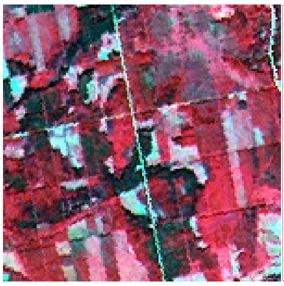
GROUND DATA PROCESSING &
PRODUCTION OF THE LEVEL 1 HIGH RESOLUTION MAPS



VALERI 2001

Järvselja site
(boreal forest)

Philippe Rossello
August 2007

CONTENTS

1. Introduction

This report describes the production of the high resolution, level 1, biophysical variable maps for the Järvselja site in June 2001 (see campaign report for more details about the site and the ground measurement campaign: annex or 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 (LAI2000 in this case).

The derived biophysical variable maps are:

- Leaf Area Index (LAI): LAI corresponds to effective LAI derived from the description of the gap fraction as a function of the view zenith angle;
- cover fraction (fCover): it is the percentage of soil covered by vegetation between 0° and 7° view zenith angle.

The site is "mostly covered by a sub-boreal mixed forest, including both conifers (Scots pine and Norway spruce) and deciduous (birch, aspen, alder). Agricultural fields are almost missing on the 3x3km region, however, a few unmanaged open areas are found". Note that the site is quite flat (for more information, see annex or campaign report: http://www.avignon.inra.fr/valeri).

The site coordinates are described in Table 1:

	Lambert-Est-92 WGS-84 (units=meters)		Geographic Lat/Lon WGS-84		UTM 35, North, WGS-84 (units=meters)	
	Easting	Northing	Lat	Lon	Easting	Northing
Upper left corner	689711.5701	6468156.8484	58.31283056	27.23791389	513940.0270	6463564.0266
Lower right corner	692711.5701	6465156.8484	58.28461754	27.28652912	516801.8817	6460433.9460
Center	691211.5701	6466656.8484	58.29872500	27.26223056	515370.9189	6461998.8340

Table 1. Description of the site coordinates: they correspond to SPOT image coordinates.

The ground measurements were carried out from 12th to 16th June 2001, while the high spatial resolution image (SPOT4, HRVIR1, resolution: 20 m) was acquired at the beginning of July.

2. Available data

2.1. SPOT image

The SPOT image was acquired the 4th July 2001 by HRVIR1 on SPOT4. The radiometric and geometric correction was performed by SPOT image (product 1B). The image was geo-referenced by TARTU Observatory. The projection is Lambert-Est-92 (Lambert Conformal Conic 2 parallel), WGS-84. The atmospheric correction was applied by TARTU Observatory. Please, refer to the campaign report for more details: annex or http://www.avignon.inra.fr/valeri.

Figure 1 shows the relationship between Red and near infrared (NIR) SPOT channels: the soil line is marked and no saturated point is observed.

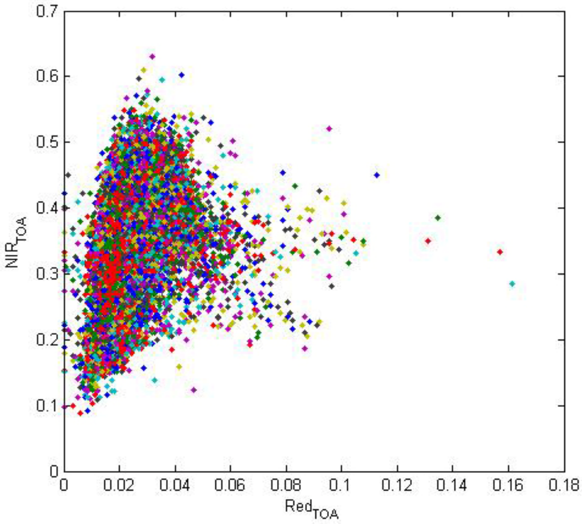


Figure 1. Red/NIR relationship on the SPOT image for Järvselja, 2001

2.2. LAI2000 measurements

For each Elementary Sampling Unit (ESU), the biophysical variables (LAI, fCover) were derived from LAI2000 instrument. The measurements have been acquired at two heights: ground level and breast height (1.30m). The two levels allow the distinction between understorey and trees. According to the sampling protocol, 48 measurements were taken at the both level for each ESU. In the VALERI context, we are interested in the whole leaf area index, therefore, the ESU biophysical variables that are used in the following were computed as:

- LAI = LAI_canopy + LAI_ground
- fCover is the percentage of soil covered by vegetation at 7° view zenith angle (ground level).

Figure 2 shows the distribution of the different measured variables over the sampled ESUs. LAI varies from 0.83 to 4.98 and fCover from 0.34 to 0.98. This range shows a heterogeneous site in terms of LAI with high biophysical variable values. To build the relationships between biophysical variables and SPOT data, the reflectance of a given forest ESU for which the height of the trees is equal or higher than 12 meters was considered as the average reflectance over the central pixel + the 8 surrounding pixels. Consequently, the fish-eye observes an area of at least $\pi \, x \, [12 \, x \, \tan(68°)]^2 \cong 2800 \, m^2$, *i.e.* close to the area of 9 SPOT pixels (=3600m²) when using a maximum view zenith angle of 68°.

