

From HyMap Imagery to Spatially Distributed Vegetation Water Contents – A Comparison of Different Estimation Approaches Based on Canopy Reflectance Modelling

MICHAEL VOHLAND¹ & SEBASTIAN MADER²

Summary: In this study, a small portion of a HyMap image was used to compare the performance of different strategies to invert a canopy reflectance model (PROSPECT + SAIL) for the retrieval of vegetation water contents. For the classical strategy of iterative optimisation, the Nelder Mead Simplex method was applied. With reference to some ground validation data, this approach provided very reliable estimates, but it also suffered from its low computation rate. As an alternative, model inversion was accomplished by means of a feed forward Artificial Neural Network (ANN). The network was trained by a backpropagation algorithm using thousands of spectra simulated by PROSPECT + SAIL in the direct mode. Simulated spectra were also used to calibrate a Partial Least Squares regression model. When applied to the HyMap image, the ANN approach provided results with a clear shift towards higher estimates, but also highly correlated with the results obtained by numerical minimisation. Furthermore, the spatial distribution of the retrieved data fields showed a clear matching in terms of identical spatial patterns that were characterised by fractal dimensions. Clearly different from this, the PLS approach did not provide consistent results especially for high water content values. The shortcomings can be traced back to the purely linear modelling approach, which provides a less efficient generalisation capacity compared to a properly trained neural network.

1 Introduction

Hyperspectral reflective data enable a detailed quantitative assessment of vegetation parameters such as leaf area index, canopy chlorophyll or canopy water content. In case of image data, this analysis benefits from the spatial coverage of the data and can provide spatially distributed vegetation data fields. This information is potentially appropriate to calibrate or validate spatially distributed process models that make use of or provide plant canopy variables by means of crop growth modelling or hydrologic and metabolic balancing.

In this study, hyperspectral image data of the HyMap sensor (Integrated Spectronics, Baulkham Hills, Australia) with a ground resolution of 5 m were used to derive the canopy water content of summer barley by means of a canopy reflectance model (PROSPECT + SAIL). For this purpose, different techniques of model inversion were applied (numerical minimisation, Artificial Neural Networks, Partial Least Squares Regression; see section 2), and the retrieved data were compared quantitatively by statistical measures. For a selective validation, ground data of a total of 12 sub-plots were collected during the overflight. Furthermore, the spatial distributions of the

¹ Remote Sensing and Geoinformatics, Remote Sensing Department, Faculty of Geography and Geosciences, University of Trier, 54286 Trier, Germany

² Remote Sensing Department, Faculty of Geography and Geosciences, University of Trier, 54286 Trier, Germany

stand variables resulting from the different approaches were studied in detail for their matching, as this allows an assessment of the applied approaches' consistency.

2 Retrieval of stand variables by canopy reflectance model inversion

Different techniques can be applied to extract canopy characteristics from measured remote sensing data by inverting a canopy reflectance model. In the classical approach of numerical minimisation, measured reflectances are iteratively compared to model simulations until the best match (global minimum) is found. Care must be taken (e.g. by using a range of initial solutions) to get a reasonable approximation and not to get trapped by local minima. This may be – depending on the performance of the respective numerical algorithm – very time consuming and therefore computationally not feasible for the inversion of e.g. large image portions. As an alternative, lookup tables (LUT) can be used to increase the rate of model inversion. In this approach, a data base is generated containing sets of input variables and corresponding reflectance values simulated in the direct modelling mode. In the inversion, reflectance measurements are compared to the stored data to identify the closest cases and to extract the appropriate canopy variables.

Conceptually different approaches are provided by estimation models which first are calibrated (to fit the response surface between reflectance values and canopy variables) and afterwards applied to retrieve estimates for the complete set of reflectance measurements of interest. For this calibration, the use of synthetic databases simulated by physically based reflectance models is recommendable as they can cover a wide range of parameter combinations (BARET et al., 2000). For hyperspectral data, Partial least Squares (PLS) regression may be used. As the PLS factoring is based on the variance of both the spectral information and the target variable, PLS regressions usually exploit the information that is inherent in the data efficiently for a powerful prediction. Even more sophisticated is the capacity of artificial neural networks (ANN) to provide estimates of the complex response surfaces when being trained on hyperspectral data. Once a network is trained, the network approach is computationally fast and therefore applicable to large spectral datasets.

