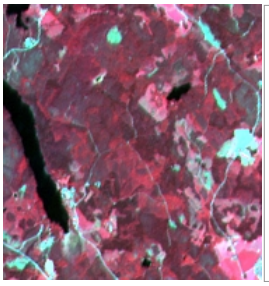
GROUND DATA PROCESSING & PRODUCTION OF THE LEVEL 1 HIGH RESOLUTION MAPS



VALERI 2008

HYYTIÄLÄ site
(evergreen forest)

Philippe Rossello

February 2009

CONTENTS

1. Introduction

This report describes the production of the high resolution, level 1, biophysical variable maps for the Hyytiälä site in July 2008 (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 (hemispherical images).

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 majority of the area is "coniferous Norway spruce or Scots pine forest, which is often mixed with deciduous species, mainly birches. Some deciduous sites, clearcut areas, agricultural fields, small water bodies and peatlands were also included". The site is relatively flat (for more information, see annex or campaign report: http://www.avignon.inra.fr/valeri).

The site coordinates are described in Table 1:

	GCS_KKJ (units = meters)		Geographic Lat/Lon WGS-84	
	Easting	Northing	Lat	Lon
Upper left corner	2514367.4037	6862545.0585	61.86916667	24.27301389
Lower right corner	2518067.4037	6858585.0585	61.83347788	24.34292290
Center	2516197.4037	6860565.0585	61.85132772	24.30760845

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

The ground measurements were carried out from 18th June to 24th July 2008. The high spatial resolution image was acquired in July.

2. Available data

2.1. SPOT image

The SPOT image was acquired the 30th July 2008 by HRVIR1 on SPOT4. It was radiometrically and geometrically corrected by SPOT image (product: SPOTView Precision 2B). The projection is GCS_KKJ. No atmospheric correction was applied to the image. 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 4 x 4 km site. Therefore, it will be taken into account everywhere in the same way.

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

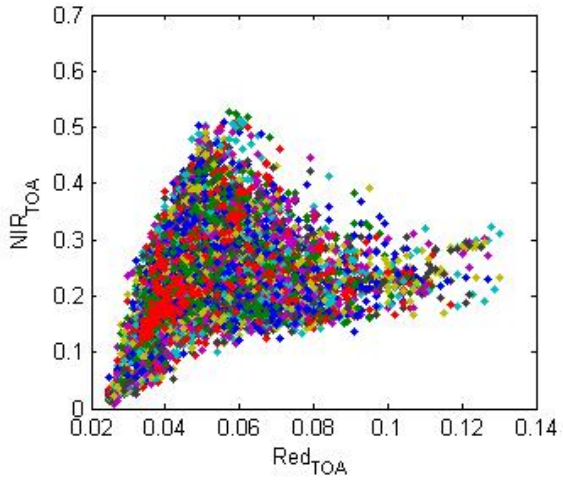


Figure 1. Red/NIR relationship on the SPOT image for Hyytiälä, 2008

2.2. Hemispherical images

For each Elementary Sampling Unit (ESU), the biophysical variables (LAI, fCover) were derived from hemispherical images. They were processed with three custom Matlab scripts. First used automatic global thresholding by Nobis & Hunziker (2005) and was used for images taken in good conditions. Second script implemented local thresholding with Ridler & Calvard's (1978) algorithm and was used where sky was not ideally overcast. Third one calculated LAI from the resulting binary images. The scripts were adjusted to match results given by LAI-2000 with separate data set. 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(above) + LAI(below).
- 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 0.15 to 3.61 and fCover from 0 to 0.74. This range shows a heterogeneous site in terms of LAI.

To build the relationships between biophysical variables and SPOT data, the reflectance of a given forest ESU (trees > 18 m) was 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 20 m and consequently the fish-eye observes an area of $\pi \times [20 \times \tan(60°)]^2 = 3770$ m², *i.e.* close to the area of 9 SPOT pixels (= 3600m²) when using a maximum view zenith angle of 60°.

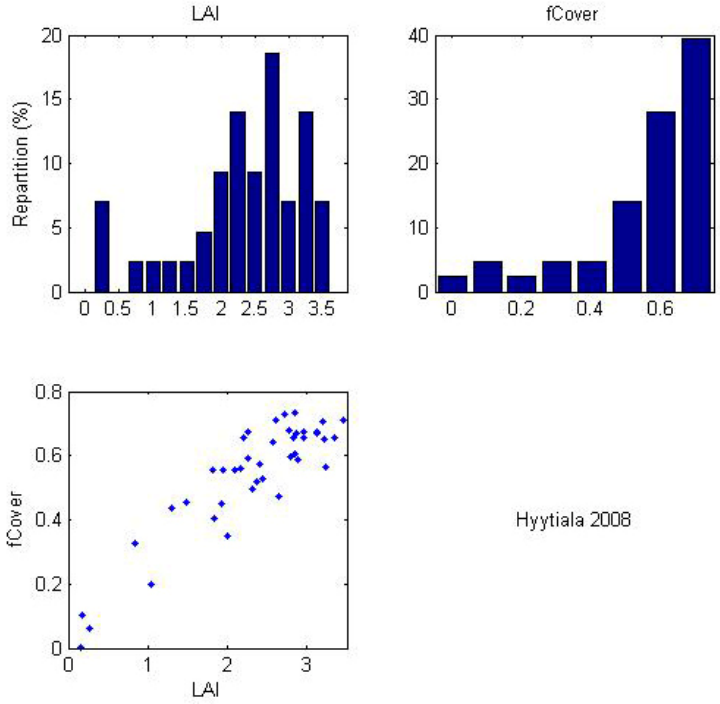


Figure 2. Distribution of the measured biophysical variables over the ESUs.

