VALERI 2003 : Haouz site (cropland)

GROUND DATA PROCESSING & PRODUCTION OF THE LEVEL 1 HIGH RESOLUTION MAPS

Marie Weiss

1 Introduction

This report describes the production of the high resolution, level 1, biophysical variable maps for the Haouz site in 2003 (see campaign report for more details about the site and the ground measurement campaign, available at http://www.avignon.inra.fr/valeri). Level 1 map corresponds to the map derived from the determination of a transfer function between reflectance values of the SPOT image acquired during (or around) the ground campaign, and biophysical variable measurements (Hemispherical Images). For each Elementary Sampling Unit (ESU), the hemispherical images were processed using the CAN-EYE software (Version 1.3) developed at INRA-CSE.

The derived biophysical variable maps are:

- Leaf Area Index: two LAI are considered, the first one corresponds to effective LAI derived from the description of the gap fraction as a function of the view zenith angle, the second one (LAI57) is derived from the gap fraction at 57.5°, which is independent on the leaf inclination and is also an effective LAI (does not take into account clumping effect).
- cover fraction (fCover) : it is the percentage of soil covered by vegetation between o° et 10° view zenith angle
- fAPAR: it is the fraction of Absorbed Photosynthetically Active Radiation (PAR=400-700nm). The fAPAR can be defined as instantaneous (for a given solar position) or integrated all over the day. Following a study based on radiative transfer model simulations, it has been shown that the root mean square error between instantaneous fAPAR computed every 30 mns and the daily fAPAR is the lowest for instantaneous fAPAR at 10h00 AM (local time, RMSE= 0.021). Therefore, the derivation of fAPAR from CAN-EYE corresponds to the instantaneous black sky fAPAR at 10h00 AM.

2 Available data

2.1 Sampling strategy

Figure 1 shows that the ESUs locations are well spatially distributed over the 3km x 3km site. According to Benoît Duchemin who organized the campaign and to the processing of the ground data, it has been shown that:

- ESUs A1 and J3 (in black on Figure 1) were located on field borders. As those fields are tight, it is difficult to shift their position in a preferential direction. Therefore, those two ESUs were eliminated
- Considering that SPOT geo-location and GPS measurements are associated to errors, we found that processed LAI for ESU B3 does not correspond to the SPOT pixel in terms of reflectance: the ESU has been shifted of 1 pixel in the wheat field (red colour).

Finally 39 ESUs have been kept for the computation of the transfer function.

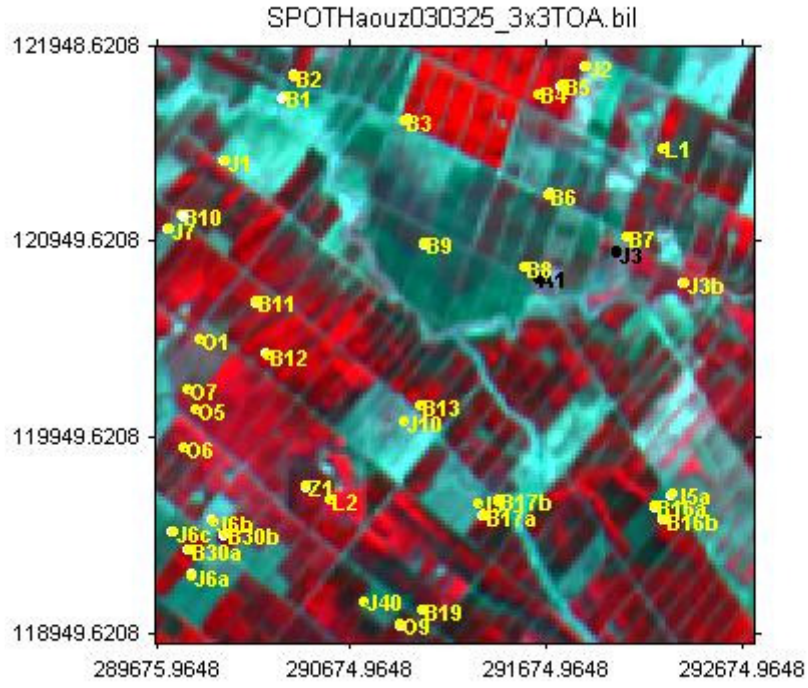

Figure 1. Distribution of the ESUs around the Haouz site. ESUs in black were eliminated for the computation of the transfer function

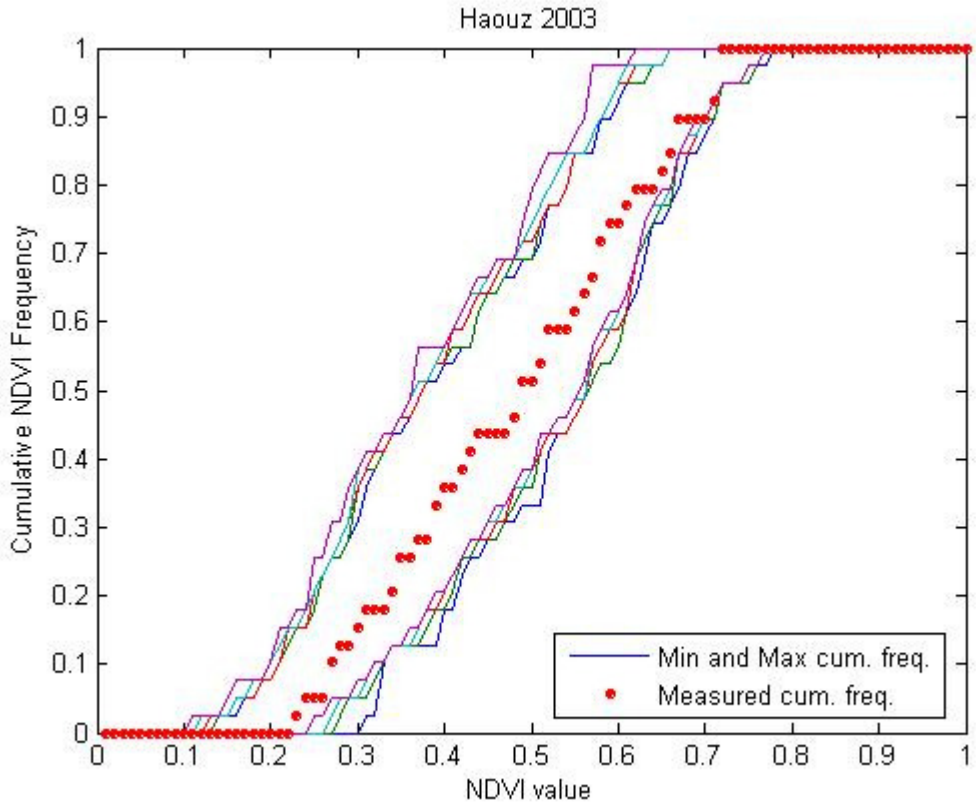

Figure 2. Comparison of the ESU NDVI distribution and the NDVI distribution over the whole image.

