VALERI 2003 : Fundulea 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 Fundulea site in 2003 (see campaign report for more details about the site and the ground measurement campaign). 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 4km x 4km site. The processing of the ground data has shown that there is an anomaly on the ESU location of U41 (in black on Figure 1) since it was described as an oak forest, same as U40 but does not correspond to a forest on the SPOT image. This ESU has been eliminated from the data. Finally, 52 ESUs are used for the computation of the transfer function. The site was sampled in May and corresponds mainly to crops (wheat U24,U28, barley U29, rape seeds U56, U5, U6, soja U10, U11, sunflower U17, maize U45, pea U8, U12) with two areas corresponding to acacia and oak forests (U38,U39, U40 bright red parts in the upper left and lower right corners of the image).

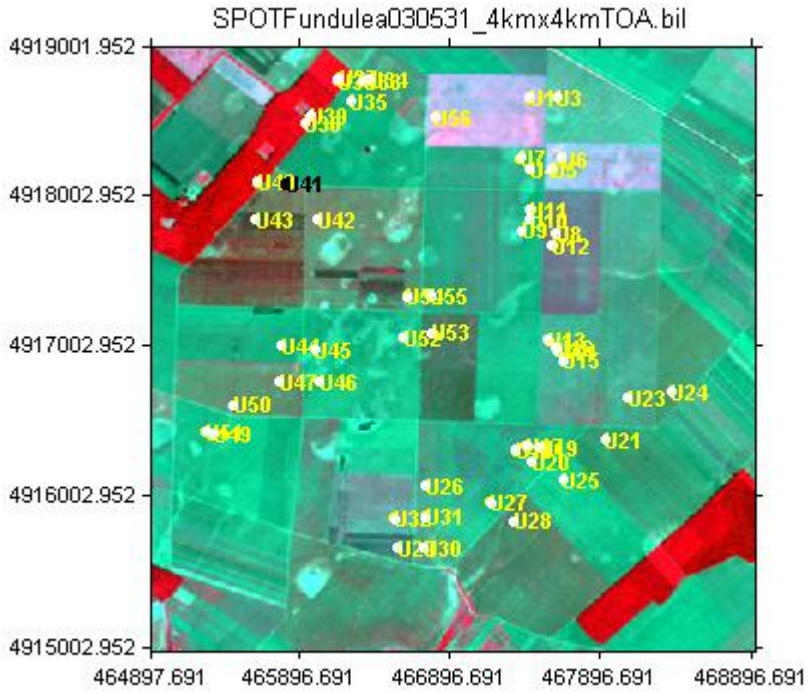


Figure 1. Distribution of the ESUs around the Fundulea site.

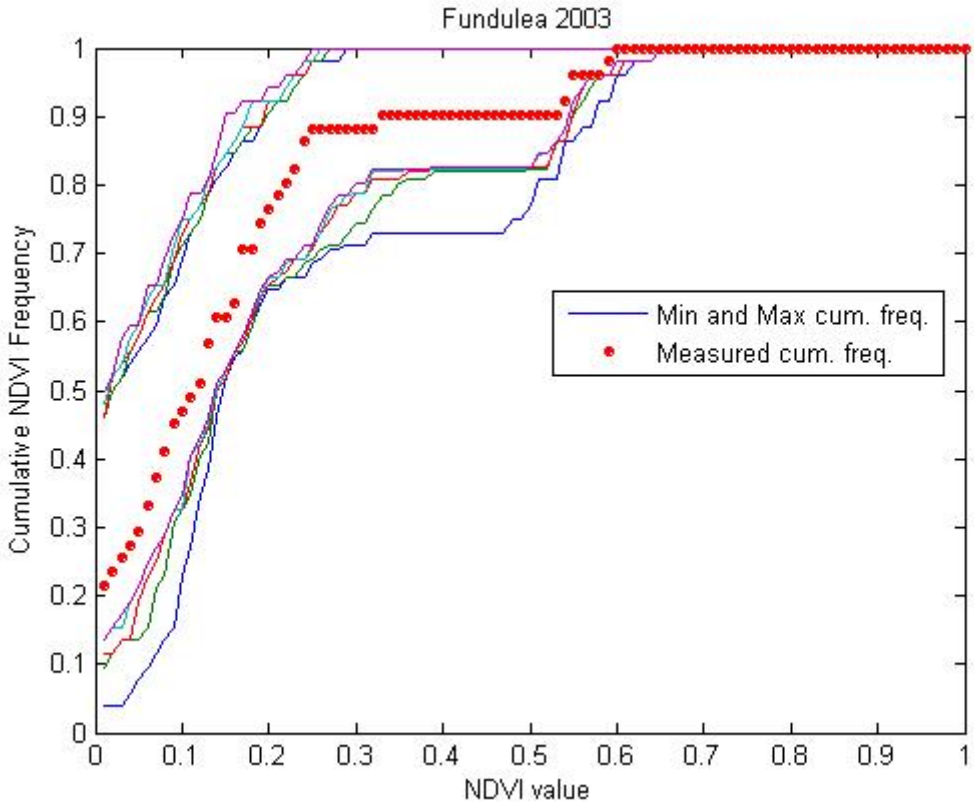


Figure 2. Comparison of the ESU NDVI distribution and the NDVI distribution over the whole image.

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 52 ESUs is quite good as compared to the NDVI distribution over the whole site since the 'ESU' curve is almost inside the 'boundary curves' and on the border for high LAI values. More points in the forest area might have improved the sampling performances. This result induces that the biophysical variable map should be accurate since little extrapolation will be required to extend the ESU measurement to the whole site.

2.2 SPOT image

The SPOT image was acquired the 31st May 2003 by HRG1 on SPOT5. It was geo-located by NOVELTIS using previous images coming from VALERI campaign in 2001. The projection is UTM 35N, WGS84, no atmospheric correction was applied to the image 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 4kmx4km site. Therefore, it will be taken into account everywhere in the same way. Figure 3 shows the relationship between RED and near infrared (NIR) SPOT channels. There is no saturation and the soil line is well plotted.

