VALERI 2002: Gilching site (crops and forest)

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

Philippe Rossello, Marie Weiss, Frédéric Baret

1   Introduction

This report describes the production of the high resolution, level 1, biophysical variable maps for the Gilching site (Table 1 gives the coordinates) in 2002 (see campaign report for more details about the site and the ground measurement campaign: 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 3.5) developed at INRA-CSE. The derived biophysical variable maps are:

• four Leaf Area Index (LAI) are considered: effective LAI (LAIeff) and true LAI (LAItrue) derived from the description of the gap fraction as a function of the view zenith angle; effective LAI57 (LAI57eff) and true LAI57 (LAI57true) derived from the gap fraction at 57.5°, which is independent on the leaf inclination. Effective LAI and effective LAI57 do not take into account clumping effect;

• cover fraction (fCover): it is the percentage of soil covered by vegetation between 0° 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.

The land cover is composed mainly of forests which are dominated by pine trees, with punctually deciduous species, and crops with a predominant presence of the winter cereals (wheat, barley, triticals and avenae). The site is generally quite flat (average altitude: 580 m), except some undulations in the middle that can reach up to 50 m amplitude (for more information, see campaign report: http://www.avignon.inra.fr/valeri). The site coordinates used for the process are described in Table 1:

	UTM, 32 North, WGS84 (units = meters)		Geographic Lat/Lon WGS84 (units = degrees)	
	Lat.	Lon.	Lat.	Lon.
Upper left corner	5329603.1210	671288.8021	48.09665408	11.30060142
Lower right corner	5326433.1210	674308.8021	48.06724294	11.33989640
Center	5328018.1210	672798.8021	48.08204003	11.32036007

Table 1. Description of the site coordinates.

The campaign report indicates different coordinates[1] but as some ESUs (V3, V7, V30…) were located outside the site (3x3 km), but close to the borders, the 3x3 km square considered during the campaign was shifted by -422 m in longitude and -37 m in latitude, but also enlarged[2] to take into account a maximum of data.

The ground measurements were carried out from 17/07/2002 to 19/07/2002, while the high spatial resolution image (SPOT2, HRV1, resolution: 20 m) was acquired on 08/07/2002.

---

[1] Please read the campaign report for more details about the coordinates: http://www.avignon.inra.fr/valeri.
[2] 3.020 x 3.170 km.

2   Available data
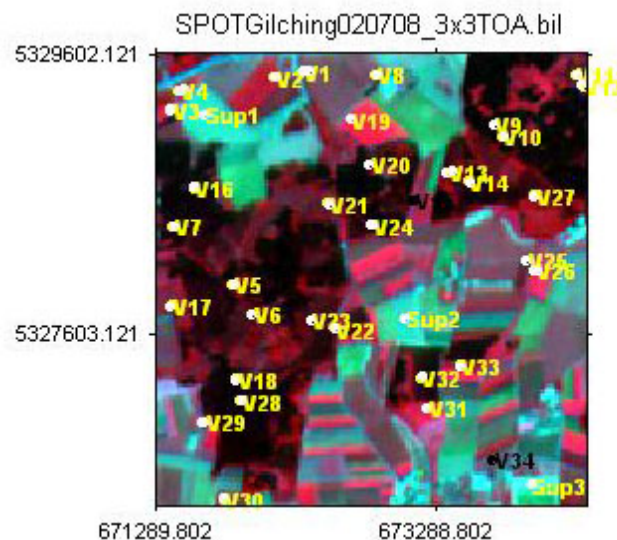
2.1   Sampling strategy

The distribution of the Elementary sampling units (ESU) was mainly designed empirically, from the impression got from the field overview. The spatial scatter between the ESUs, the representativeness of the land cover variability and some accessibility criterions were used to choose the location of 34 ESUs[3] (Figure 1). See campaign report for more details about the sampling strategy: http://www.avignon.inra.fr/valeri.

Figure 1 shows that the ESU locations are evenly distributed over the site (3x3 km). The processing of the ground data has shown that:

- ESUs V15 and V34 (in black on Figure 1) were located on a small plot with a strong heterogeneity on the borders. These two ESUs were eliminated;
- considering that SPOT geo-location and GPS measurements are associated to errors, we found that processed LAI for ESUs V1, V3, V4, V13, V24, V26 and V29 did not correspond to the SPOT pixel in terms of reflectance as compared to the knowledge of the land use: according with the people who acquired the data, they have been shifted by 1 or 2 pixels.

Three ESUs (Sup1, Sup2 and Sup3) were added to improve the representativeness of the land cover (§2.2). Sup1 and Sup2 are located in mature colza fields, while Sup3 is located in a winter cereal field. For these three ESUs, the LAI value is equal to 0. The aim is to refine the results of the robust regression.

Finally 35 ESUs have been kept for the computation of the transfer function (Figure 1).

Figure 1. Distribution of the ESUs around the Gilching site. ESUs in black (V15 and V34) were eliminated for the computation of the transfer function.

Figure 2 shows the location of the ESUs and the land cover. A landscape of forest and various crops characterize the Gilching site.

---

[3] 19 forest, 5 winter cereals, 5 grassland, 2 maize, 3 potatoes.
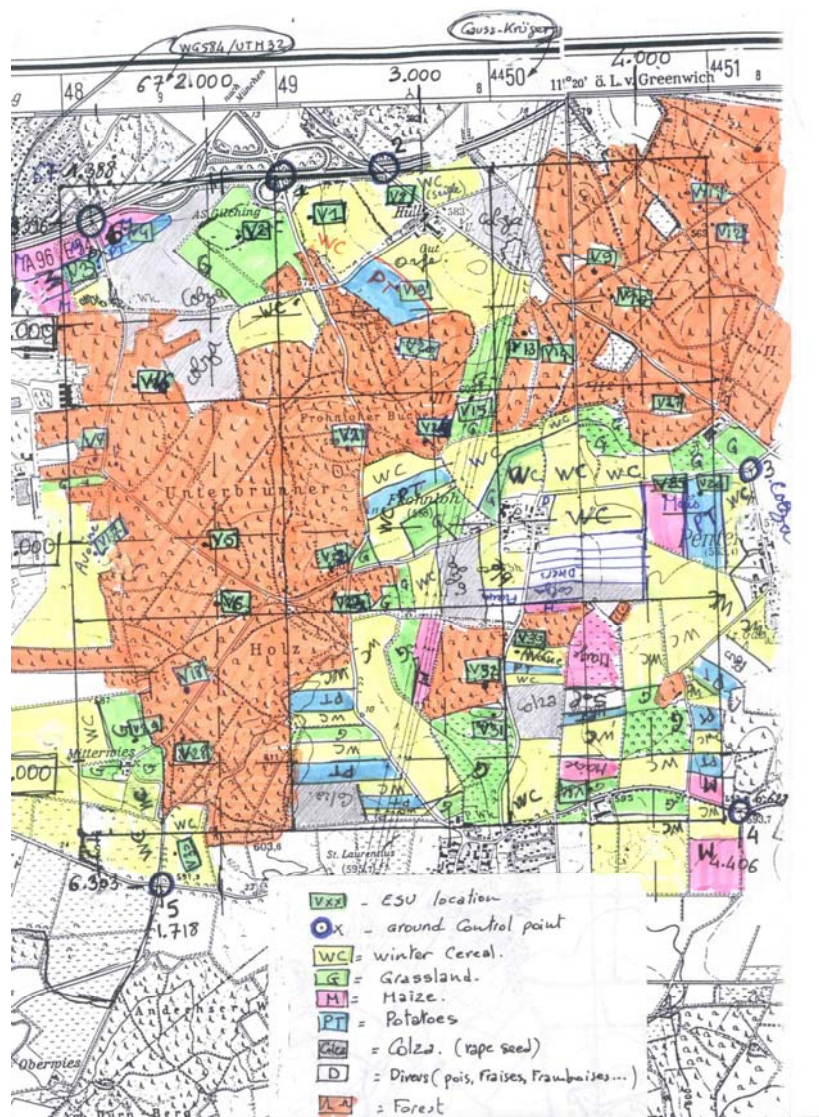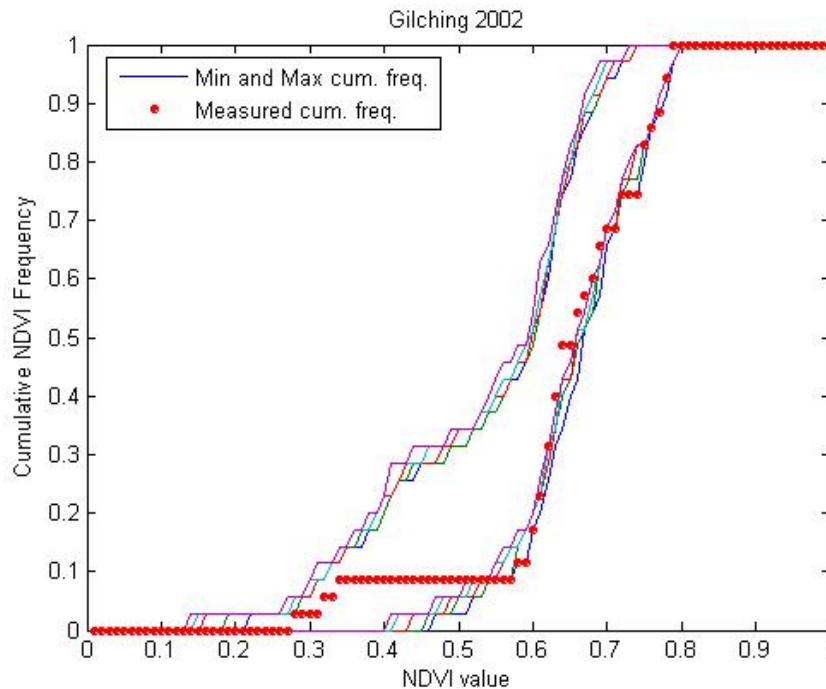
Figure 2. Land cover map[4] of the Gilching site (2002)