2.3. Sampling strategy

2.3.1. Principles

The sampling strategy is defined in the campaign report: http://www.avignon.inra.fr/valeri or in annex. It was attempting to represent as much as possible the range of variation of canopy types and conditions.

Figure 3 shows that the 43 ESUs are evenly distributed over the site (4 x 4 km). The processing of the ground data has shown that: considering that SPOT geo-location and GPS measurements are associated to errors, we found that F1 was too close to the road: it has been shifted by 1 pixel.

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

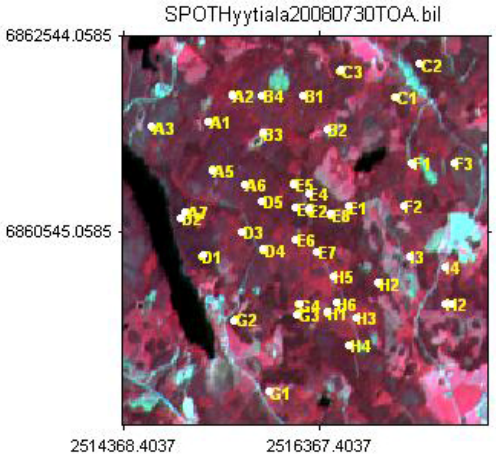


Figure 3. Distribution of the ESUs around the Hyytiälä site.

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 = 40000 in case of a 4 x 4 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.

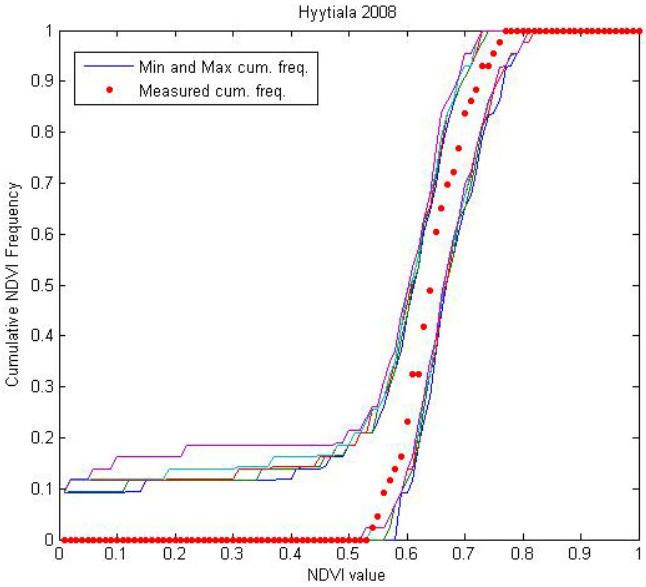


Figure 4. Comparison of the ESU NDVI distribution and the NDVI distribution over the whole image.

Figure 4 shows that the NDVI distribution of the 43 ESUs is good over the whole site even if NDVIs lower than 0.54 have not been sampled although they are present in the image. They may correspond to bare soil (roads, paths…), water, open areas, agricultural fields…

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 4 classes was chosen (Figure 5). The distribution of the classes on the image and on the ESUs is rather different. The classes 1 and 3 are under-represented, while the class 4 appears to be over-sampled. The class 2 (green) corresponds to water bodies.

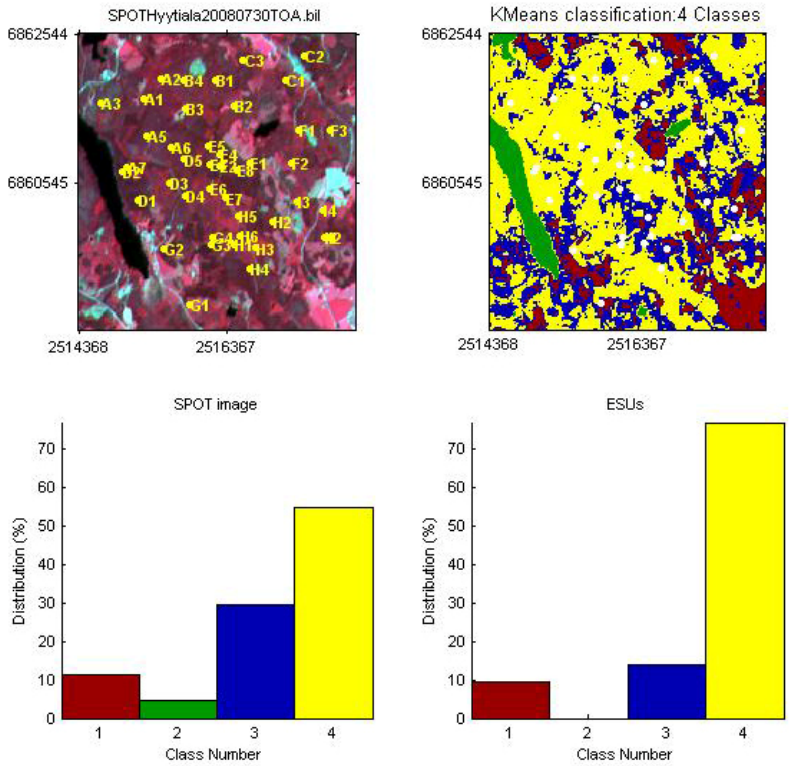


Figure 5. Classification of the SPOT image and comparison of the class distribution between the satellite image and sampled ESUs.

