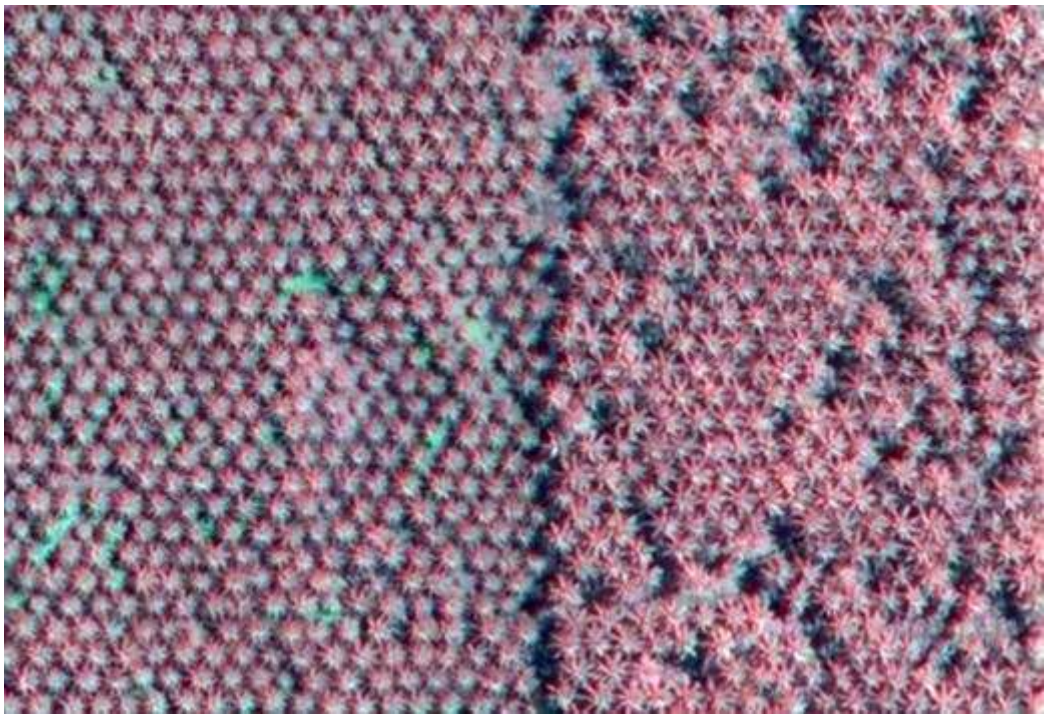
Potential of optical remote sensing for the evaluation of oil-palm agronomic status:

(PT-SMART 2006-2010 Project Final Report)



Camille LELONG

CIRAD ES – UMR TETIS

January 2012

Table of content

1. Introduction

This project objective was the evaluation of the optical remote sensing potential (either by satellite, airplane or from in-situ platforms) to provide any information about oil-palm trees and groves physiological status. Especially, it focused on three agronomical indicators:

1) the nutrient concentration, and the nutritional deficiencies,
2) the canopy structure, through the LAI,
3) the ganoderma disease presence, and severity level.

In practical, it consists in the analysis of the optical properties of oil-palm trees at different scales, to establish whether there exists any relationship between, on one hand, their reflectance in the visible and near-infrared domains and, on the other hand, one of these three characteristics of oil-palm physiological status.
Indeed, some field and satellite surveys were achieved to get data that are representative of the largest range of organic and mineral deficiencies, density and pruning management, and of ganoderma disease levels, to determine if this type of information is actually available in such kind of data. For the purpose of this project, innovative dedicated protocols of data acquisition experimentations were developed in the oil-palm context, using either the very high spatial resolution of imagery, either the very high spectral resolution of a field spectroradiometer.

2. Test-sites

This study was focused on three oil-palm groves managed by PT-SMART in Sumatra (Indonesia):

1) Kandista, near SMART-RI station (Riau)
2) Naga Mas and Naga Sakti trials (Riau)
3) Padang Halaban Estate (Sumatra Utara)

The two first sites were selected for the project considering the opportunity of having a documented variability of nutritional status through the existence of fertility trials (N, P, K, and Mg), or the availability of a large range of natural organic and mineral deficiencies (Fe, K, N, and Mg) due to poor sandy soils.
The third one was selected on the basis of two criterions:

a) a large variability of tree ages and sizes, density of plantation, management practices, and cultivars, resulting in various canopy structures and thus in a wide range of LAI values
b) a widespread and intense contamination by the Ganoderma disease.

Three field surveys, each during three months, were achieved over these sites by French MSc internship students (M. Lanore in 2006, F. Roussel in 2008, F. Dubertret in 2009) and their advisor, C. Lelong (CIRAD agent, staying herself two to three weeks in the fields), with the local support of J.-P. Caliman (head of Smart-Ri), Prasetya and N.A. Sitorus (SMART-RI PHLE), A.R. Syakharosie (SMART-RI Libo), and the help of Indonesian Grd. Student internship/contactors (D. A. Raharjo, N. A. Prabowo).

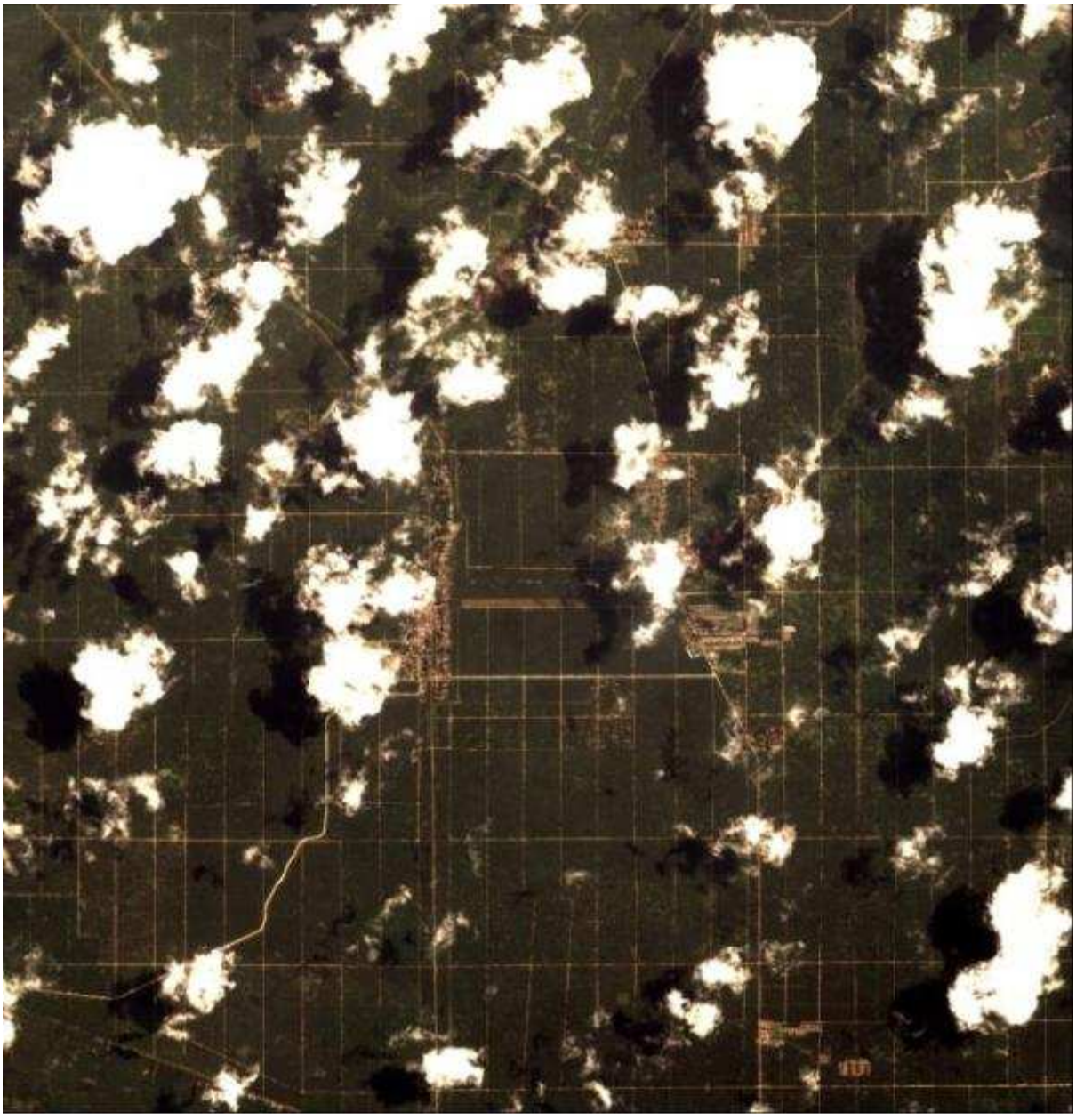
1) The first survey (2006) was based in Kandista, Naga Mas, and Naga Sakti, and was dedicated to establish the protocols of measurements of hyperspectral reflectance data at the tree canopy level and at the leaf scale, along with foliar laboratory analyses. It was the first contact with the object of study and with the acquisition conditions, to fix the ideas to be developed during the project and test the feasibility of the experiments. It resulted in a large data base concerning nutrition deficiencies, ganoderma disease, and other characteristics (e.g. age, genetic selection).
2) The second survey (2008) was based in Padang Halaban and was only dedicated to the LAI measurement of individual trees and of whole blocks, with different techniques and protocols, over the large range of variability available inside the estate.
3) The last survey (2009) was also based in Padang Halaban and was dedicated to the acquisition of hyperspectral reflectance data of both canopies and leaves of sick trees only, plus sane individuals as a reference.

3. Remote sensing data

3.1. Satellite images

Two satellite images were acquired for this study, by Quickbird very high spatial resolution sensor, providing a multispectral dataset in four spectral bands (blue, green, red and near-infrared) at 2.5m spatial resolution per pixel, plus a panchromatic image at 0.6m/pixel. Those two products have been calibrated and merged using the Gram-Schmidt spectral pan-sharpening to provide a multispectral data-set in four bands at 0.6m/pixel.

- The first one was shot over NagaMas and NagaSakti area in june 2004, covering 8km in latitude and 10km in longitude. Some of the blocks are masked by clouds and their shadows. Due to the context of the preliminary feasibility studies in this site, and a lack of time, this image was neither deeply processed nor analyzed. This is thus only a radiance data set, not orthorectified.
- The second one was shot over Padang Halaban Estate (PHLE) in june 2008, covering 8.5km in latitude and 9km in longitude, and comprising a lot of big clouds and their shadows that mask some areas (cf. Figure 1). GPS measurements of the altitude and geographical coordinates of a large sample of referable points in the estate allowed to produce a local Digital Terrain Model and thus to orthorectify the image. In addition, spectral measurements on invariant points were used to calibrate the data in reflectance using the indirect method.



