THE USE OF STATISTICAL TREE METHODS ON RICE FIELD MAPPING

Dyah R. Panuju[1], Bambang H. Trisasongko[2]
[1]Department of Soil Sciences and Land Resources, Bogor Agricultural University. Jalan Meranti. Kampus Darmaga. Bogor 16680. Indonesia. Email: d.panuju@hotmail.com
[2]Center for Regional Development (P4W), Bogor Agricultural University. Jalan Pajajaran. Bogor. Indonesia.

ABSTRACT

Java, as the largest populated island in Indonesia, has experienced tremendous agricultural land use conversion, mostly to industrial and settlement uses. In order to optimize current rice fields, 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 extents of current agricultural fields and estimation of yields. This paper discusses first theme of the role using Landsat multi-spectral data coupled with statistical tree analyses. Two test sites were selected covering different land tenure. Although the rate of classification accuracy was similar, we found that the decision tree approaches consistently provided acceptable accuracy on around 90%.

Keywords: rice, paddy, statistical tree, CRUISE, QUEST

ABSTRAK

Jawa, sebagai pulau terpadat di Indonesia, telah mengalami perubahan atau konversi penggunaan lahan pertanian yang sangat besar, yang kebanyakan menjadi kawasan industri dan perumahan. Untuk mengoptimalkan lahan pertanian padi yang ada, teknik intensifikasi pertanian telah diperkenalkan termasuk di antaranya teknologi pembibitan dan estimasi jarak jauh. Untuk estimasi jarak jauh, data penginderaan jauh mempunyai peran yang penting dalam menyediakan informasi yang terkini untuk ketahanan pangan. Pada umumnya data penginderaan jauh menyediakan dua informasi dasar yaitu cakupan keruangan dari lahan pertanian dan estimasi produksi pertanian. Makalah ini membahas pertama-tama peran dari citra multi spektral Landsat yang digabung dengan statistical tree analyses. Dua lokasi test dipilih yang mempunyai kepemilikan yang berbeda. Meskipun tingkat ketelitiannya serupa, telah ditemukan bahwa pendekatan decision tree secara konsisten menghasilkan tingkat ketelitian sekitar 90%.

INTRODUCTION

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

Remotely sensed imageries have 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. In a temperate agricultural field, 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 in tropical Asia. Synthetic Aperture Radar (SAR) data have been used for some extent, for instance see a study by Wang et al. (2005) in China or, in case of Indonesia, a publication of Raimadoya et al. (2007).

To date, limited reports were published on comparing various algorithms to extract information on the extent and growing stages of paddy field, in particular on tropical sites. Casanova et al. (1998) investigated rice reflectance to estimate Leaf Area Index (LAI) and biomass. Apparently, estimation of biomass from remote sensing data was more reliable than of LAI. The result promised a possibility to predict yield and monitor rice production using remote sensing data.

Indonesian agricultural fields particularly irrigated paddy fields (*sawah*) are characterized by small land parcel and heterogeneous management system. Previous studies by van der Kroef (1963) noted average farmers land holding was only 0.47 ha per capita for irrigated land in Java just after World War I. Another study by Panuju and Trisasongko (2008) found that the number was reduced to less than 0.3 hectares per capita in 2004 for agricultural area including upland (*gogo*) and *sawah*. The nature of land ownership implies difficulty in data processing and urges evaluation of various presently-available methods. This paper discusses the use of two decision tree analyses to provide growing stage map of rice, which in turn may be guide further exploration of classification methodologies.

TEST SITES

Two test sites, which are widely recognized as the main production centers in Indonesia, were used to demonstrate the problem. Both sites represent different land tenure, hence have different land parcel. The first site was located in West Java province, located on north coastal region. The site was previously studied on aspects of land conversion and agricultural sustainability (Winoto et al., 1996). Despite its importance, high-rate conversion has been witnessed, in particular on its alteration to industrial uses (Firman, 1997). The second location was situated on the Brantas River delta 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 due to industrial expansion) as reported by Damayanti (2003) but also has been vulnerable to recent mud flooding (for details, see www.eastjavamud.net). Figure 1 shows the site locations.