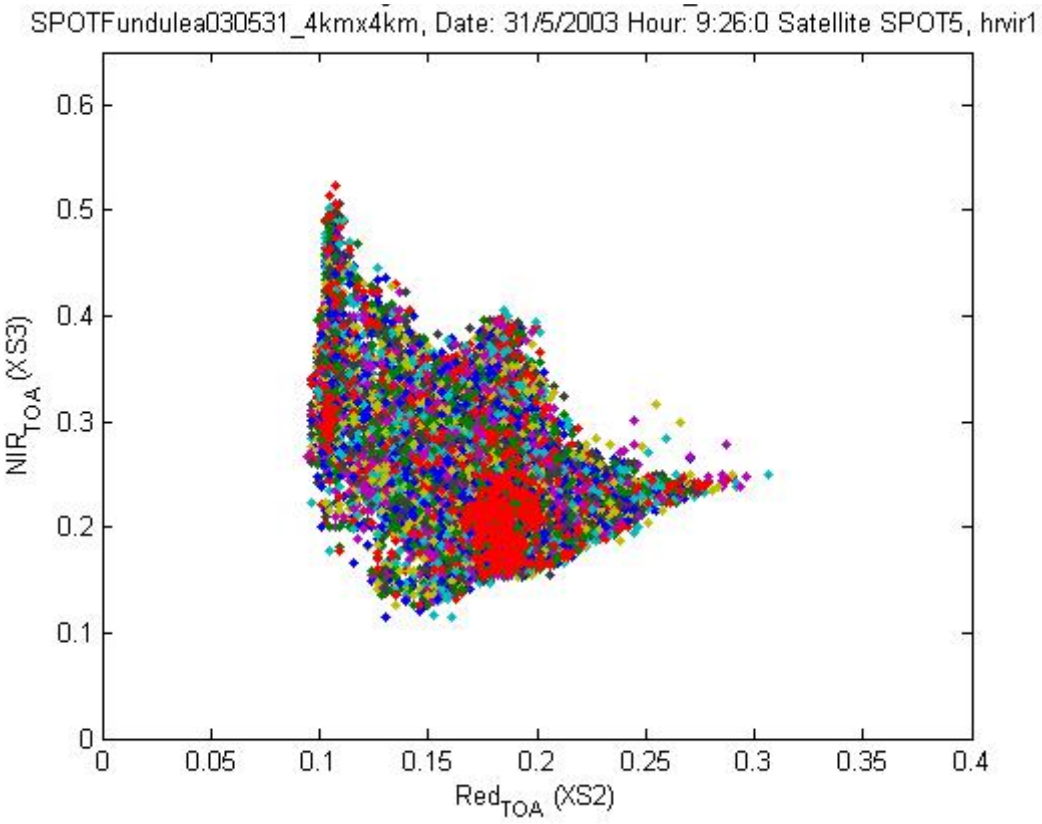


Figure 3. Red/NIR relationship on the SPOT image for Fundulea, 2003.

A non supervised classification based on the k_means method (matlab statistics toolbox) was applied to the NDVI of the SPOT image to distinguish if different behaviours on the image for the biophysical variable-reflectance relationship exist. A number of 5 classes was chosen (Figure 4). The repartition of the classes on the image and on the ESUs are very similar.

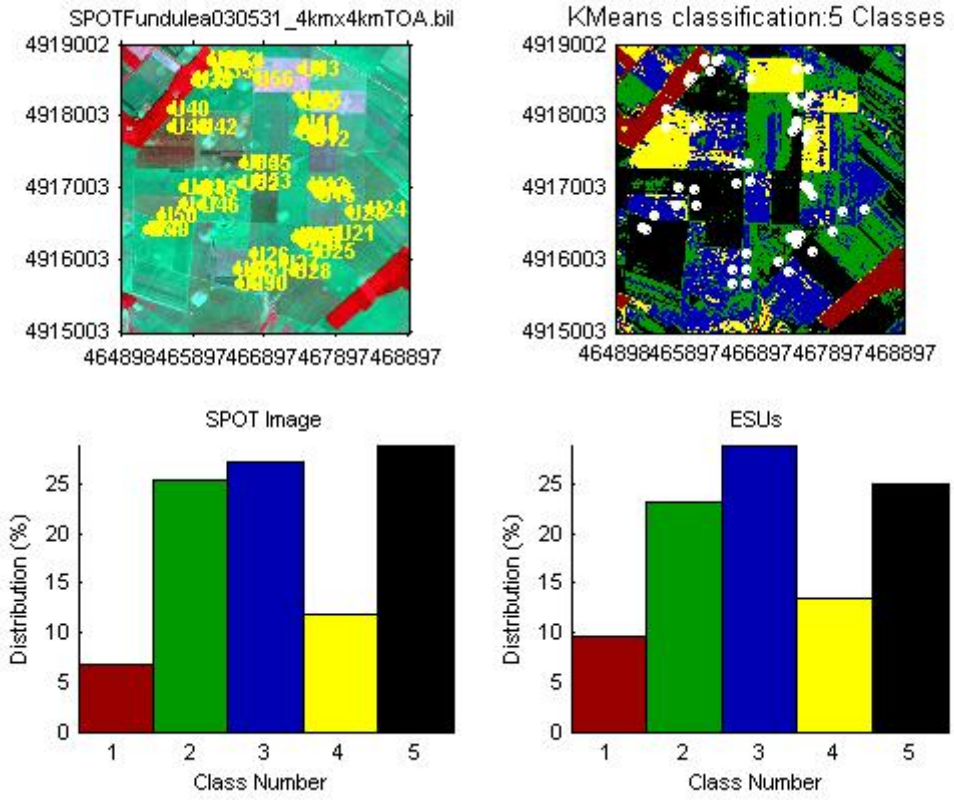


Figure 4. Classification of the SPOT image. Comparison of the class distribution between the SPOT image and sampled ESUs.