The sampling strategy is evaluated using the SPOT image by comparing the NDVI distribution over the site with the NDVI distribution over the ESUs (Figure 2). As the number of pixels is drastically different for the ESU

and whole site (WS=22500 in case of a 3x3km SPOT image), it is not statistically consistent to directly compare the two NDVI histograms. Therefore, the proposed technique consists in comparing the NDVI cumulative frequency of the two distributions by a Monte-Carlo procedure which aims at comparing the actual frequency to randomly shifted sampling patterns. It consists in,

1. Computing the cumulative frequency of the N pixel NDVI that correspond to the exact ESU locations.
2. Then, applying a unique random translation to the sampling design (modulo the size of the image).
3. Computing the cumulative frequency of NDVI on the randomly shifted sampling design
4. Repeating steps 2 and 3, 199 times with 199 different random translation vectors.

This provides a total population of N=199+1(actual) cumulative frequency on which a statistical test at acceptance probability $1-\alpha = 95\%$ is applied: for a given NDVI level, if the actual ESU density function is between two limits defined by the $N\alpha/2 = 5$ highest and lowest values of the 200 cumulative frequencies, the hypothesis assuming that WS and ESU NDVI distributions are equivalent is accepted, otherwise it is rejected.

Figure 2 shows that the NDVI distribution of the 39 ESUs is quite well representing the NDVI distribution on the 3kmx3km image. The cumulative frequency curve is outside but close to the boundaries for high NDVI values. Therefore the sampling is quite representative of the whole site, which implies that few extrapolations will be needed for the derivation of the biophysical variable maps.

2.2 SPOT image

Two SPOT images were acquired the 4th and 25th March 2003 by HRVIR2 and HRVIR1on SPOT4. They have been geo-located by Benoît Duchemin (IRD) who organized the campaign. The projection is Lambert Nord Maroc (Merchich datum) and is fully described in the campaign report that can be downloaded at http://www.avignon.inra.fr. No atmospheric correction was applied to the images since no atmospheric data were available. However, as the SPOT image is used to compute empirical relationships between reflectance and biophysical variable, we can assume that the effect of the atmosphere is the same over the whole 3kmx3km site. Therefore, it will be taken into account everywhere in the same way. Biophysical variable maps were first generated using each image separately, and then using the average value of the two images since the campaign was achieved between the 10th and 14th March. The two images were weighted by their distance in days to the middle of the campaign. As the results were very closed using the three methods (and as there was no difference between the variable maps obtained at 500m resolution), we decided to use the average value between the two images. Note that the geo-location between the two images is very good so that the images superpose well. Figure 3 shows the relationship between RED and near infrared (NIR) SPOT channels: the soil line is well marked, and no saturated points are observed.

A non supervised classification based on the k_means method (matlab statistics toolbox) was applied to the 4 bands of the SPOT image to distinguish if different behaviours on the image for the biophysical variable-reflectance relationship exist. A number of 5 classes were chosen (Figure 4). The repartitions of the classes on the image and on the ESUs are quite different. Class 3 and 5 appears to be over-sampled whereas class 4 is under-represented.

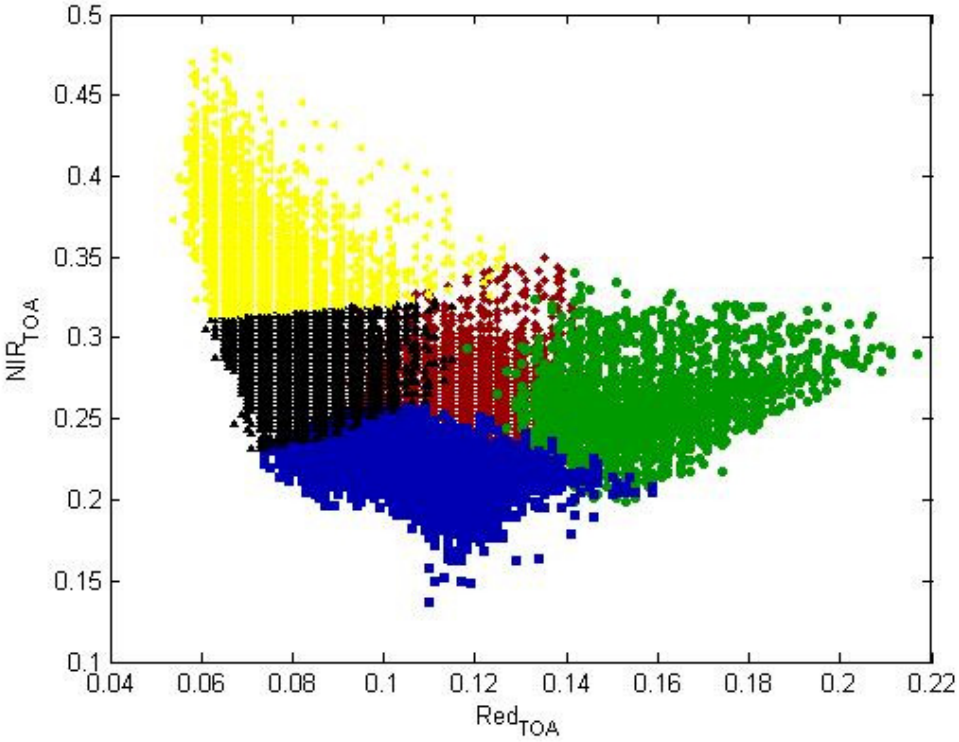

Figure 3. Red/NIR relationship on the SPOT image for Haouz, 2003. Note that the colours corresponds to the different classes that are described in the following.