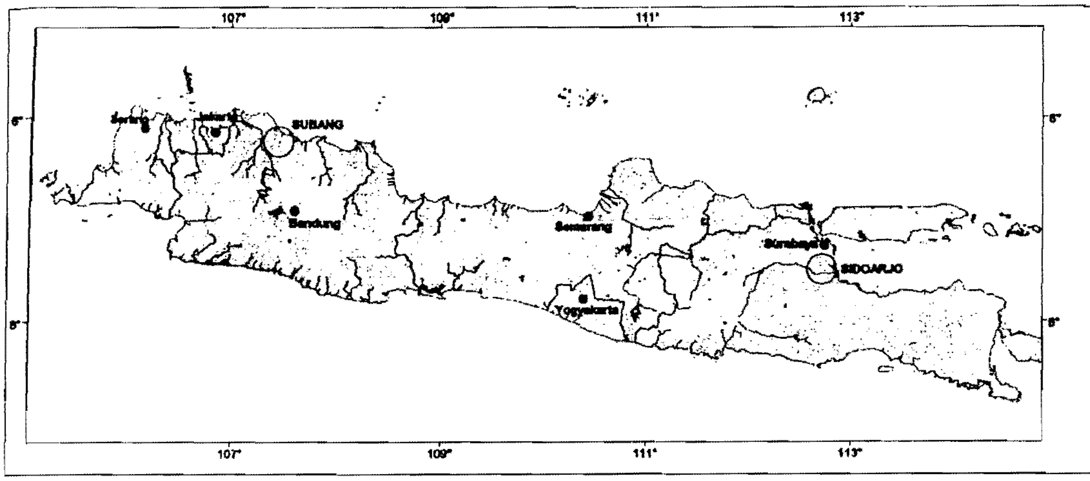

Figure 1: Test sites.

Both sites were studied by means of Landsat data. We selected fairly clear scenes from our database to avoid atmospheric effects on the images since no atmospheric correction was taken. 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. All bands were used, except thermal bands and panchromatic band for ETM data. Four classes of rice production stages were identified for both sites: fallow, fallow in wet condition, vegetative and generative. Wet-conditioned fallow is an indicative for the start of planting season. About 2 or 3 planting seasons were observed on both locations every year.

STATISTICAL TREE ANALYSES

Statistical tree (also known as classification or decision tree) has been a popular approach for data mining or segmentation of 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). Another popular approach named C4.5 was presented by Quinlan (1993). Pal and Mather (2003) provided an example of the application of C4.5 on multispectral Landsat and DAIS hyperspectral data.

In this research, we evaluated two contemporary multivariate decision tree approaches. The first is known as QUEST (Quick, Unbiased, Efficient Statistical Trees) which was introduced by Loh and Shih (1997). The model allows cutting down tree size, develops class prediction and builds up data visualization. Reduction of tree size can be accomplished through the use of discriminant models. In addition, the models enhance estimation accurateness. The analysis allows obtaining better decision on classification problems. The use of QUEST on remote sensing data was previously reported, for instance by Pal and Mather (2003).

Another version of multivariate decision tree using unbiased multiway splits was presented by Kim and Loh (2001) and enhanced by incorporating bivariate linear discriminant node models (Kim and Loh, 2003), widely known as CRUISE (Classification Rule with Unbiased Interaction Selection and Estimation). Limited reports, if any, were presented dealing with the application of CRUISE on remotely-sensed data.

In this research, sampling data were divided into two groups, training and testing datasets in order to reduce bias. Training data were employed to build the classification model and testing dataset was utilized to evaluate the classifier. Both methods were then assessed based on the original tree and the pruned version. Pruning was taken to minimize bias only if the surviving node was less than 6%. Comparison on both versions of pruning scheme was appraised based on Kappa statistics.

RESULTS AND DISCUSSION

Performance of classification using statistical tree methods is presented in Table 1. In general, CRUISE produced higher accuracy than QUEST observable from Kappa statistics. Furthermore, results on West Java datasets tend to be slightly higher than of East Java dataset.

Table 1: Error rates

	CRUISE		QUEST	
Class	West Java	East Java	West Java	East Java
Fallow	12.55	3.77	0.00	3.35
Wet	0.00	0.10	0.00	0.83
Vegetative	10.72	11.65	8.51	23.94
Generative	2.19	24.21	22.42	14.18
Kappa coefficient	0.91	0.87	0.89	0.87

Apparently, the difference of land parcels generated differences on error rates and accuracy as well. On complex land parcels in East Java, both techniques produced similar accuracy which have Kappa coefficient of 0.87. On the other hand, on bigger and more homogeneous patches in West Java, both techniques generated higher accuracy. CRUISE was able to achieve greater accuracy (Kappa=0.91) than of QUEST.