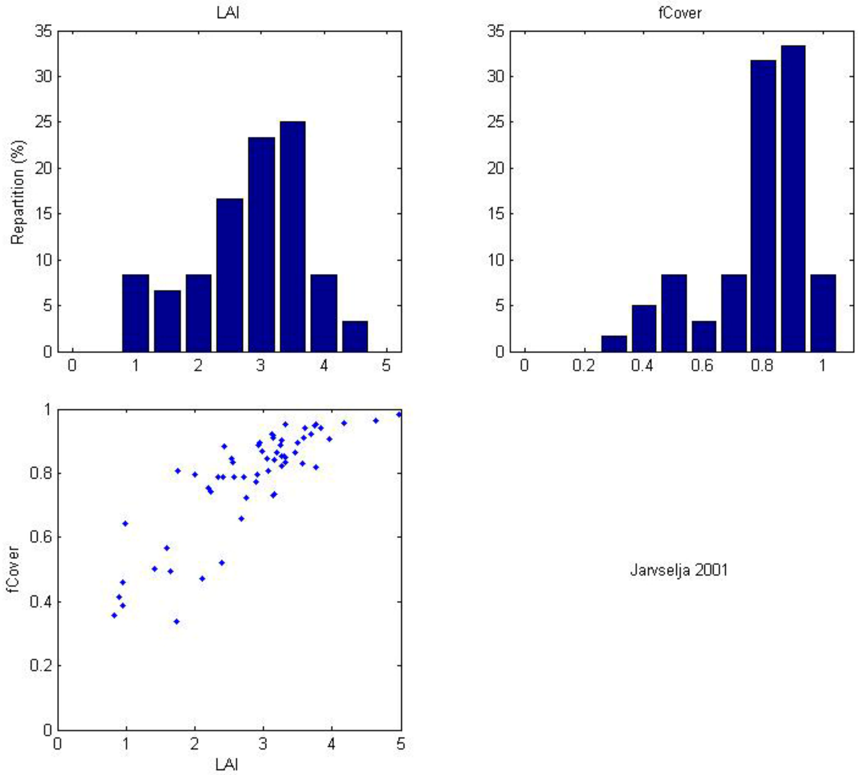


Figure 2. Distribution of the measured biophysical variables over the ESUs.

2.3. Sampling strategy

2.3.1. Principles

The sampling strategy is defined in the campaign report: http://www.avignon.inra.fr/valeri. It was attempting to represent as much as possible the range of variation of canopy types and conditions. In addition, some ESUs were organised within a cross pattern at the centre of the site to be able to get geostatistic estimates.

Figure 3 shows that the 69 ESUs are evenly distributed over the site (3 x 3 km). The processing of the ground data has shown that:

- ESUs t13612a, t13612b, t13612c, t13605a, t13605b, t13608, t13704b, t13704c, t13705a (in black on Figure 3) were located on a small plot with a strong heterogeneity on the borders or very close to the road. They were eliminated;

- considering that SPOT geo-location and GPS measurements are associated to errors, we found that processed LAI for e7604 did not correspond to the SPOT pixel in terms of reflectance as compared to the knowledge of the land use: it has been shifted by 1 pixel.

Finally, 60 ESUs have been kept for the computation of the transfer function:

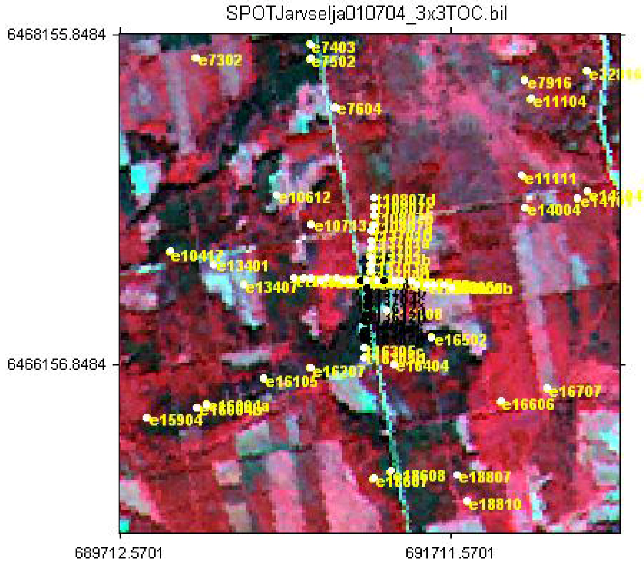


Figure 3. Distribution of the ESUs around the Järvselja site.

2.3.2. Evaluation based on NDVI values

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 4). As the number of pixels is drastically different for the ESU and whole site (WS = 22500 in case of a 3 x 3 km image at 20 m resolution), 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.

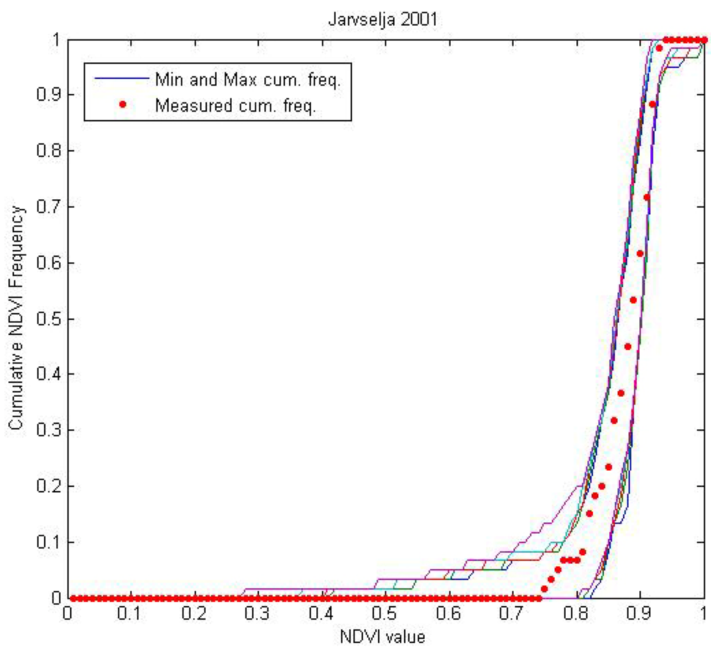


Figure 4. Comparison of the ESU NDVI distribution and the NDVI distribution over the whole image.