2.3 Hemispherical images

The hemispherical images were processed by the CAN-EYE software (Version 1.3) to derive the biophysical variables. Figure 5 shows the distribution of the different measured variables over the sampled ESUs. As there was understorey in forest ESU (U38 to U40), hemispherical images were acquired from above the understorey and from below the canopy (trees). For these ESUs, the two sets of acquisition were processed separately to derived LAI, fCover, and fAPAR. The ESU biophysical variable was then computed as:

- LAI, LAI57 : LAI(above)+LAI(below).
- fCover :1-(1- fCover(above))*(1-fCover(below)). This assumes that independency of the gaps inside the understorey and the gaps inside the trees which is not true but it the only way to get the total fCover.
- fAPAR: [1-(1-fAPAR(below))*(1-fAPAR(above))], since 1-fAPAR can be considered equivalent to a gap fraction.

LAI derived from directional gap fraction and LAI derived from gap fraction at 57.5° are consistent (Figure 5). To build the relationships between biophysical variables and SPOT data, the reflectance of the forest ESU considered as the average reflectance over the central pixel + the 8 surrounding pixels. This takes into account the fact that the height of the trees are about 20m which makes the fisheye observing an area of $\pi x[20x\tan(60°)]^2 \cong 3800m^2$, *i.e.* the area of 9 SPOT pixels (=3600m²), when using a maximum view zenith angle of 60°. For the other ESUs (crops), images were taken from above the canopy (looking downward) and the area of the SPOT pixel corresponds roughly to the ESU area.

Figure 6 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. It appears that two clusters of classes can be distinguished: ESUs corresponding to forest (class 1) and ESUs corresponding to crops

(classes 2 to 5). Therefore, two transfer functions will be estimated, one for forest and one for the other classes. Note that the NDVI can be negative for class 5, which corresponds to ESU were very low vegetation can be observd (youg maize, 3 leaves 20cm height, or young soja 10cm height). The fAPAR of forest ESUs is very hig (near 0.99) and might be overestimated by the CAN_EYE software.

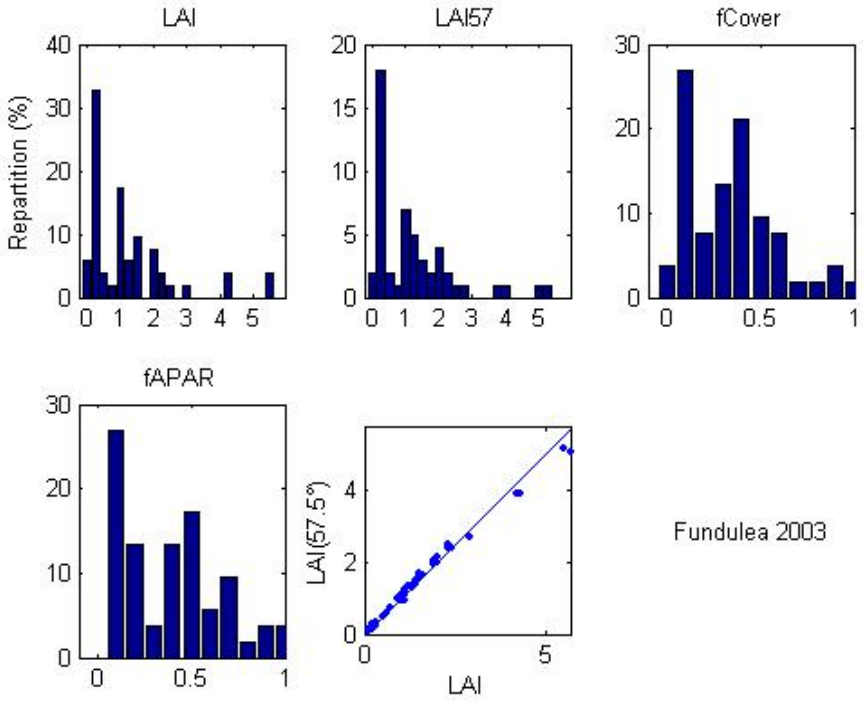


Figure 5. Distribution of the measured biophysical variables over the ESUs.