It is interesting to see that sensitivity of the technique on land cover classes was also quite different. For West Java dataset, error rate generated by CRUISE on fallow class was 12.55 while QUEST perfectly learned the training data and classified the test dataset. Both CRUISE and QUEST were fairly adaptive to vegetative cover and delivered reasonable accuracy on big and homogenous region, nonetheless the difference in performance was rather obvious when it was applied on a complex landscape. In this case, CRUISE was preferable to maintain around 90% accuracy. Again, CRUISE maintained good accurateness on West Java case, nonetheless, when East Java was applied, both

algorithms were unable to construct acceptable accuracy. Complexity in rice field mapping was observed here, suggesting more thorough experiment in the future. This was somewhat in agreement with previous results on variability of rice field, especially due to variable varieties (Casanova et al. 1998), although Le Toan et al. (1997) mentioned that tropical rice field generally had 110- to 120-day cycle.

The difference and inconsistency on generating error rate on vegetative and generative classes was possibly related to the variability on age. Small ownership observable from small patches in color composite image implies the different management and likely has different time on plantation time. Furthermore, heterogeneity of plant vigor within the land parcel might also affect classification accuracy, hence generating different error rate for specific area. We need to rationalize here that the performance could also be different on different varieties which have distinctive responses to visible or near infrared wavelength. Furthermore, lowest errors were consistently found in wet class, due to similar soil moisture.

Evaluation of different tree pruning produced by both methods is presented in Figure 2 and 3. Only West Java dataset is used to illustrate the effect of pruning. From the original tree, it is shown that CRUISE produced more compact tree structure. The technique generated tree with seven levels of branch, while QUEST produced slightly more complex structure with nine stages of stem. Both algorithms used Landsat Band 4 as starting node, which is understandable for being the most sensitive band over vegetative covers. CRUISE employed band 4 for initial assessment to separate wet (including water bodies) from dry regions, while QUEST did not exclusively separate wet class from others in its initial step.

From the original tree structure, it is apparent that the structure was fairly complex with many small leaves. Those leaves could be regarded as bias in the tree structure, mostly associated with noise forms inherently found in the remote sensing images. It suggests that using proper pruning methods, more efficient tree structure could be obtained. In this research, pruning was taken by analyzing tree structure, and we found that critical value of 6% of total tree structure might be feasible. After pruning, the tree produced by CRUISE had five branches and QUEST had only four nodes. Again, an evaluation of the pruning effect was performed through accuracy analysis. The difference of accuracy between original and pruned version is presented in Figure 4.

The figure indicates that pruning could improve or conversely reduce classification accuracies. For West Java dataset, pruning increased accuracy about 1% on CRUISE, and almost 5% on QUEST. In contrast, pruning on East Java dataset reduced accuracy for almost 4% on CRUISE and less than 1% on QUEST. The results suggest that pruning should be carefully taken on any cases. It also implies that further analyses should be explored on more diverse location. The use of two datasets such as indicated in this research, was fairly insufficient to conclude which algorithms or processing techniques reliable for operational monitoring. It requires a properly-designed experiment to investigate the complexity. Possible considerations include number of land cover classes (content of thematic maps), size and structure of land parcels, image quality, dominant land use in the region and also spatial autocorrelation in sampling design.

Applying both original and pruned statistical trees on all dataset, thematic maps of rice field were obtained. Figure 5 shows the maps acquired using CRUISE and QUEST algorithms on West Java dataset. Different performance is observable in the figure, in

particular on location 1 and 2 where different performance between CRUISE and QUEST is fairly noticeable. From location 1, apparently QUEST could differentiate better on fallow (including road and built up area) from other classes than of CRUISE. This accomplishment is very important in identifying borders of each parcel since land possessions of farmers tend to be small, as noted previously (van der Kroef, 1963; Panuju and Trisasongko, 2008). Furthermore from location 2, it seems that QUEST performed better in discriminating wet class from the surrounding classes than of CRUISE. However, we observed that both techniques tended to overestimate generative class.