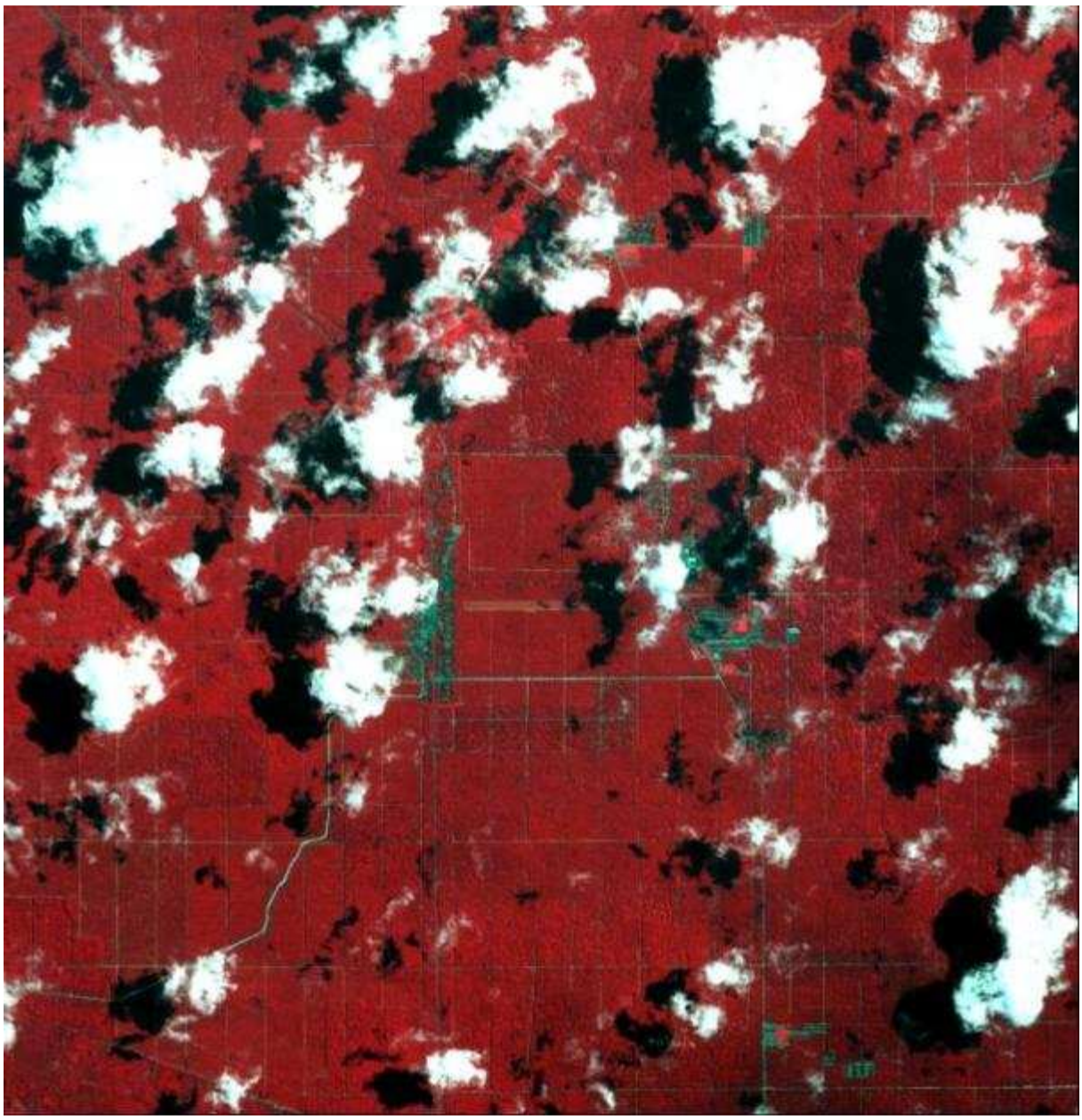


Figure 1 : global view of the Quickbird image acquired over Padang Halaban Estate in june 2008. Top: true colour (blue-green-red) composition, bottom: false-colour (green-red-near infrared) composition.

Figure 2 displays an example of zoom-in view inside several oil-palm blocks, and illustrates the gain in spatial information that allows the pan-sharpening providing multispectral images at the resolution of 0.6m/pixel. The individual trees are clearly identified, and separated. Even the shape of the tree crown and its shadow are distinguishable. Small details are observable and plot limits appear clearly. This level of details allow the understanding of colour heterogeneities seen in the basic image, like the variations of planting vs. canopy density, or the delimitation and discrimination of clearings due to missing tree(s) and smaller gaps due to tree crown reduction. It also makes the pixels spectrally purer.

The counter part of this higher amount of information is the introduction of a wider range of spectral heterogeneity, especially for a single tree itself, mostly due to directional effects, shadowing contribution, and parasite signals, and an appearance of spatial incoherence of the reflectance. In this context, traditional methodologies of image signal processing are not suitable and we had to test several new directions to extract, for instance, the individual oil-palm tree mean reflectance. In the limited time dedicated for this project, no automatic processing was found performing to delimitate properly the oil-palm crown and we finally digitalized the trees as 9m diameter circles, with a computer-assisted photo interpretation referring to different scales to manage the centre location fixing. Considering the difficulty and time-consumption associated to this

processing, we are not able to provide an operational tool for that purpose, and this digitalization was only achieved on sampled trees useful for the methods development. Examples are shown at Figure 3.

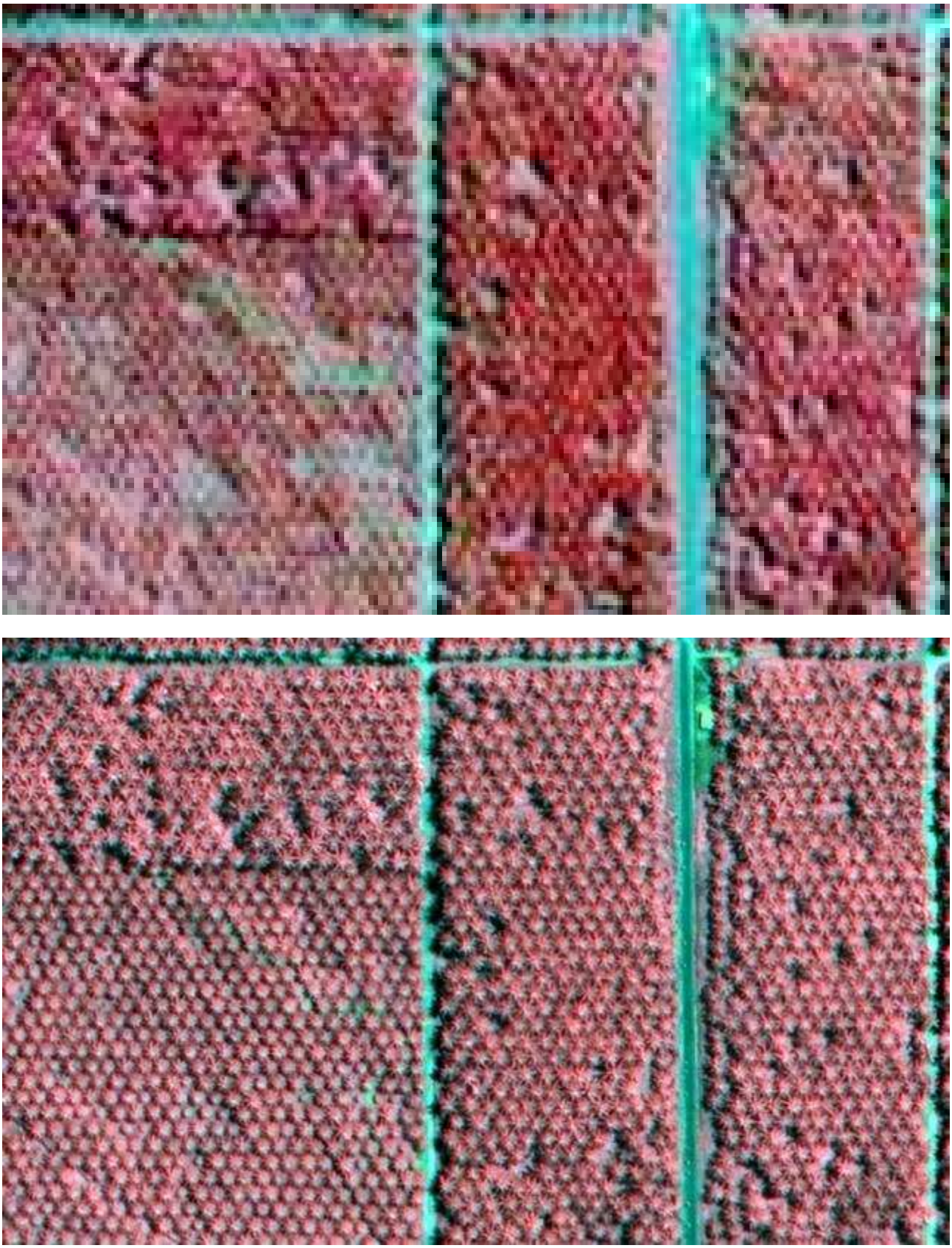


Figure 2 : zoom in the Quickbird image acquired over Padang Halaban Estate in june 2008, showing in false colours (green-red-near infrared) the gain of information from the basic multispectral data at 2.5m/pix (top) to the pan-sharpened multispectral data at 0.6m/pix (bottom). Individual oil-palm trees are clearly indentified.