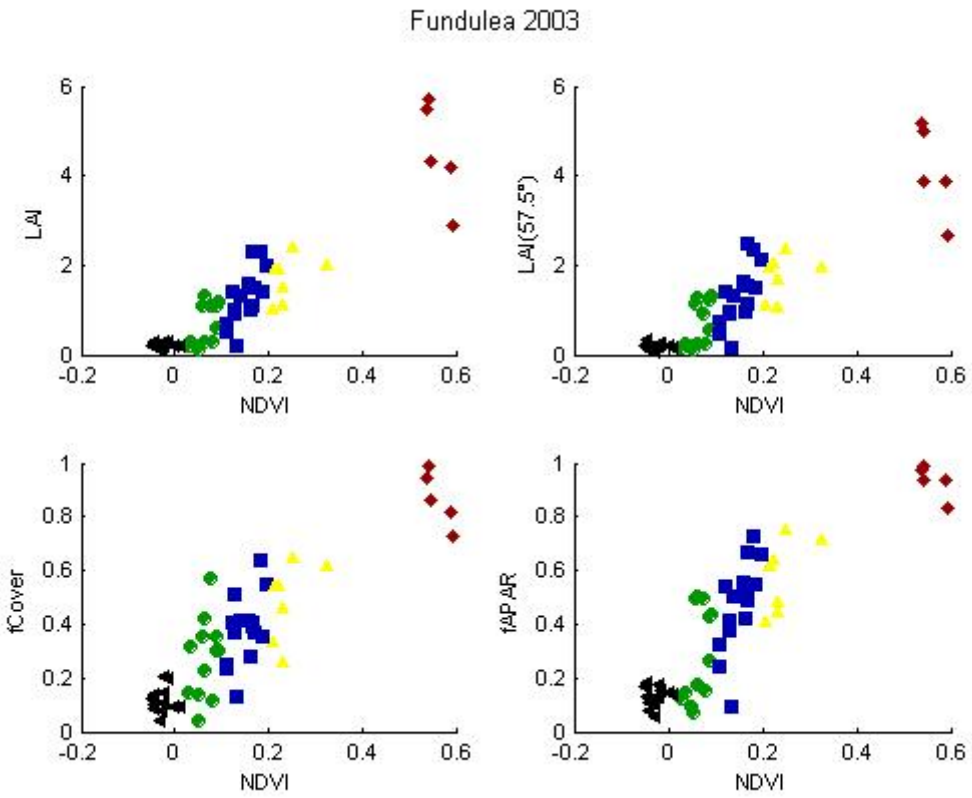


Figure 6. NDVI-Biophysical Variable relationships as a function of SPOT classes

3 Determination of the transfer function for the 4 biophysical variables: LAI, LAI57, fAPAR, fCover

3.1 Tested Transfer functions

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

- AVE: If the number of ESUs belonging to the class is too low, the transfer function consists only in attributing the average value of the biophysical variable measured on the class to each pixel of the SPOT image belonging to the class.

- 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. In §2.1, 2.2it has been shown that the two distribution are quite different because of the presence of bare soil areas. Therefore, the results of this method will be shown for Fundulea site but will not be applied to derive the biophysical variable maps.

Both regression and Look-Up-Tables are tested using either the reflectance or the logarithm of the reflectance for any band combination. 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 (not AVE method), 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 Fundulea site

For all the ESUs that do not correspond to forest, we have tested REG and LUT methods using all the classes together to keep a reasonable number of data for the regression. For forest ESUS, the AVE method was applied.

3.2.1 Choice of the method for classes 2,3,4,5

Figure 7 and Figure 8 show the results obtained for all the possible band combinations using either the reflectance or the logarithm of the reflectance. As the NDVI distribution of the ESUs (Figure 2) represents well the whole site, REG and LUT methods are both tested. The REG method provides similar results for all the

variables as compared to LUT.Using either the logarithm of the reflectance or the reflectance itself provides very similar results in terms of cross-validation RMSE, for the best band combinations. For all the variables, the number of points with a weight lower than 0.7 in the robust regression is higher (for the best combination) when using the reflectance itself than the logarithm. Therefore we have chosen to use the robust multiple regression using the ESUs reflectance value for LAI, LAI57, fCover, and fAPAR.

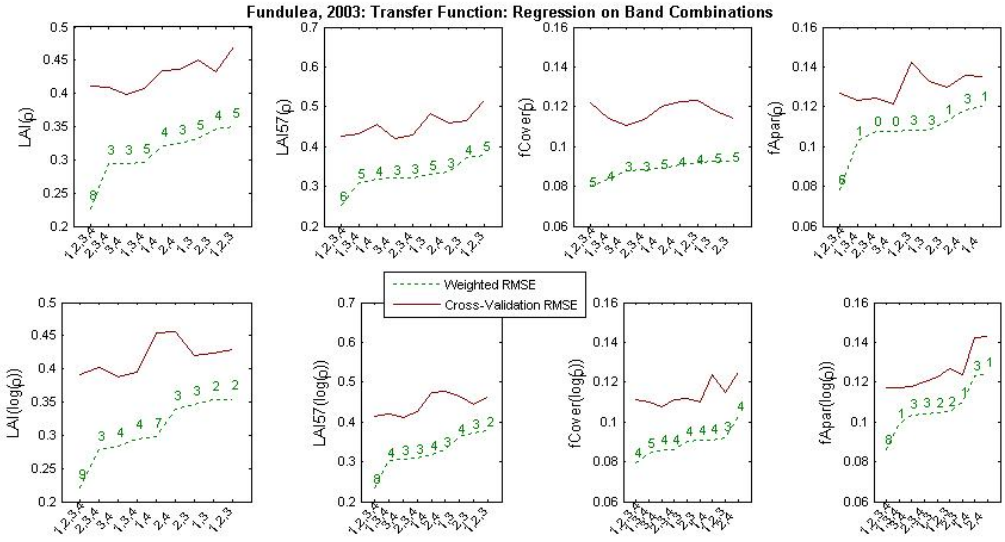


Figure 7. Transfer function: test of multiple regressions applied on different band combinations. Band combinations are given in abscissa. The estimated biophysical variable is given in ordinate. Top graphs correspond to regression made on reflectance (ρ): the weighted root mean square error (RMSE) is 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.