Figure 4 shows that the NDVI distribution of the 60 ESUs is very good over the whole site even if the cumulative frequency curve is close to the boundaries for high NDVI values comprised between 0.9 and 0.92. Note that NDVIs lower than 0.75 and higher 0.94 have not been sampled although they are present in the image. They may correspond to clear cuts, open areas, dense forest areas…

2.3.3. Evaluation based on classification

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 mainly different at level of the classes 2 and 4. The class 2 is under-represented, while the class 4 appears to be over-sampled.

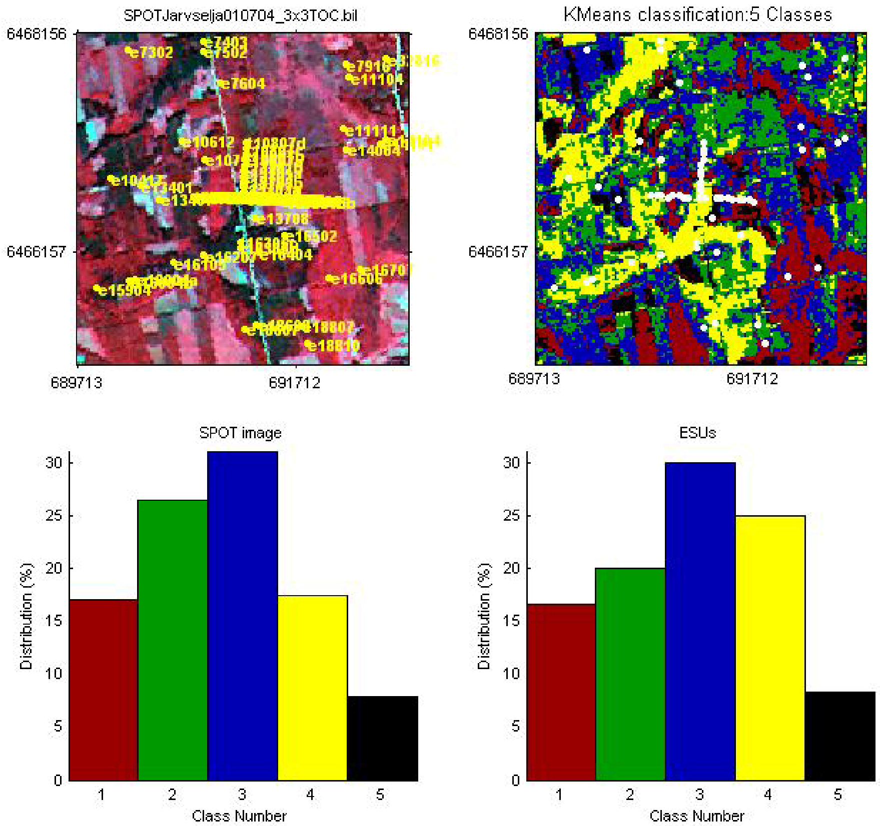


Figure 5. Classification of the SPOT image and comparison of the class distribution between the satellite image and sampled ESUs.

Figure 6 shows the different relationships observed between the biophysical variables and the corresponding NDVI on the ESUs, as a function of the SPOT classes determined from non supervised classification.

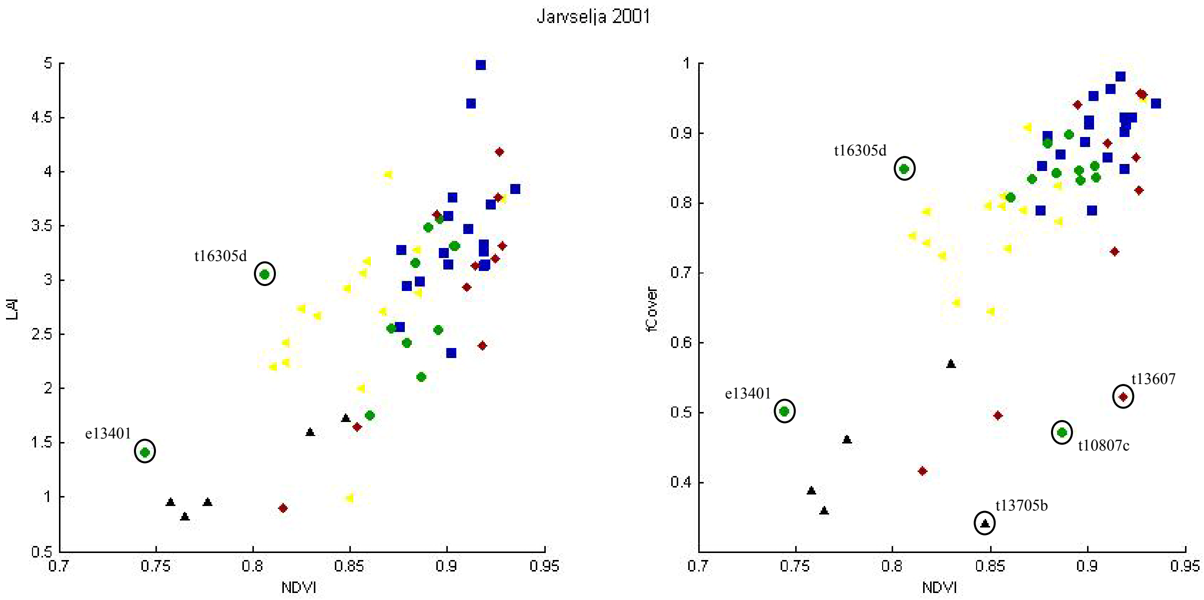


Figure 6. NDVI-biophysical variable relationships as a function of SPOT classes

Even if no different behaviour between the classes can be observed, a few ESUs (Figure 6) differ from the others: their biophysical variable values are generally high while NDVIs are low (or conversely). A single transfer functions will be generated.