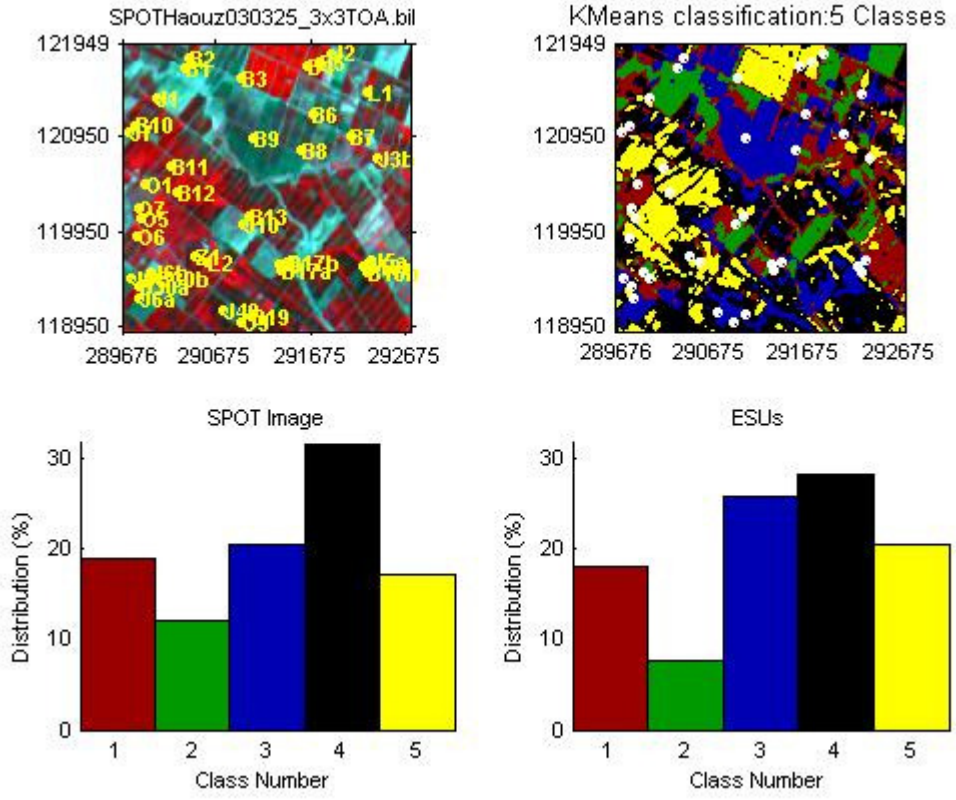

Figure 4. Classification of the SPOT image. Comparison of the class distribution between the SPOT image and sampled ESUs.

2.3 Hemispherical images

The hemispherical images were processed by the CAN-EYE software (Version 2.1) to derive the biophysical variables. Figure 5 shows the distribution of the different measured variables over the sampled ESUs. Green LAI derived from directional gap fraction and LAI derived from gap fraction at 57.5° are consistent and varies from 0 to 3.5 (wheat fields). Figure 6 shows the comparison between the LAI results of the hemispherical images processing using two different softwares. There is no significant difference between the two processing results, which gives confidence in the measurements.

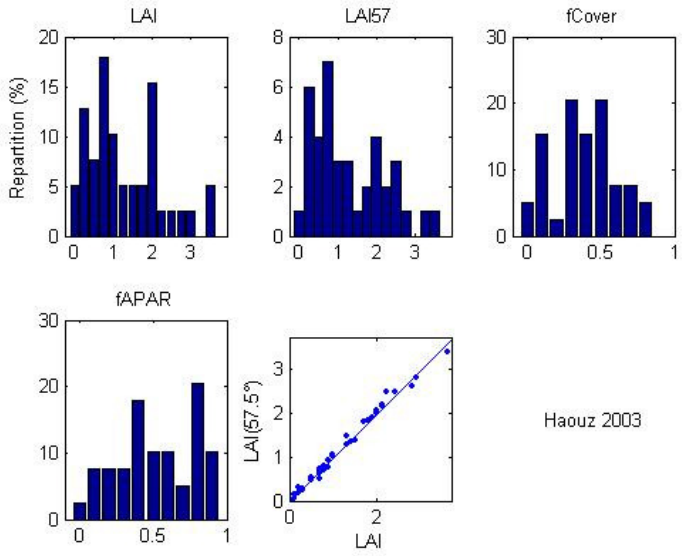

Figure 5. Distribution of the measured biophysical variables over the ESUs.

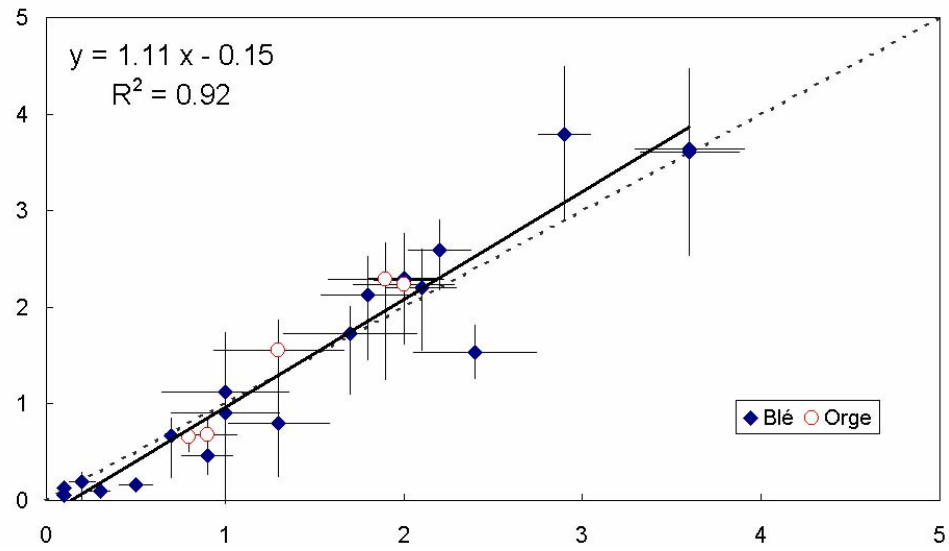

Figure 6. Comparison between the processing of Haouz hemispherical images using two softwares : CAN-EYE developed at INRA (abscissa) and ALICE (ordinate) developed at IRD, for wheat (blue) and barley (red)

Figure 7 shows the different relationships observed between the biophysical variables and corresponding NDVI or the ESUs, as a function of the SPOT classes determined in §2.2. No different behaviours between the classes can be observed, therefore a single transfer function per variable will be generated. The scattering is higher for low NDVI values (bare soil, or low vegetation), especially for fCover and fAPAR.

