DISCRIMINATION OF FUNGAL DISEASE INFESTATION IN OIL-PALM CANOPY HYPERSPECTRAL REFLECTANCE DATA.

Camille C.D. LELONG[(1)], Jean-Michel ROGER[(2)], Simon BRÉGAND[(1)], Fabrice DUBERTRET[(1)], Mathieu LANORE[(1)], Nurul A. SITORUS[(3)], Doni A. RAHARJO[(4)], Jean-Pierre CALIMAN[(5)].

(1) CIRAD / UMR TETIS, Maison de la Télédétection, Montpellier, France
(2) Cemagref / UMR ITAP, Montpellier, France
(3) P.T. SMART / SMARTRI, Padang Halaban, North Sumatra, Indonesia
(4) P.T. SMART / SMARTRI, Pekanbaru, Riau, Sumatra, Indonesia
(5) CIRAD / UR34 & P.T. SMART / SMARTRI, Pekanbaru, Riau, Sumatra, Indonesia

ABSTRACT

This study focuses on the calibration of a statistical model of discrimination between different stages of a fungal disease attack on oil palm, based on field hyperspectral measurements at the canopy scale. Combinations of pre-processing, partial least square regression and factorial discriminant analysis are tested on a hundred of samples to prove the efficiency of canopy reflectance to provide information about the plant sanitary status. A robust algorithm is thus derived, allowing classifying oil palm in a 4-level typology, based on disease severity levels from the sane to the critically sick tree with a global performance of more than 92%. Applications and further improvements of this experiment are discussed.

***Index Terms*—**Reflectance spectroscopy, in-situ measurements, Partial Least Square, discrimination, phytopathology.

1. INTRODUCTION

Early and non-destructive diagnostic of crop disease is a major issue in precision farming and sustainable agriculture in general. Hyperspectral reflectance spectroscopy theoretically meets these requirements, thanks to strong relationships existing between the plant optical properties on one hand, and leaf pigment concentration, and foliar and canopy structures on the other hand (eg.[1][2][3][4][5]). Some authors have even shown that hyperspectral data acquired by satellite or airborne remote sensing might be actually relevant to detect crop diseases or to assess crop damage severity (eg.[6][7][8]). However, these studies are always crop and/or disease-dedicated and new experiments need to be performed to validate the detectability of a different pathology of another crop.

At least, some robust methodologies have proven to be efficient whatever the context, to classify spectra into different groups as long as a good sample helps training a statistical model of discrimination. Cluster analyses, for instance, allow good classifications of plant stress levels when combined with Partial Least Square Regression (PLS) [9][106] or Principal Component Analysis (PCA) [11].

In this paper, we propose to apply these approaches to validate the efficiency of hyperspectral reflectance spectroscopy to discriminate several levels of Ganoderma fungus contamination on oil palm trees. Indeed, this disease is one of the major issues in oil-palm crop management [12], which will benefit from an appropriate remotely sensed diagnosis tool. We will thus evaluate different statistical models for the classification of spectra acquired at the canopy level depending on the number of attack degrees. Then, we will analyze the possibility of developing a remote sensing tool on this basis, in the aim of precision farming applications.

2. MATERIAL

Field measurements were achieved in an oil palm plantation located in North Sumatra, Indonesia: Padang Halaban Estate, which has been drastically attacked by the Ganoderma fungus for years. It thus provides a wide variety of disease severity. We have surveyed more specifically a hundred of oil palm trees, geo-localized and spotted in the plantation grid for easy subsequent identification, and we have assigned them a score in a four-level disease typology: 0 for sane (not sick) trees, 1 for a light attack, 2 for a medium one, and 3 for a severe (close to death) infestation. Among the different sampled trees, even the sane ones, some showed symptoms of nutritional stress like nitrogen, iron, bore or magnesium deficiencies.

Then, we performed hyperspectral reflectance measurements above the canopy of these trees with a Unispec (http://www.ppsystems.com/Literature/EDSUniSpec-SC.pdf) from PP-SYSTEMS, equipped with a Cosine Receptor and fibre optics with 20° of field of view. This spectroradiometer covers 256 spectral bands in the range 310-1130nm. Climbing on scaffoldings to reach the top of each tree (up to six to ten meters high), we made six to ten radiance acquisitions distributed around the crown (about nine meters in diameter), each one integrating a surface of about nine square-meters. Each canopy reflected radiance measurement was directly followed by a diffuse incident light radiance acquisition for a scaling in reflectance. Then, we averaged these intermediate reflectance values to derive the mean reflectance of the whole tree crown.