2.3.4. Using convex hulls

A test based on the convex hulls was also carried out to characterize the representativeness of ESUs. Whereas the evaluation based on NDVI values uses two bands (red and NIR), this test uses the 4 bands (green, red, NIR and SWIR in this case) of the SPOT image. A flag image, is computing over the reflectances (Figure 7). The result on convex-hulls can be interpreted as:

- pixels inside the 'strict convex-hull': a convex-hull is computed using all the SPOT reflectance corresponding to the ESUs belonging to the class. These pixels are well represented by the ground sampling and therefore, when applying a transfer function the degree of confidence in the results will be quite high, since the transfer function will be used as an interpolator;
- pixels inside the 'large convex-hull': a convex-hull is computed using all the reflectance combination (±5% in relative value) corresponding to the ESUs. For these pixels, the degree of confidence in the obtained results will be quite good, since the transfer function is used as an extrapolator (but not far from interpolator);
- pixels outside the two convex-hulls: this means that for these pixels, the transfer function will behave as 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.

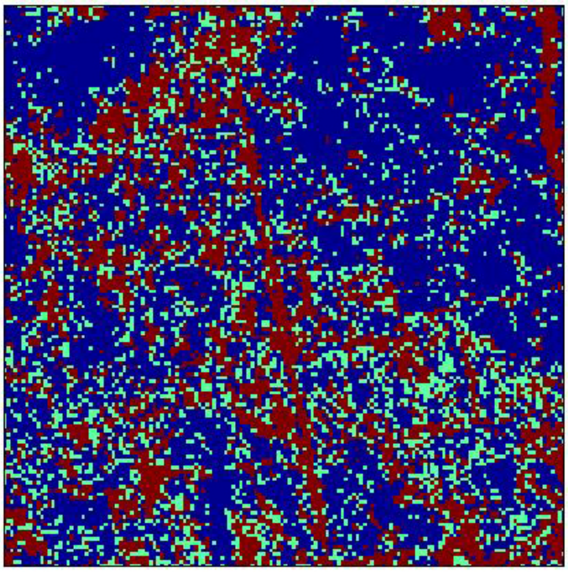


Figure 7. Evaluation of the sampling based on the convex hulls. The map is shown at the bottom: blue and light blue correspond to the pixels belonging to the 'strict' and 'large' convex hulls and red to the pixels for which the transfer function is extrapolating.

The flag map shows that the representativeness of the ESUs is good, even if pixels are outside the two convex-hulls. They mainly correspond to lowest NDVI values: open areas, paths, recent clear cuts…

3. Determination of the transfer function for the two biophysical variables: LAI, fCover

3.1. The transfer function considered

Two types of transfer functions are usually tested in the frame of the VALERI project:

- 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 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).

For all the classes, the 'REG' function is tested using either the reflectance or the logarithm of the reflectance for any band combination as well as the simple ratio or NDVI. As the method has poor extrapolation capacities, a flag image, based on the convex hulls is computing over reflectances.

3.2. Results

3.2.1. Choice of the method

For all the ESUs, a single transfer function is computed. Figure 8 shows the results obtained for all the possible band combinations using either the reflectance (ρ) or the logarithm of the reflectance (log(ρ)): even if the regression made on the log(ρ) provides slightly better results (only LAI variable), the results using the reflectance were selected for LAI and fCover. The transfer function using the log(ρ) indeed creates coplanar points which do not allow the determination of the 'strict' and 'large' convex hulls.

The Red*NIR ('+' or RN) combination is added to all the band combinations (except NDVI and SR). Please read the document (http://www.avignon.inra.fr/valeri/table_methods/new_linear.pdf): "A method to improve the relation between the biophysical variables".

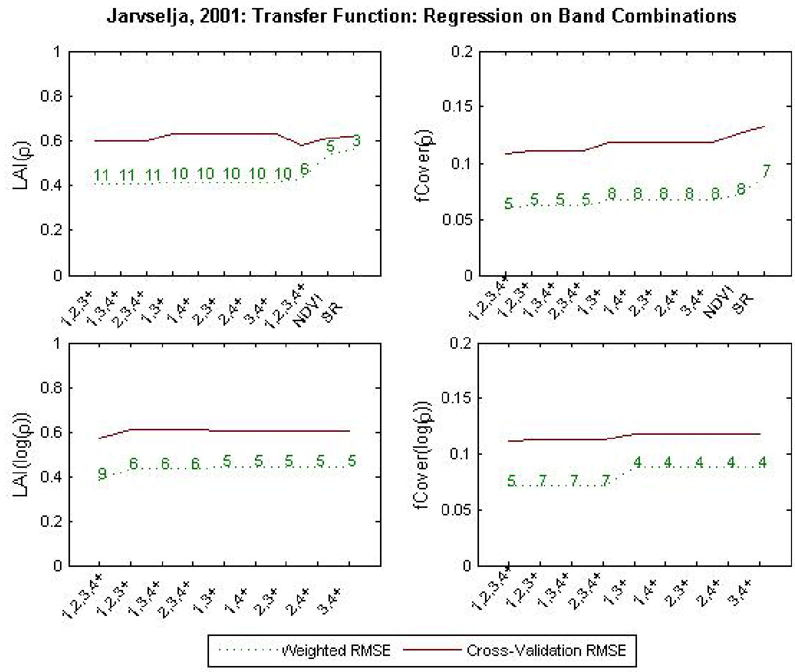


Figure 8. Transfer function: test of multiple regression applied on different band combinations. Band combinations are given in abscissa. The estimated biophysical variable is given in ordinate. Top graphs correspond to regression made on reflectance (ρ): the weighted root mean square error (RMSE) is 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 that could be considered as outliers. Bottom graphs correspond to regression made on the logarithm of the reflectance.

3.2.2. Choice of the band combination

For the effective LAI, the XS1, XS2, XS3, XS4, RN combination on reflectance (Figure 9 and Figure 10) was selected since it provides a good compromise between the cross-validation RMSE (lowest value), the RMSE (lowest value) and the weighted RMSE. Six weights are lower than 0.7. Note that the results are close between REG on ρ and REG on log(ρ).