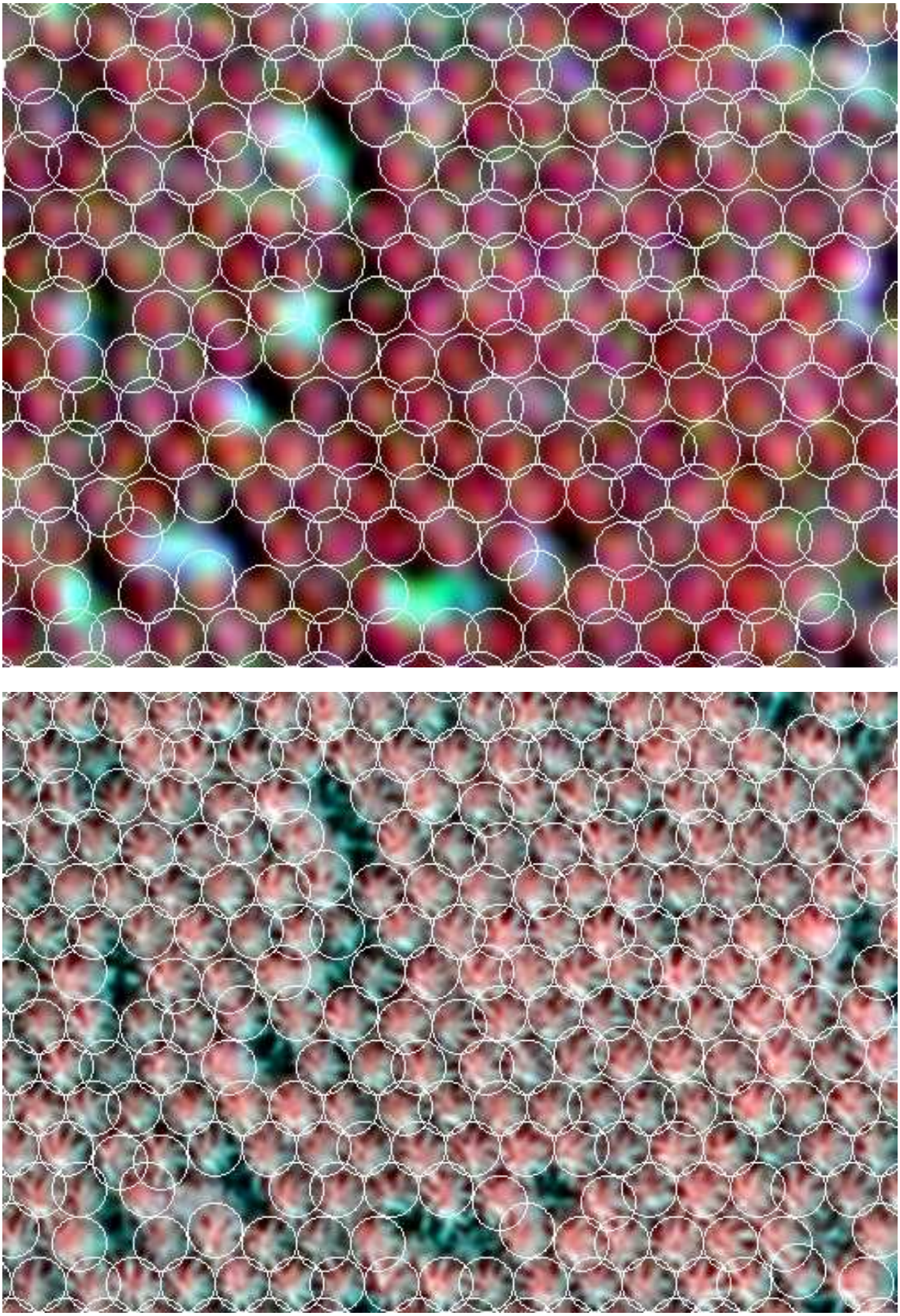


Figure 3 : example of digitalized oil-palm tree crowns, respectively displayed in the basic multispectral image at 2.5m (top) and in the pan-sharpened multispectral image at 0.6m (bottom).

3.2. Hyperspectral Reflectance data

The instrument used for the measurements of the oil-palm tree and leaf reflectance spectra was a field spectroradiometer called UNISPEC'2001, designed by PP-System. It acquires a spectrum in 256 wavelengths comprised between 303nm and 1131nm, with the help of a fibre-optics.
We have acquired the spectra at two different scales: the canopy and the leaf, designing step by step an innovative protocol for each scale that was adapted to the oil-palm context.

3.2.1. Tree canopy reflectance

The fibre optics(FOV: 26°) is fixed with an incident angle of 40° on a home-made inertial device mounted on a wooden stick (perch) along with a cosine receptor dedicated to measure the solar diffuse incident light simultaneously with the reflected light to allow the reflectance computing. The operator climbs up thanks to scaffoldings to reach the top of the canopy and to maintain the fibre-optics at 1m above the canopy to get an integrated surface reflectance. Then he makes an acquisition close to the centre of the tree crown, and repeats the acquisition 6 times all around the tree (with the obligation to move the scaffoldings). These repetitions aim at avoiding the directional effects, decreasing the small local heterogeneities, and giving a mean spectrum of the total crown surface.

Figure 4 : view of canopy-reflectance measurement operators climbed on scaffoldings to reach the top of the oil-palm canopy, and closer view of the fibre-optics mounted on the inertial device.

3.2.2. Leaf reflectance

The fibre optics is fixed on a leaf-clip (cf. Figure 5), allowing the use of an internal calibrated light source and a close measurement of the reflected light in a little spot of about 5 square millimetres. During the first experiments, and for all the deficient trees, 5 leaves were cut and measured: L1, L5, L9, L13, and L17 (cf. Figure 4), to encompass the main variability of composition and structure gradients through the canopy layers. After some preliminary spectral analysis, it appeared that ganoderma discrimination worked better on younger leaves and since then only L5 and L9 are cut to avoid too much destruction of the tree and to fasten the protocol. These leaves were selected due to several reasons:

- stress symptoms are visible on at least one of those leaves,
- they are young and are not covered by lichen (as old leaves are) that would interfere in the reflectance measurement,
- they are located on the top of the tree and are visible from above, so the measurements can be closer to satellite remote sensing data,
- they are located on opposite sides of the tree, whatever could be the angle between leaves,
- they belong to two different but adjacent spires,

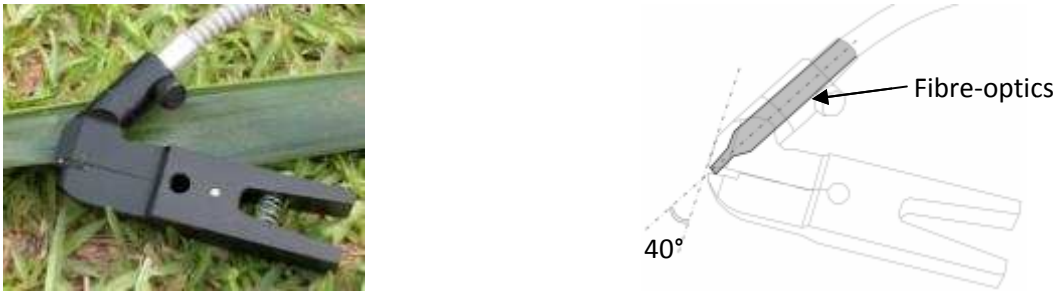


Figure 5 : Leaf-clip ready for an oil-palm leaflet reflectance measurement (left) and schematic view of the mounting of the fibre-optics in the leaf-clip (right).

In the objective to get a mean and representative spectrum of the whole leaf, several leaflets are sampled on each cut leaf: f10, f35, f55 (close to the B point where foliar analysis are used to be processed), f70, and f85 (cf. Figure 6) to cover the most of the composition and structure variability along the leaf, especially due to mineral gradients and insertion angles. Indeed, we observed an actual variation of the single spectrum of individual leaflets, which has to be averaged along the whole leaf.

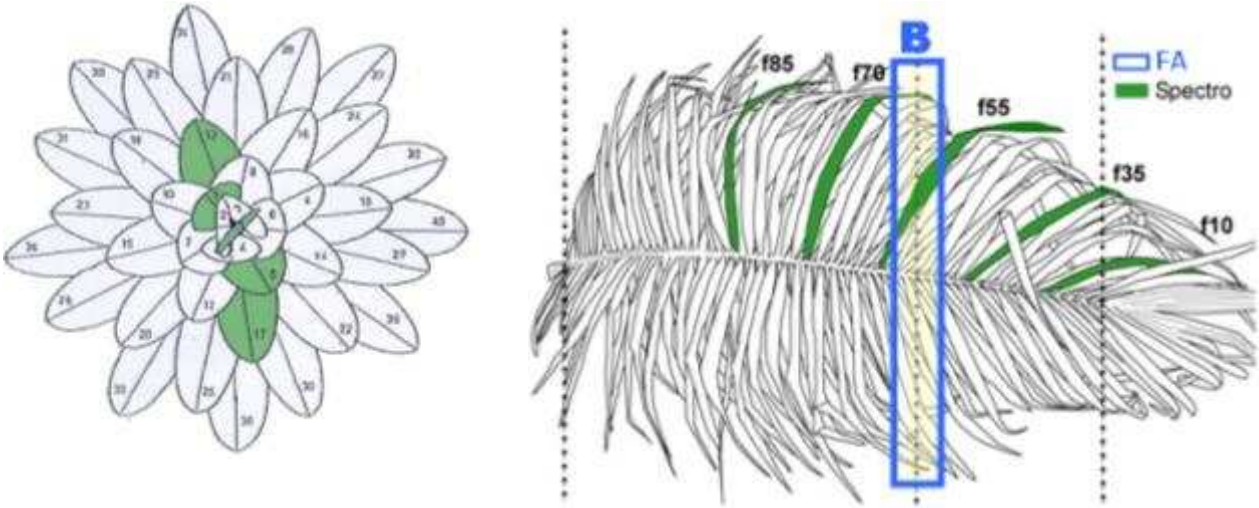


Figure 6 : location of the selected leaves (left) and leaflets (right) for the spectral measurements

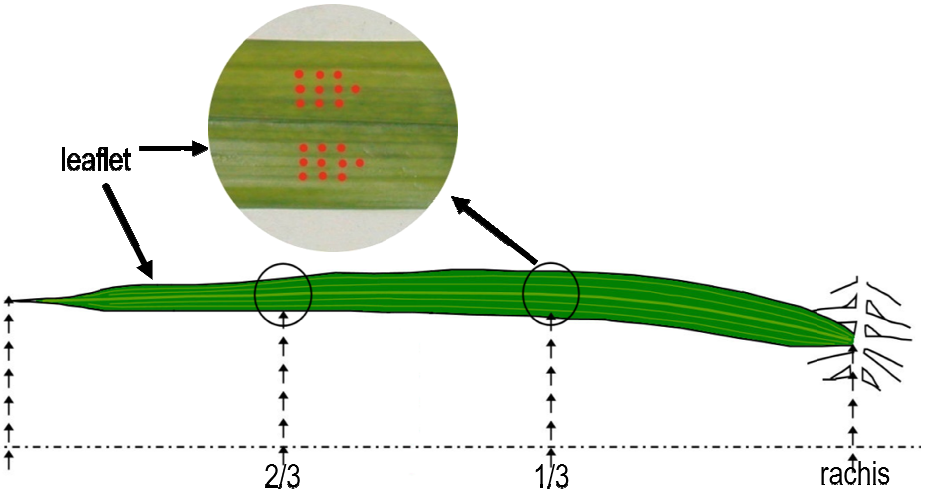


Figure 7 : location of the spectral acquisitions on the leaflets (20 at the first and 20 at the second thirds of its length), averaged to provide the representative leaflet reflectance

On each leaflet, a set of 40 spectral acquisitions are operated and average to provide the leaflet representative spectrum. These acquisitions are equally distributed inside an area of 5 square centimetres, at the first and the second third of the leaflet length and on both sides of its central nerve (cf. Figure 7). This protocol was established after a statistical analysis of the spectral variability and representativeness of a wider quantity of acquisitions scattered over the whole leaflet. Finally, these five leaflet spectra are averaged to provide the mean leaf spectrum.

3.2.3. Database

The database includes the tree canopy reflectance spectrum and the L5 and L9 (plus L1, L13, and L17 for the deficient trees and the trials) leaf reflectance spectra for all the trees which characteristics are listed in Table 1.