Due to high level of noise in the resulting spectra in the two extreme domains (310-450nm) and (1100-1130nm), only the range 450-1100nm was actually analyzed in this study.

At the end of the campaign, the data base contained the canopy reflectance spectra of 36 palm trees belonging to score 0, 20 to score 1, 36 to score 2, and only 3 to score 3, for a total of 95 trees.

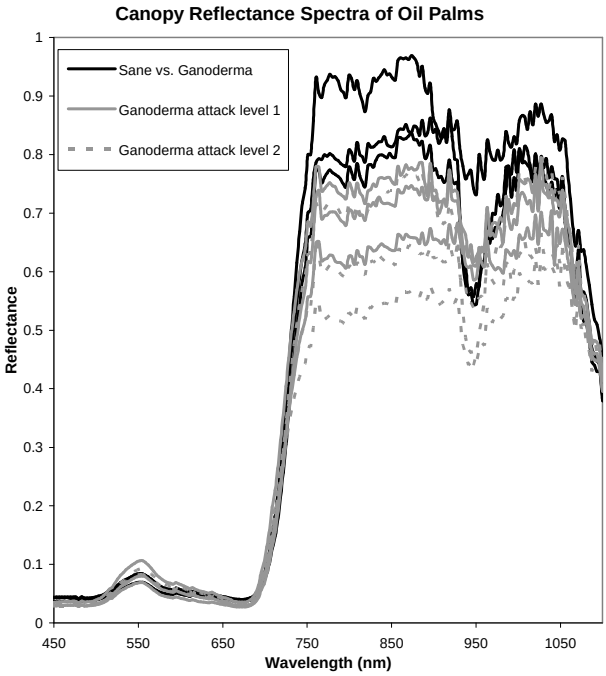

Figure 1: Examples of reflectance spectra for oil palm canopies when the tree is sane (black solid lines) or attacked by the Ganoderma disease at the respective levels 1 (grey solid lines) and 2 (grey dotted lines).

3. METHOD

3.1. Spectra preprocessing

As spectral signatures associated with Ganoderma disease symptoms might be very faint, it seems necessary to avoid any noise source or signal contamination due, for instance, to variations of sunlight and skylight illumination, soil and other backgrounds reflectance or even instrumentation itself. The Savitzky–Golay filtering [13] consisting in a polynomial fitting followed by a derivative computation is commonly performed to meet this requirement [14][15][16].

However, its major constraint is the choice of, on one hand, the smoothing window size and, on the other hand, the degree of the polynomial fit [17], along with the derivative order [14]. We have thus chosen to test a large set of combinations of these parameters, calculating derivative-spectra at the null, first, and second order of derivation for polynomials of second and third degrees, each smoothed at nine different window sizes selected to broom the spectral bins from 10 nm to 160 nm. It thus results in 54 databases of derivative spectra. Higher orders of derivation were not tested because the canopy architecture effects on reflectance is blurred at higher orders, and ganoderma symptoms largely appear on the tree canopy structure. The original unprocessed reflectance data was also tested to evaluate the actual gain of preprocessing.

3.2. Partial-Least-Square Discrimination Analysis

We have applied on each database the Partial-Least-Square Discrimination Analysis (**PLS-DA**) [18][19]. It consists in:

1. Partial Least Square Regression (PLS) applied to the preprocessed derivative spectra, reducing the spectral data to little or not correlated latent variables that are fewer than the number of variables in the sample. To perform a simple PLS, we have chosen to set a specific scale fixing a predictable value for each analyzed class in the bin [0,1].
2. Discriminant Analysis (DA) applied to the most significant latent variables, enhancing the interclass variability while minimizing the intraclass variability of the sample to build a classification model. The selection of the number of PLS-variables is guided by the compromise between minimization of Root Mean Square Error of Prediction (RMSEP) and gain in correlation coefficient (R^2) between predicted and reference values, on one hand, and stability of the model thanks to the fewer number of implied variables, on the other hand.