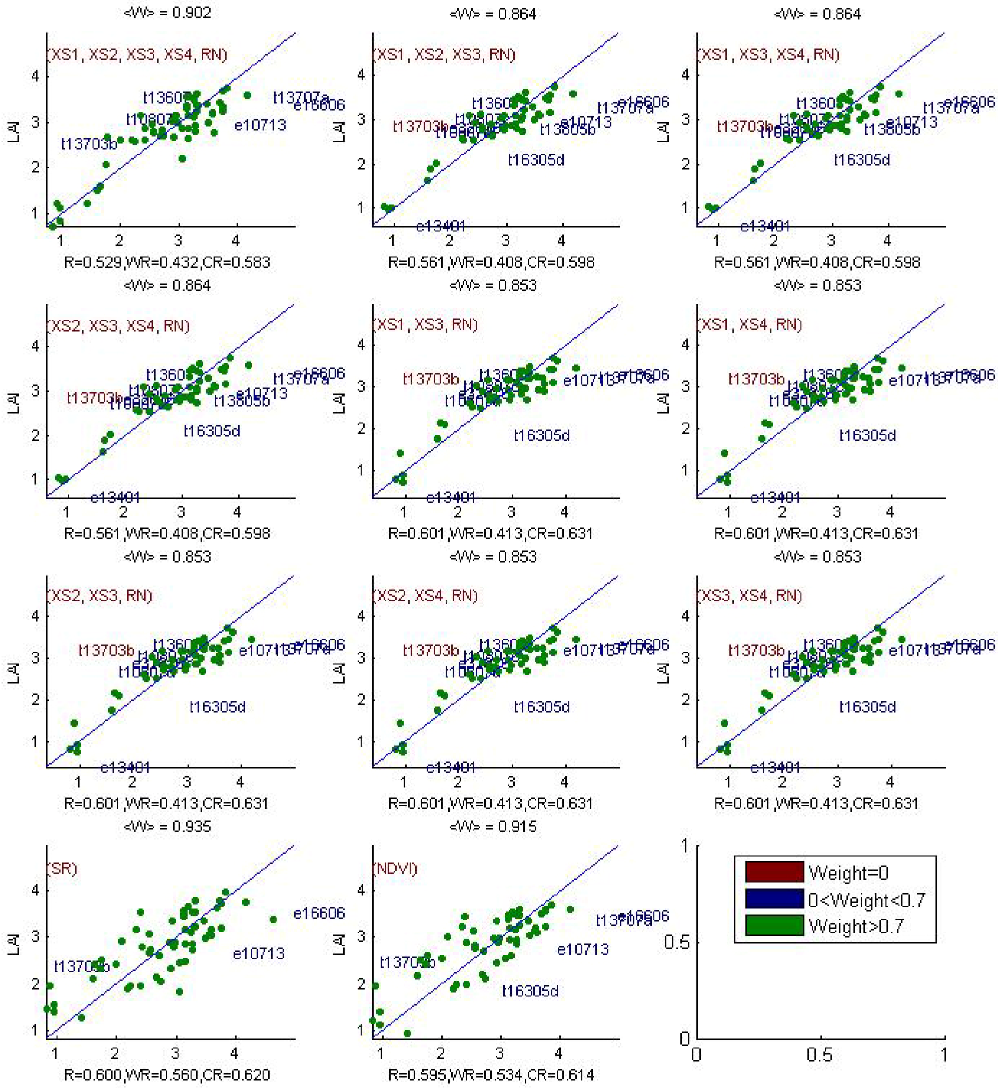


Figure 9. Leaf Area Index: results for regression on reflectance using different band combinations. R is the root mean square error computed between LAI and estimated LAI. WR is the weighted root mean square error and CR is the cross validation root mean square error.

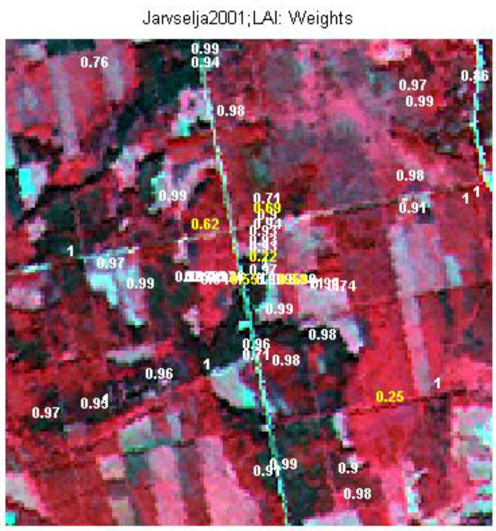


Figure 10. Weights associated to each ESU for the determination of LAI transfer function.

For the fCover, the XS1, XS2, XS3, XS4, RN combination on reflectance (Figure 11 and Figure 12) was selected since it provides the best results. Five weights are lower than 0.7 (one weight = 0).