3 Materials

The study site is located near Newel in the Bitburger Gutland (Rhineland-Palatinate, Germany), where one field (4.7 hectares) cropped with summer barley in 2005 was selected for further investigation. For the Bitburger Gutland, a data set of the HyMap airborne imaging sensor was acquired on the 28th of May (acquisition time: 9:01:20 UTC) during the HyEurope 2005 campaign. The HyMap data provide 126 bands with bandwidths ranging from 12.9 to 21.3 nm that cover a spectral range from 434 to 2486.5 nm (central wavelengths of the first and last band); the ground resolution was approximately 5 m.

In the preprocessing, an across-track illumination correction was performed for a spatial subset excluding forested areas. The FLAASH (Fast Line of-sight Atmospheric Analysis of Spectral Hypercubes) module of ENVI™, based on the MODTRAN4 radiation code (MATTHEW et al., 2000), was used for the atmospheric correction. A parametric geometric correction was per-

formed by integrating a high resolution digital elevation model, GPS ground control points and flight navigation data provided with the HyMap data using the PARGE™ software (SCHLÄPFER & RICHTER, 2002). Some noisy bands (e.g. those near the water vapour absorption features) were eliminated; thus, a total of 114 spectral bands were finally used for the image inversion.

On the 27th and 28th of May, spectroradiometer measurements using an ASD Field Spec II instrument (Analytical Spectral Devices, Boulder, CO) were performed for a total of 12 sub-plots with a size of 50 cm × 50 cm. The spectral readings were taken integratively over the sub-plots' area with nadir view in the principal plane. Afterwards, the exact position of each sub-plot was located using a differential GPS (GPS Pathfinder Pro XRS; Trimble, Sunnyvale, CA). The sub-plots' above ground plant material was harvested, the fractions of leaves and stalks were separated from each other and weighed separately in the laboratory. After dehydrating the material at 105°C for 24 hours, the total water contents of both fractions were calculated by subtracting the dry weights from the previously measured fresh weights.

4 Methodology – PROSPECT + SAIL modelling approach and applied inversion techniques

For this investigation, the PROSPECT model describing the optical properties of plant leaves (BARET & FOURTY, 1997) has been coupled with the SAIL model (VERHOEF, 1984), a 1 D turbid medium radiative transfer model that is suited for homogeneous vegetation canopies. Leaf reflectance and transmittance from 400 to 2500 nm is described by PROSPECT as a function of the leaf mesophyll structure parameter N , the chlorophyll a+b concentration (C_{ab}), the leaf equivalent water thickness (C_w) and the leaf dry matter content (C_m). For a given illumination and viewing geometry, SAIL calculates the canopy bidirectional reflectance using the leaf optical properties, canopy structure (leaf area index LAI, mean leaf inclination angle θ_l , hot spot size parameter s) and the reflectance of the underlying soil (ρ_s). The parameter $skyl$ refers to the atmospheric conditions and describes the fraction of diffuse illumination.

However, several sets of model parameters may produce very similar or even fully identical spectral signatures. This parameter equifinality is one main reason for the ill-conditioned nature of model inversion and may be reduced by reasonable constraints in the inversion procedure (COMBAL et al., 2002). For this purpose, we fixed some parameters (N to 1.3, s to 0, and $skyl$ to 0.1) and coupled the equivalent water thickness and the leaf dry mass at a ratio of 4:1; this ratio equals a foliage moisture content (FMC) of 400 % and can be assumed to be a typical value for fresh plant material (for a detailed discussion refer to VOHLAND & JARMER, 2007). To define ρ_s , soil samples were taken in the field to identify a typical mean background soil reflectance that was afterwards used in the inversion.