PLS-DA was achieved on the entire sample of 95 individuals by cross-validation based on the “leave-one-out” method. The potential of the method was tested following two objectives independently:

1. potential discrimination between the healthy ("score 0"; 36 individuals in the sample) and the sick ("scores 1", "2", and "3" together; 56 individuals in the sample) trees,
2. potential of classification of a given tree in the 4-level scoring of disease severity.

Classification results were then compared on the basis of the confusion matrix and the global precision values.

4. RESULTS AND DISCUSSION

The first discrimination objective was perfectly met (100% of good classification for each of the two classes) with a PLS-DA applied on the first-order derivative of a second-degree polynomial fitted on a smoothing window of 32 nm. PLS-predictable values were set to 0 for "sane" and 1 for "sick". DA was applied on the ten first latent variables derived from the PLS with a root mean square error of prediction of 0.27 and a correlation coefficient of 0.7.

The second objective was best met, corresponding to a global accuracy of 92.6%, while using the second-order derivative of a third-degree polynomial fitted on a smoothing window of 26 nm. PLS-predictable values were set on the basis of a simple unmixing of mean spectra of each class between the two endmembers "score 0" and "score 4": 0 for "score 0", 0.4 for "score 1", 0.6 for "score 2", and 1.0 for "score 3". DA was applied on the seven first latent variables derived from the PLS with a root mean square error of prediction of 0.13 and a correlation coefficient of 0.8. The two first discriminant factors are then able to split the space into four clusters. The corresponding confusion matrix is given Table1.

Score	***0***	***1***	***2***	***3***	***% of good classification***
0	34	2	0	0	94 %
1	0	17	3	0	85 %
2	0	2	34	0	94 %
3	0	0	0	3	100 %
Global precision					92.6%

Table 1: Confusion matrix obtained for the classification in four levels of disease severity.

As a factor of comparison, the best result obtained on the original (not filtered and not derived) reflectance spectra only gave a global precision of 63%, with strong confusion between sane and sick trees and bad assignation of individuals inside distant classes.

Both models thus allow a very good discrimination between the sane and the sick trees, even if 6% of false alarms can be expected in the 4-level classification process. Even though, these alarms concern only two individuals that are classified as lightly attacked by the disease; it is still possible that the visual symptoms on which was based the ground-truth diagnostic were not yet observed while the reflectance spectrum already features some changes compared to sane individuals.

Moreover, errors occurring in this classification all shift the trees from one class to the direct next one. Considering that the limits between one score evaluated in-situ and the closer one is very fuzzy, these errors can be either due to the classification or to the field diagnosis without clue to conclude. Even if it is purely a classification error, they are very rare and allow a good confidence in the overall results. The cross validation process also allows a good confidence in the robustness and stability of the model.

The tests also show that the preprocessing of the data has a considerable impact on the detectability of the spectral features associated with the disease presence, and its level of severity too.

Considering the loadings of the PLS, no privileged spectral range seems to contribute more than others do, and no range seems uninformative. This proves the essential need of the entire spectral richness to detect discriminating features in the canopy reflectance. Applications in remote sensing seems thus limited to hyperspectral sensors only, but this still has to be checked by dedicated studies. Indeed, airborne acquisitions would be of great help for fast mapping of the Ganoderma infestation in plantations, compared to field measurements that are still long, difficult, and dangerous to set up at this canopy height.

5. CONCLUSION

Statistical algorithms like PLS-DA applied to preprocessed hyperspectral reflectance data acquired in the fields over oil palm canopies are thus efficient to detect the Ganoderma fungal disease attack. They can even classify it into four levels of severity from sane to highly damaged trees with more than 92% accuracy. Even sane trees that present nutritional deficiencies are correctly are not misclassified as sick.

It proves the potential of hyperspectral reflectance spectroscopy for tree-crop sanitary status evaluation and pushes for further improvements towards remote sensing applications. Nevertheless, new protocols must be fitted to airborne or satellite-borne hyperspectral data to calibrate a dedicated model that would take into account the imaging specificities (e.g. noise, scale, spatial contributions, etc…).

10. AKNOWLEGDEMENTS

This project was funded by SMART-TBK Company and CIRAD Institute. The authors would also like to thanks Dr. Prasetya, Head of SMART-RI antenna in Padang Halaban Estate, and Dr. Wahyu, phytopathologist at SMART-RI office in Pekanbaru, for their help in the ganoderma diagnostic survey and operations in the fields.