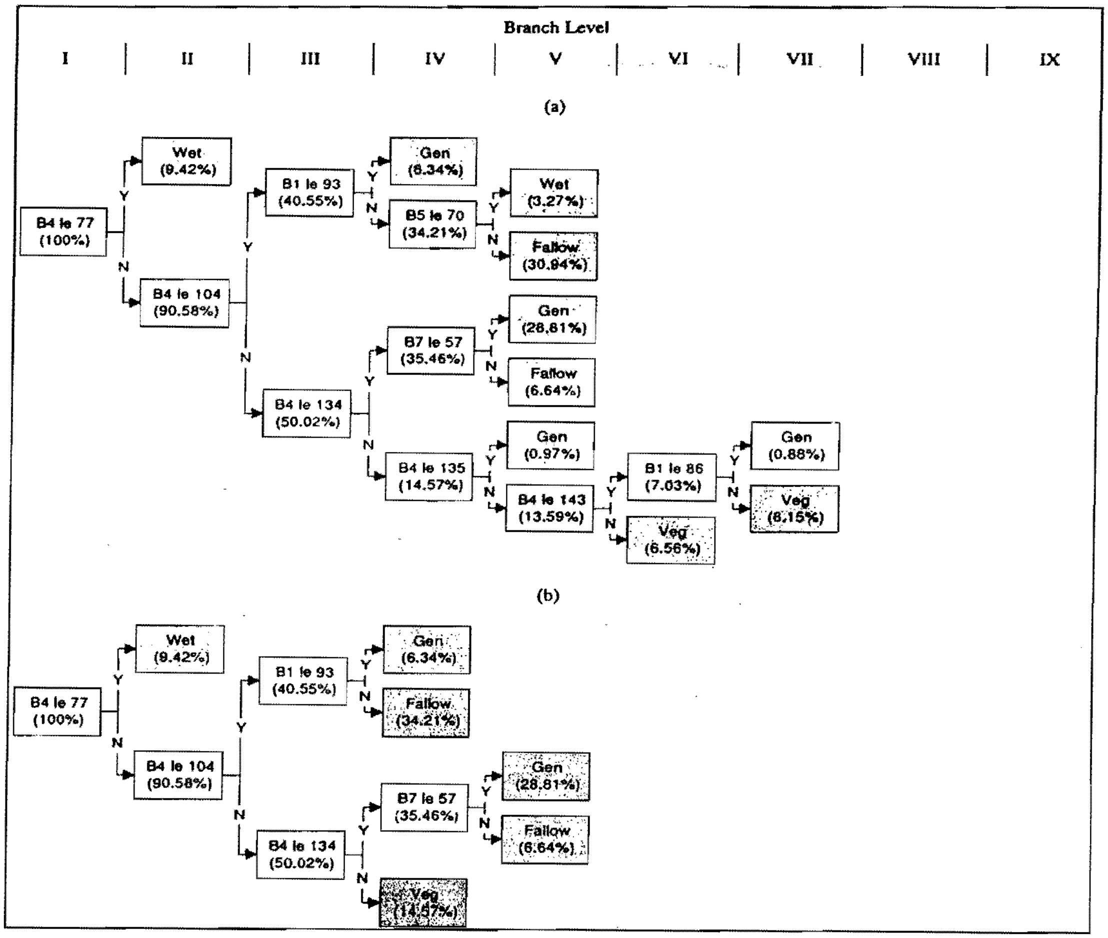

Figure 2: Decision trees from original (a) and pruned-version (b) of CRUISE on Subang dataset.