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.

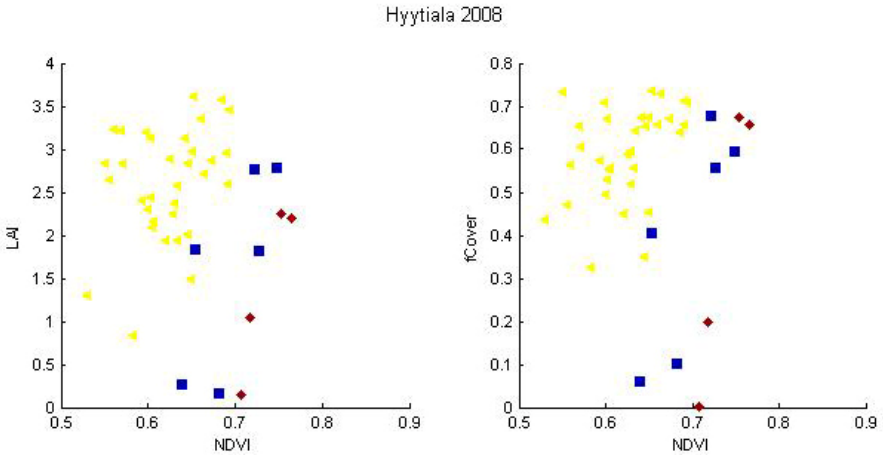


Figure 6. NDVI-biophysical variable relationships as a function of SPOT classes

The relationship between NDVI and biophysical variables is not good. Class 4 which corresponds to forest is distinguishable from others classes. The average value of the biophysical variable measured will attribute to it (§3.1). Therefore, two different 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 and NIR, 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.

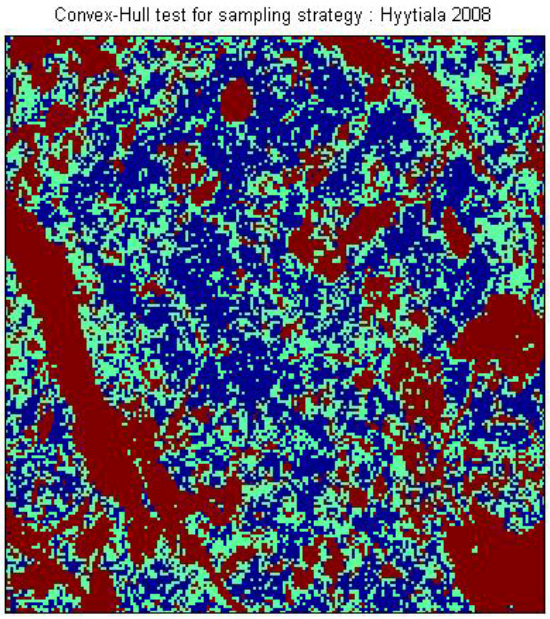


Figure 7. Evaluation of the sampling based on the convex hulls. The map is shown at the bottom: blue and light blue correspond to the pixels belonging to the 'strict' and 'large' convex hulls and red to the pixels for which the transfer function is extrapolating.

The flag map shows that the representativeness of the ESUs is rather good, even if the forest is over-sampled. Pixels outside the two convex-hulls mainly correspond to water bories, recent clear cuts, open areas, bare soil…

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).

3.2. Results

3.2.1. Choice of the method

For classes 1 and 3, '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. For class 4, 'AVE' function is applied even if the number of ESUs belonging to the class is sufficient. The value 0 is attributed to class 2 (water).

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)): the regression made on the reflectance provides better results. The results using the reflectance are thus selected for LAI and fCover.

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".

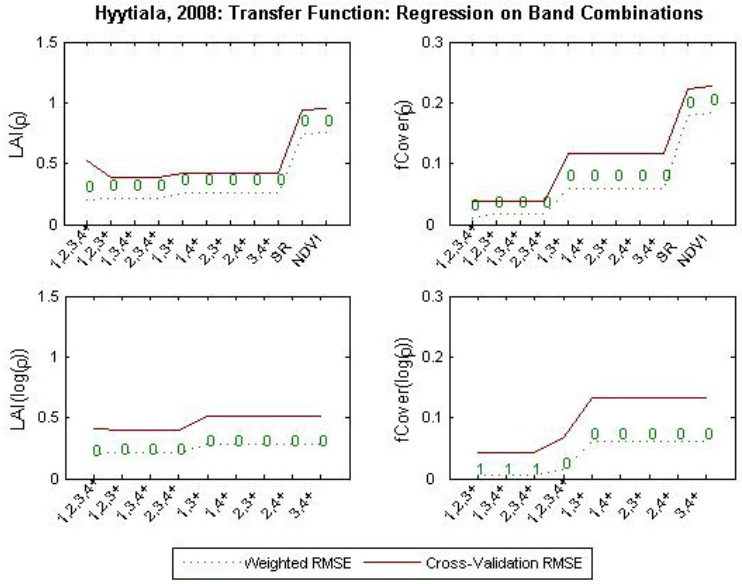


Figure 8. Transfer function: test of multiple regression applied on different band combinations. Band combinations are given in abscissa. The estimated biophysical variable is given in ordinate. Top graphs correspond to regression made on reflectance (ρ): the weighted root mean square error (RMSE) is 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 LAI, the 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 weighted RMSE and the RMSE. No weight is lower than 0.7.