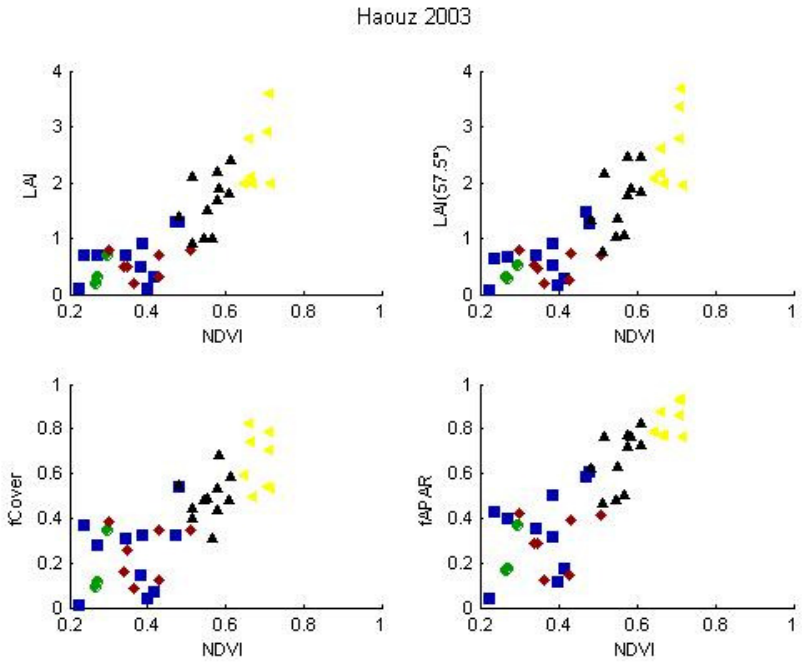

Figure 7. NDVI-Biophysical Variable relationships as a function of SPOT classes

3 Determination of the transfer function for the 4 biophysical variables

3.1 Tested Transfer functions

For each class determined in §2.2, two types of transfer functions are tested:

- REG: If the number of ESUs is sufficient, multiple robust regression between ESUs reflectance (or Single Ratio) and the considered biophysical variable can be considered: we used the 'robustfit' function from the matlab statistics toolbox. It uses an iteratively re-weighted least squares algorithm, with the weights calculated at each iteration by applying the bisquare function to the residuals from the previous iteration. This algorithm gives lower weight to ESUs that do not fit well. The results are less sensitive to outliers in the data as compared with ordinary least squares regression. At the end of the processing, three errors are computed: classical root mean square error (RMSE), weighted RMSE (using the weights attributed to each ESU) and cross-validation RMSE (leave-one-out method).

- LUT: If the number of ESUs is sufficient, Look-Up-Tables are also enviewed : a look-up table is build using ESUs reflectances and corresponding measured biophysical variable. For a given pixel, a cost function is computed as the sum square difference between the pixel reflectances and the ESU reflectances over the 4 bands, divided by the standard deviation computed on ESU reflectances. The result of the cost function is sorted in ascending order, and the biophysical variable estimated for the given pixel is computed as the mean value of the first *n* ESUs providing the lowest value of the cost function. Different values of *n* are considered to get the lowest cost function. This method is reliable only if the ESU NDVI distribution is quite comparable with the whole site NDVI distribution.

Both regression and Look-Up-Tables are tested using either the reflectance or the logarithm of the reflectance for any band combination, plus the simple ratio. As both methods have poor extrapolation capacities, a flag image, based on the computation of convex hull over reflectances, is computed showing:

- Pixels inside the 'strict convex-hull': for each class, a convex-hull is computed using all the reflectance combination used for the transfer function, and corresponding to the ESUs belonging to the class. For those

pixels, the transfer function is used as an interpolator, and the degree of confidence in the results obtained is quite high.

- Pixels inside the 'large convex-hull': for each class, a convex-hull is computed using all the reflectance combination (±5% in relative value) used for the transfer function, and corresponding to the ESUs belonging to the class. For those pixels, the transfer function is used as an extrapolator (but not far from interpolator), and the degree of confidence in the results obtained is quite good.
- Pixels outside the two convex-hulls: this means that for these pixels, the transfer function acted like an extrapolator which makes the results less reliable. However, having *a priori* information on the site may help to evaluate the extrapolation capacities of the transfer function.

3.2 Results on the Haouz site

3.2.1 Choice of the method

Figure 8 and Figure 9 show the results obtained for all the possible band combinations using either the reflectance or the logarithm of the reflectance. The LUT method provides systematically higher RMSE value than for REG (weighted RMSE for REG). therefore the REG method will be chosen for all the variables. It can be noticed that using the logarithm of the reflectance provides a cross-validation RMSE which increases as the weighted RMSE increases as a function of the band combination. Moreover, the number of points with weights lower than 0.7 is lower using the logarithm than the reflectance itself.

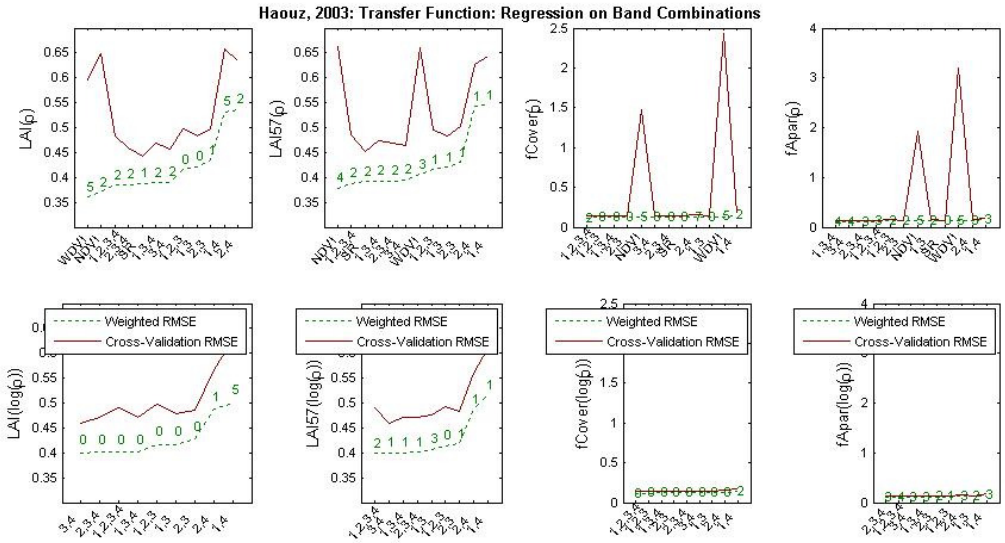

Figure 8. Transfer function: test of multiple regressions applied on different band combinations. Band combinations are given in abscissa. The estimated biophysical variable is given in ordinate. Top graphs correspond to regression made on reflectance (ρ): the weighted root mean square error (RMSE) is presented in green along with the cross-validation RMSE in red. The numbers indicate the number of data used for the robust regression with a weight lower than 0.7. Bottom graphs correspond to regression made on the logarithm of the reflectance.