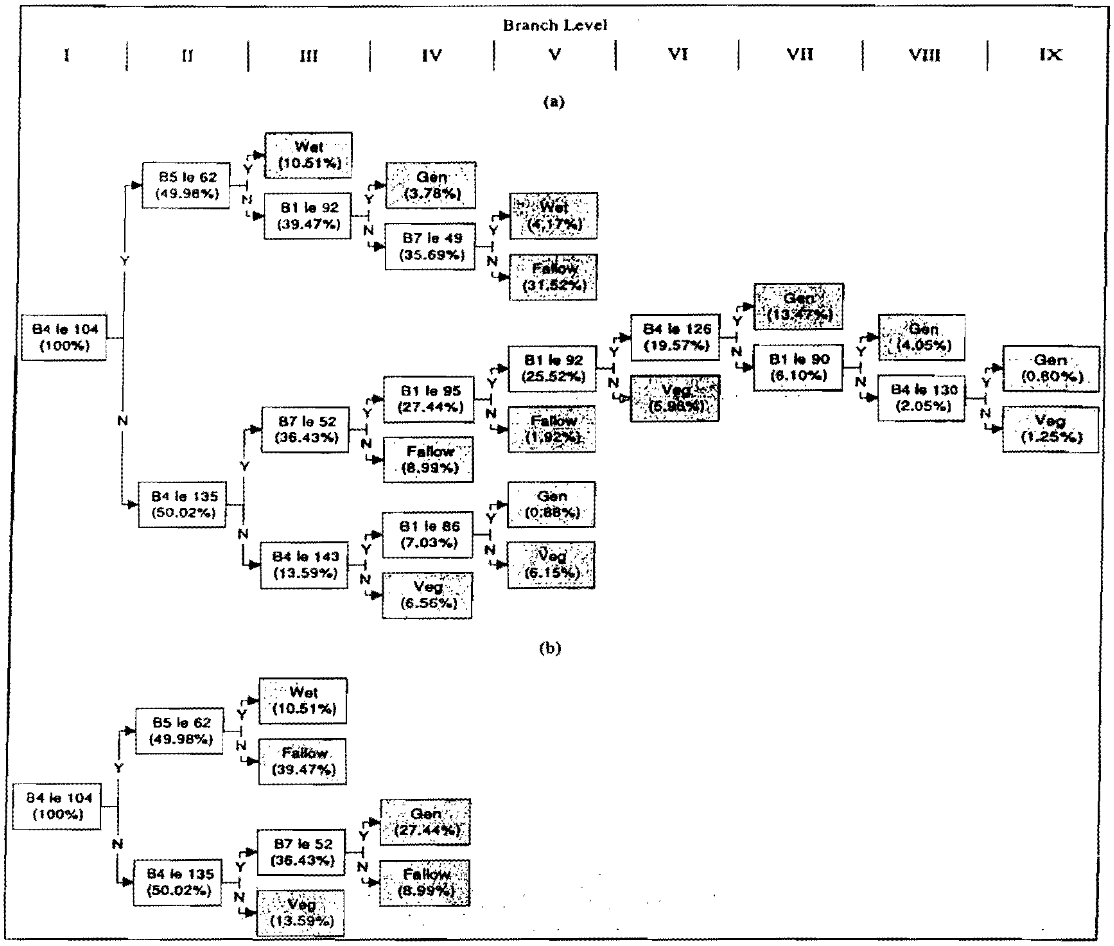

Figure 3: Tree structure of original (a) and pruned-version (b) using QUEST on Subang dataset.

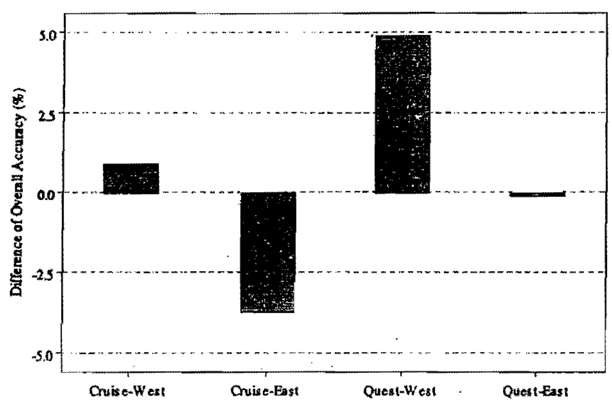

Figure 4: Difference of accuracy between original and pruned version.