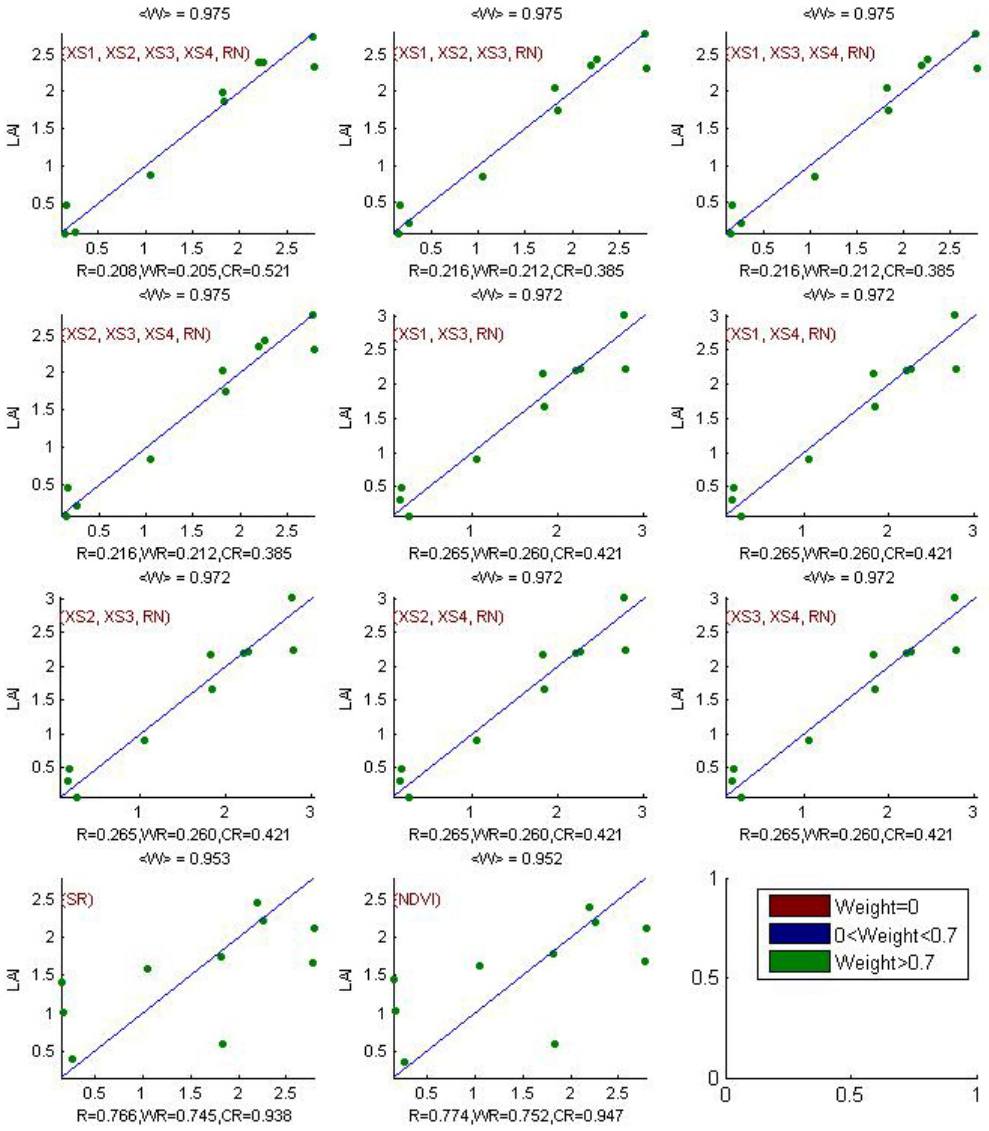


Figure 9. Leaf Area Index: results for regression on reflectance using different band combinations. R is the root mean square error computed between LAI and estimated LAI. WR is the weighted root mean square error and CR is the cross validation root mean square error.

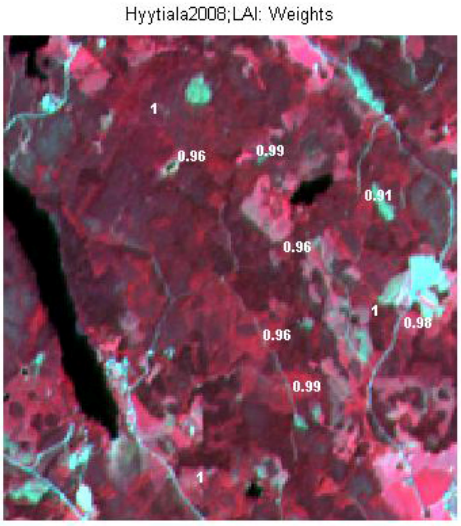


Figure 10. Weights associated to each ESU for the determination of LAI transfer function.