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 3). As the number of pixels is drastically different for the ESU and whole site (WS=22500 in case of a 3x3 km 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.

---

[4] This map gives an approximate indication of the land cover (document given for information).

Figure 3. Comparison of the ESU NDVI distribution and the NDVI distribution over the whole image.
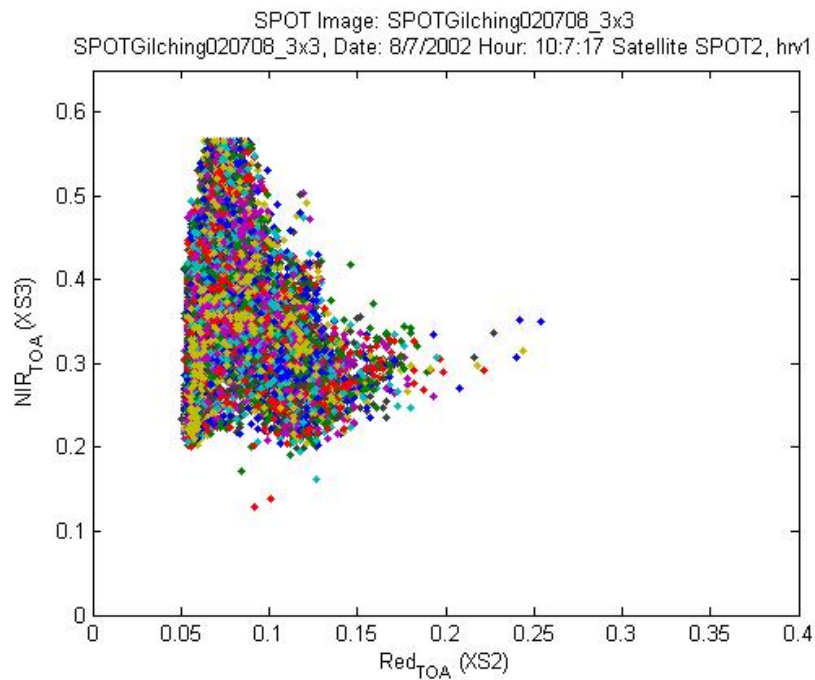
Figure 3 shows that the NDVI distribution of the 35 ESUs is quite good over the whole site (comprised between the 5 highest and lowest cumulative frequencies) even if the cumulative frequency curve is very close to the boundaries for high NDVI values. It reaches even the boundaries on several occasions since NDVI between 0.33 and 0.57 have not been sampled. Note that NDVIs lower than 0.28 have not been sampled either although they are present in the image. Moreover, the site is quite homogeneous in terms of NDVI since the highest and lowest distributions are close.
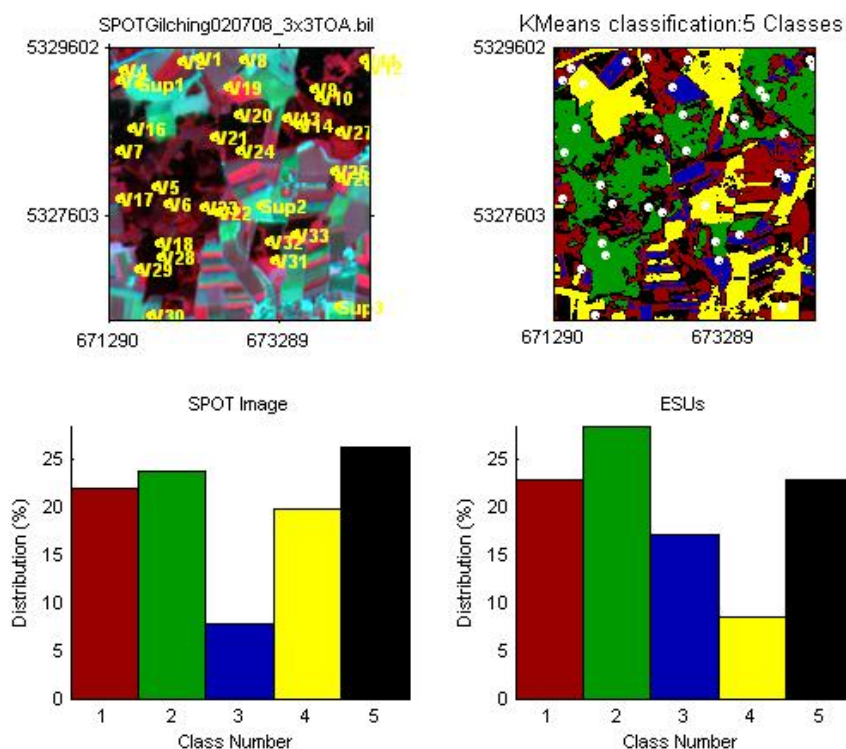
2.2   SPOT image

The SPOT image was acquired the 8th July 2002 by HRV1 on SPOT2. It was geo-located by SPOTimage (SPOTView basic). The projection is UTM 32N, WGS84 (please, refer to the campaign report for more details: http://www.avignon.inra.fr/valeri) and 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 3x3km site. Therefore, it will be taken into account everywhere in the same way.

Figure 4 shows the relationship between RED and near infrared (NIR) SPOT channels: the soil line is rather well marked; the saturation in NIR is around 0.57 (the high values in forest are certainly at the origin of saturation); a few bare soil points are also visible.

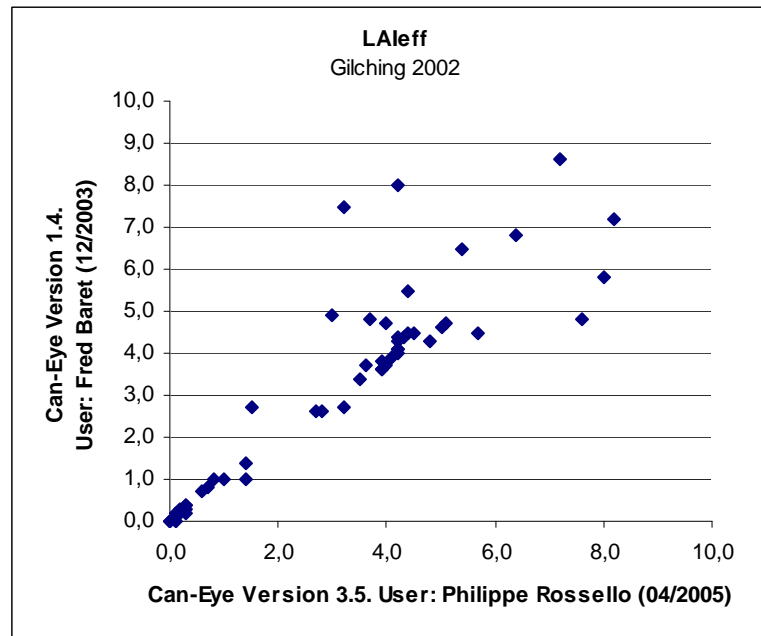**Figure 4. Red/NIR relationship on the SPOT image for Gilching, 2002**

Figure 5. Classification of the SPOT image. Comparison of the class distribution between the SPOT image and sampled ESUs.