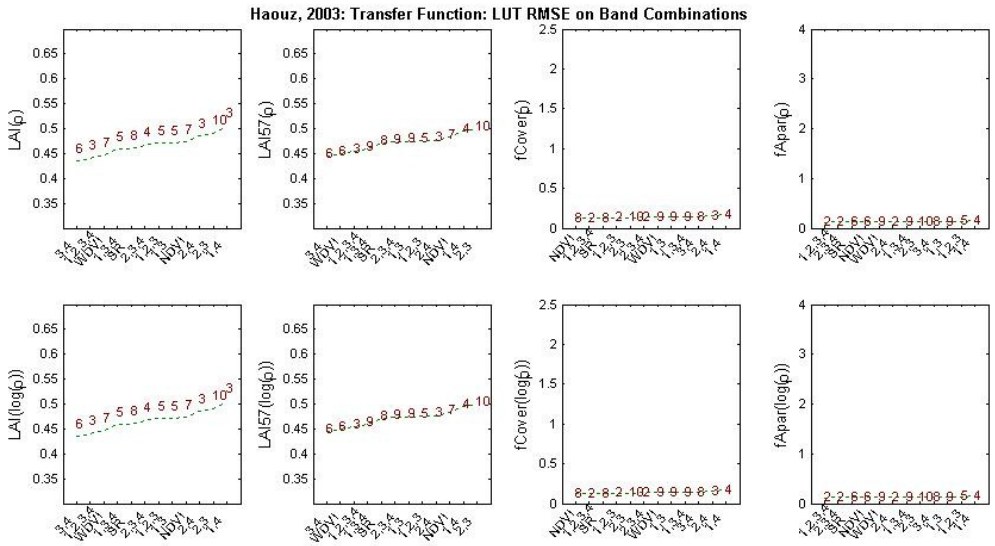

Figure 9. Transfer function: test of LUT applied on different band combinations. Band combinations are given in abscissa. The estimated biophysical variable is given in ordinate. Top graphs correspond to regression made on reflectance (ρ): the root mean square error is presented in green. The numbers indicate the number of elements selected in the LUT to compute the resulting biophysical variables. Bottom graphs correspond to LUT using the logarithm of the reflectance.

3.2.2 Choice of the band combination

For the effective LAI and LAI at 57.5°, Figure 11 and Figure 12 show the (log(XS3),log(XS4)) combination performs the best for both LAI and LAI57 selected since it presents a good compromise between the RMSE values and the number of points with a weight lower than 0.7 (Figure 10). The associated RMSE is lower than 0.5 for a LAI varying between 0 and 3.5. It should be noted also that ESU B3, which has been manually shifted following our knowledge of the ground sampling, present a weight near 0.7 for the two variables, which may indicate that this ESU is still not well located.

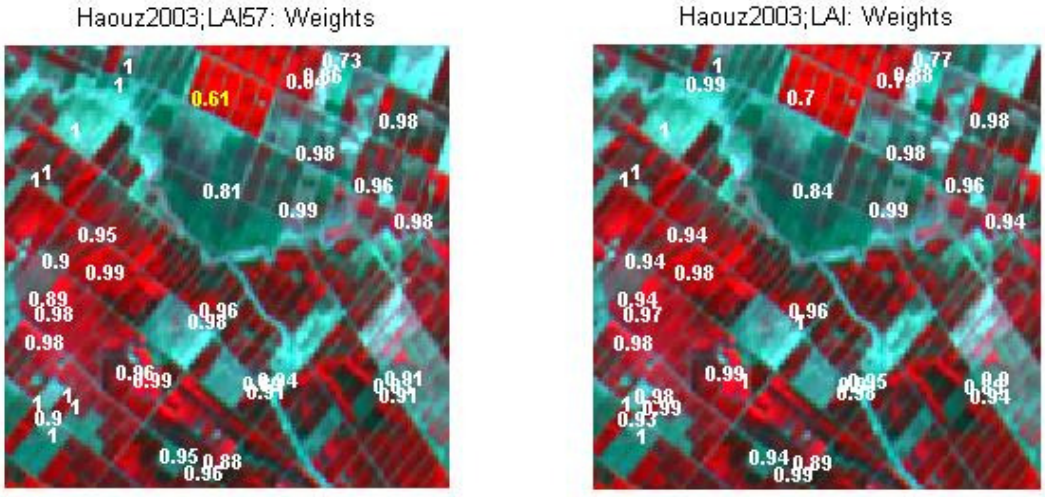

Figure 10. Weights associated to each ESU for the determination of LAI57 (left) and LAI (right) transfer function.