Oil-palms with visible deficiency	K (Kandista)	8
	Mg (Kandista)	5
	Fe (Kandista)	5
	N (Kandista)	5
Oil-palm in a nutritional trial	N and P (NM&NS)	12 (6 trials)
	Mg and K (NM&NS)	12 (6 trials)
Identified genotypes	1.5 times the standard mineral treatment (NM&NS)	16 (8 different genotypes)
	0 times the standard mineral treatment (NM&NS)	16 (8 different genotypes)
Ganoderma disease attacked oil-palm	Different unidentified levels (Kandista)	6
	Sane trees (Kandista)	6
	Sane trees (PHLE)	36
	Level 1 (PHLE)	18
	Level 2 (PHLE)	38
	Level 3 (PHLE)	3

Table 1 : List of sampled trees for reflectance spectrum measurements.

3.3. LAI measurements

LAI was measured at two different scales: the individual tree, and the block. Different tools were used to be compared or to help their calibration: the destructive method for the individual tree LAI evaluation, and LICOR-LAI2000 Plant Canopy Analyser (LAI2000-PCA) measurements and hemispherical photographs acquisitions for the two different scales.

3.3.1. Direct method, or destructive method

32 trees were sampled, representing 4 vegetal material and 20 different plantation dates (years), as shown in the Table 2. The destructive method of the individual oil-palm tree LAI estimation consists in six steps.

1. the canopy projected area is estimated on the basis of the measurement of the mean distance from the stem to the leaf extremity and the stem perimeter to derive its diameter. The total number of leaves is also counted and the stem height is measured.
2. 3 leaves of distant ages are cut, to encompass the possible but small variability of leaf area, and provide a value for the mean leaf area.
3. one leaflet is cut each 6 leaflets, respectively on each side of the rachis (so a total of about 40 to 50 leaflets per leaf), and each one is numbered and associated with its insertion distance on the rachis.
4. the leaflet five specific widths and heights are measured for each leaflet, and the leaflet area is computed on the basis of Tailliez and Koffi (1992) geometrical model.

5. the total leaflet area is then estimated for each leaf with an interpolation for the unsampled leaflets depending on their location along the rachis, and is average over these three leaves to provide the mean leaf area.
6. the mean leaf area is finally multiplied by the number of leaves in the tree and divided by the crown projected area to give the tree LAI.

Genetic material	Number of sampled trees	Number of different plantation dates / number existing for the genetic material
Socfindo	16	9 years/ 13
Marihat	10	5 years / 7
Dami	4	2 years / 4
Costa Rica	2	1 year / 1

Table 2 : Description of sampled trees for the LAI destructive measurements

3.3.2. LAI2000-PCA

For the estimation of the canopy LAI with the LICOR instrument called LAI2000-PCA, two devices are required (cf. Figure 8): one being fixed in an open area closed to the fields of interest and continuously acquiring the diffuse incident light (A), and the second being mobile and used for the diffuse light below the canopy (B). "A" acquisition are programmed each 15 seconds. The two data sets are then merged afterwards, using the dedicated C2000 software to associate each B acquisition to the closer A acquisition in time.

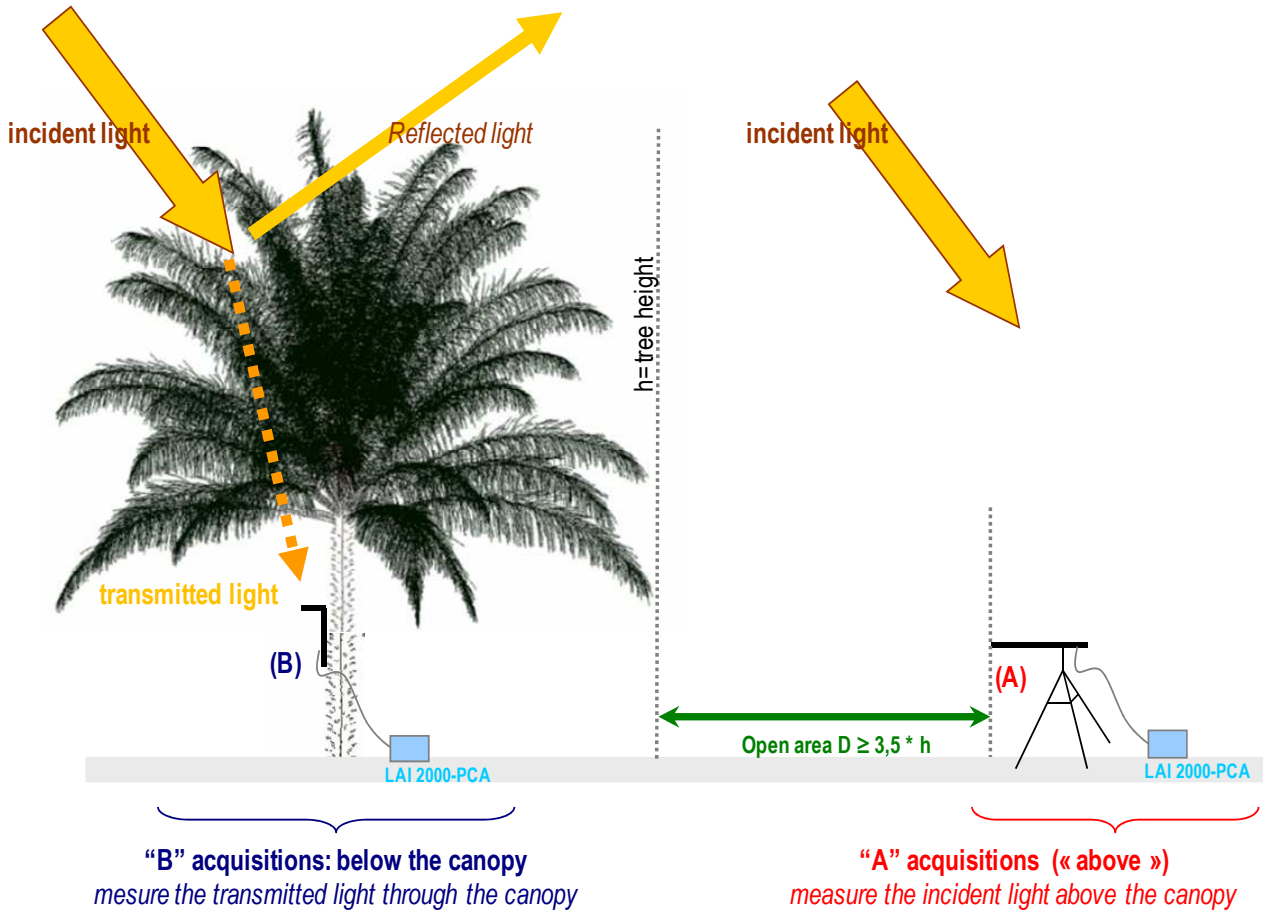
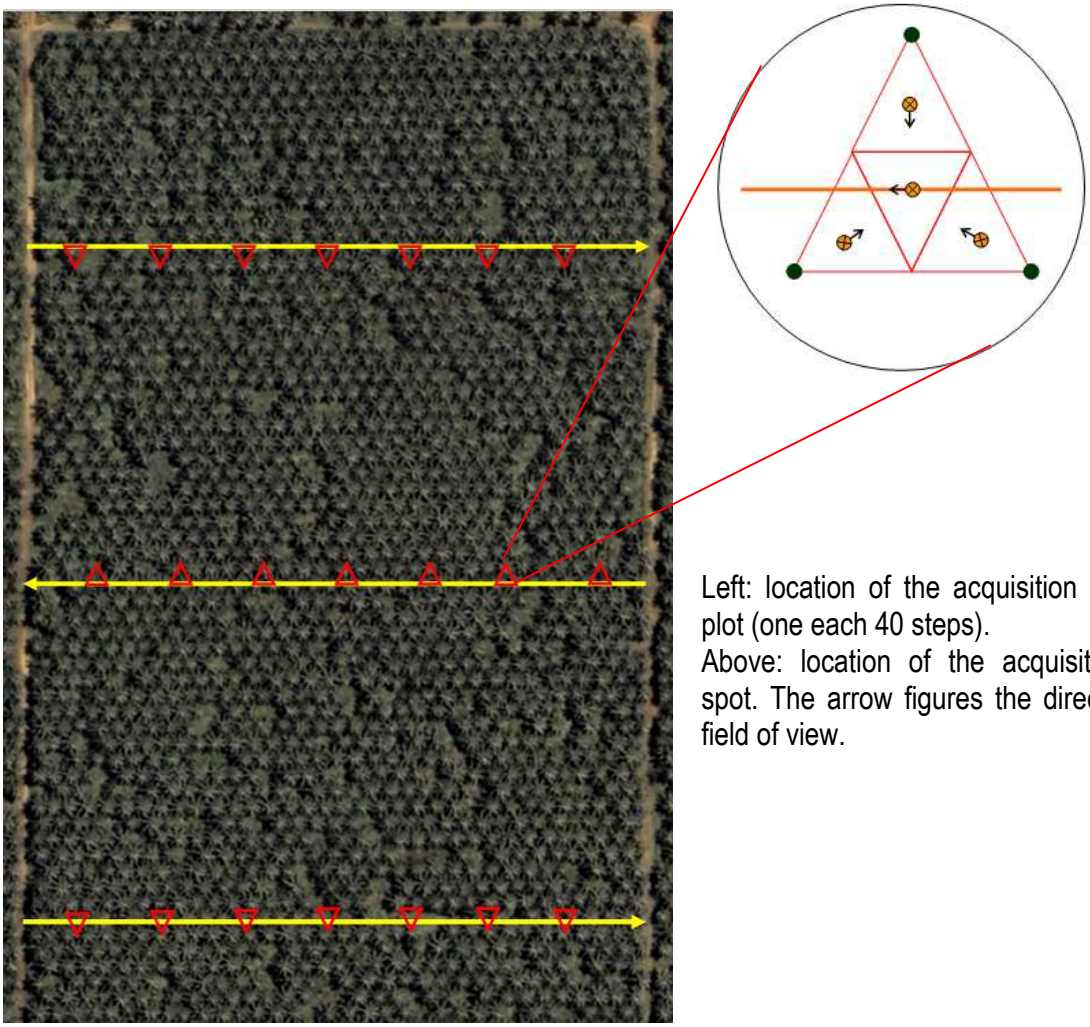


Figure 8 : principle of the LAI2000-PCA measurements for an individual oil-palm tree

A cap masks the lens of the instrument so that the field of view is only 45°, to discard the contribution of the operator (and of the tree stem for the individual tree measurements) during the acquisition.

1) For the estimation of the individual tree LAI, the operator places the stick of the instrument against the stem and turns around the stem in order to make 12 regularly spaced “B” acquisitions, covering the whole field of the tree crown (cf. Figure 8). This has been achieved for the 32 trees referenced in the Table 2, before the leaves were cut, plus 388 others scattered over the different blocks, so a total of 420 individual oil-palm trees were measured.

2) For the estimation of a given block mean LAI, several protocols were imagined, tested on a given set of plots, and analyzed. Only the more relevant is detailed here, being the easiest to achieved, with 150 acquisitions per block in average and about 90 minutes of work, and allowing fixing the block variability in any configuration of the canopy structure, including with the presence of large gaps (clearings). 40 blocks were measured that way.