A non supervised classification based on the k_means method (Matlab statistics toolbox) was applied to the reflectance 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 5). The distribution of the classes on the image and on the ESUs is rather different. Classes 1, 2 (forests) and 3 appear to be over-sampled whereas classes 4 and 5 are under-represented. Class 4 is only represented by the three added ESUs (Sup1, Sup2 and Sup3, §2.1).

2.3   Hemispherical images

The hemispherical images were processed by the CAN-EYE software (Version 3.5) to derive the biophysical variables. A first process was carried out with the Version 1.4. The major improvement between these two versions is that version 3.5 provides estimation of true LAI and clumping effect whereas version 1.4 does not.

Figure 6 shows the comparison between the LAIeff results of the hemispherical images processing using two different CAN-EYE versions.

Figure 6. Comparison of CAN-EYE processing results from two versions used by different users. Points in blue correspond to 35 ESUs (please, refer to the campaign report: http://www.avignon.inra.fr/valeri).

The relationship is consistent, even if differences are visible especially when the soil is covered with leafs. The way of processing the hemispherical images explains also the variations. The "user effect" is probably a determinant factor.

Figure 7 and Figure 8 show the distribution of the different measured variables over the sampled ESUs. As there was understorey in most of the ESUs, hemispherical images were acquired from above the understorey and from below the canopy (trees). The two sets of acquisition were processed separately to derived LAI (effective and true), LAI57 (effective and true), fCover, and fAPAR. The ESU biophysical variable was then computed as:
- LAIeff, LAI57eff, LAItrue, LAI57true: 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 at all the scales but it is the only way to get the total fCover. However, for the local scales considered, this might be true as a first order approximation.
- fAPAR: [1-(1-fAPAR(below))*(1-fAPAR(above))], since 1-fAPAR can be considered equivalent to a gap fraction. Here again, the same independency between the two layers has to be assured.

LAI (effective and true) derived from directional gap fraction and LAI derived from gap fraction at 57.5° (effective and true) are consistent (Figure 8). Effective LAI (LAIeff, LAI57eff) varies from 0 to 7, while true LAI (LAItrue, LAI57true) varies from 0 to 10.

This range shows a quite heterogeneous site in terms of LAI. The ESUs have actually effective LAI (LAIeff, LAI57eff)>2 and true LAI (LAItrue, LAI57true)>3 since the value 0 corresponds to Sup1, Sup2 and Sup3. For values, LAIeff and LAI57eff are lower than LAItrue and LAI57true. This is due to the clumping observed for several ESUs. Note that the relationship between fAPAR and LAI is in agreement with what is expected (beer lambert law) while the fCover-LAI relationship is more noisy.

To build the relationships between biophysical variables and SPOT data, the reflectance of a given forest ESU was considered as the average reflectance over the central pixel + the 8 surrounding pixels, whereas, for crops, we took the reflectance of the pixel corresponding to the ESU. This takes into account the fact that the height of the trees are about 20 m and consequently the fish-eye observes an area of $\pi \; x \; [20 \; x \tan(60°)]^2 \cong 3770$ m², *i.e.* close to the area of 9 SPOT pixels (=3600m²) when using a maximum view zenith angle of 60°.
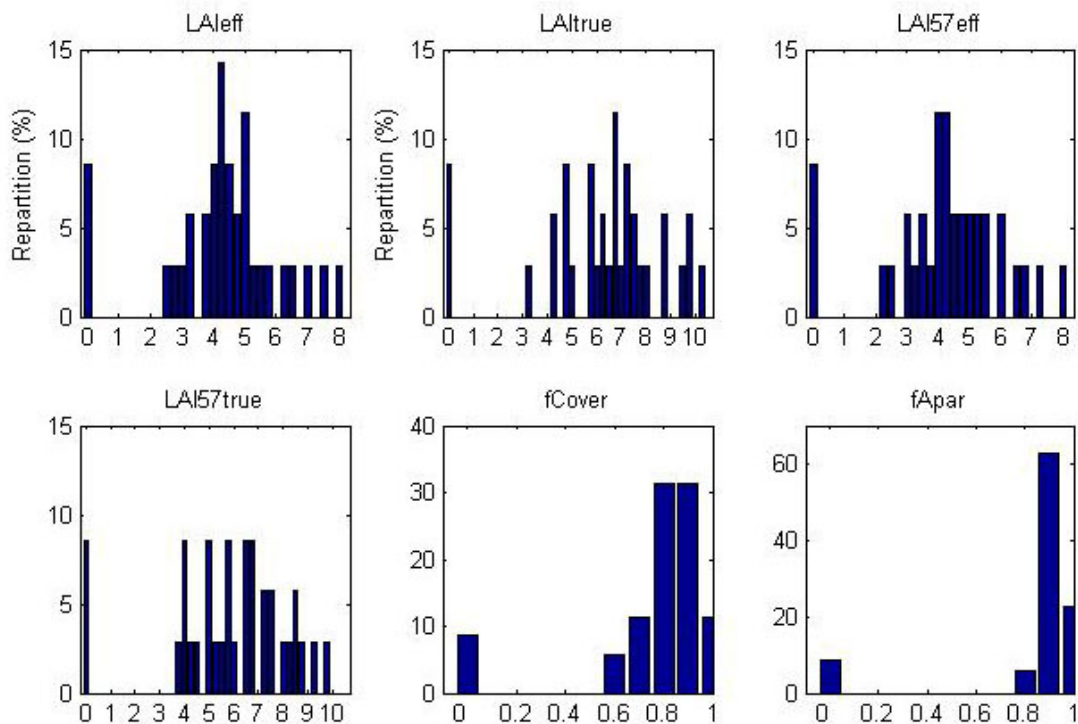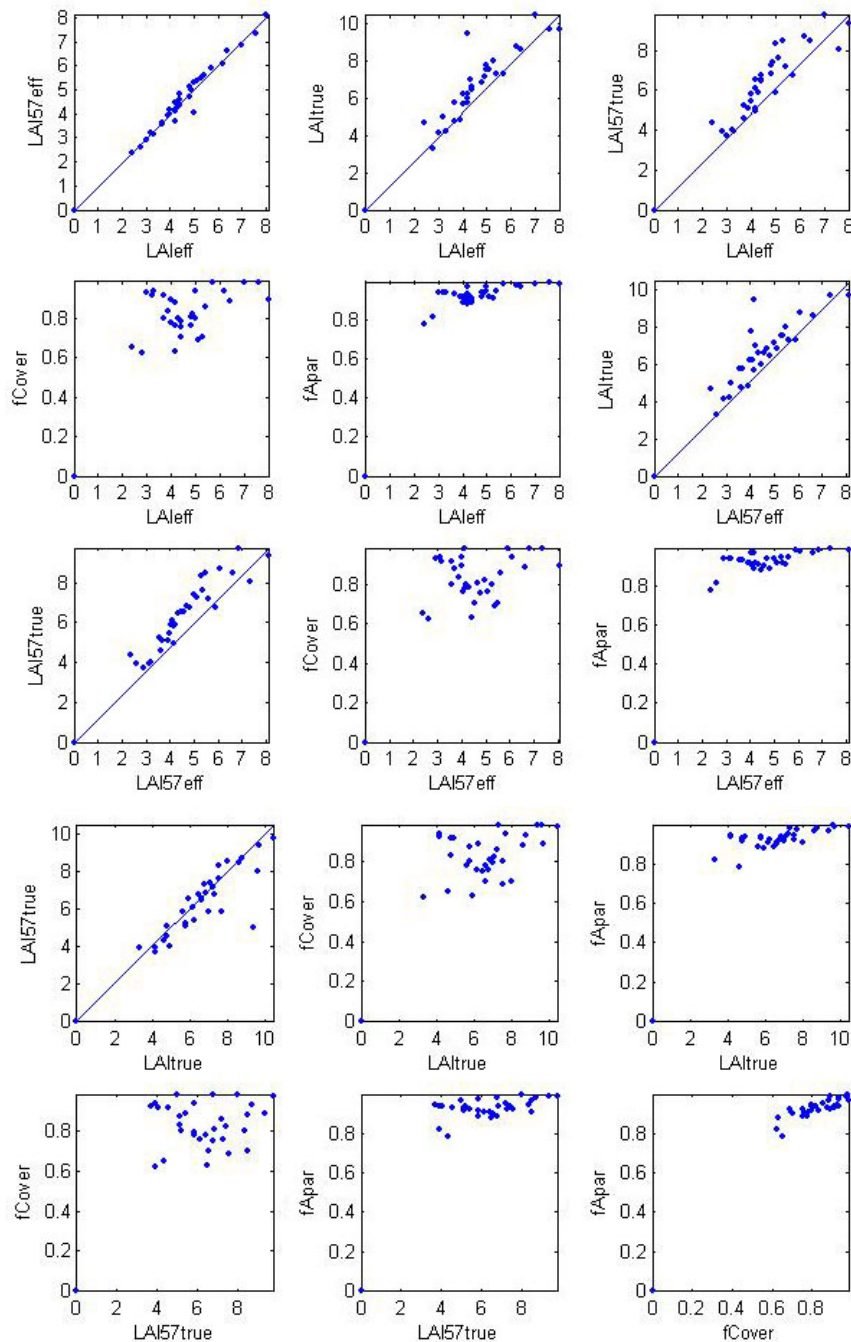