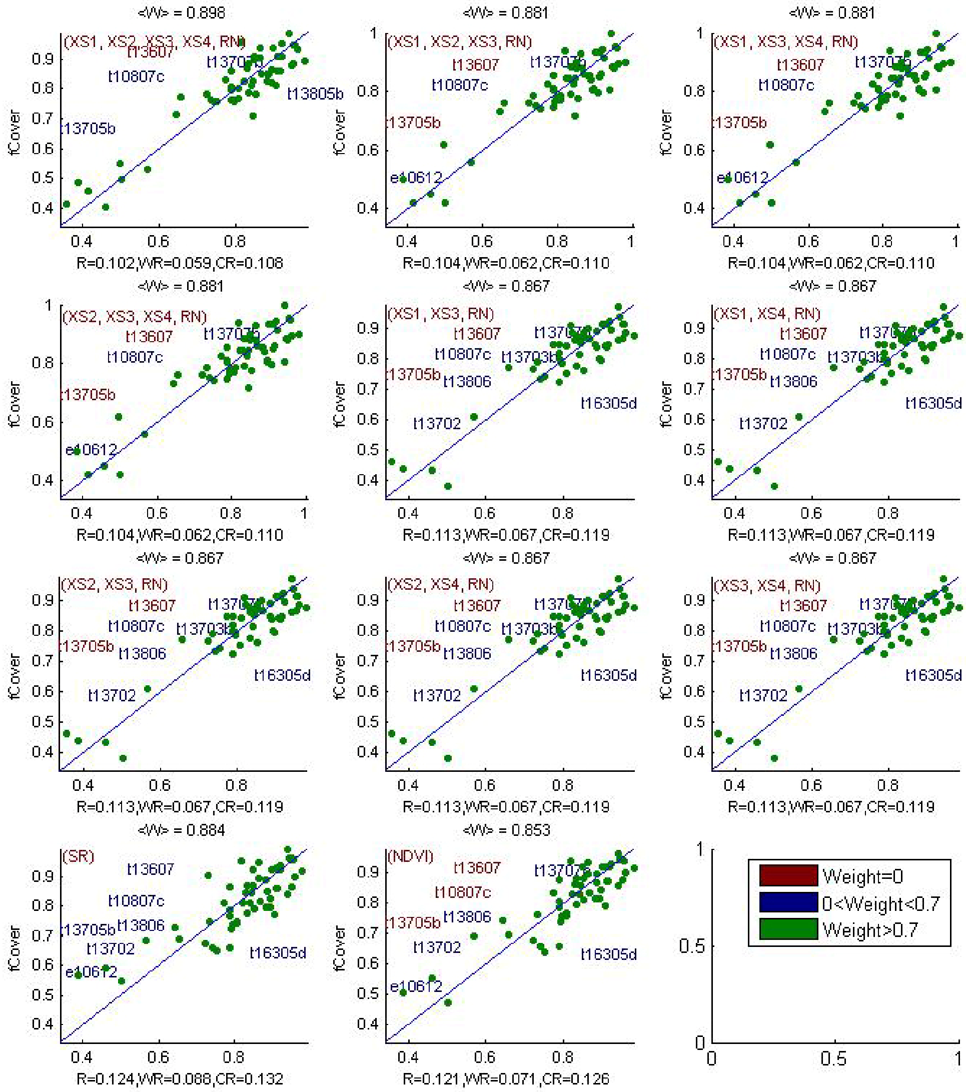


Figure 11. fCover: results for regression on reflectance using different band combinations. R is the root mean square error computed between fCover and estimated fCover. WR is the weighted root mean square error and CR is the cross validation root mean square error.

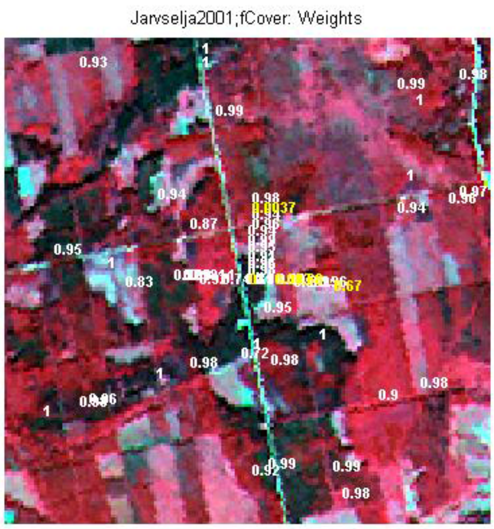


Figure 12. Weights associated to each ESU for the determination of fCover transfer function.

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

Variable	Band Combination	RMSE	Weighted RMSE	Cross-valid RMSE
LAI	2.3788 + 4.8326(XS1) - 19.325(XS2) + 8.1176(XS3) - 7.3509(XS4) - 124.3508(RN)	0.529	0.432	0.583
fCover	0.6006 + 3.4677(XS1) - 0.4441(XS2) + 1.7672(XS3) - 1.3578(XS4) - 40.4624(RN)	0.102	0.059	0.108

RN = Red*NIR

Table 2. Transfer function applied to the whole site for LAI and fCover and corresponding errors

3.3. Applying the transfer function to the Järvselja SPOT image extraction

Figure 13 presents the biophysical variable maps obtained with the transfer function described in Table 2 for all the classes. The maps obtained for the two variables are consistent, showing similar patterns: low LAI values where low fCover are observed and conversely…