Left: location of the acquisition spots inside the plot (one each 40 steps).
Above: location of the acquisitions inside this spot. The arrow figures the direction of the lens field of view.

Figure 9 : illustration of the protocol followed for the estimation of a transect mean LAI with the LAI2000-PCA

3.3.3. Hemispherical Photographs

A Nikon Coolpix digital camera equipped with a fish-eye lens was used to acquire the hemispherical photographs in 360°. This camera must be horizontal, with its lens exactly at the zenith. Several pictures are acquired for the same object (either tree or transect), and processed simultaneously with the CanEye software to provide the mean gap-fraction and its standard deviation.

1) For the individual trees, 6 pictures are acquired, regularly spaced all around the stem like for the LAI2000-PCA measurements, for any of the 420 trees.
2) For the transects, 1 picture is acquired at each location of a LAI2000-PCA “B” acquisition (4 for each acquisition spot, refer to the previous section), for each of the 40 blocks.

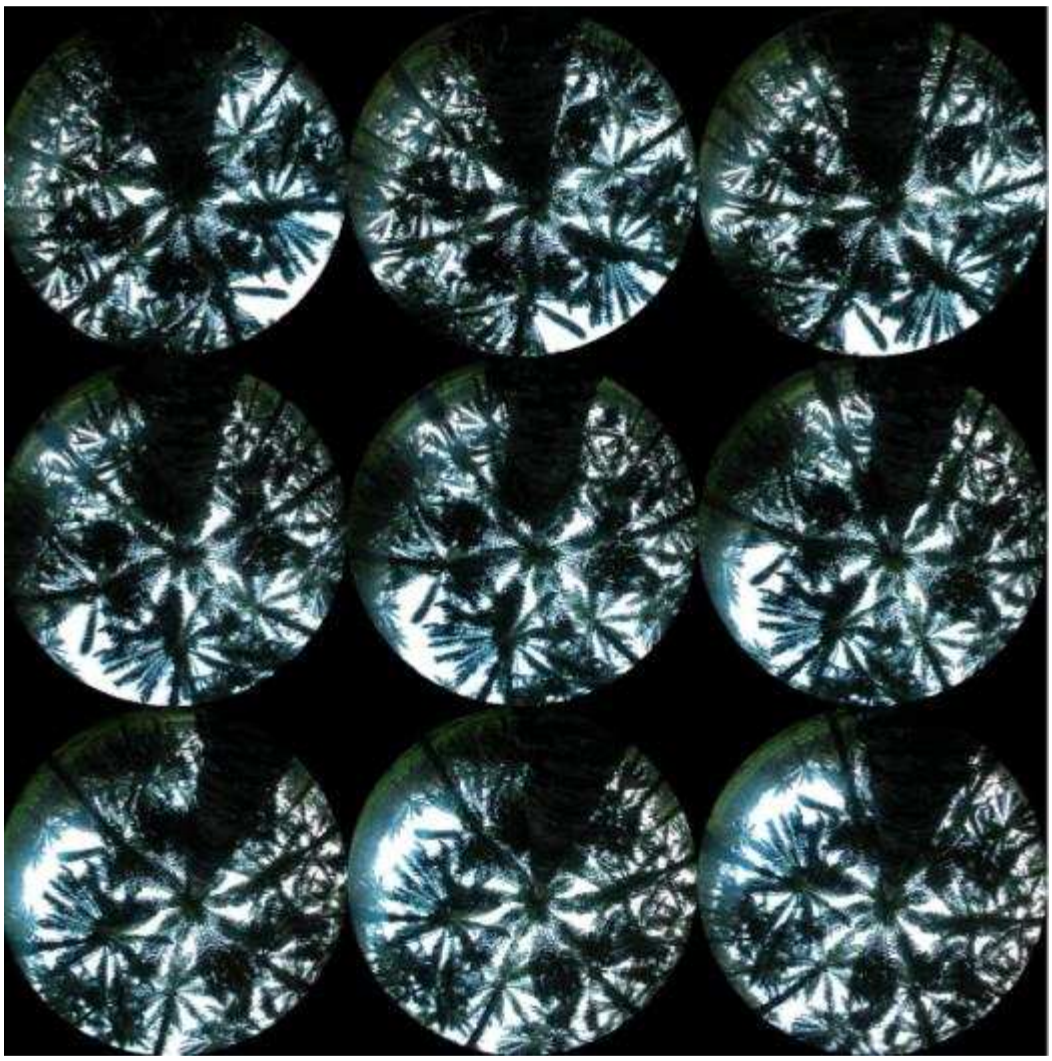


Figure 10 : Example of a set of 6 hemispherical photographs acquired for n individual oil-palm tree.

4. Results for LAI estimation

Among the 420 measured individual oil-palm trees, only 60 are actually visible on the satellite image, due to the presence of clouds and their shadow that mask parts or the image. These 60 trees were digitalized to extract their mean radiance in respectively the red and the near-infrared bands, to derived their mean Normalized Difference Vegetation Index NDVI=(R_{PIR}-R_{red})/(R_{PIR}-R_{red}), known in remote sensing to be correlated to the LAI. Then, a relationship between the LAI and the NDVI was derived by cross-validation on a set of 40 individuals as the calibration data, and 20 individuals as the validation data, randomly selected by 28 successive draw-lots. This modelling was achieved four times, using respectively the destructive LAI, the Lai obtained thanks to the gap-fraction estimated on the hemispherical photographs, the LAI2000-PCA LAI, and this later corrected by a clumping factor derived from the hemispherical photographs analysis. The model converges with a solution without a large error only for the basic LAI2000-PCA measurements. The final relationship is then the following**: $LAI_{(tree)} = 21.3 \times NDVI_{(tree)} - 9.3$**, having a determination coefficient of 0.34, and a correlation coefficient of 0.58, with a root mean square error of 0.9.

Among the 40 blocks measured with the transects, only 21 are visible in the satellite image. The mean NDVI of each of these blocks was thus computed on the basis of the mean reflectance in the red and the near infrared, respectively. A relationship was then derived between the LAI2000-PCA values and the NDVI**: $LAI_{(block)} = 25.16 \times NDVI_{(block)} - 11.6$**, having a determination coefficient of 0.57 and a correlation coefficient of 0.76, with a root mean square error of 0.5. No validation was performed on this relationship due to the small amount of samples.

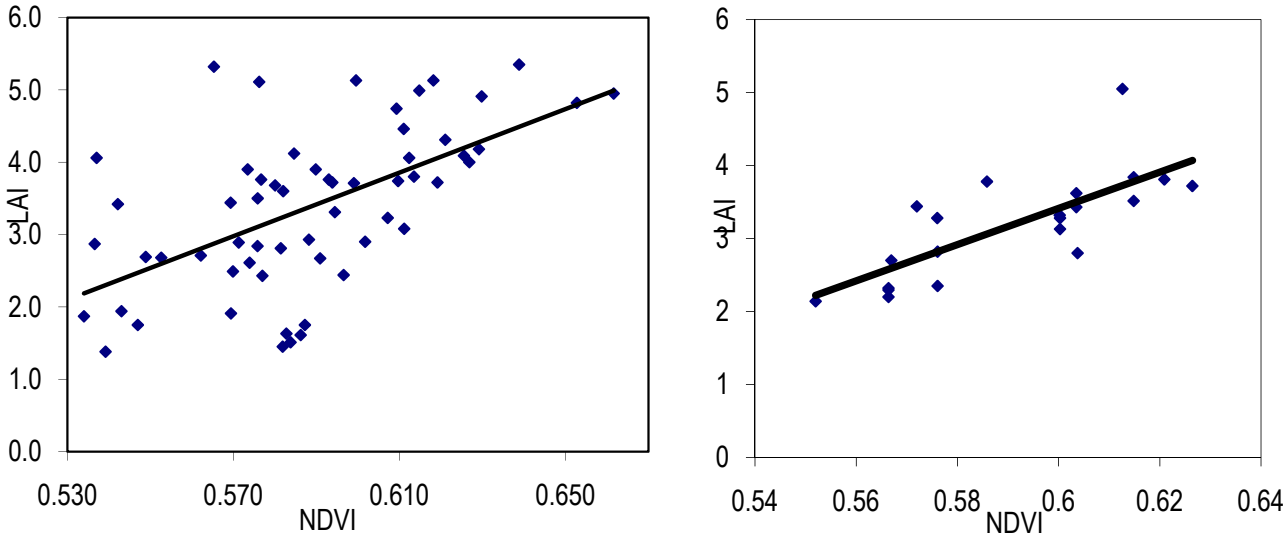


Figure 11 : graphs displaying the measured LAI (with the LAI2000-PCA) as a function of the NDVI (derived from the satellite image) for the individual trees (left) and for the blocks (right).

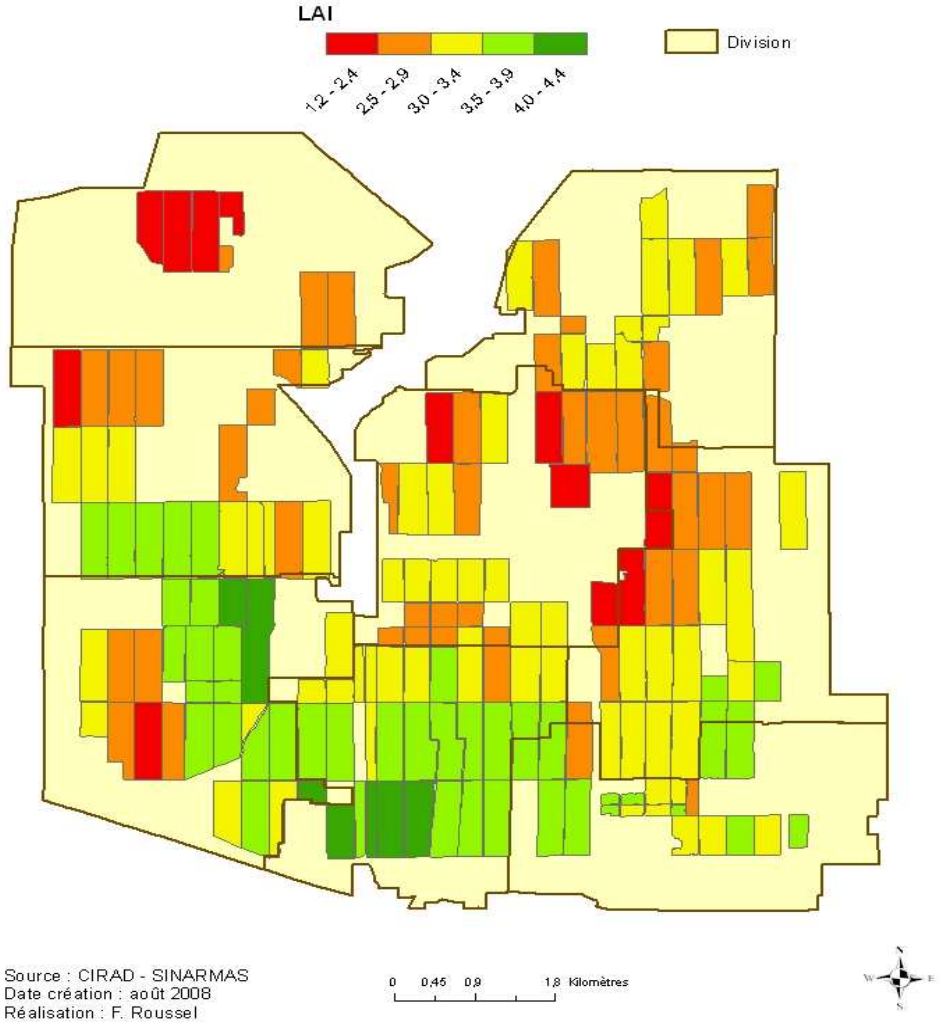


Figure 12 : Map of the mean block LAI of Padang Halaban Estate in june 2008.