Figure 7. Distribution of the measured biophysical variables over the ESUs.

Figure 8. Relationships between the different biophysical variables

Figure 9 shows the different relationships observed between the biophysical variables and the corresponding NDVI on the ESUs, as a function of the SPOT classes determined in §2.2. The additional ESUs (Sup1, Sup2 and Sup3) improve the relationships between the biophysical variables and corresponding NDVI, even if the results are not very good. The under-represented class 4 (bare soil for example) is in question.

Following this figure, it is clear that the evaluation of the transfer function will be affected by the lack of data for NDVI between 0.33 and 0.57. No different behaviour between the classes can be observed; therefore a single transfer function per variable will be generated.
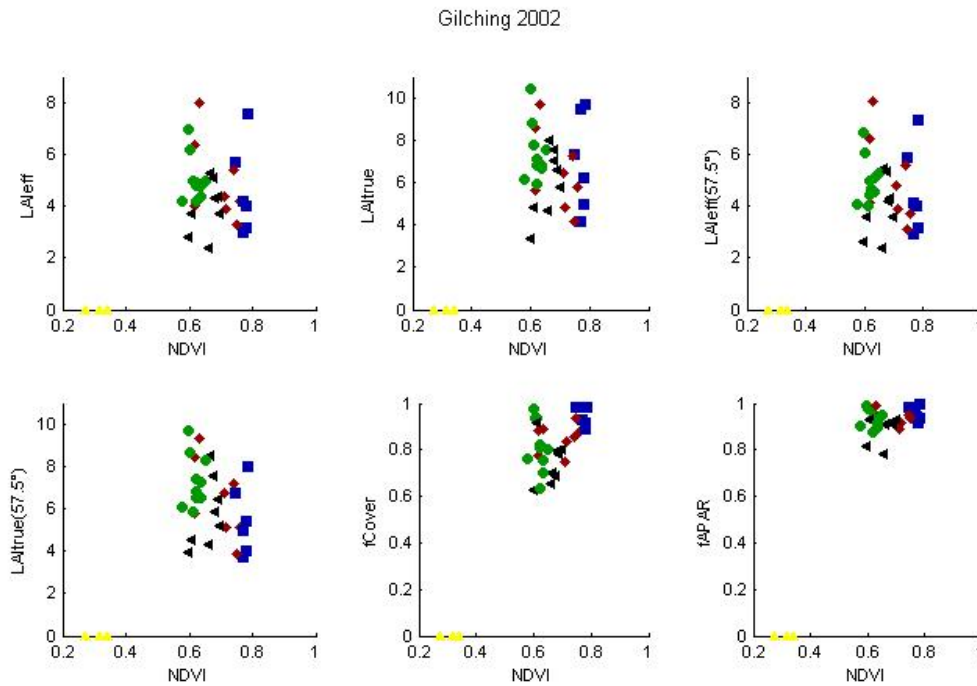
Figure 9. NDVI-Biophysical Variable relationships as a function of SPOT classes

3   Determination of the transfer function for the 6 biophysical variables: LAIeff, LAI57eff, LAItrue, LAI57true, fCover, fAPAR

3.1   The Transfer functions considered

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

• REG: If the number of ESUs is sufficient, multiple robust regression between ESUs reflectance (or Simple Ratio) and the considered biophysical variable can be applied: we used the 'robustfit' function from the matlab statistics toolbox. It uses an iteratively re-weighted least squares algorithm, with the weights at each iteration computed by applying the bisquare function to the residuals from the previous iteration. This algorithm provides 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 the corresponding measured biophysical variable. For a given pixel, a cost function is computed as the sum of the 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, which was quite the case for Gilching.

Both regression and Look-Up-Tables are tested using either the reflectance or the logarithm of the reflectance for any band combination as well as 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.
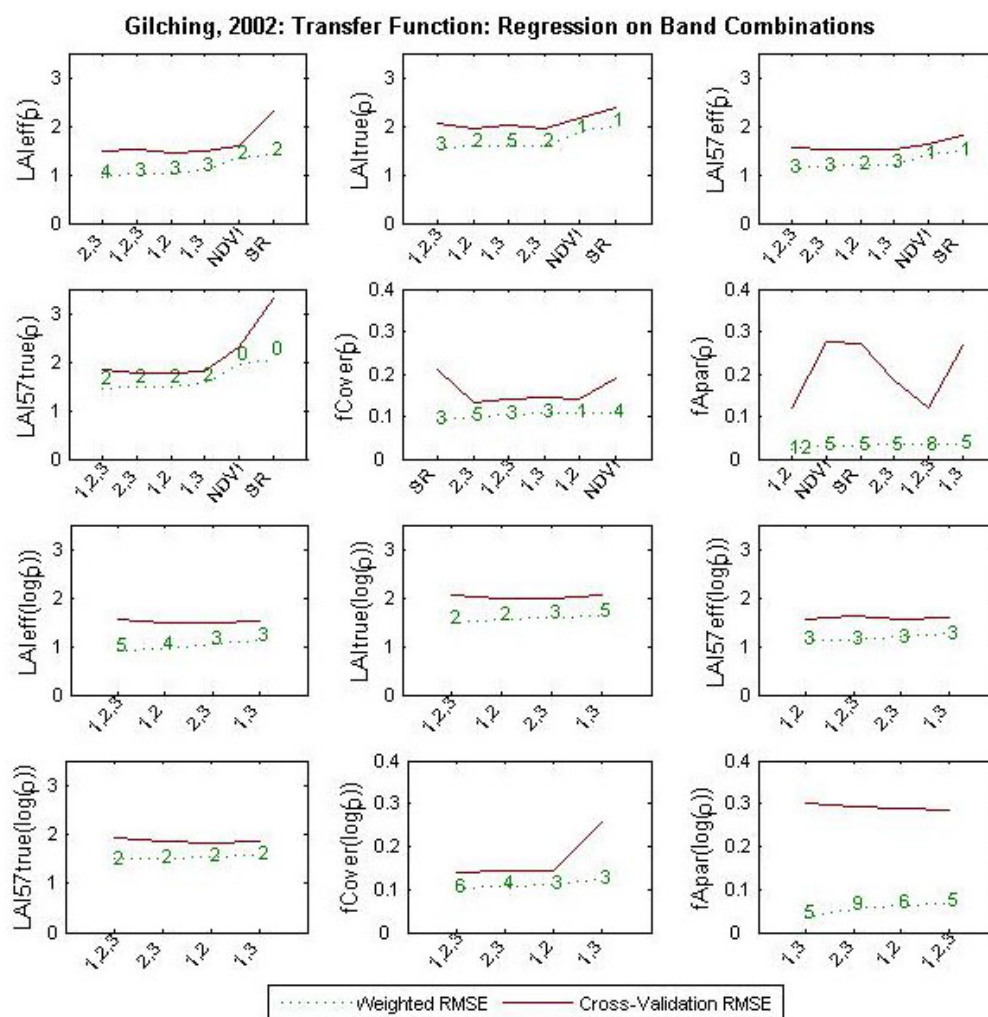
- Pixels inside the 'large convex-hull': for each class, a convex-hull is computed using all the reflectance combination (±5% in relative value) used for the transfer function, and corresponding to the ESUs belonging to the class. For those pixels, the transfer function is used as an extrapolator (but 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 Gilching site