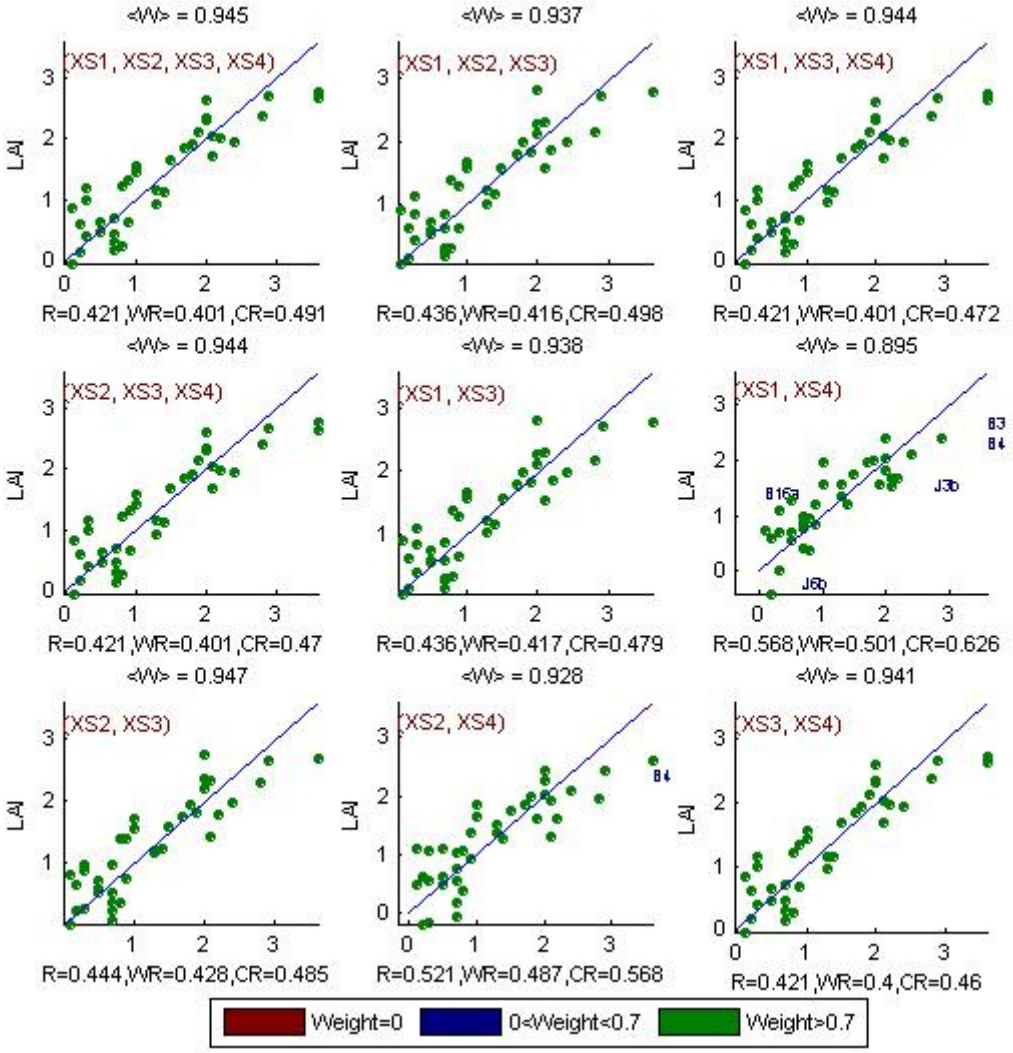

Figure 11. Effective Leaf Area Index: results for regression using different band combinations. R is the root mean square error computed between LAI and estimated LAI. WR is the weighted root mean square error and CR is the cross validation root mean square error.

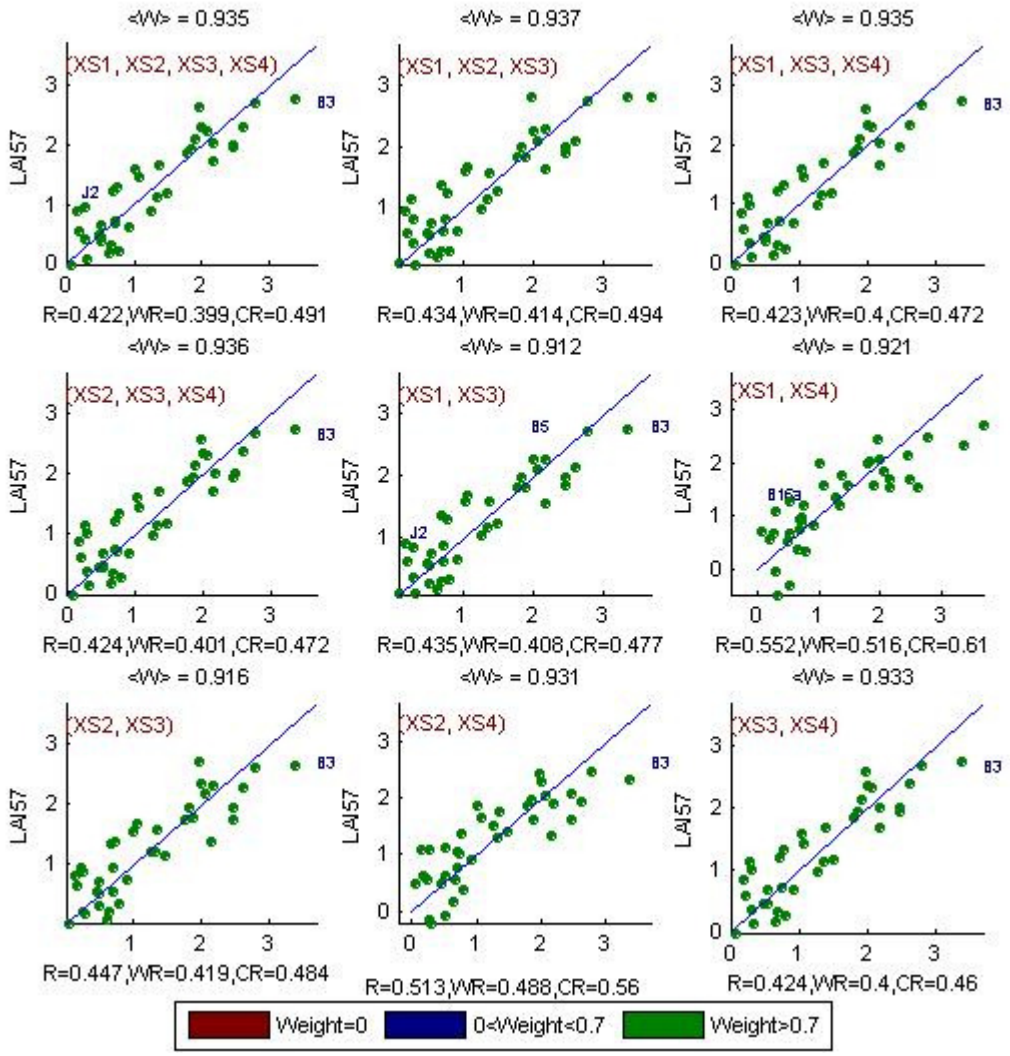

Figure 12. Effective Leaf Area Index at 57.5°: results for regression using different band combinations

For fCover (log(XS1), log(XS2), log(XS3), log(XS4)), we have chosen the best combination (best RMSE value) since no point has a weight under 0.7 for the variable. **For fAPAR (log(XS2), log(XS3), log(XS4)),** we made a compromise between RMSE values and the number of ESUs with weight lower than 0.7.

Variable	Band Combination	RMSE	Weighted RMSE	C-valid RMSE
Effective LAI	-0.0784+ 2.8698log(XS3) -2.7996log(XS4)	0.421	0.400	0.460
Effective LAI (57.5°)	-0.3745+2.7701 log(XS3) -2.8884 log(XS4)	0.424	0.400	0.460
fCover	2.7798+ 2.6454log(XS1)-1.2567log(XS2)+ 0.3633log(XS3) -0.6882log(XS4)	0.121	0.115	0.140
fAPAR	-0.1640-0.0745log(XS2)+ 0.6574log(XS3)- 0.7540log(XS4)	0.126	0.108	0.138

Table 1. Transfer function applied to the whole site for the different biophysical variables, and corresponding errors