The relationship obtained for the blocks was then applied to any of the blocks contained in the image, to provide the whole estate LAI map (cf. Figure 12). This product is very easy to obtain from any new satellite acquisition, but is not fully validated. To be operational, additional measurements should be performed

following the same transect protocol, to validate the derived LAI value on several blocks that were not used in the calibration of the relationship. This map allows for instance to quickly analysing the interblock variability of LAI, to be related to any agronomical data (e.g. soil, planting density, topography, and so on) to be understood, or to comparing different plantations.

The relationship obtained for the individual trees was applied to any of the trees digitalized in a test-block: Blok5-Divisi5 (Socfindo genetic material, planted in 1996) (cf. Figure 13). 2748 oil-palm trees were manually digitalized using photo-interpretation: this is far from an operational tool and other techniques have to be developed to make it easier and quicker to perform. But this is a very useful product, giving with a good precision the local values of LAI and thus displaying the intra block heterogeneity, to be analysed for precision farming for instance.

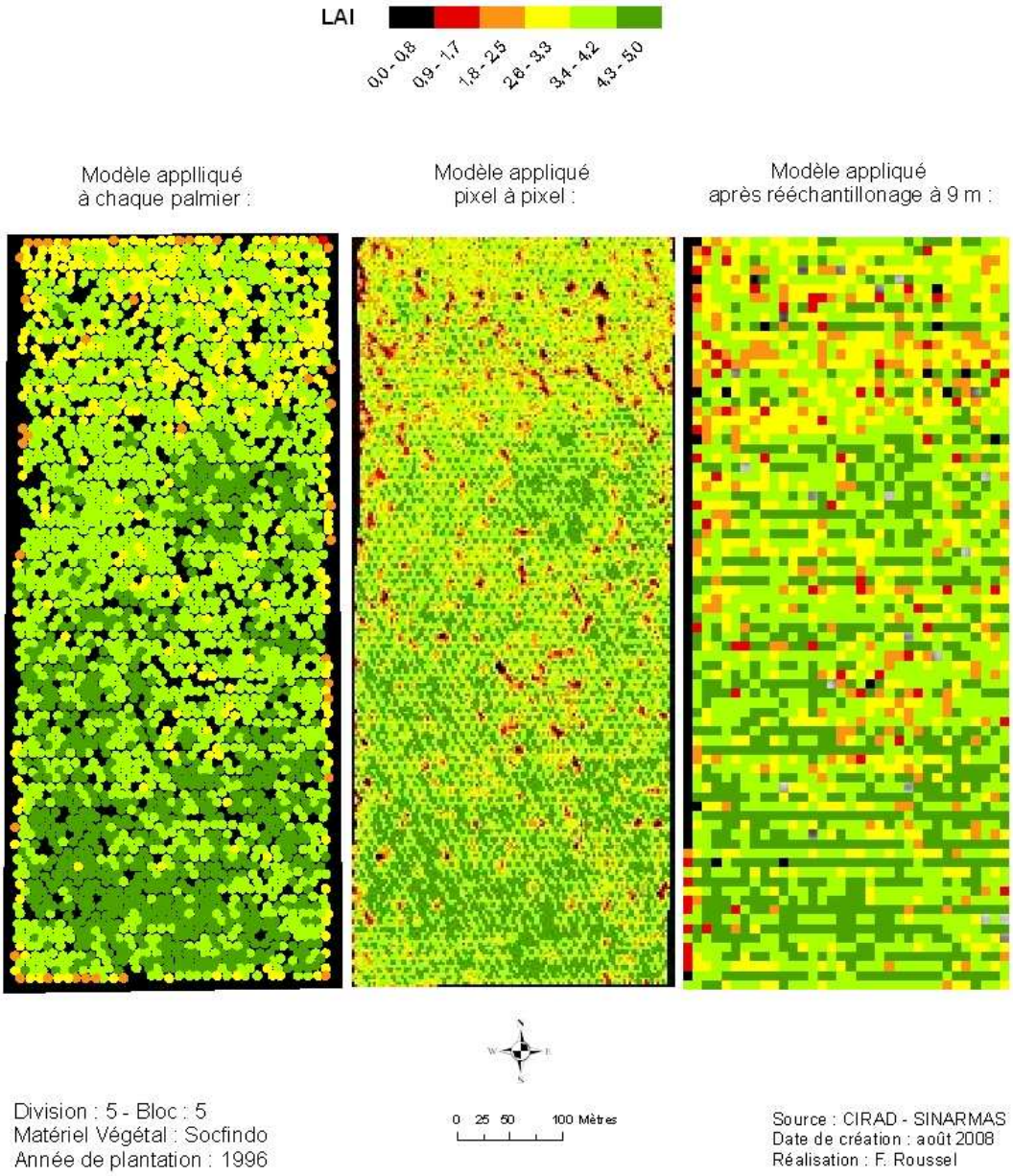


Figure 13 : Map of the oil-palm tree LAI in Blok5-Divisi5 of Padang Halaban Estate in june 2008.

So, we studied the robustness of this relationship, calibrated on objects being the whole trees, if applied to single pixels of 2.5m in dimension, i.e. objects smaller than the one it was designed for. Figure 13 shows the result of the application of the model to each pixel of the block, and for pixels over-sampled to the size of a tree (ie. 9m/pix). It shows that the variability and heterogeneity areas are stable from one map to the other, and that LAI values are quite well preserved even if a little bit overestimated closed to missing trees. This

overestimation is even more critical when dealing with over-sampled pixels at 9m, along with the loss of spatial information too: this sampling is thus definitely not valuable. Therefore, the individual tree relationship between LAI and NDVI can be reasonably applied to each pixel of an image, making the process considerably quicker and more operational, even if it should be more validated to be transferable. It was thus applied over the whole image, as presented at Figure 14, to provide a LAI map at 2.4m resolution.

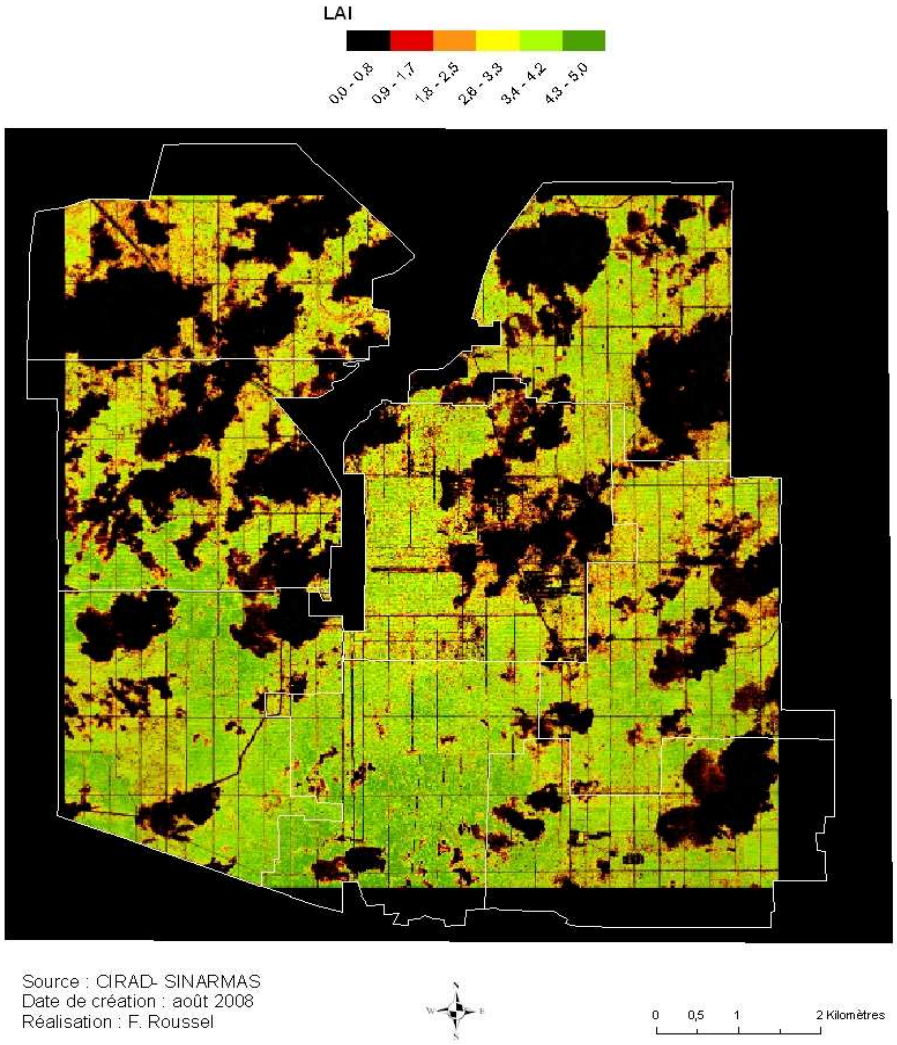


Figure 14 : LAI map at 2.5m spatial resolution of Padang Halaban Estate in june 2008

This map (cf. Figure 14) shows the limit of this application in the tropical areas, were the cloud cover is difficult to avoid and thus a large part of the plantation might be missing in the map. But this is also a digital product that can be integrated in a Geographical Information System of the plantation. It allows the fast and accurate location of any problem affecting the canopy density or the accurate identification of the under producing focuses, and can help the stratification of a field sampling. In addition, it can be integrated in a functioning model to estimate the yield at the tree scale.

5. Results on nutrition deficiency

This work lays on the observation that strong and isolated deficiencies provoke a noticeable change in the spectral signature of the oil-palm leaf. Figure 15, for instance, shows the reflectance measured during this project on oil-palm leaves that present a single deficiency, established by the foliar analysis, compared to a leaf without any deficiency. A strong decrease of the absorption in the green to red domain (520-670nm), corresponding to a higher reflectance, clearly appears, sign of a fainter photosynthetic activity. Also, the shape of the absorption band in this domain is very different from one deficiency to another, with a shifting of the green maximum around 550nm and a variation of slopes. So, it might be possible to detect or even to

discriminate the different deficiencies, especially if they are pure. Nevertheless, the problem is: how to decipher this information to provide an efficient diagnostic tool? And will it be performing also when the deficiencies are mixed, which most often occurs?