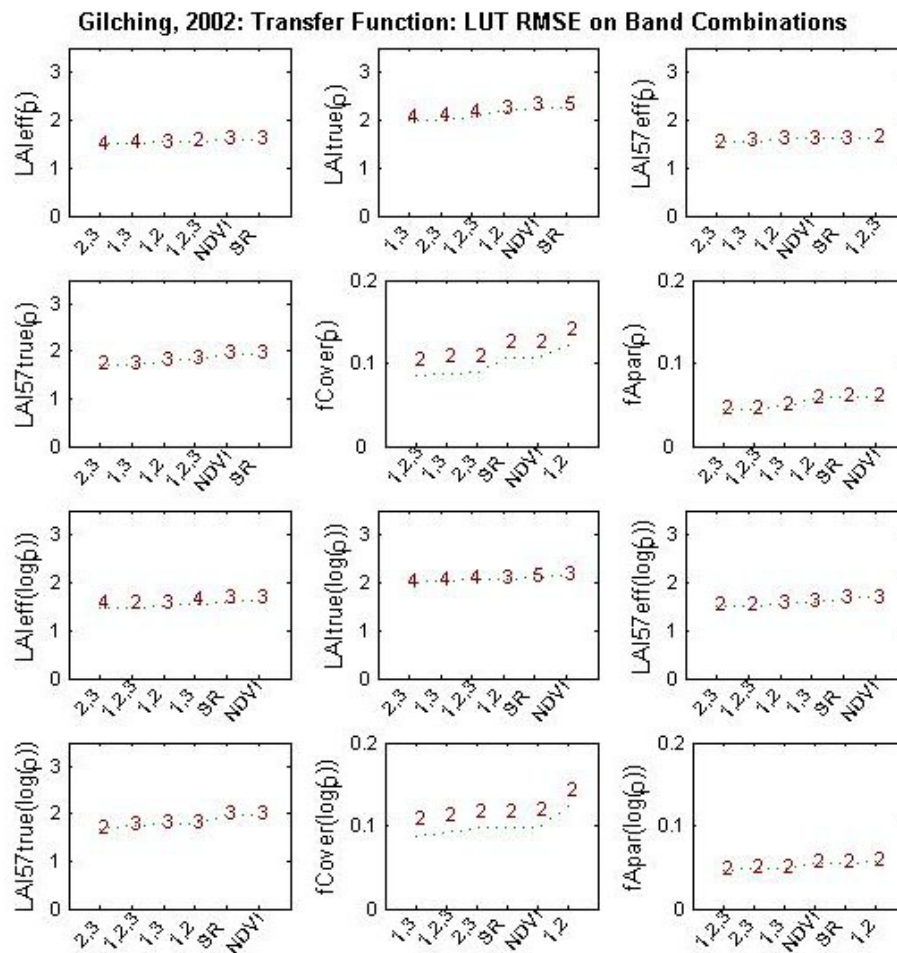### 3.2.1   *Choice of the method*

For the 5 classes, a unique transfer function was computed. Figure 10 and Figure 11 show the results obtained for all the possible band combinations using either the reflectance or the logarithm of the reflectance:

- The REG method provides better results in terms of cross-validation RMSE for all the variables and is therefore selected as the transfer function instead of the LUT;
- For LAIeff, LAItrue, LAI57eff, LAI57true, fCover and fAPAR, the results using the reflectance are better. According to the biophysical variable, the choice of the method proves to be difficult because the results are close.

Figure 10. Transfer function: test of multiple regression 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 11. 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 LAIeff, the XS2, XS3 combination on reflectance was selected since it provides a good compromise between the number of weights lower than 0.7 (four), the cross-validation RMSE value (among the lowest values) and the weighted root mean square error (Figure 12).

For all the biophysical variables, V8, V9, V30 and V31 very often have weights lower than 7. The forest characterizes V9 while V31 is covered with dense grassland. V8 and V30 correspond to mature wheat which is considered as "green vegetation" by the CAN-EYE software (choice of the user). This method is questionable even if it does not pose problem for V1. The LAI evaluated by the software is quite high (respectively 6.4 and 8). The four ESUs are also located in edge of fields or in small fields. Moreover, the estimation of LAI for V30 and V31 strongly differs according to the version of the CAN-EYE software. These different parameters probably explain the behaviour of the four ESUs.

Note that the RMSE is quite high (of the order of 1.5) since the LAI values over the site are quite high which means that corresponding reflectances are in the saturation domain. Moreover, the acquired image (SPOT2) does not provide reflectances in the middle infrared domain which could have improved the results (especially for forests).
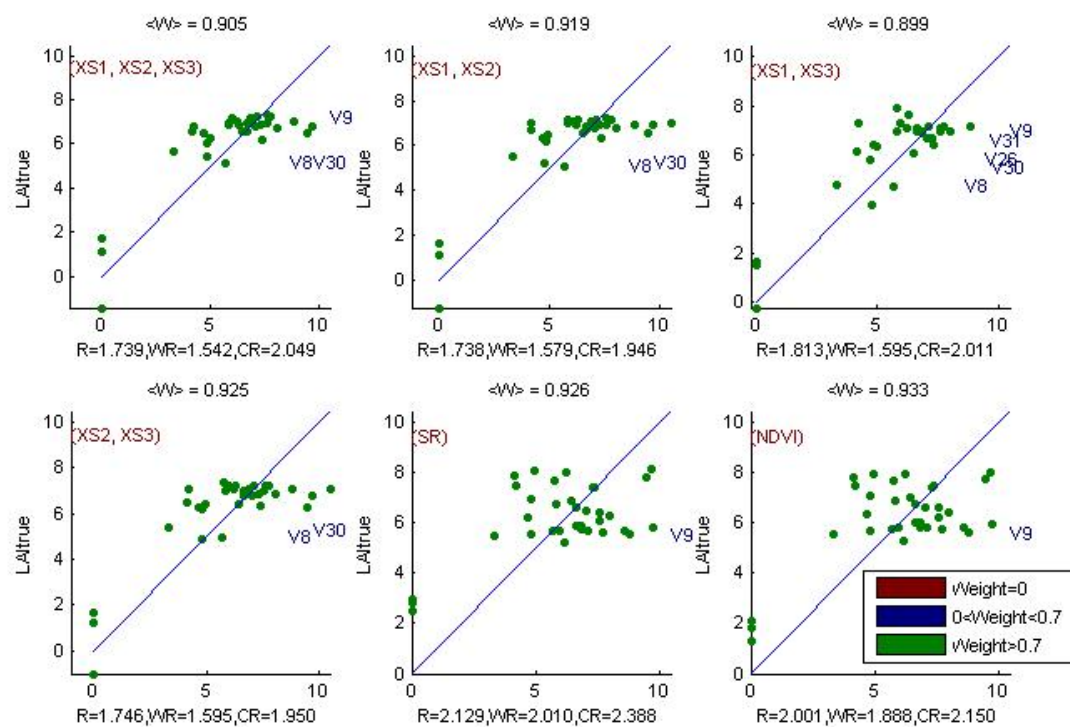
Gilching, 2002: Regression on reflectance:LAIeff