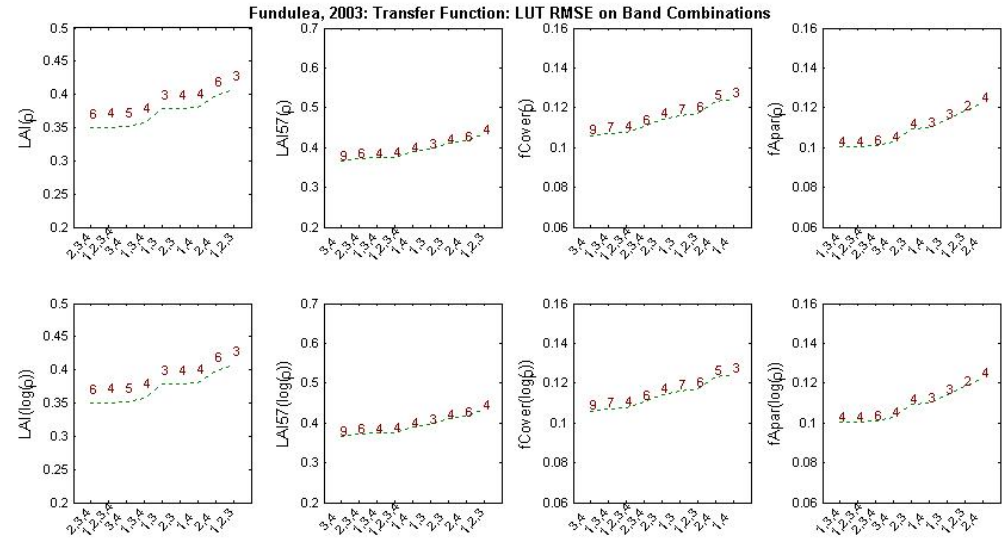


Figure 8. Transfer function: test of LUT applied on different band combinations. Band combinations are given in abscissa. The estimated biophysical variable is given in ordinate. Top graphs correspond to regression made on reflectance (ρ): the root mean square error is presented in green. The numbers 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 classes 2,3,4,5

For the effective LAI and LAI at 57.5°, Figure 9 and Figure 11 show that some ESUs have systematically a weight lower than 0.7 for almost all tested band combination: ESU U28, U32 and U50 (wheat and barley). No indication in the GPS file, as well as in the CAN-EYE processing report can provide an explanation of why these ESUs have systematically a low weight as compared to the other ESU.

The (XS2,XS3,XS4) combination for LAI and (XS1,XS3,XS4) for LAI at 57.5° were selected since they present a good compromise between the RMSE values and the number of points with a weight lower than 0.7 (Figure 10). Note that the RMSE is quite low (around 0.4 for LAI varying between 0 and 2.5).

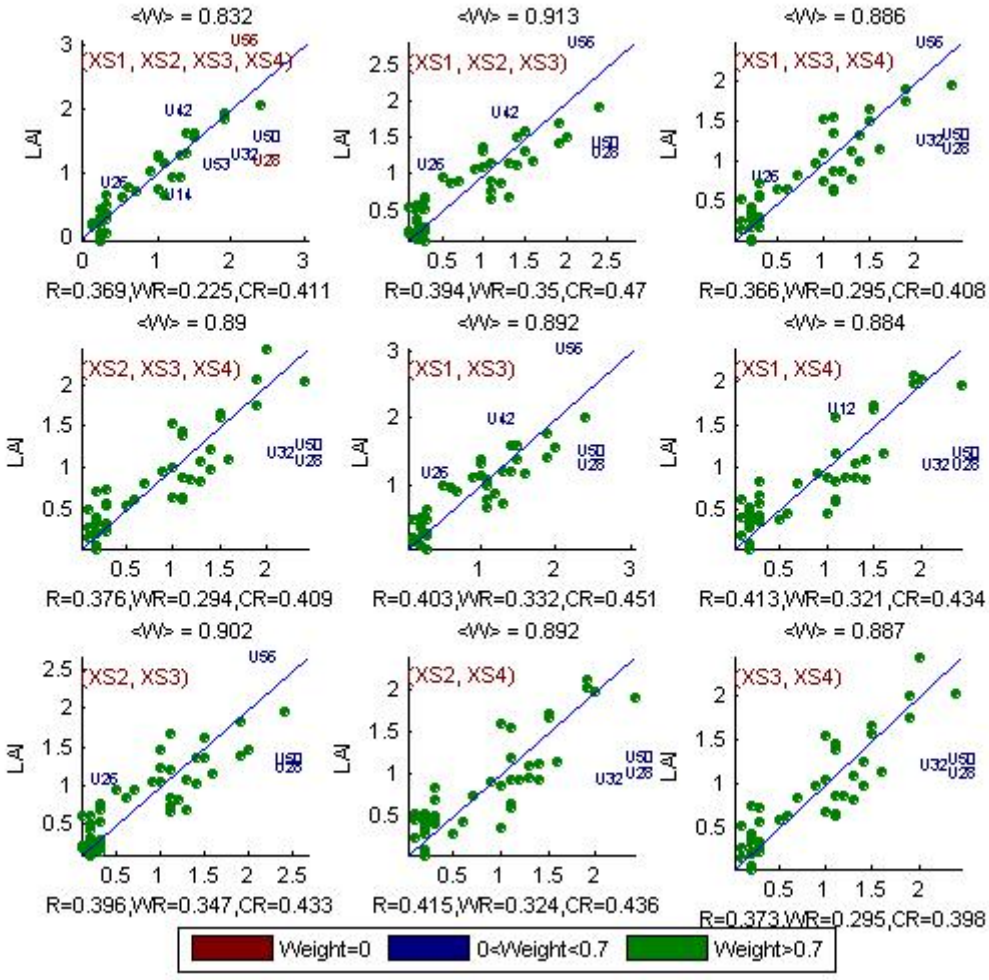


Figure 9. Effective Leaf Area Index: results for regression using different band combinations. R is the root mean square error computed between LAI and estimated LAI. WR is the weighted root mean square error and CR is the cross validation root mean square error.