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. No weight is lower than 0.7.

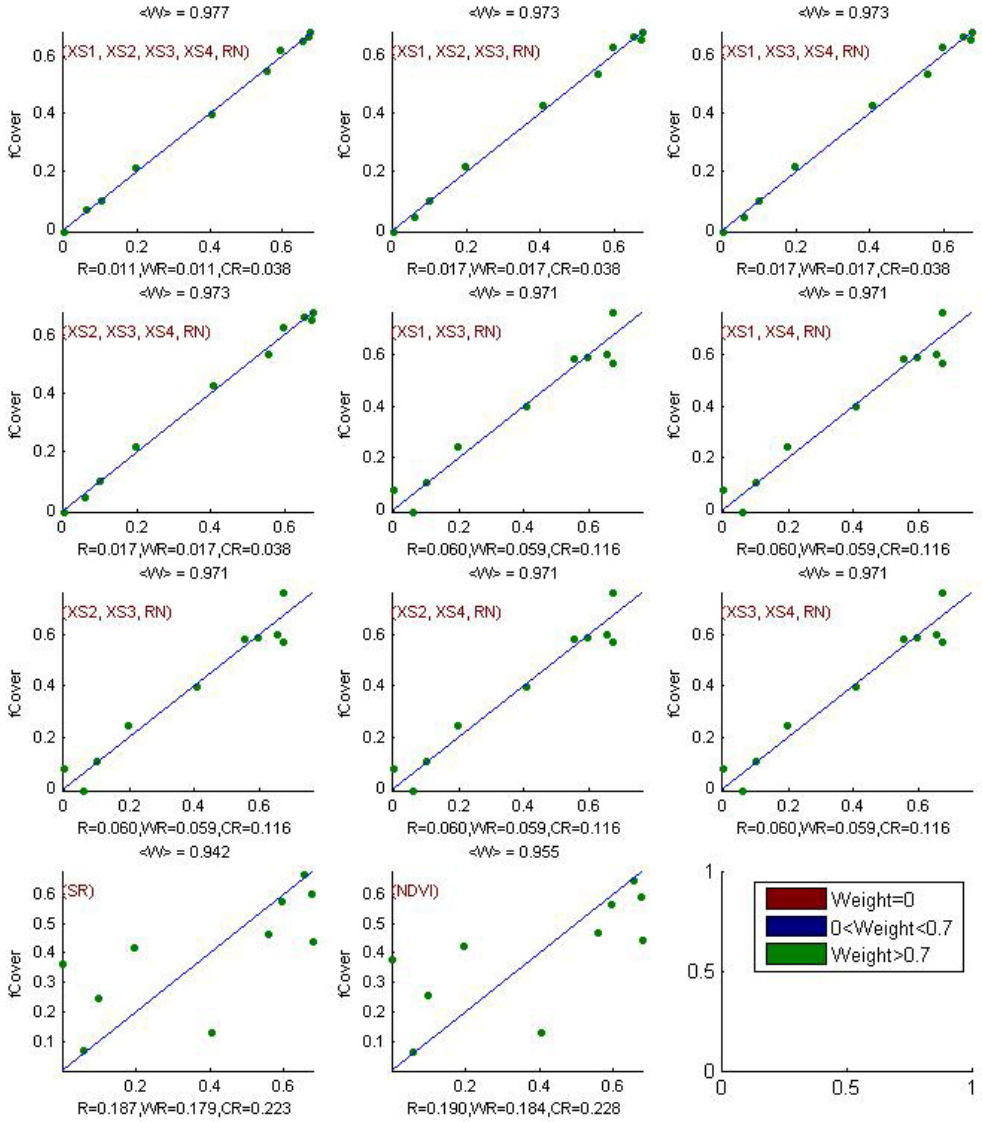


Figure 11. fCover: results for regression on reflectance using different band combinations. R is the root mean square error computed between fCover and estimated fCover. WR is the weighted root mean square error and CR is the cross validation root mean square error.

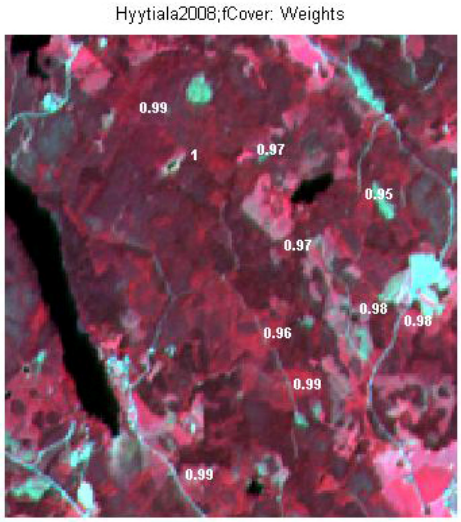


Figure 12. Weights associated to each ESU for the determination of fCover transfer function.