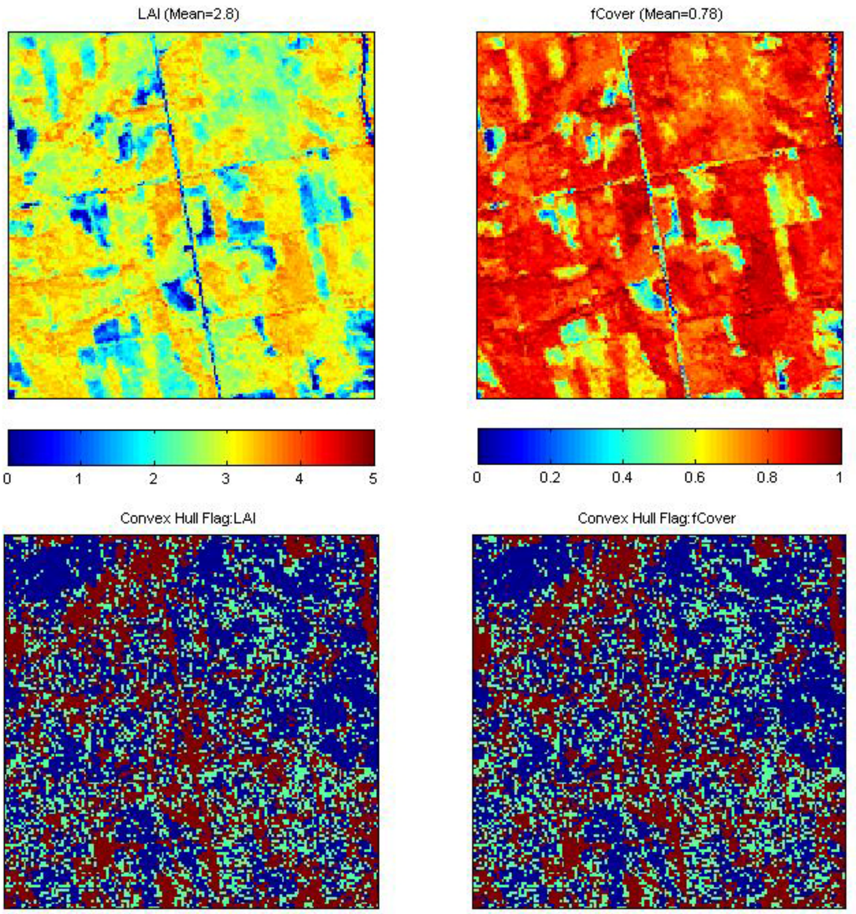


Figure 13. High resolution biophysical variable maps applied on the Järvselja site (top). Associated Flags are shown at the bottom: blue and light blue correspond to the pixels belonging to the 'strict' and 'large' convex hulls and red to the pixels for which the transfer function is extrapolating

The flag maps are comparable between the two biophysical variables. Note that the pixels outside the two convex-hulls mainly correspond to open areas, paths, recent clear cuts (§2.3.4)…

4. Conclusion

The Järvselja site is heterogeneous in terms of LAI. The relationship between LAI and NDVI is consistent. The representativeness of the land cover of the different ESUs is good. The 'REG' method (§3.1) is applied to all the classes. The results of the robust regression are good and the maps obtained for the biophysical variables are consistent. The flag associated to each map shows that the extrapolation of the transfer function is mainly bounded to open areas, clear cuts… For LAI and fCover, the regression coefficients are computed by relating the variable itself to reflectance.

The biophysical variable maps are available in Lambert-Est-92 (datum: WGS-84) projection coordinates at 20m resolution.

5. Acknowledgements

We want to thank: **Andres Kuusk**, **Mait Lang**, **Tõnu Lükk**, **Tiit Nilson** (Tartu Observatory) for the organisation and participation to the campaign.

ANNEX

REPORT ON VALERI ESTONIAN CAMPAIGN

12-16 JUNE, 2001

By Marie WEISS and Tiit NILSON

Participants: Andres Kuusk, Mait Lang, Tõnu Lükk, Tiit Nilson (TARTU Observatory).

A] Location and description of the test area

This Valeri test site is located in the so-called forest of Järvselja, in the eastern part of Estonia. It corresponds to a 10 km x 10 km square, centered at the geographical coordinates 58°15'N - 27°28'E, corresponding more or less to the POLDER pixel in this area. The location of the test site is shown in the following maps (figure 1). In 2001, a 3x3 km sub-area was chosen for the LAI-campaign.

This area is mostly covered by a sub-boreal mixed forest, including both conifers (Scots pine and Norway spruce) and deciduous (birch, aspen, alder). Agricultural fields were almost missing on the 3x3km region, however, a few unmanaged open areas are found. At the south-east and north-east extremities, some bogs and mires (peatland) are taking place. The whole "pixel" is thus very heterogeneous at first sight.

Figure 1: Localisation of the Järvselja Estonian site.

B] Sampling protocol and measurement plots selection.

The area sampling is based on a non-supervised classification out of a Landsat image acquired on 10.07.99. A posteriori interpretation achieved by the Estonian team with data base and field observation led to the extraction of sixteen dominant classes. In each of these classes, several candidates are selected in relation with existent ancillary data in the forestry data base (species fractional distribution, age, understorey description, height and width of trunks, management practices…), and also taking into account the accessibility of the parcels. These candidates are also chosen so that they are spread all over the area to maximize the spatial sampling of the field data collection.

The center of the plot (CP) is selected as close as possible to the center of the parcel, in an homogeneous area. This CP is then located by a GPS measurement. All the GPS positions are given in the "Position.txt" file. Most of them (42 plots) have been post-processed with differential data provided by Tartu fix station, and are marked in the file with a 1 in column "dif". This allow a precision of about one meter for the corrected positions, and about 10 meters for the others. These data are georeferenced in LAMBERT-EST projection (Lambert Conformal Conic 2 parallel) described in table 1.

Datum	ETRS-89(GRS-80)
1st Standard Parallel	58°00'N
2nd Standard Parallel	59°20'N
Central Meridian	24°00'00''E
Coordinates of Origin	57°31'03.19415''N , 24°00'E
False northing	6375 000 m
False easting :	500 000 m

Table 1: Description of Lambert-Est projection characteristics.

C] Protocol of measurement in a given plot, and abbreviations used in this document.

The same LAI measurement protocol is drawn for almost each forest plot. On each of the 1x1km sub-area, at least 3 stands were chosen for measurements. A preliminary choice of the measurement stands was made according to the existing forestry database of the region, to guarantee more or less adequate representatively of different forest types. The region is divided into more or less homogeneous forest parcels – stands. Each stand has a code number according to the forestry database. The code number consists of two parts: large compartment number and smaller parcel number. Such as number 225_05 means that the large compartment number is 225 and stand number 5. With these numbers, the stands can be identified, and the forestry data extracted from the respective forestry database. Also, stand borders as digital vector files can be identified with the similar number (e.g. 22505) and the respective borderlines superimposed on georeferenced raster images.