Figure 12. Effective Leaf Area Index: results for regression using different band combinations. R is the root mean square error computed between LAIeff and estimated LAIeff. WR is the weighted root mean square error and CR is the cross validation root mean square error.
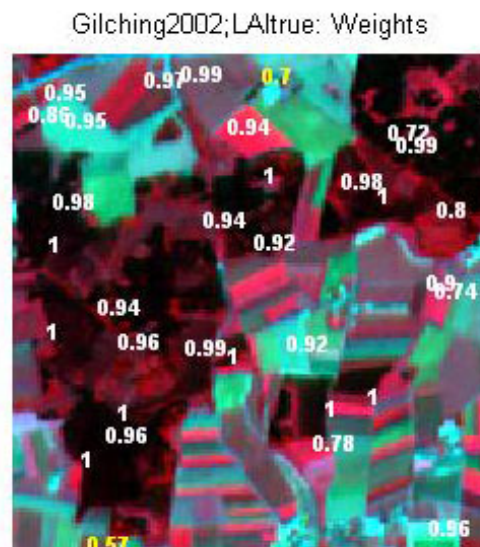

Figure 13. Weights associated to each ESU for the determination of LAIeff transfer function.

For the LAItrue, the XS2, XS3 combination on reflectance was selected since it provides a low cross-validation RMSE value, a weighted root mean square error value among the lowest and two weights lower than 0.7 (Figure 14).
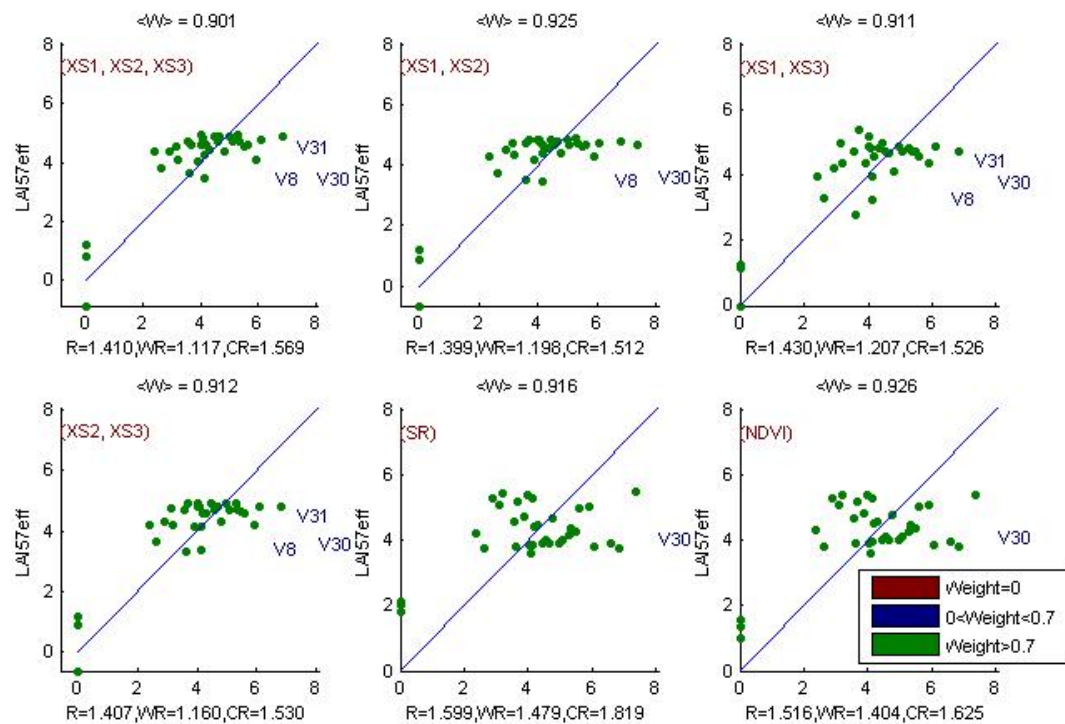
Gilching, 2002: Regression on reflectance:LAItrue

Figure 14. True Leaf Area Index: results for regression using different band combinations. R is the root mean square error computed between LAItrue and estimated LAItrue. WR is the weighted root mean square error and CR is the cross validation root mean square error.

Figure 15. Weights associated to each ESU for the determination of LAItrue transfer function.

For the LAI57eff, the XS2, XS3 combination on reflectance was selected since it provides a good compromise (Figure 16) between the number of weights lower than 0.7 (three), the cross-validation RMSE value (among the lowest values) and the weighted root mean square error (among the lowest values).

Gilching, 2002: Regression on reflectance:LAI57eff

Figure 16. Effective Leaf Area Index at 57.5°: results for regression using different band combinations. R is the root mean square error computed between LAI57eff and estimated LAI57eff. WR is the weighted root mean square error and CR is the cross validation root mean square error.
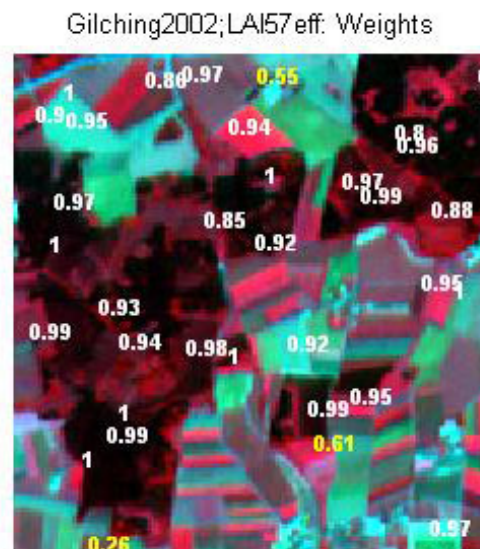

Figure 17. Weights associated to each ESU for the determination of LAI57eff transfer function.

For the LAI57true, the XS2, XS3 combination on reflectance was selected since it provides a good compromise (Figure 18) between the number of weights lower than 0.7 (two), the cross-validation RMSE value (the lowest value) and the weighted root mean square error (among the lowest values).
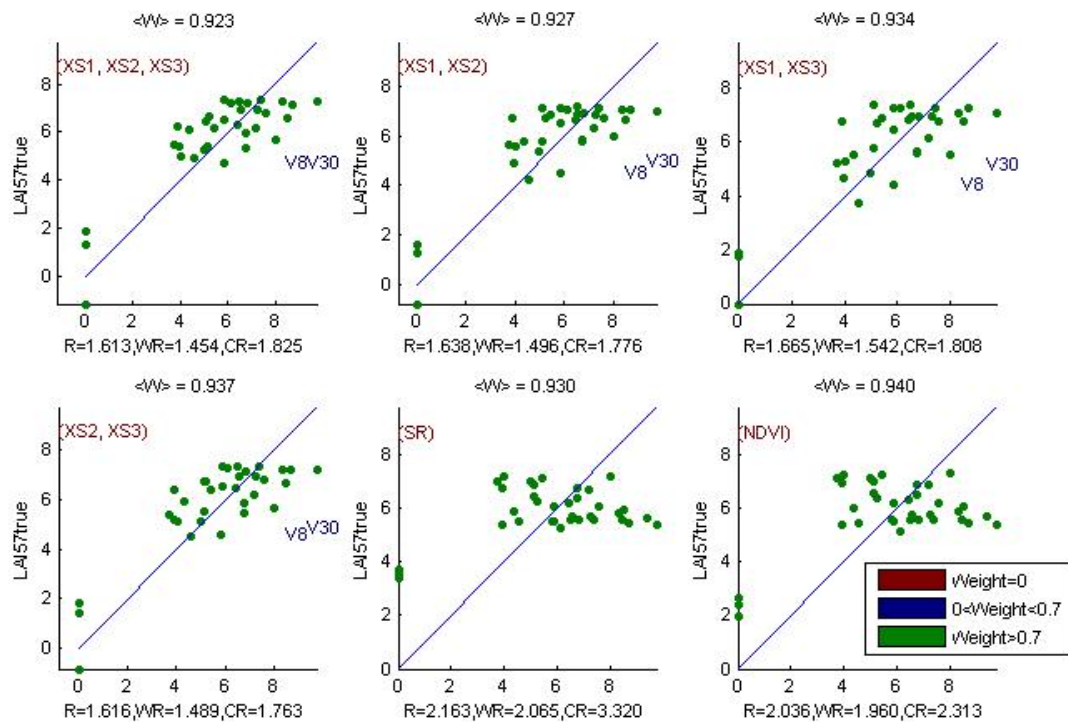
Figure 18. True Leaf Area Index at 57.5°: results for regression using different band combinations. R is the root mean square error computed between LAI57true and estimated LAI57true. WR is the weighted root mean square error and CR is the cross validation root mean square error.
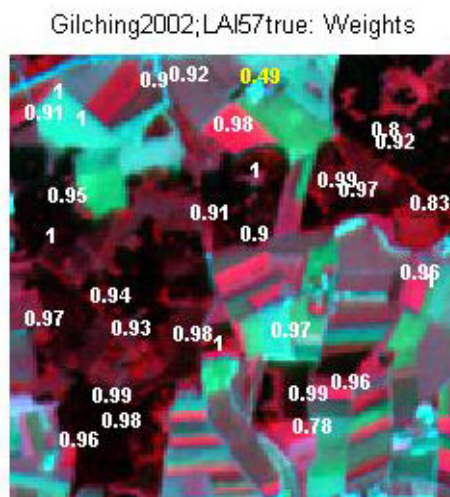