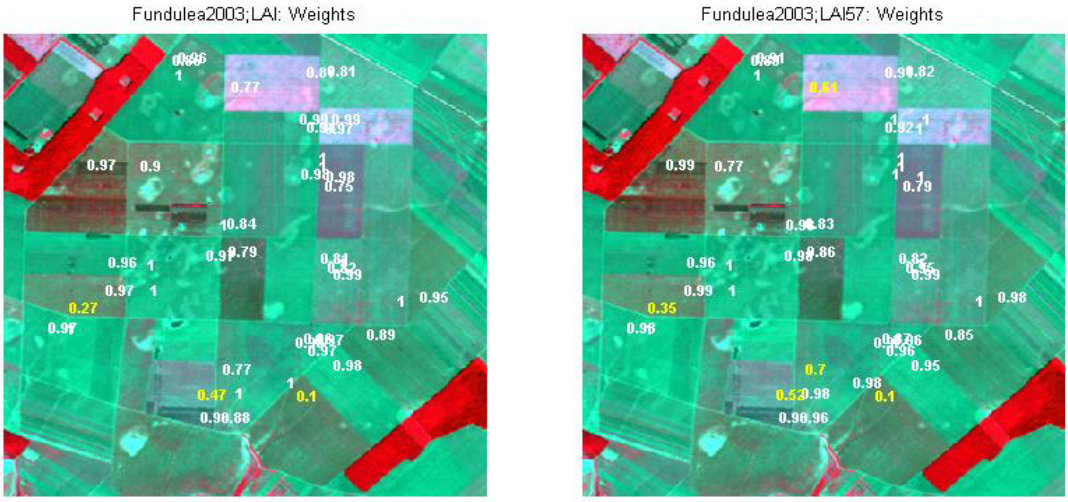


Figure 10. Weights associated to each ESU for the determination of LAI (left) and LAI57 (right) transfer function.

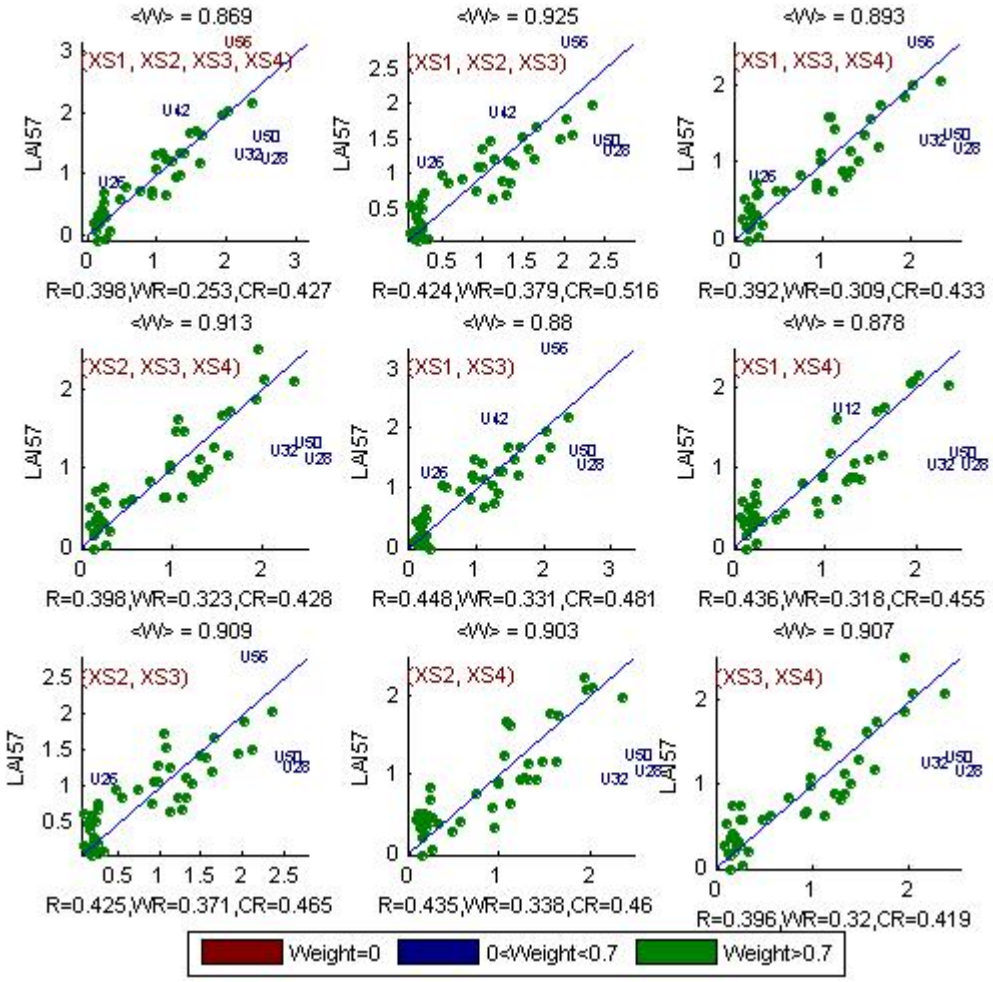


Figure 11. Effective Leaf Area Index at 57.5°: results for regression using different band combinations. <W> is the average weight value computed over all the ESUs.

For fCover (XS3, XS4), the same remarks can be done for ESUs U14, U30, and U50 as for LAI. The associated RMSE is around 0.1.