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

Variable	Band Combination	RMSE	Weighted RMSE	Cross-valid RMSE
LAI	-7.9776 + 131.9032(XS2) + 53.7219(XS3) + 46.5566(XS4) - 1179.5106(RN)	0.216	0.212	0.385
fCover	-2.7979 + 29.0349(XS1) + 35.4386(XS2) + 15.8149(XS3) + 1.4761(XS4) - 396.3734(RN)	0.011	0.011	0.038

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 Hyytiälä 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…

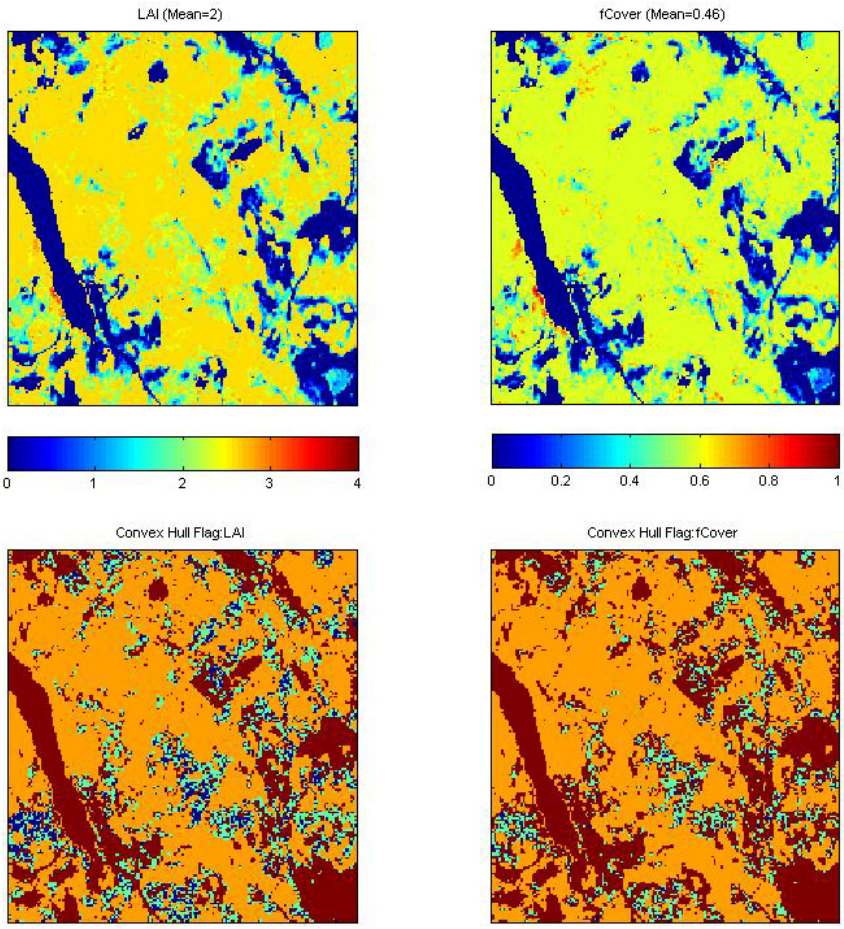


Figure 13. High resolution biophysical variable maps applied on the Hyytiälä site (top). Associated Flags are shown at the bottom: blue and light blue correspond to the pixels belonging to the 'strict' and 'large' convex hulls, red to the pixels for which the transfer function is extrapolating and orange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

The flag maps are comparable between the two biophysical variables. The pixels outside the two convex-hulls mainly correspond to water bories, clear cuts, open areas, bare soil (§2.3.4)… Note that the 'AVE' transfer function is applied to most of the area (forest). The extrapolation is also relatively large. In theory, the more the number of bands increases, the larger the extrapolation is.

4. Conclusion

The Hyytiälä site is heterogeneous in terms of LAI. The representativeness of the land cover of the different ESUs is very good in forest while clear cuts, open areas, bare soil… are under-sampled. 'REG' method (§3.1) is applied to two classes. The results of the robust regression are good. As class 4 (forest) is distinguishable from others classes (§2.3.3), 'AVE' method is applied. 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 water bories, clear cuts, open areas, bare soil… For LAI and fCover, the regression coefficients are computed by relating the variable itself to reflectance.

The biophysical variable maps are available in GCS-KKJ projection coordinates at 20m resolution.

5. Acknowledgements

We want to thank: **Pauline Stenberg** (University of Helsinki) and **Lauri Korhonen** (University of Joensuu) for the organisation and participation to the campaign.

ANNEX

Ground measurement acquisition report for the VALERI site **Hyytiälä**

Organization: University of Helsinki

sampled from 18.6.2008 to 24.7. 2008

Complilation of report: Lauri Korhonen
Organization: University of Joensuu
email: lauri.korhonen@joensuu.fi

Date of report 13.12.2008

People participating to the field experiment:

Name	Organization
Lauri Korhonen	University of Joensuu
Pauline Stenberg	University of Helsinki

Site coordinates

	Lat-Long WGS84 (Deg min.00)	
	Lat.	Long.
Upper left corner	61 51.655	24 16.603
Lower right corner	61 50.668	24 61.844

Ground control points

# Name	Easting(m)	Northing(m)	Comments on the vegetation status, condition of acquisitions, etc...
GCP1	24.287133	61.839918	Place where road leaves lakeside
GCP2	24.295926	61.860035	NW corner of a small pond
GCP3	24.329111	61.863235	Crossing of two major roads
GCP4	24.277211	61.863969	Roadcrossing leading to a gravel pit