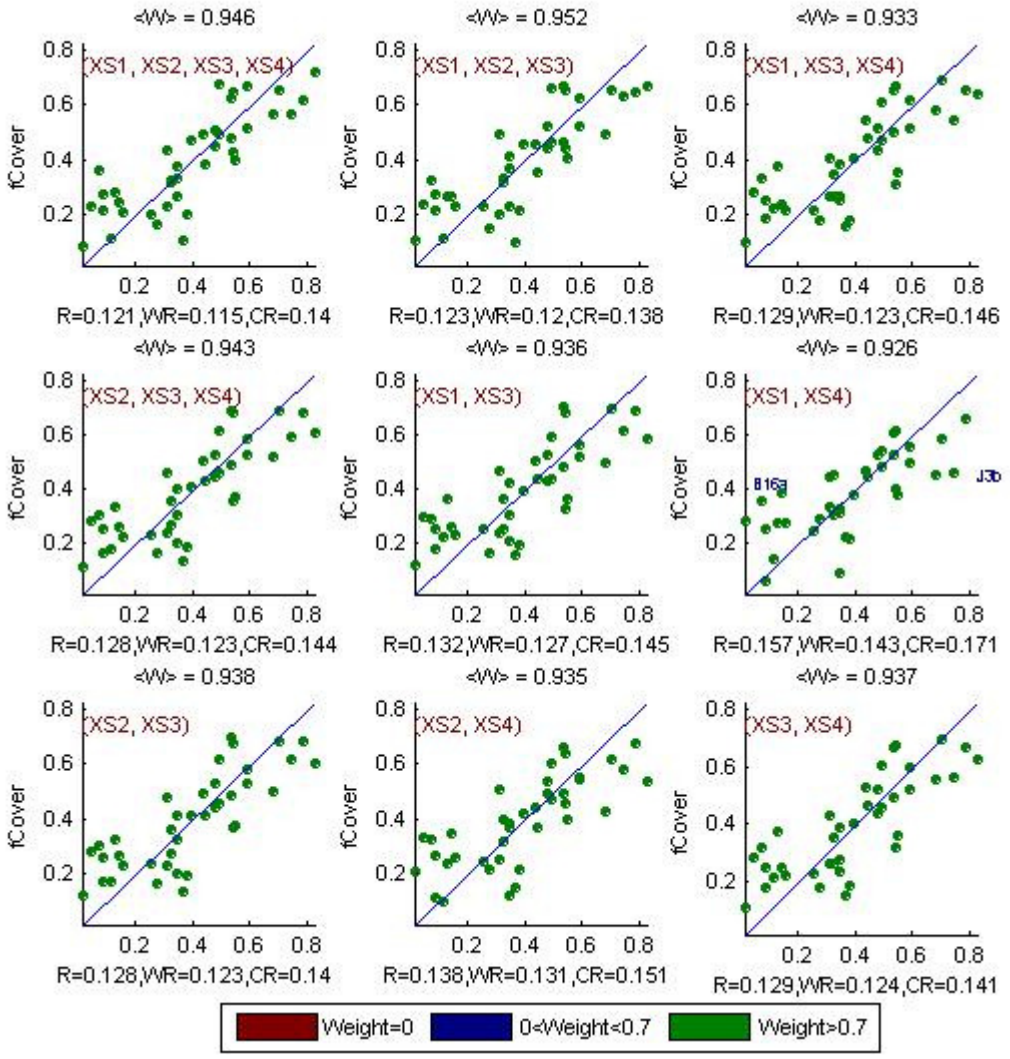

Figure 13. : fCover: results for regression using different band combinations.