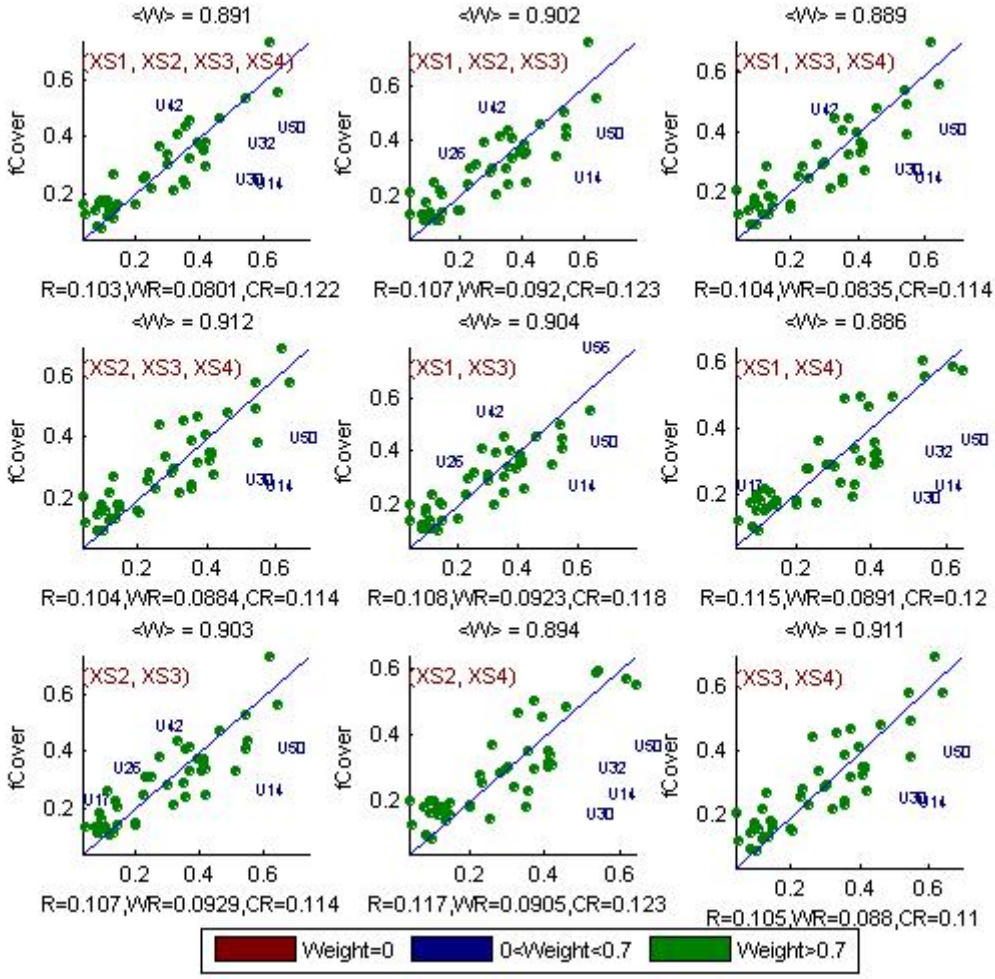


Figure 12. fCover : Results for regression using different band combinations

For fAPAR (XS2,XS3,XS4), all the ESUs have a weitgh higher than 0.7 and are taken into account in the regression . The results are quite good (RMSE of 0.16 with fAPAR between 0 and 1).

Variable	Band Combination	RMSE	Weighted RMSE	C-valid RMSE
Effective LAI	2.2265+ 2.6240 XS2+ 6.6817 XS3-10.8958 XS4	0.589	0.475	0.531
Effective LAI (57.5°)	2.5338 -3.5792XS1+ 8.8941XS3 -9.6068XS4	0.571	0.474	0.655
fCover	0.5327+ 1.9580XS3 -2.2099XS4	0.105	0.088	0.110
fAPAR	0.7439 -0.7936XS2+ 2.5563XS3-2.6841XS4	0.114	0.107	0.124