GPS system used: Trimble GeoXH
Typical uncertainty of GPS position: 1 m

Description of the site and land cover

Category according to IGBP classification

Needle-leaved evergreen forest

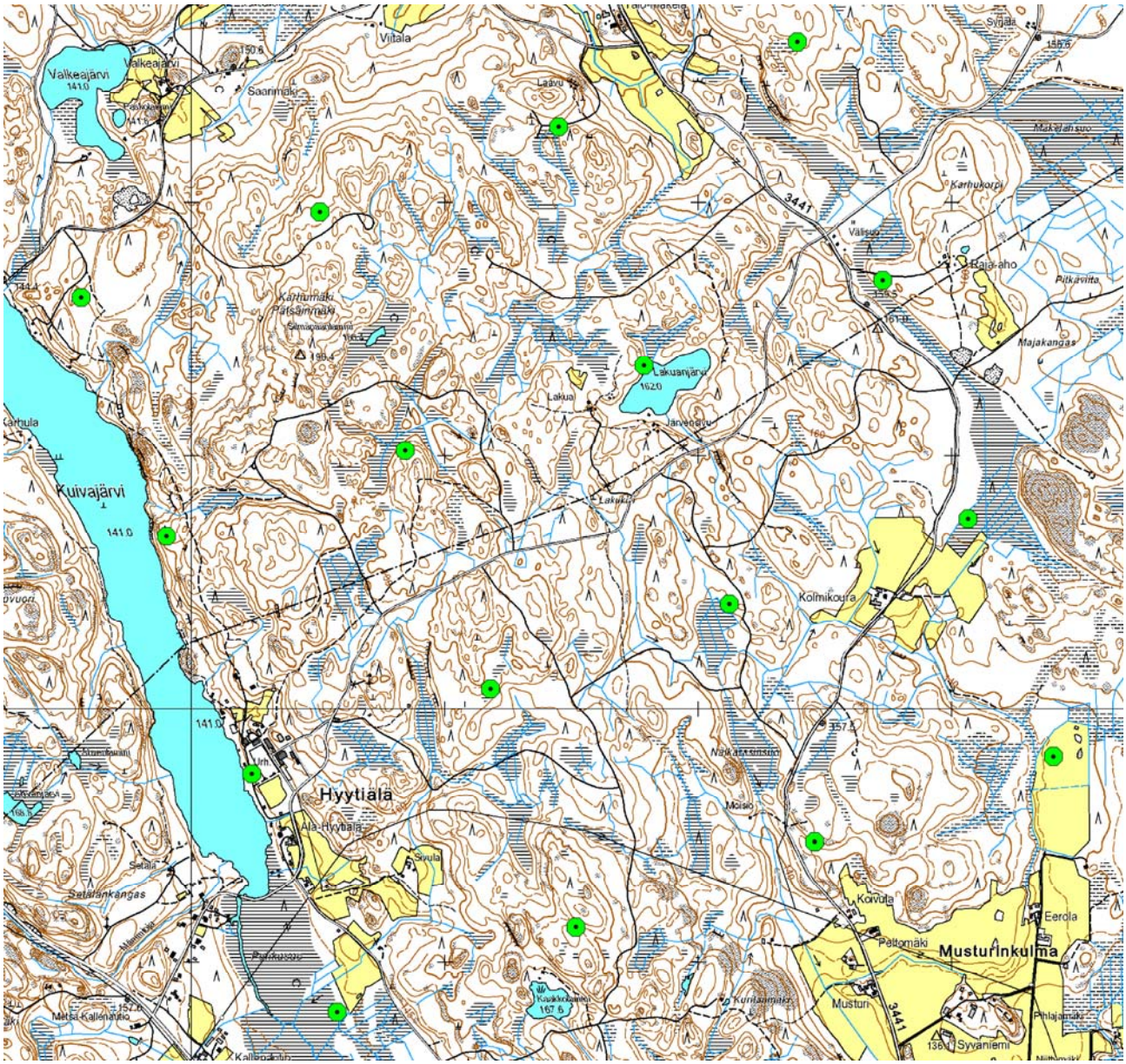
Comments on the land cover

Majority of the area is coniferous Norway spruce or Scots pine forest, which is often mixed with deciduous species, mainly birches. Some deciduous sites, clearcut areas, agricultural fields, small water bodies, and peatlands were also included.

Topography

Relatively flat

Land cover map

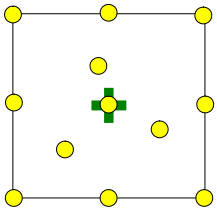


Spatial Sampling scheme

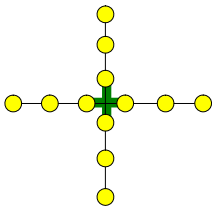
Sensors used for sampling the ESUs

	Method	**Comments**
x	Hemispherical photographs	at breast height (1.3 m)
	LAI-2000	
	TRAC	
	Ceptometer	
	Direct measurements	
	Other	basic stand inventory