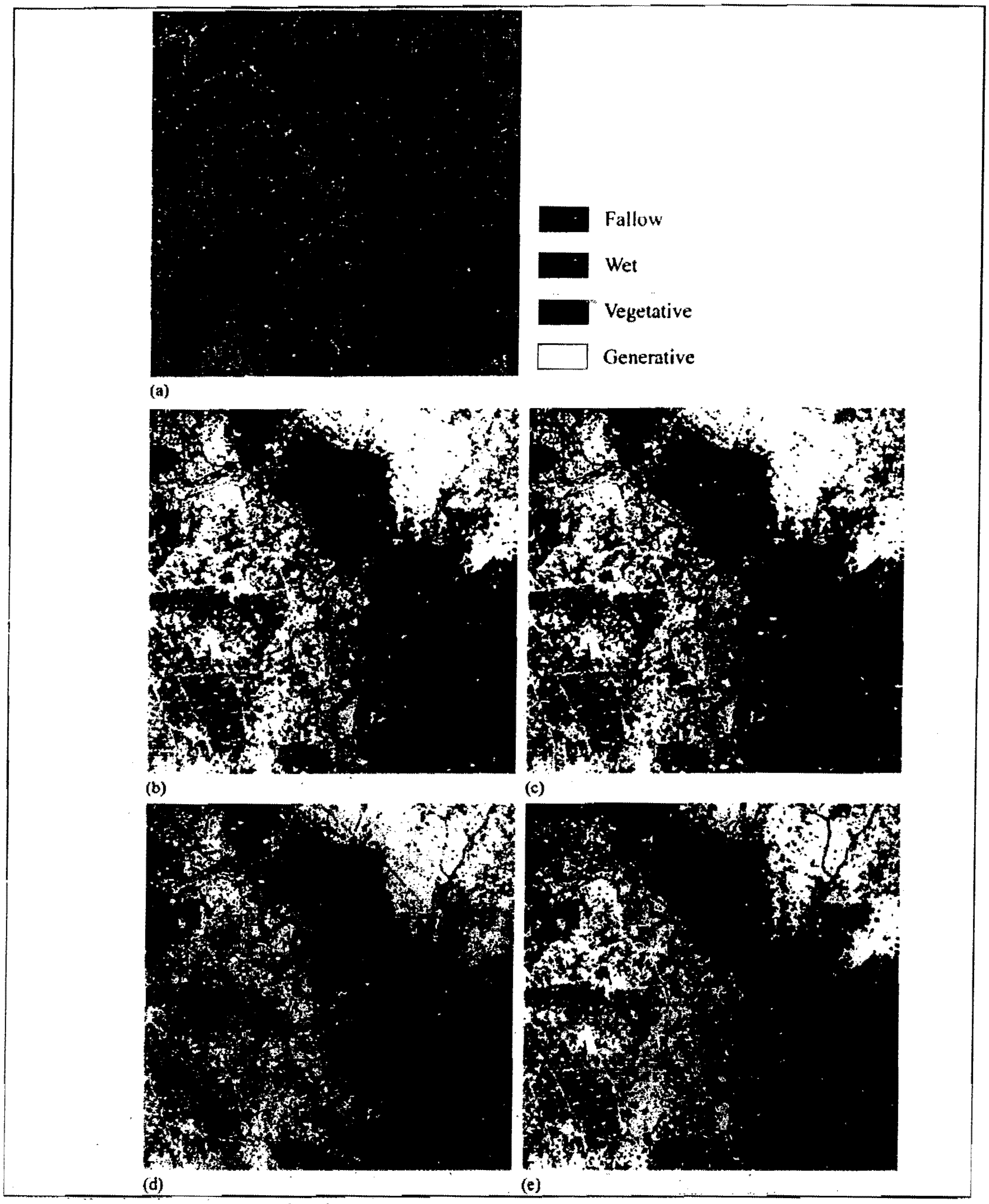

Figure 5. Color composites (a) and CRUISE original (b) and pruned tree (c); QUEST original (d) and pruned version (e) of West Java dataset

CONCLUSION

Monitoring agricultural fields have been a national priority to support food security. As remote sensing technology becomes mature and readily available, data exploitation requires testing and validation of currently-available data processing. Numerous classification methods have been presented in the literature, therefore further examination on their performance is required for operational purposes, especially on tropical environment.

In this research, two contemporary statistical tree algorithms were evaluated to attain better understanding of the techniques in specific application. In general, acceptable error rate was obtained using both algorithms indicating that the techniques may be employed for further tests or in operational basis. Nonetheless several indications were observed in the analysis that should be considered in advance.

CRUISE algorithm generated slightly higher classification accuracy than of QUEST on simpler landscape structure, therefore it is arguable that the results could be replicable on other agricultural sites. On small land tenure, hence complex landscape, both techniques suffered and higher rate of error was observable. It suggests that structure of land ownership (or the size of land parcel) should be considered in the analysis. Therefore comparison of classification approach should be assessed in fairly similar land parcels.

Variability within land cover classes could be one of important factors in producing thematic maps. This is especially correct on vegetative and generative stages when many issues including relationship between plant vigor and varieties are apparent. Within the same stage, different varieties reflect different responses on radiometric data especially on visible/near infrared data.

Specifically on data analysis, pruning should be carefully taken. Using pruning approach, simpler tree structures can be obtained, hence requires shorter computation time. Nonetheless, pruning had variable responses. It seems that effect of pruning depends on landscape structure more than classification algorithms.