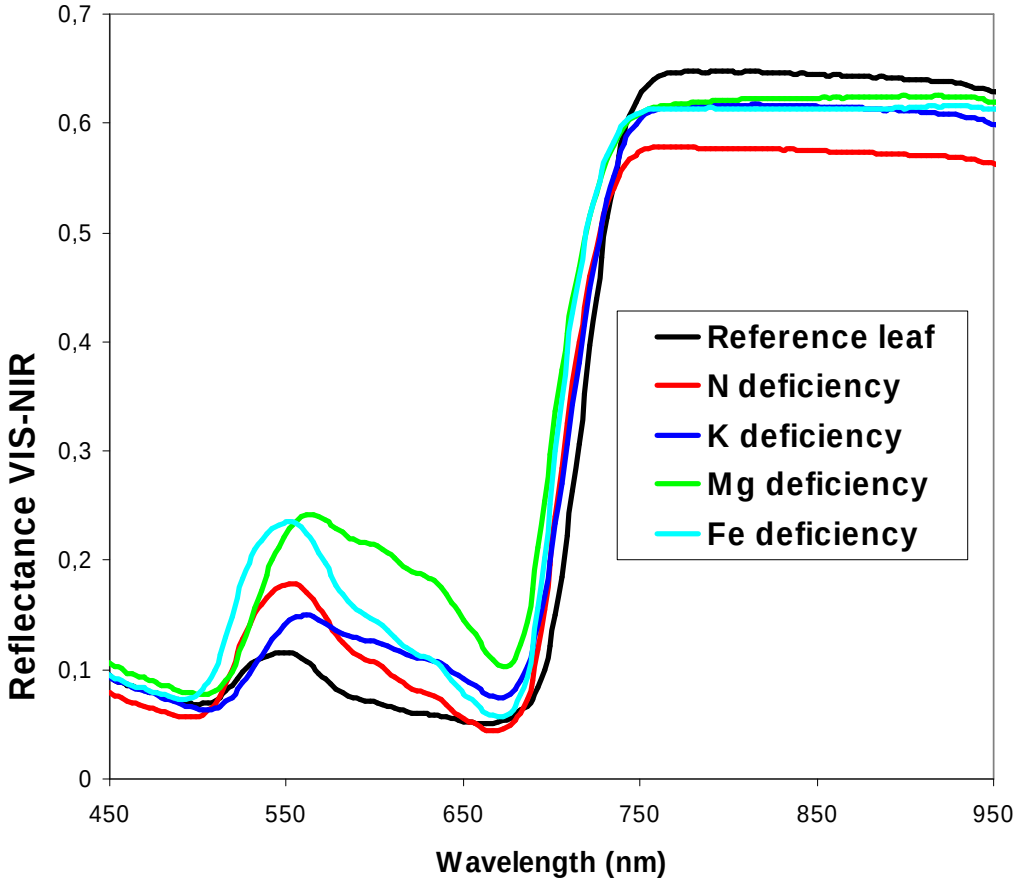


Figure 15 : Reflectance spectra of oil-palm leaves affected by a single nutritionnal deficiency.

To answer to these questions, several chemiometrics methodologies were tested on the whole data set of leaf reflectance acquired in the fields and associated to foliar analysis providing the whole description of organic and mineral composition of the leaf. The tested methods include, for instance, partial least square regressions, principal component analysis, factorial discriminant analysis, stepwise procedure, multiple linear regressions.
Different levels of data pre-processing were also tested: raw spectra, filtered spectra, derivative spectral, and 50 descriptive variables derived from the spectra (e.g. local extremes, feature area, slopes, curvature indices, spectral indices...). Best results were obtained on the second derivative spectra pre-processed by the Savitsky-Golay filter.
No model of [K], [Mg], or [Fe] prediction converged to a solution: it is thus impossible to estimate the concentration in these three components.
For [N] and [P], the best model was obtained using the partial least-square regression (PLS) as a concentration predictive tool, using five components, and resulting in a linear combination of the reflectance in 98 spectral bands. The root mean square error remaining with the model of prediction of [P] is 0.02%. At this date, no information was provided by the agronomists on the [P] level considered as a deficiency to evaluate if this error is acceptable or not, so we couldn't conclude on the potential use of this relationship. For [N], the root mean square error is 0.34% (so about 10% of relative error), which is quite accurate in a general point of view. Nevertheless, as communicated by agronomists, the acceptable nitrogen concentration in oil-palm leaves is about 2.9%, while it is considered as stressed when it reaches 2.7%. The oil-palm nitric stress threshold is thus of 0.2%, which is lower than the reached 0.34%. Therefore, it is possible to estimate [N] based on reflectance data, but not accurate enough to detect a deficiency or a stress.

Another direction of research was to evaluate the possibility to detect a nutritional deficiency, whatever the cause. The PLS was then used as a first step in a discrimination analysis, consisting in a factorial discriminant

analysis applied on the scores of the PLS with the reflectance spectra as the inputs and the scores 0=no deficiency and 1=existence of deficiency as the outputs, merging all the occurring deficiencies. The result is less than 70% of correct classification of leaves with a systematic false alarm, i.e. non deficient leaves are always classified as deficient. In conclusion, it is not possible to make the diagnostic of a nutritional deficiency based on the leaf reflectance.

Finally, the potentiality of detecting any nutritional deficiency out of the multispectral image was tested, digitalizing the sampled trees and observing their relative radiance. But no statistical model converged to estimate the nutrient concentration. And no discrimination was possible between the different stresses.

6. Results on Ganoderma

We first analyzed the potential of reflectance data processing to detect the disease, and then if it is also able to discriminate between different levels of severity of the disease.
At the leaf or leaflet scale, no statistical model converges, with negative R^2 and a RMSEP highly superior to 1, which shows that no relationship can be found between the disease presence or level of infection and the leaf spectrum. Indeed, at this scale, the spectral signature is not too variable with the increase of the disease level (cf. Figure 16). This might be due to faint leaflet compositional modifications caused by the disease, in regards with the canopy structure alteration that can be provoked by cumulated stresses caused by the disease. So, it is not surprising that no information can be got at this scale while it is available at the tree scale, like exposed hereafter.

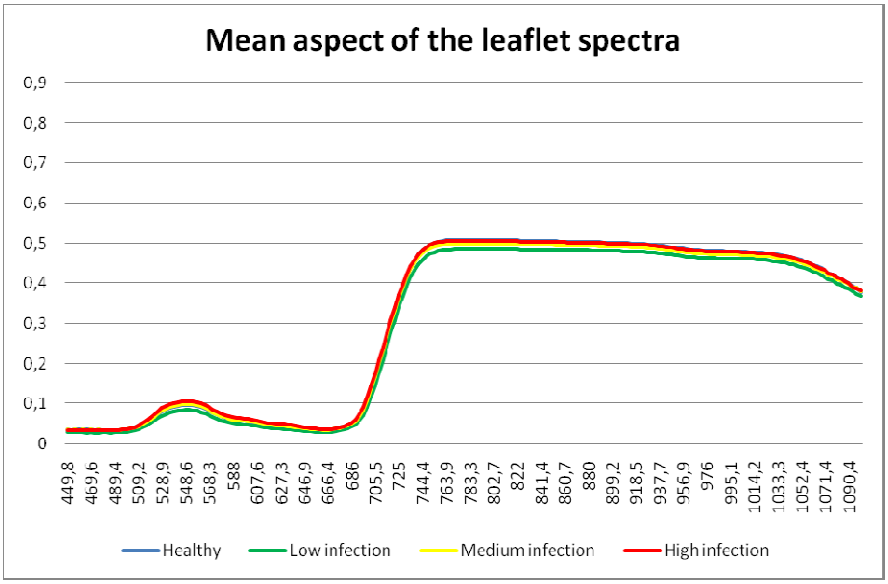


Figure 16 : Comparison of the mean leaflet spectra of each ganoderma infection level class

At the tree canopy scale, the analysis focused on the 96 trees sampled in PHLE, fully documented in terms of ganoderma infestation. In the detection procedure, the score 0 was attributed to sane trees and the score 1 attributed to sick trees, whatever the level of the symptoms. In the infestation level discrimination procedure, the score 0 was given to sane trees, and 1, 2, and 3 respectively to the three identified levels.
In both classifications, the second order derivative of a third degree polynomial fitted on a smoothing window of 26nm Savitsky-Golay filtering gave the best result. The small smoothing window size indicates that spectral features associated with the disease severity are located in small spectral ranges, showing the great advantage of hyperspectral data compared to broadband data, inside which the Ganoderma signature might be faded out.

1) The resulting model for sane/sick status classification is a linear discriminant analysis applied on the 6-latent-variables-PLS scores. It allows a very good discrimination between the healthy and the sick trees, with only 2% of error (98% of global precision), corresponding to one “false alarm” that means that a healthy tree is found sick, and one “detection failure”, i.e. a sick tree is considered as sane. This is very performing.

2) The resulting model for the 4 status classification is a linear discriminant analysis applied on the 7-latent-variables-PLS scores. It presents almost 94% of global accuracy, and no failure dedetection like shown in the confusion matrix (cf. Table 3). The resulting model thus allows a very good discrimination between the healthy and the sick trees, with only 2% of error, corresponding to "false alarm" that means that a healthy tree is found sick. In this case, results are very good because no sick tree is missed, which is the most important issue in the context of disease control. It is thus even better than the previous model considering only sane and sick status. Even though, these false alarms concern only two individuals that are classified as lightly attacked by the disease; it is 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 healthy individuals. In this case, our field estimation of the Ganoderma level of attack was wrong while the hyperspectral reflectance analysis is already able to detect the disease. However, this should be confirmed with a sampling and chemical analysis of some stem tissues before a strong conclusion. On the other hand, these two oil palms did not show any nutritional or water deficiency that could be visually detectable and that could have been a source of modification of the reflectance independent of Ganoderma attack. Errors occurring in the determination of disease severity only corresponds to shifts from level 1 to level 2 and inversely. Considering that the limit between these two scores for *in situ* evaluation is very fuzzy, these errors can be either due to the classification or to the field diagnosis, which is impossible to argue without a chemical proof. Even in the case of an actual classification error, these misclassifications are very few and it allows a good confidence in the overall results. The cross validation process also insures the stability of the model.

The convolution of the transfer functions respectively determined by the eigenvectors of the LDA and the PLSR hence transposes the second derivative reflectance of any newly sampled tree, initially measured as a vector $\begin{pmatrix} R_1 \\ \vdots \\ R_{202} \end{pmatrix}$ in the space defined by the 202 wavebands, into new coordinates (x,y) such as:

$$(x, y) = \begin{pmatrix} b_{1,1} & b_{1,2} \\ \vdots & \vdots \\ b_{7,1} & b_{7,2} \end{pmatrix} \times \begin{pmatrix} a_{1,1} & \cdots & a_{1,5} \\ \vdots & \ddots & \vdots \\ a_{202,1} & \cdots & a_{202,7} \end{pmatrix} \times \begin{pmatrix} R_1 \\ \vdots \\ R_{202} \end{pmatrix}$$

where $a_{i,j}$ are the PLSR coefficients, and $b_{m,n}$ are the LDA coefficients. The first coordinate x then allows estimating the tree degree of sickness (cf. Figure 7):

- if x< –2, the tree is healthy;
- if x > 6 or 7, the tree is dramatically sick, almost dead;
- if –2 < x <6, the threshold between Level1 and Level2 of disease severity is fuzzier and lays between 1.2 and 1.5.