Due to extremely difficult weather conditions just before the measurement series (non typical heavy rains for that season), some of the preselected stands were inaccessible. For that reason, new more easily accessible stands were chosen instead. Location of the sample plots for the measurements can be seen on the SPOT image of the region. Measurements were made at 5m, 10m and 15m distances from the center of the plot (CP) in each geographic direction, in the fixed sequence N-S-E-W. At each point, 4 records were made. In total, on a sample plot 3x4x4 = 48 LAI readings were taken.

One record consisted of a measurement below the understorey (at the ground level) plus one at the shoulder level (mode G+S). So, on each plot 48 readings were taken at the ground level and 48 at the shoulder level. The sensor was partly screened with the 180° mask, and always oriented with the sun in the back of the operator.

The reference measurements (above canopy) were acquired in an automatic procedure (each 30s) by a LAImeter displayed in an open area close to the measured plots. The reference sensor in the open was also screened and oriented away from the sun direction. The LAImeters have been intercalibrated and the given files provide operational data, which means intercalibrated and mixed below/above canopy records.

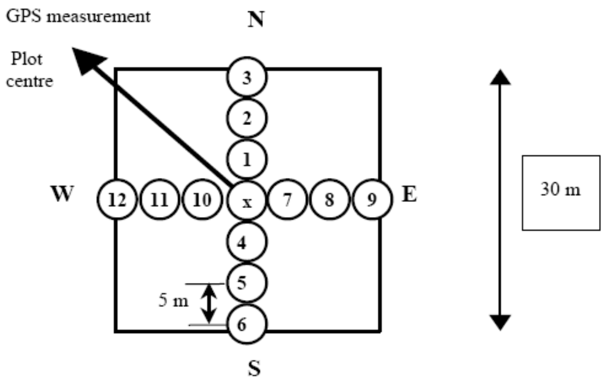


Figure 3: Schematic representation of plot measurement protocol.

Another series of measurements was conducted on so-called 'large cross'. As the centre of the cross, the centre of the 3x3 km region served. First of all, the GPS coordinates of the centre were found. Measurement points were set at each 50m, i.e. at 50m, 100m, 150m, 200m, 250m, 300m, 350m, 400m, 450m and 500m from the centre to N, S, E, and W directions. All points were located by a 50m measuring tape and compass and marked. Afterwards, the GPS coordinates of all points were measured. Four LAI-2000 measurements were made at each point at the ground level and four at shoulder level.

D] Atmospheric correction of SPOT images

The atmospheric correction of SPOT images was made by using the 6S package (Tanre et al., 199). First, the DN-s of the images were transformed into satellite-level radiance units by means of SPOT sensors calibration data containing in the auxiliary files of the images. Since no sunphotometer data were available in Estonia, we had to use indirect methods to determine the needed atmospheric data. Some of the input data we estimated from measurements, like the ozon amount from the TOMS data. Water vapour was estimated via a linear regression of total water vapour content on ground-measured absolute humidity (partial water vapour pressure, hPa) derived for Estonia (Okulov et al. 2001). Aerosols are the weakest part of the procedure. We assumed that certain dark objects (a selection of spruce forests) had a given ground-level reflectance factor in the red band (xs2) The reflectance factor of spruce forests was calculated by our forest reflectance model (Kuusk and Nilson, 2000) depending on the sun angle and to some extent on phenology at the moment of acquiring the image. We assumed that the lacking radiance signal in the red band was caused by aerosol, and thus its amount was estimated. The adjacency effect was neglected, so a linear relation between digital numbers of the image (DN) and ground-level reflectance factors (RF) was obtained:

$RF= b\ DN + a.$

The procedure resulted in the following relations:

SPOT2 image of 25 June 2001:

Band	Slope b	Intercept a
XS1	0.002471	-0.0776
XS2	0.002620	-0.0398
XS3	0.003031	-0.0285

SPOT4 image of 4 July 2001

Band	Slope b	Intercept a
XS1	0.000805	-0.0515
XS2	0.000694	-0.0244
XS3	0.002254	-0.0109
XS4	0.001615	-0.0008

Raw Data: file names and format (directory RawData0)

File: Summary_data.xls

Gap fraction data are given as averaged over all measurements on the same plot on zenith angle rings 7, 23, 38, 53 and 68º, respectively. Gap fractions are given for the breastheight level (1.3m) LAI-2000 measurements (labelled as B, columns from Z to AD) and ground-level (G, columns from AH to AL). The gap fraction for the understorey vegetation (columns from AO to AS), only, are calculated as the ratio of mean values of gap fraction at the ground level and breast-height level (G/B). In column AE, the inverted values of the canopy LAI by using Nilson's method are given. In column AF, the canopy LAI values as determined from the data containing in the forestry database are given. In column AT the inverted LAI value of the ground vegetation obtained from the gap fraction data in columns AO-AS by the standard method of LAI-2000 are given. LAI_total (column AV) is the sum of LAI_canopy and LAI_ground-veget both estimated by the inversion method.

LAI-2000 data processing was made with the same methods as described in 2000.

For Nilson's algorithm, see also the report of 2000. The only difference in the algorithm from that of in 2000 was that crown and canopy closures were not measured in 2001. Canopy closure estimates were derived from the average for the whole plot LAI-2000 estimated gap fraction in the highest ring (7º).

Comments on the inverted data in the file summary_data.xls

Forestry database for the region was created in 1994/1995. By now the information containing in the database is, to some extent, out of date mainly beacuse of management treatments made in the forests – clear cutting and thinning. Also in some very young stands, during the time interval 5-6 years substantial changes may occur. Although we avoided ignoring very drastic changes, like clear cutting, nevertheless in many cases the forestry data used in calculating the initial guess of the canopy LAI did not correspond to the present situation. This is the main reason why in several cases there are considerable differences between the LAI estimates obtained by the two methods.