Acknowledgements

We would like to express our gratitude to the Dr. Kukuh Murtilaksono (Head of the Department) and Dr. Ernan Rustiadi (Head of the Center) on their permission to use laboratory facilities. The authors thank colleagues from LAPAN (Mrs Ita Carolita and Mr. Susanto) for fruitful discussions and acknowledge Prof. Brian Lees and D. David Paull of The University of New South Wales at Australian Defence Force Academy for their assistances during literature reviews.

REFERENCES

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. 1984. Classification and regression trees. Monterey: Wadsworth.

Casanova D., Epema, G.F., Goudrian J. 1998. Monitoring rice reflectance at field level for estimating biomass and LAI. *Field Crops Research*, 55(1-2), 83-92.

Damayanti, R. 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.
Firman, T. 1997. Land conversion and urban development in the northern region of West Java, Indonesia. *Urban Studies*, 34(7), 1027-1046.
Herold, N.D., Koeln, G., Cunningham, D. 2003. Mapping impervious surfaces and forest canopy using classification and regression tree (CART) analysis. ASPRS 2003 Annual Conference, Anchorage.
Kim, H., Loh, W-Y. 2003. Classification trees with bivariate linear discriminant node models. *Journal of Computational and Graphical Statistics*, 12(3), 512-530.
Kim, H., Loh, W-Y. 2001. Classification trees with unbiased multiway splits. *Journal of American Statistical Association*, 96(454), 589-604.
Le Toan, T., Ribbes, F., Wang, L., Floury, N., Ding, K., Kong, J-A., Fujita, M. Kurosu, M. 1997. Rice cropping mapping and monitoring using ERS-1 data based on experiment and modeling results. *IEEE Transactions on Geoscience and Remote Sensing*, 35(1), 41-56.
Loh, W-Y., Shih, Y-S. 1997. Split selection methods for classification trees. *Statistica Sinica*, 7(4), 815-840.
Maxwell, S.K., Nuckols, J.R., Ward, M.H., Hoffer, R.M. 2004. An automated approach to mapping corn from Landsat imagery. *Computers and Electronics in Agriculture*, 43(1), 43-54.
Pal, M., Mather, P.M. 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, 86(4), 554-565.
Panuju, D.R., Trisasongko, B. 2008. The dynamic of rice production in Indonesia 1961-2004. In review, *Singapore Journal of Tropical Geography*.
Quinlan, J.R. 1993. C4.5: programs for machine learning. San Mateo: Morgan Kaufmann.
Raimadoya, M.A., Trisasongko, B.H., Nurwadjedi. 2007. Exploration of radar imageries for food security intelligence. Second Indonesian Geospatial Technology Exhibition. Jakarta (in Indonesian).
Van der Kroef, J.M. 1963. Indonesia rice economy: problems and prospects. *American Journal of Economics and Sociology*, 22(3), 379-392.
Waheed T., Bonnell, R.B., Prasher, S.O., Paulet, E. 2006. Measuring performance in precision agriculture: CART – a decision tree approach. *Agricultural Water Management*, 84(1-2), 173-185.
Wang X., Wang, Q., Ling, F., Shi, X., Chen, Y., Zhu X. 2005. Envisat ASAR data for agriculture mapping in Zhangzhou, Fujian Province, China. 2005 Dragon Symposium Mid-Term Results, Santorini.
Winoto J., Selari, M., Saefulhakim, S., Santosa, D.A., Achsani, N.A., Panuju, D.R. 1996. Agricultural land conversion to non agricultural uses: performance, institutional effect and coordination mechanism on land conversion controlling. National Seminar of National Land Agency, Jakarta (in Indonesian).
Xiao X., Boles, S., Liu, J., Zhuang, D., Frolking, S., Li, C., Salas, W., Moore III, B. 2005. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. *Remote Sensing of Environment*, 95(4), 480-492.

ISSN: 0854-2759

Jurnal Ilmiah GEOMATIKA

Kajian Ilmiah dan Hasil Penelitian dalam Bidang Survei dan Pemetaan

Vol. 14 No. 2, Desember 2008

Nomor Akreditasi 128/Akred-LIPI/P2MBI/06/2008

Diterbitkan Oleh:

Badan Koordinasi Survei dan Pemetaan Nasional

Jl. Raya Jakarta – Bogor Km 46 Cibinong

Telp. 021-8752062 Fax. 021-8752064 Homepage: http://www.bakosurtanal.go.id