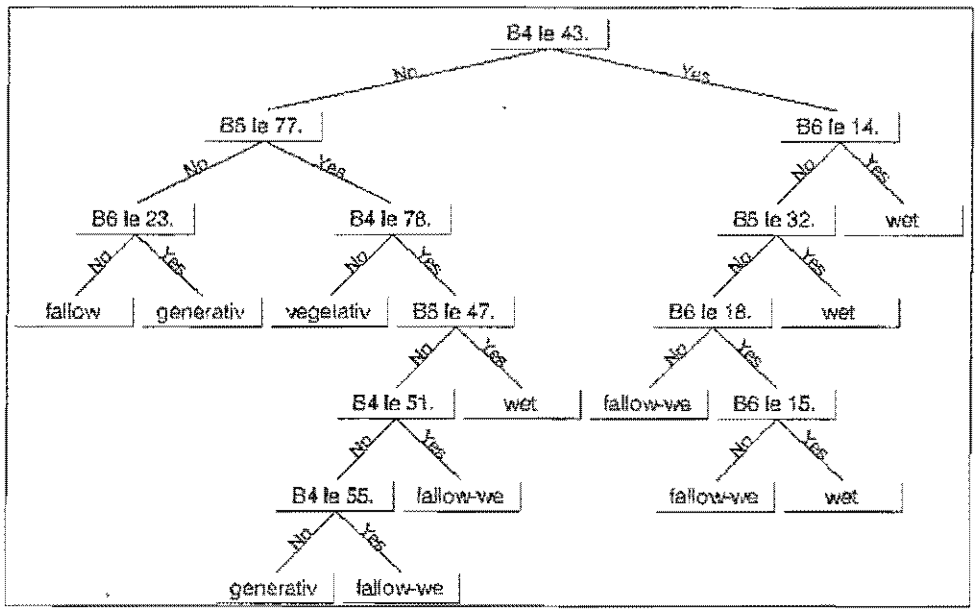


Figure 3. Decision tree of East Java dataset

In this tree model, starting node was band 4 with a threshold of 43. This value discriminated wet regions (with DN less or equal than 43) from their surroundings. The regions were then separated into nodes using infra-red regions to retrieve purely-wet area and partly-wet (in our case, wet fallow class). This separation was reasonable since infra-red spectrum (especially band 7) provides information on underlying soil. Vegetative area was directly governed by band 4 and 5, which was showing distinct classes of vegetative and generative stages. Table 4 shows importance of band 4 in this model, similarly to band 5. Band 7 was used in some cases if similarity of characteristics on bånd 4 and band 5 were detected. Comparison of accuracy assessment between classification algorithms is presented in Table 5.

Table 5 shows that the overall accuracy of DTA was slightly better than of MLC. This result shows similar finding on West Java dataset. We observed that DTA was better not only on differentiating vegetative and generative stages, but also better in discriminating wet and wet fallow.

Table 4. Alternative combination DN characteristics of DTA result of East Java Dataset

Classes	Alternative	DN characteristics		
		Band4	**Band5**	**Band7**
Wet	1	DN<43	-	DN<14
	2	DN<43	DN<32	DN≥14
	3	DN<43	DN≥32	14≤DN<15
	4	43≤DN<78	DN<47	-
Fallow wet	1	DN<43	DN≥32	14≤DN<18
	2	DN<43	DN≥32	DN≥18
	3	43≤DN<51	47≤DN<77	-
	4	51≤DN<55	47≤DN<77	-
Fallow	1	DN≥43	DN≥77	DN≥23
Vegetative	1	DN≥78	DN<77	-
Generative	1	55≤DN<78	47≤DN<77	-
	2	DN≥43	DN≥77	DN<23

Table 5. Comparison of algorithms using East Java dataset

Class	QUEST	MLC
Fallow	100.00	100.00
Fallow-wet	98.31	98.09
Wet	99.55	97.77
Vegetative	98.25	98.50
Generative	99.44	98.87
Overall accuracy	**99.13**	**98.65**

Apparently, DTA of East Java dataset were relatively more reasonable in line with deducted knowledge about band characteristics than of West Java dataset. Despite of its complexity in taking proper training regions, our understanding on the East Java case was coherent with common knowledge in spectral sensitivity. It indicates that an autonomous model derived from

QUEST algorithm was reasonable and can be linked with common spectral response of surface covers.

5. Conclusion and Outlook

Monitoring agricultural fields have been a priority to support food security. As remote sensing technology becomes mature and readily available, exploitation of the data requires major breakthrough on data processing. To date, most users employed a remote sensing toolbox available on-the-shelf which eventually uses maximum likelihood approach.

In this research, we discuss a comparison of a version of decision tree analyses (QUEST) with mostly-used maximum likelihood on two different agricultural sites. Although the difference was insignificant, we showed that on both test sites, QUEST consistently outperformed maximum likelihood.

Mentioning the consistency, it is strongly suggested that further explorations on decision tree analysis should be conducted. The decision tree has been widely known for its simplicity, therefore is easily to be implemented. In near future, we would like to implement a wider comparison study employing various approach of statistical tree, including CRUISE or C-5. To avoid site bias, we suggest an exploration on fairly difficult region with significant class overlap.