For the retrieval of the five parameters C_{ab} , C_w , C_m , LAI and θ_l , we started with the classical strategy to minimise the merit function numerically (Figure 1) using the Nelder-Mead Simplex method (NELDER & MEAD, 1965). To increase the probability of identifying the merit function's global minimum, we restarted the algorithm several times with a repeated recovery of the initial

values with previous results. In the last minimisation loop, the coupling of C_w and C_m was removed to allow a comprehensible and moderate variation of the foliage moisture content.

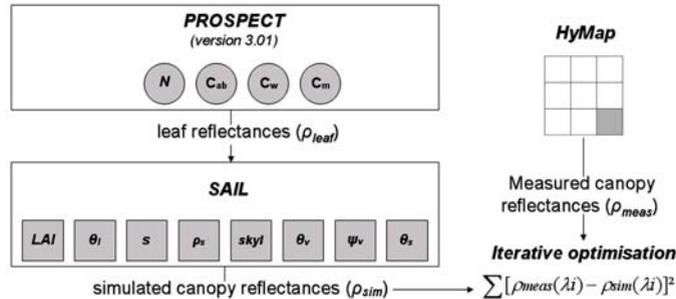


Figure 1: Numerical minimisation to invert PROSPECT + SAIL on the HyMap image.

In total, the numerical optimisation was performed for each pixel of the selected barley plot ($n = 1899$). A validation of these results was accomplished with the plant samples collected in the field. For each sub-plot, the GPS measured positions were used to centre a window of 3×3 pixels in the HyMap data. Within this window, we extracted the best-fitting pixel with the smallest spectral deviations from the ground-measured spectra (see VOHLAND et al. 2006 for a detailed description). For the 12 pixels finally selected, the inverted values of $C_w \times LAI$ were compared to the measured water contents of the leaf fraction.

As an alternative approach, a three-layer feed forward ANN was trained on a number of sample spectra simulated by PROSPECT + SAIL in the direct mode. For the data of this study, the number of input and output neurons was given by the number of spectral bands (114) and the model parameters that were neither fixed nor known (LAI , θ_l , C_{ab} , C_w , C_m). One critical point in this approach is to select an appropriate number of hidden neurons, as it decides about the net's capability to represent the reflectance model's complexity as well as the net's generalization power. We followed the rough guide line of the geometric pyramid rule (MASTERS, 1993), by which the number of hidden neurons is computed as the square root of the product of the number of input and output neurons ($n = 24$ in this case). As activation function, the hyperbolic tangent was used according to the study of KALMAN & KWASNY 1992.

The network was trained with an improved version of the RPROP algorithm (RIEDMILLER & BRAUN, 1993; IGEL & HÜSKEN, 2000). To prepare the training samples, PROSPECT + SAIL were parametrised as described above; again, C_w and C_m were coupled and their ratio was allowed to vary between 3.8:1 and 4.2:1. In total, 152,000 parameter combinations were generated by random to simulate canopy reflectances. From these data, 76,000 samples were selected randomly for the training procedure, and the remaining samples were used as independent data to test the network's performance. Since the errors obtained with the test data did not differ significantly from the errors that remained after the training, the ANN, now trained and validated, was then applied to the image data.

Different from ANN, PLS regression provides a purely linear model for estimating canopy variables. Similar to the principal component analysis (PCA), it produces non-correlated factor scores from the original and highly correlated predictor variables (X, HyMap reflectance values) to estimate Y (total water content of the leaf fraction). But it differs from PCA in that this extraction also reflects the covariance structure between X and Y. Thus, the selected components (latent variables) are supposed to provide an optimum for explaining both X and Y. This step is followed by a regression where the decomposition of X is actually used to predict Y (ABDI, 2003). One critical point might be the lack of generalization power due to an over-fitting in the calibration. This can be avoided by model validation performed internally (cross-validation) or externally with an independent dataset. In our study, we randomly extracted 5,000 samples from the ANN training data (PROSPECT + SAIL simulations) for the PLS model calibration, and another 5,000 samples were used for an external validation. After a successful validation (see section 5), the PLS approach was also applied to the HyMap image.