Figure 19. Weights associated to each ESU for the determination of LAI57true transfer function.

For the fCover, the XS1, XS2, XS3 combination on reflectance was selected since it provides a good compromise (Figure 20) between the number of weights lower than 0.7 (three), the cross-validation RMSE value (among the lowest values) and the weighted root mean square error (among the lowest values).

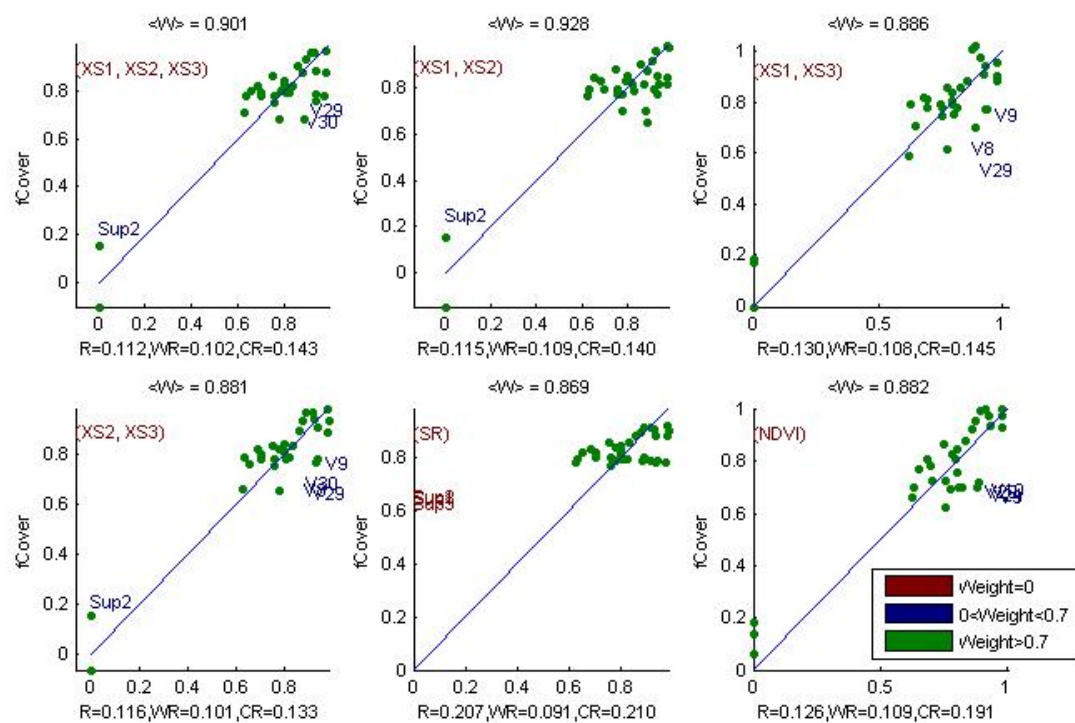**Gilching, 2002: Regression on reflectance:fCover**

Figure 20. fCover: results for regression using different band combinations. R is the root mean square error computed between fCover and estimated fCover. WR is the weighted root mean square error and CR is the cross validation root mean square error.
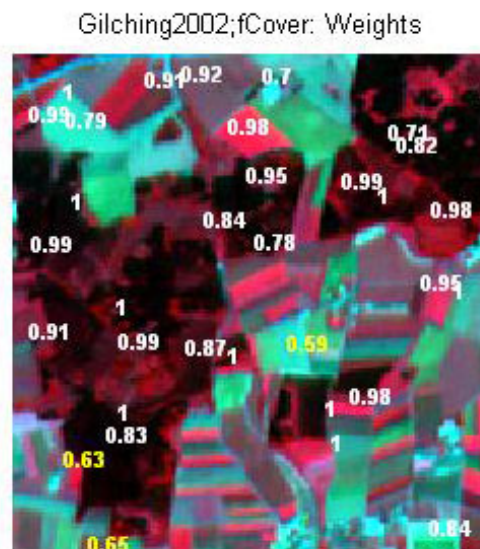

Figure 21. Weights associated to each ESU for the determination of fCover transfer function.

For the fAPAR, the XS1, XS2, XS3 combination on reflectance was selected since it provides the lowest cross-validation RMSE value and a weighted root mean square error value among the lowest. However, eight weights are lower than 0.7 (Figure 22).

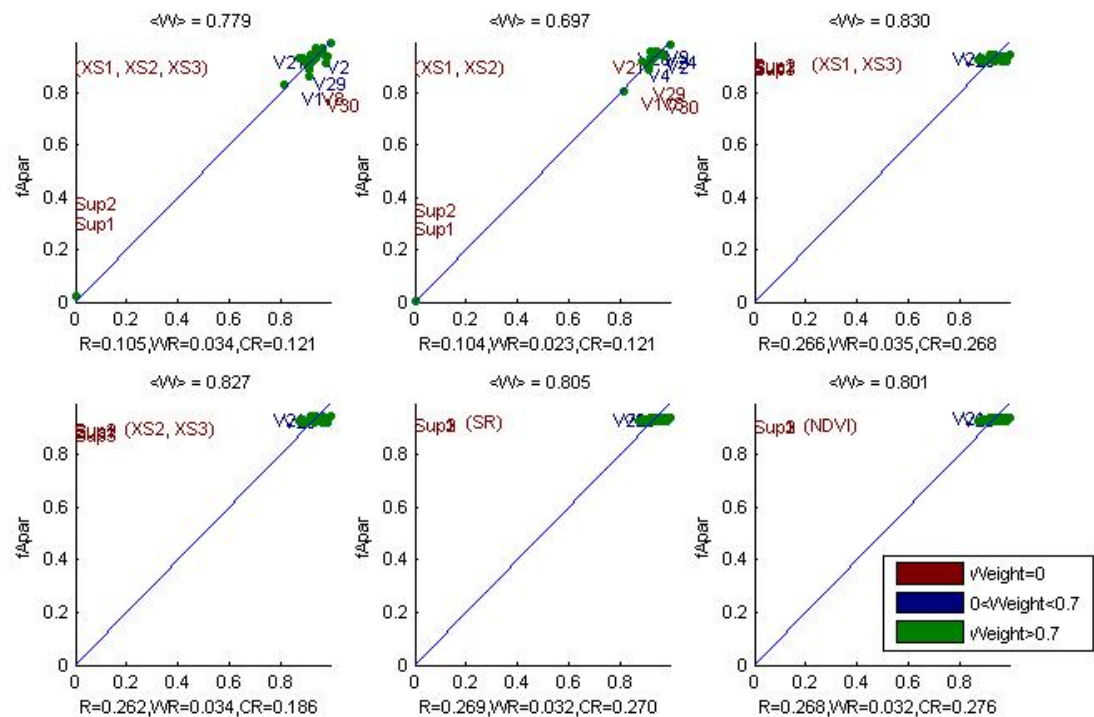**Gilching, 2002: Regression on reflectance:fAPAR**