6. Acknowledgements

We would like to express our gratitude to the Head of Department of Soil Sciences and Land Resources of IPB for the support. The authors thank Prof. Brian Lees and Dr. David Paull (School of PEMS, UNSW at ADFA) for their support and AusAID (APS scholarship) for funding assistance.

7. References

L. Breiman, J.H. Friedman, R.A. Olshen and C.J. Stone. 1984. *Classification and regression trees.* Chapman and Hall, New York.

R. Damayanti. 2003. *Land use change in an area surrounding an industrial estate: a case study of Surabaya Industrial Estate Rungkut (SIER), Indonesia.* Thesis, Curtin University of Technology, Australia.

T. Firman. 1997. Land conversion and urban development in the northern region of West Java, Indonesia. *Urban Studies*, 34: 1027-1046.

N.D. Herold, G. Koeln and D. Cunningham. 2003. Mapping impervious surfaces and forest canopy using classification and regression tree (CART) analysis. *ASPRS 2003 Annual Conference Proceedings,* Anchorage, Alaska, May 2003.

H. Kim and W.-Y. Loh. 2003. Classification trees with bivariate linear discriminant node models. *Journal of Computational and Graphical Statistics*, 12: 512-530.

H. Kim and W.-Y. Loh. 2001. Classification trees with unbiased multiway splits. *Journal of American Statistical Association*, 96: 598-604.:

T.-S. Lim, W.-Y. Loh and Y.-S. Shih. 2000. A comparison of prediction accuracy, complexity and training time of thirty-three old and new classification algorithms. *Machine Learning*, 40: 203-228.

W.-Y. Loh and Y.-S. Shih. 1997. Split selection methods for classification trees. *Statistica Sinica*, 7: 815-840.

S.K. Maxwell, J.R. Nuckols, M.H. Ward and R.M. Hoffer. 2004. An automated approach to mapping corn from Landsat imagery. *Computers and Electronics in Agriculture*, 43: 43-54.

M.A. Raimadoya, B. Trisasongko and Nurwadjedi. 2007. Exploration of radar imageries for food security intelligence. *Proceedings Second Indonesian Geospatial Technology Exhibition*. Jakarta, Indonesia, 29 August – 1 September 2007. [in Indonesian]

T. Waheed, R.B. Bonnell, S.O. Prasher and E. Paulet. 2006. Measuring performance in precision agriculture: CART – a decision tree approach. *Agricultural Water Management*, 84: 173-185.

X. Wang, Q. Wang, F. Ling, X. Shi, Y. Chen and X. Zhu. 2005. Envisat ASAR data for agriculture mapping in Zhangzhou, Fujian Province, China. *Proceedings 2005 Dragon Symposium Mid-Term Results*, Santorini, Greece, 27 June – 1 July 2005. ESA SP-611 (January 2006).

J. Winoto, M. Selari, S. Saefulhakim, D.A. Santosa, N.A. Achsani and D.R. Panuju. 1996. Agricultural land conversion to non agricultural uses: performance, institutional effect and coordination mechanism on land conversion controlling. *Proceedings National Seminar of National Land Agency*, Jakarta, 28 May 1996. [in Indonesian]

X. Xiao, S. Boles, J. Liu, D. Zhuang, S. Frolking, C. Li, W. Salas, B. Moore III. 2005. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. *Remote Sensing of Environment*, 95: 480-492.

BIOGRAPHY

Dyah Retno Panuju is a junior lecturer of Laboratory of Regional Development, Department of Soil Sciences and Land Resources, and a researcher of P4W, Bogor Agricultural University. Her research interest has been spatio-temporal modeling of land use and environmental problems. Email: d.panuju@hotmail.com

Ernan Rustiadi is a senior lecturer and the director on the same institutions. He received a master and a doctorate degree in regional planning from Kyoto University, Japan. Previously, he was a principal investigator for many research funded by national and international donors, such as RUT and Sumitomo. His research covers various aspects of regional analysis span from spatial analysis to institutional development. Email: ernan@indo.net.id

Ita Carolita completed a master degree in soil science with specialization in remote sensing and GIS analysis after conducting undergraduate study on statistics. She is now a senior researcher with National Institute of Aeronautics and Space (LAPAN) in Jakarta. She was a principal investigator for research on optical and microwave remote sensing and is currently involved with Japanese ALOS research. Email: ita_carolita@lapanrs.com

Bambang Hendro Trisasongko is a researcher on remote sensing applications at P4W, Bogor Agricultural University. His research interest has been exploitation of SAR data on forest and agricultural landscapes. Recently, he is a student at The University of New South Wales at The Australian Defence Force Academy. Email: b.trisas@yahoo.it

Susanto is a senior researcher of LAPAN. He holds a bachelor degree in physics from Gadjah Mada University and a master degree in soil science (remote sensing) from Bogor Agricultural University. Email: susanto@lapanrs.com