11. REFERENCES

[1] Chaerle, L.; Van Der Straeten, D. Imaging techniques and the early detection of plant stress. *Trends in Plant Science* **2000**, *5*, 295-500.

[2] Larsolle, A.; Muhammed, H.H. Measuring crop status using multivariate analysis of hyperspectral field reflectance with application to disease severity and plant density. *Precision Agriculture* **2007**, *8*, 1385-2256.

[3] Liew, O.W.; Chong, P.C.J.; Li, G.; Asundi, A.K. Signature optical clues: emerging technologies for monitoring plant health. *Sensors* **2008**, *8*, 3205-3239.

[4] Muhammed, H.H. Hyperspectral Crop Reflectance Data for characterising and estimating Fungal Disease Severity in Wheat *Biosystems Engineering* **2005**, *91*, 9-20.

[5] Wang, W.; Zhang, M.; Zhu, J.; Geng, S. Spectral prediction of Phytophtora infestans infection on tomatoes using artificial neural netword (ANN). *International Journal of Remote Sensing* **2008**, *29*, 1693-1706.

[6] Nilsson, H.-E. Remote sensing and image analysis in plant pathology. *Annual Review of Phytopathology* **1995**, *15*, 489-527.

[7] Apan, A.; Held, A.; Phinn, S.; Markley, J. Detecting sugarcane 'orange rust' disease using EO-1 hyperion hyperspectral imagery. *International Journal of Remote Sensing* **2004**, *25*, 489-498.

[8] Goodwin, N.; Coops, C.C.; Stone, C. Assessing plantation canopy condition from airborne imagery using spectral mixture analysis and fractional abundances. *International Journal of Applied Earth Observation and Geoinformation* **2005**, *7*, 11-28.

[9] Jorgensen, R.N.; Christensen, L.K.; Bros, R. Spectral reflectance at sub-leaf scale including the spatial distribution discriminating NPK stress characteristics in barley using multiway partial least square regression. *International Journal of Remote Sensing* **2007**, *28*, 943-962.

[10]Huang, J.F.; Apan, A. Detection of Sclerotinia rot disease on celery using hyperspectral data and partial least squares regression. *Journal of Space Science* **2006**, *51*, 129-142.

[11]Zhang, M.; Liu, X.; O'Neill, M. Spectral discrimination of Phytophtora infestans infection on tomatoes based on principal component and cluster analyses. *International Journal of Remote Sensing* **2002**, *23*, 1095-1107.

[12]Flood, J.; Bridge, P.D.; Holderness, M. *Ganoderma diseases of perennial crops*. CABInternational Ed. - Wallingford Press **2000**, 245 pages.

[13]Savitzky, A.; Golay, M.J.E. Smoothing and differentiation of data by simplified least-squares procedures. *Analytical Chemistry* **1964**, *64*, 1627-1639.

[14]Tsai, F.; Philpot, W. Derivative analysis of hyperspectral data. *Remote Sensing of Environment* **1998**, *66*, 41-51.

[15]Estep, L.; Carter, G.A. Derivative analysis of AVIRIS data for crop stress detection. *Photogrammetric Engineering & Remote Sensing* **2005**, *71*, 1417-1421.

[16]Ruffin, C.; King, R.L.; Younani, N.H. A combined derivative spectroscopy and Savitzky-Golay filtering method for the analysis of hyperspectral data. *Giscience &Remote Sensing* **2008**, *45*, 1-15.

[17]Browne, M.; Mayer, N.; Cutmore, T.R.H. A multiscale polynomial filter for adaptive smoothing. *Digital Signal Processing* **2007**, *17*, 69-75.

[18]Roger, J.-M.; Palagos, B.; Guillaume, S.; Bellon-Maurel, V. Discriminating from highly multivariate data by Focal Eigen Function discriminant analysis; application to NIR spectra. *Chemiometrics and Intelligent Laboratory Systems* **2005**, *79*, 31-41.

[19]Gorretta, N.; Roger, J.-M.; Aubert, M.; Bellon-Maurel, V.; Campan, F.; Roumet, P. Determining vitreousness of durum wheat kernels using near infrared hyperspectral imaging. *Journal of Near Infrared Spectroscopy* **2006**, *14*, 231-239.