Figure 22. fAPAR: results for regression using different band combinations. R is the root mean square error computed between fAPAR and estimated fAPAR. WR is the weighted root mean square error and CR is the cross validation root mean square error.
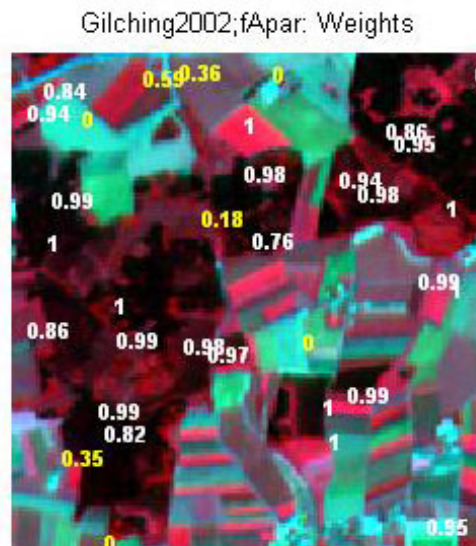

Figure 23. Weights associated to each ESU for the determination of fAPAR transfer function.

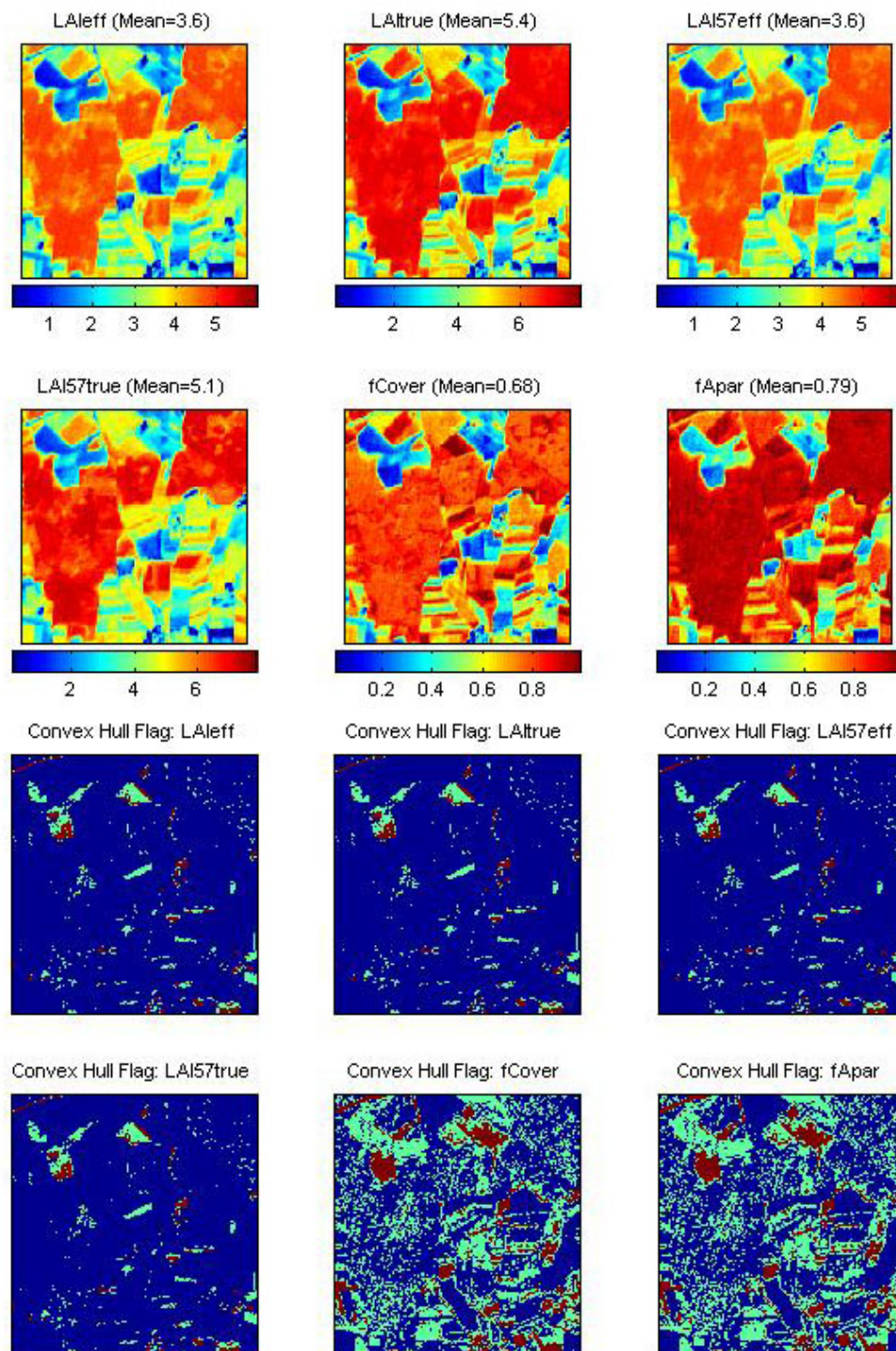Following, the results of the transfer function (Table 2):

Variable	Band Combination	RMSE	Weighted RMSE	Cross-valid RMSE
LAIeff	7.7154 - 48.131(XS2) - 0.79514(XS3)	1.376	0.957	1.479
LAItrue	10.891 - 70.648(XS2) + 0.66255(XS3)	1.746	1.595	1.950
LAI57eff	7.4365 - 47.14(XS2) + 0.060872(XS3)	1.407	1.160	1.530
LAI57true	11.817 - 67.755(XS2) - 3.6058(XS3)	1.616	1.489	1.763
fCover	0.60128 + 8.8759(XS1) - 13.169(XS2) + 0.54846(XS3)	0.112	0.102	0.143
fAPAR	0.72659 + 12.094(XS1) - 14.965(XS2) - 0.1155(XS3)	0.105	0.034	0.121

Table 2. Transfer function applied to the whole site for the different biophysical variables, and corresponding errors

3.3   Applying the transfer function to the Gilching SPOT image extraction

Figure 24 presents the biophysical variable maps obtained with the transfer function described in Table 2. The maps obtained for the six variables are consistent, showing similar patterns: low LAIeff values where low fCover or fAPAR are observed and conversely… The difference between effective LAI and true LAI is significant, due to the clumping effect in forest (see the average values in Figure 24). This was expected when looking the LAIeff/LAI true relationship in Figure 8, showing that for high LAI the difference between the two can be significant. For example, an effective LAI of 5 may correspond to an effective LAI of 8.

Figure 24. High resolution biophysical variable maps applied on the Gilching site (top). Associated Flags are shown at the bottom: blue and light blue corresponds to the pixels belonging to the 'strict' and 'large' convex hulls and red to the pixels for which the transfer function is extrapolating.

The flag maps are not very different between the biophysical variables. The results are very comparable between LAIeff, LAItrue, LAI57eff and LAI57true, but also between fCover and fAPAR. The extrapolation of the transfer function is little all over the site. It corresponds to the plots where the LAI values (effective or true) are low or null. The sampling of the bare soil and the number of bands are also in question. Note that the extrapolation is minimal on the LAI flag maps. Few pixels are outside the strict convex hull.

4   Conclusion

The transfer functions are obtained by using 35 ESUs. As the representativeness of the land cover (§2.1 and 2.2) by the different ESUs was not optimal, the first results of the multiple robust regression were not very good. Three ESUs were added to improve it. They are located in mature colza fields and their reflectance value is low. Finally, the maps obtained for the biophysical variables are consistent and the flag associated to each map show that the transfer function is used as an extrapolator in little areas. For all the variables, the regression coefficients are computed by relating the variable itself to the reflectance.

The biophysical variable maps are available in UTM, 32 North, projection coordinates (Datum: WGS-84) at 20m resolution.

5   Acknowledgements