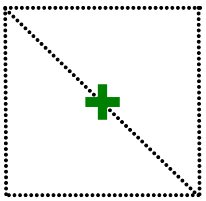
Sampling strategy for the ESU



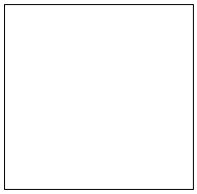
a
☐



b
x

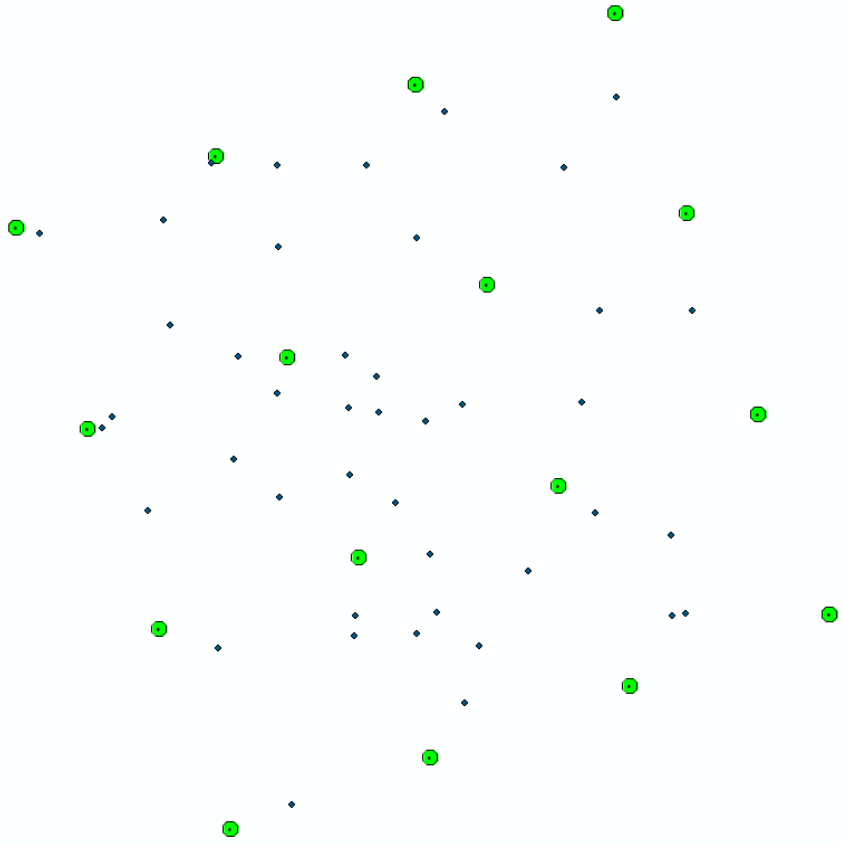


c
☐



d
☐ (specify)

Distribution of the Elementary sampling units



The high spatial resolution image

Satellite

Satellite used: SPOT 4, HRVIR1
Level of processing: 2B
Projection type: GCS-KKJ
Acquisition date: 30th July 2008.

List of the ESUs

Plot	Easting(m)	Northing(m)	Vegetation
A1	24.289830	61.861242	Pine
A2	24.294186	61.863623	Spruce, 10% birch
A3	24.278776	61.860681	Pine
A5	24.290380	61.856802	Pine
A6	24.296514	61.855445	Pine, 15% birch & spruce
A7	24.285181	61.852935	Old growth, pine 60%, spruce 30%, birch 10%
B1	24.308020	61.863496	Pine, 5% spruce
B2	24.312526	61.860407	Spruce
B3	24.300131	61.860058	Open bog, some pines and birches
B4	24.300083	61.863504	Pine
C1	24.325684	61.863387	Pine, 30% spruce
C2	24.330424	61.866324	Spruce
C3	24.315040	61.865757	Pine, 10% spruce
D1	24.288280	61.848973	Old growth, 60% spruce, 40% pine
D2	24.284228	61.852465	Old growth, 75% spruce, 25% pine
D3	24.296045	61.851139	Pine, understorey spruce & birch
D4	24.300135	61.849510	Spruce, 25% birch
D5	24.299978	61.853918	80% Spruce, some pines & birches
E1	24.316506	61.853399	Birch, spruce understorey
E2	24.308979	61.853076	90% spruce
E3	24.306343	61.853275	Old growth, 65% pine, 30% spruce, 5% birch)
E4	24.308819	61.854575	75% spruce, 20% pine, 5% birch)
E5	24.306086	61.855510	50% spruce, 50% birches
E6	24.306409	61.850457	50% spruce, 40% birch, 10% pine)
E7	24.310446	61.849263	Spruce
E8	24.313186	61.852681	40% pine, 50% spruce, 10% birches
F1	24.328871	61.857325	Birch, spruce understorey
F2	24.327192	61.853478	Pine
F3	24.337081	61.857300	Pine
G1	24.301048	61.836568	Spruce seedling stand, overtaken by bushes
G2	24.294608	61.843160	60% spruce, 35% pine, 5% birch
G3	24.306729	61.843641	Spruce
G4	24.306842	61.844490	Very dense spruce, 15% birch, 10% pine
H1	24.312307	61.843737	Pine, 10% birch
H2	24.322349	61.846366	60% pine, 40% spruce
H3	24.317908	61.843234	Birch
H4	24.316578	61.840799	Spruce, 10% birch
H5	24.313505	61.847064	Birch
H6	24.314087	61.844627	65% spruce, 35% pine
I1	24.335175	61.844464	Spruce, 10% birch
I2	24.336395	61.844525	Birch 65%, spruce 35%
I3	24.328360	61.848795	Thinned spruce sapling stand
I4	24.335111	61.847820	Bushy pine/spruce seedling stand

Acknowledgements

Support from Finnish Graduate School in Forest Sciences (GSForest) is gratefully acknowledged.