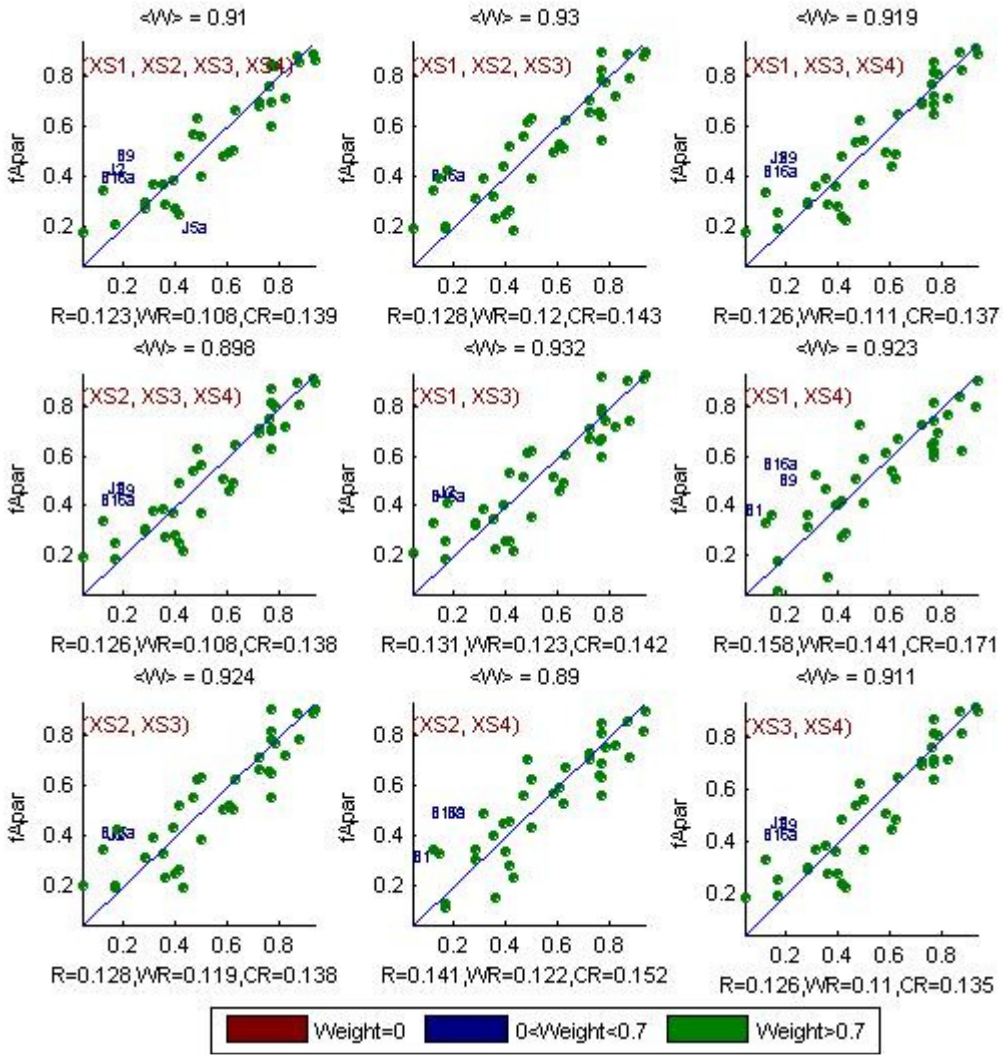

Figure 14. fAPAR : Results for regression using different band combinations

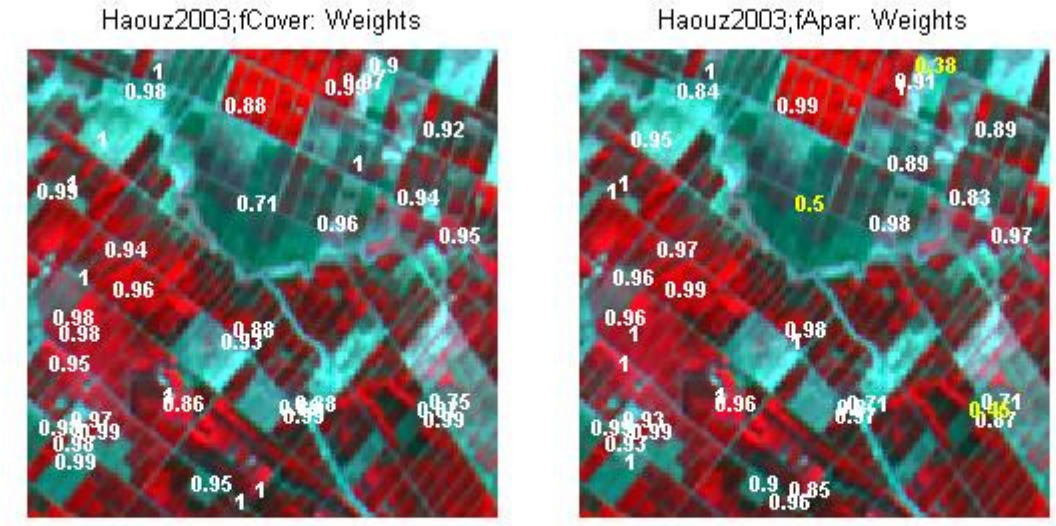

Figure 15. Weights associated to each ESU for the determination of fCover (left) and fAPAR (right) transfer function

3.3 Applying the transfer function to the Haouz SPOT image extraction

Figure 16 presents the biophysical variable maps obtained with the transfer function (REG on the logarithm of the reflectance) described in Table 1 for all the classes. The maps obtained for the different variables are consistent, showing similar patterns, low LAI values where low fAPAR and fCover are observed and conversely. Note that the average value for LAI and LAI57 are consistent (1.2 for both). It can also be noted that the fAPAR values are quite high as compared to LAI and fCover. This may be due to the fact that the fAPAR was computed for 10h00AM solar hour, which induces an instantaneous fAPAR computed at a solar zenith angle at 47°. For high zenith angles, the gap fraction is lower than at nadir (since more vegetation is seen), which may explains the quite high values that are observed.

The flag maps are quite different between LAI (LAI57), fCover and fAPAR. This is explained by the fact that the summits of the convex-hulls are defined using 2 (LAI) or 4 (fCover) bands. Therefore, the use of four bands in the regression induces a convex-hull which is stricter than when using two bands. For LAI and LAI57, the main part which is flagged in red corresponds to bare soil areas, which is consistent with the LAI map where it is equal to 0. For fCover, almost all the images are outside the large convex-hull. However, when looking to the resulting map and comparing it to the LAI map, it is very consistent. Therefore, we can think that the pixels are very close to the boundaries of the convex-hull.

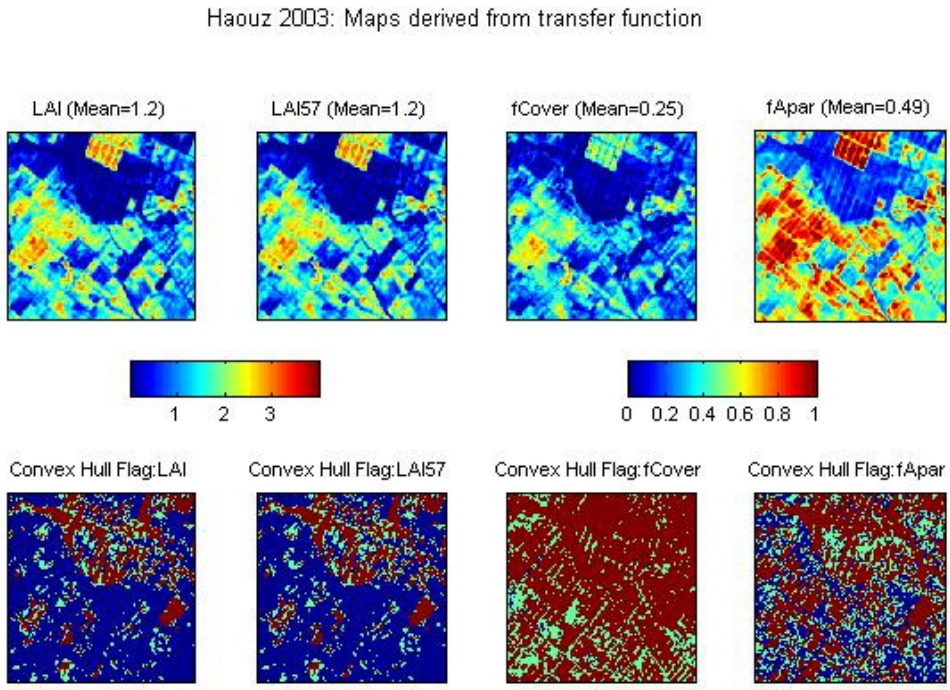

Figure 16. High resolution biophysical variable maps applied on the Haouz site (top). Associated Flags are shown at the bottom: orange corresponds to the bare soil class for which we applied the averaging method. dark blue and green corresponds to the pixels belonging to the 'strict' and 'large' convex hulls, and red to the pixels for which the transfer function is extrapolating.

4 Conclusion

The transfer functions are finally obtained by using Reg on the logarithm of the reflectance and 39 ESUs together. The band combinations are different from one regression to another. Results show good consistency between the variables and the flag associated to each map show that the transfer function is used as an extrapolator in quite large areas for fCover and fAPAR, but appears to be very close to the boundaries of the convex-hull.

The biophysical variable maps are available in UTM, 29 North, projection coordinates (Datum: WGS-84) at 20m resolution. The transfer function was applied on an image which results from the averaging of two SPOT images which were geo-located in Lambert nord Maroc, Merchich datum.

5 Acknowledgements

We want to thank:

- Benoît Duchemin for the organization and participation to the campaign, his help for the processing of these data, as well as having provided the comparison graph between results from the software developed at IRD and CAN-EYE (INRA)
- Roland Bosséno (INRA-CSE) for the processing of CAN-EYE data
- All the people who have participated to the campaign: Abourida Ahad, Cheggour Aouatif, Er-Raki Salah, Errouane Sadik, Gentine Pierre, Hadria Rachid, Khabba Said, and Lahrouni Abdelrahman.
- Albert Olioso and Olivier Marloie (INRA-CSE) who were in charge of the transport of the devices for the campaign.