It might still need some improvements to be able to fix the exact edge between Level1 and Level2, but let us remember that even in the field or in the laboratory this limit is not very well defined too.

		Classification result				
	Level	0	1	2	3	% of good classification
Actual status	0	**34**	**2**	**0**	**0**	94 %
	1	**0**	**16**	**2**	**0**	89 %
	2	**0**	**2**	**36**	**0**	95 %
	3	**0**	**0**	**0**	**3**	100 %

Table 3 : Confusion matrix for the classification of oil-palm trees in the 4-level classification.

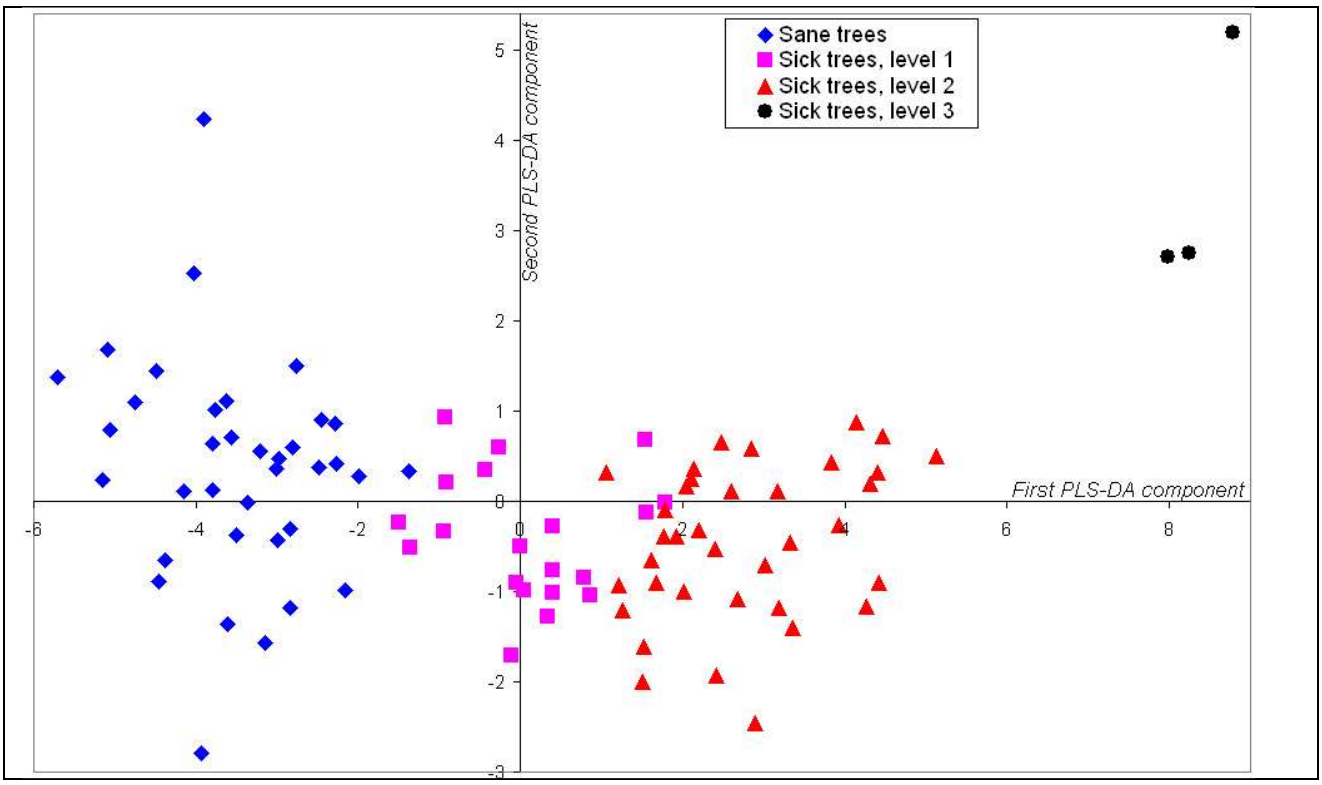


Figure 17 : Representation of oil-palm trees in the plane defined by the two first eigenvectors of PLS-DA: healthy palms are displayed in diamond, Level 1 in square, Level 2 in triangle, and Level 3 in circle symbols. It appears obvious that the first PLS-DA component (abscissa) is an indicator of disease severity, and some thresholds can be fixed to classify any new individual.

Sampled trees in the field were digitalized in the multispectral image, and the same statistical procedures were applied to their reflectance in 4 bands, testing several types of filtering, derivative levels, and PLS models before the DA. But no model converged, either to detect the disease or to discriminate between the different levels of severity. It proves the need of hyperspectral data for providing a diagnostic on the disease.

7. Conclusions

This project showed that multi- or hyperspectral optical remote sensing data can provide with some information on the oil-palm physiological status in some conditions.

1) It is possible to estimate the LAI of either an individual tree or of a whole block using multispectral imagery at very high spatial resolution (<1m) like Quickbird data, with root mean square errors of 0.9 (~20% of relative error) and 0.4 (~4% of relative error), respectively. It lays on a relationship between the NDVI, extracted from the image, and LAI, that can be applied pixel per pixel.

2) A statistical model, applied on preprocessed hyperspectral data acquired in the fields at the tree canopy scale, allows the discrimination of ganoderma-infested trees from sane trees (~98% of accuracy) and even the classification of a sick tree into three levels of disease severity (~94% of accuracy). It thus gives a near-operational diagnostic tool, that will be transferred to PT-SMART during the stay of Doni Artanto Raharjo on March 2012. But it is not possible to make this discrimination on the basis of a 4-band multispectral image like those acquired by Quickbird.

3) A validated model gives the nitrogen and the phosphorus concentrations in the oil-palm leaf based on its hyperspectral reflectance spectrum measured in the field (or laboratory). Errors on these estimated concentrations are respectively 0.34% (~10% of relative error) and 0.02 (~relative error). But it seems to be insufficiently robust to be used as a predictor of stress in the context of oil-palm management. This information is not reachable at the oil-palm scale, and definitely not from 4-band multispectral imagery.

4) No information can be extracted out of hyperspectral data, or out of multispectral images, about the mineral concentration in oil palm trees of leaves.

In conclusion, this study clearly shows the feasibility of developing some non-destructive tools, cheaper than stem tissue or leaflet chemical analyses, and faster to be applied at larger scale, but it still needs some validation and improvement to be actually and efficiently operational. Indeed, present measurements using field spectroradiometer on top of oil palm canopies is still very hard to set up and somehow dangerous, especially when dealing with mature and older trees. It might also be long to perform with a good quality. Acquiring such hyperspectral data from the air would be of major interest to cover a larger area in less time and better conditions. It also pushes for further improvements towards remote sensing applications such as airborne or satellite-borne images analysis. In addition, imagery would add the spatial information, and thus the opportunity to map quickly the location of attacked trees for disease control or to estimate the nitrogen/phosphorus lower concentrations. Nevertheless, the use of imagery is not yet close to operational. New dedicated models would then have to be calibrated for airborne or satellite-borne hyperspectral images, taking into account the imaging specificities (mainly the transfer of scales from trees to canopy). The automatization of the tree segmentation in an image would have to be developed with adapted enhanced image processing tools. The experiments show that simple multispectral images, with only 4 spectral bands, are not efficient to provide any information neither on nutrition nor on ganoderma status. And, finally, even if airborne hyperspectral images are available, they are still very expensive and depend of very good, stable, without wind, and clear from clouds atmospheric conditions that are very difficult to fulfil in Indonesia. Satellite-borne hyperspectral data at a spatial resolution small enough to identify the individual oil-palm trees, on their part, won't be available before 2020 or later!

8. Annex: publication references

8.1. Scientific article published in an Impact Factor Journal:

2010: C. Lelong*, et al.*, "Evaluation of oil-palm fungal disease infestation with canopy hyperspectral reflectance data," *Sensors,* vol. 10, 734-747. OpenAccess at http://www.mdpi.com/1424-8220/10/1/734.

8.2. Conference communications:

- "Evaluation of hyperspectral remote sensing relevance to estimate oil palm trees nutrition," Lelong et al., in *2nd Int. Symposium on "Recent advances in quantitative remote sensing"*, Valencia (ESP), 147-155, 2006..
- "Towards precision agriculture for oil palm mineral nutrition management: Relationships between the Reflectance Spectrum of Oil Palm Leaves and Nutrient Deficiencies," Lelong, Lanore, Caliman, Roger, Syakharosie, in *PIPOC'07: International Palm Oil Congress*, Kuala Lumpur (Malaysia), Paper 1A6, 15 pages, 2007.
- "Discrimination of fungal disease infestation in oil palm canopy hyperspectral reflectance data," Lelong*, et al.*, in *WHISPERS: First Workshop on hyperspectral image and signal processing: Evolution in Remote Sensing*, Grenoble (FR), 4 pages, 2009.
- "Potentialities of very high resolution remote sensing for the estimation of oil palm leaf area index (LAI)," Lelong*,* Roussel, Sitorus, Raharjo, Prabowo and Caliman, in *PIPOC'2009: International Palm Oil Congress*, Kuala Lumpur (Malaysia), Paper AP7, 10 pages, 2009.
- "Developing non-destructive tools for the diagnostic of ganoderma attack-level on oil palms trees: potentialities of reflectance spectroscopy," Lelong, Roger, Brégand, Dubertret, Lanore, Sitorus, Raharjo, Caliman, in *PIPOC'2009: International Palm Oil Congress*, Kuala Lumpur (Malaysia), Paper AP6, 9 pages, 2009.

8.3. Student internship reports:

- M. Lanore, "Evaluation de la pertinence de la télédétection pour estimer l'état physiologique des plantations de palmiers à huile : analyse des relations entre le propriétés optiques des palmes et leur statut nutritionnel," DAA Sciences de l'Environnement, Ecole Nationale Supérieure Agronomique de Toulouse. 2006.

- S. Brégand, "Analyse des comportements statistiques de la réponse spectrale des palmiers à huiles aux stress nutritionnels et sanitaires," Master Pro en Mathématiques Appliquées, UFR Sciences et Techniques de Besançon. 2007.
- F. Roussel, "Estimation du LAI du palmier à huile par télédétection à très haute résolution spatiale," DAA Agrogeomatique, INP/ Ecole Nationale Supérieure Agronomique de Toulouse. 2008.
- F. Dubertret, "Prediction of palm-tree ganoderma affection degree by reflectance spectroscopy: proposed methodology," Césure cycle ingénieur, SupAgro Montpellier. 2009