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

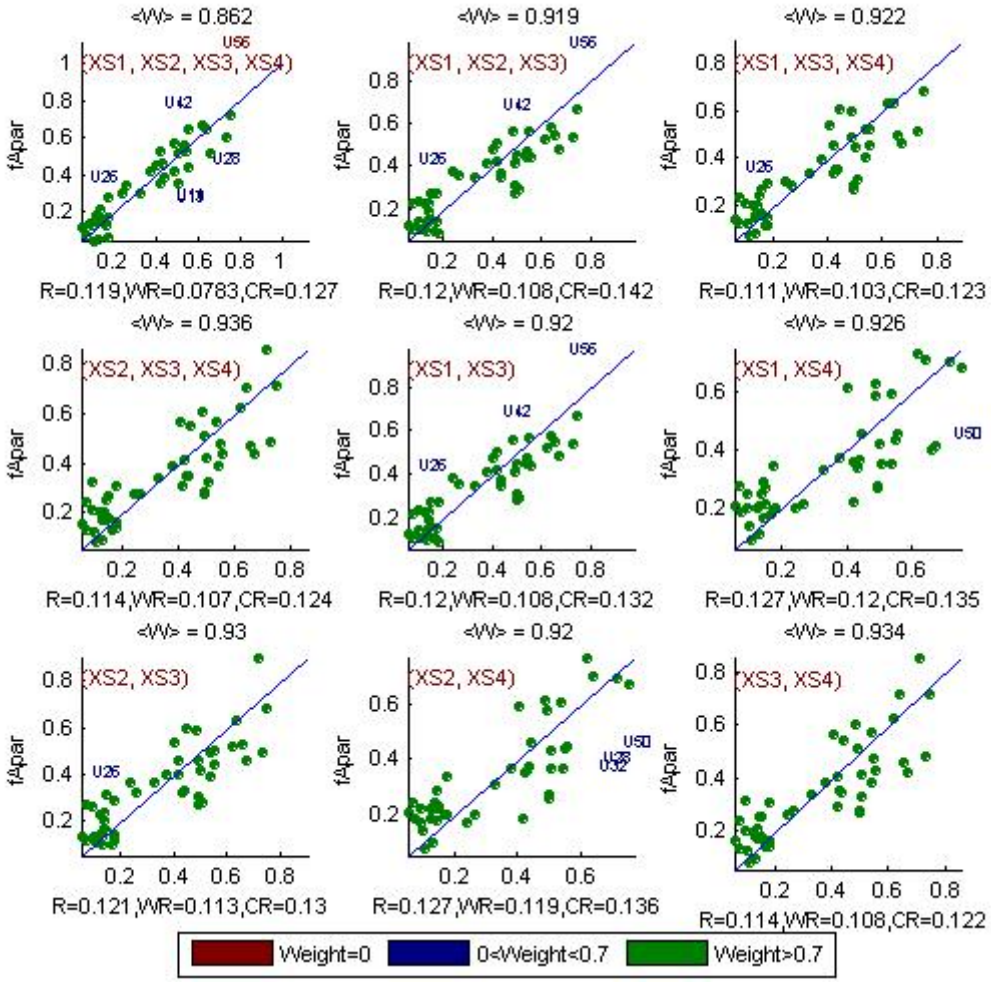


Figure 13. fAPAR : Results for regression using different band combinations

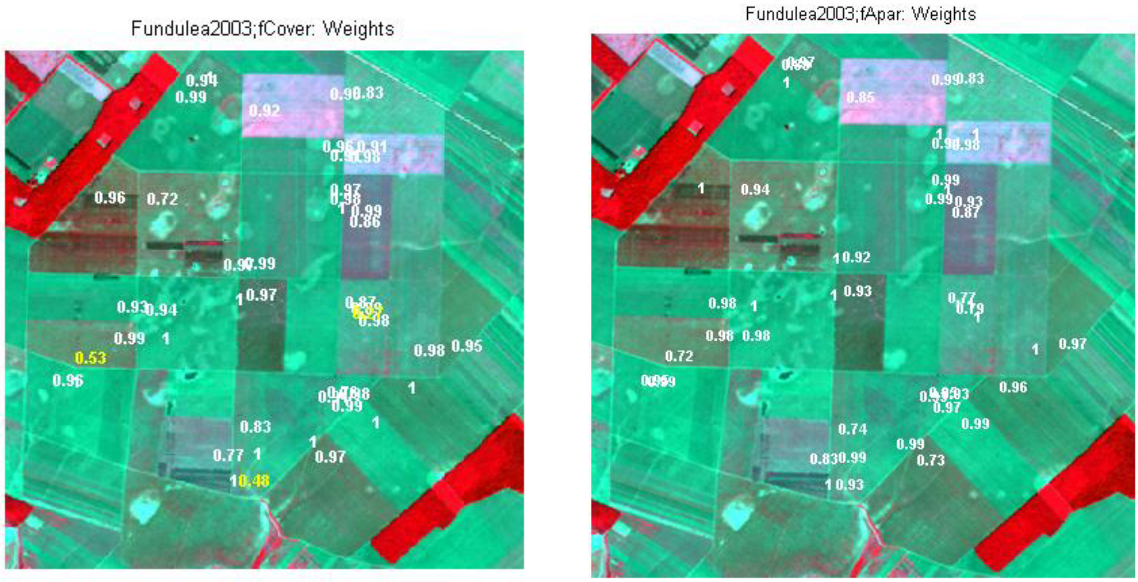
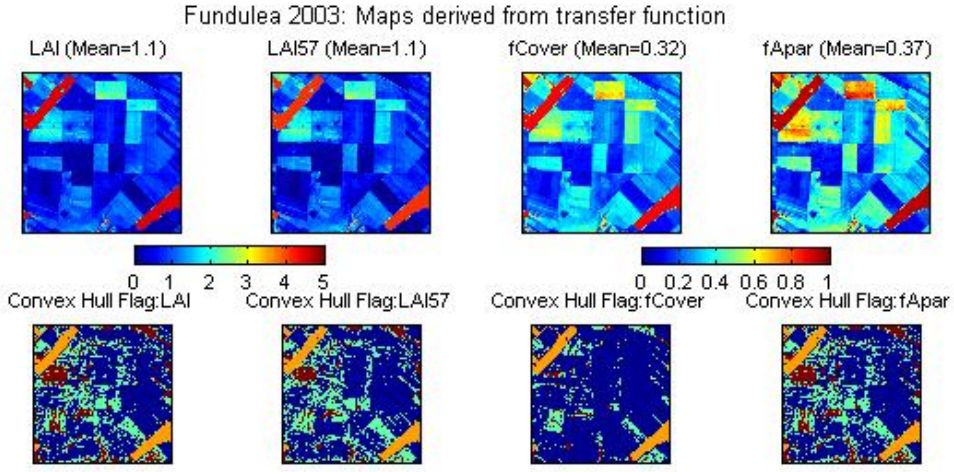


Figure 14. Weights associated to each ESU for the determination fCover and fAPAR transfer functions

3.3 Applying the transfer function to the Fundulea SPOT image extraction

Figure 15 presents the biophysical variable maps obtained with the transfer function (REG) described in Table 1 for all classes 2,3,4,5. The average value of the ESU belonging to class 1 was applied to the pixels belonging to this class (orange colour in the flag image). Although different bands were used to estimate the variables, the maps obtained for the different variables are consistent, showing similar patterns, low LAI values where low fAPAR (or fCover) are observed and conversely. Note that the average value for LAI and LAI57 are consistent (1.1 for both).



The flag maps show also that there is little extrapolation of the transfer function all over the site as shown by the test on the sampling strategy : 13% for LAI and fAPAR, 9.5% for LAI57 and 4% for fCover. Obviously, when using two bands to describe the summits of the convex hull (fCover) the convex_hull is less strict than when using 3 bands (other variables).

Figure 15. High resolution biophysical variable maps applied on the Fundulea site (top). Associated Flags are shown at the bottom: 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 and 48 ESUs together for all classes 2 to 5, and the AVE method for class 1 which corresponds to forests. For all the variables, the regression coefficients are computed by relating the variable itself to the reflectance. The band combinations are different from one regression to another. Results show good consistency between the variables and the available land cover map and the flag associated to each map show that the transfer function is used as an extrapolator for less than 15% of the cases.

The biophysical variable maps are available in two projections:

- Plate carrée: latitude/longitude in WGS84 at 1.7857142857e-004° resolution
- UTM, 35N, WGS84 at 20m resolution.

The transfer function was applied on the SPOT image which was geo-located in UTM35N, WGS84. The resampling to obtain the plate carrée projection was performed using the nearest neighbour convolution to get consistent flags.

5 Acknowledgements

We want to thank all the people who participated to the campaign: Frédéric Baret, Claire Lauvernet, Franck Oro. Special thanks to Roxana Vintila for the organization of the campaign,her participation, as well as all her useful advices for the SPOT image geo-referencing.