The data fields retrieved for the summer barley plot applying the ANN and PLS approaches to the HyMap image were then compared to the results of numerical minimization in terms of absolute values and spatial patterns (see Figure 2 with an overall view of the selected procedure).

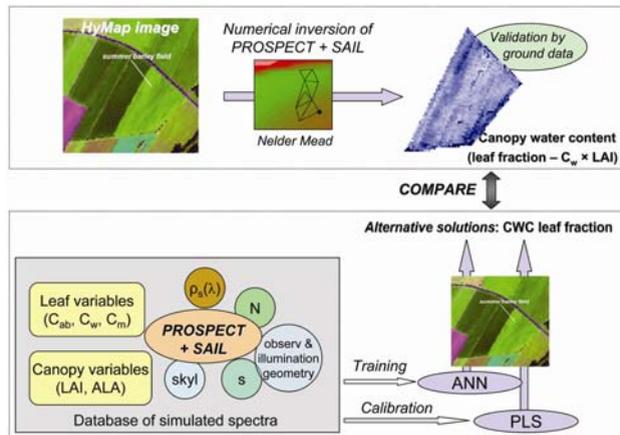


Figure 2: Workflow to compare the different inversion techniques in this study.

5 Results & Discussion

In the numerical minimization that was performed for each pixel, we retrieved the water content in g per cm² leaf area, which can be scaled up to the canopy (g × cm⁻² ground area) by multiplication with the LAI. For a pointwise validation, the ground measured water contents of the leaf fraction were utilised as described above. This validation proved reliable estimates in terms of

the coefficient of determination (r^2), the root mean squared error (RMSE) and the percentage RMSE ($\text{RMSE} \times \text{measured mean}^{-1}$) (Figure 3).

The retrieved values for the selected summer barley plot ($n = 1899$) were afterwards compared to the statistics of the ANN- and PLS-approach. For the neural network, we found results for $C_w \times \text{LAI}$ to be significantly higher than for the Nelder Mead inversion (Table 1). For a detailed comparison, the other canopy variables obtained by the neural network have also to be analysed. Here, estimates for the LAI and the canopy dry mass were also clearly higher, whereas the results for θ_l were very low and showed only little variation.

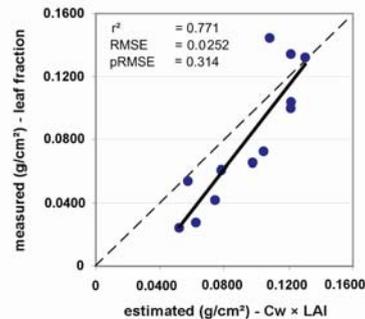


Figure 3: Scatterplot of pixel-based numerical inversion results for $\text{LAI} \times C_w$ against measured field data ($50 \times 50 \text{ cm}^2$ sub-plots, $n = 12$).

Table 1: Statistics of inversion results ($n = 1899$) obtained by numerical optimisation and ANN.

	Numerical Minimisation (Nelder Mead)			Artificial Neural Network		
	mean	stdv ^a	varcoeff ^b	mean	stdv ^a	varcoeff ^b
$\text{LAI} \times C_w \text{ (g cm}^{-2}\text{)}$	0.1154	0.0168	14.55	0.1823	0.0177	9.69
LAI	5.93	0.53	8.67	6.71	0.43	6.40
$\text{LAI} \times C_{ab} \text{ (}\mu\text{g cm}^{-2}\text{)}$	263.6	33.7	12.88	258.9	26.3	10.17
$\text{LAI} \times C_m \text{ (g cm}^{-2}\text{)}$	0.0278	0.0041	15.36	0.0425	0.0043	10.12
$\theta_l \text{ (}^\circ\text{)}$	34.1	3.86	11.32	25.7	0.63	2.45
$C_w \times C_m^{-1}$	4.15	0.09	2.12	4.29	0.09	2.07
mean RMSE (reflectance) ^c	0.0158			0.0264		

^a standard deviation

^b coefficient of variation (%)

^c inverted variables are used for spectra reconstruction in the direct mode of PROSPECT + SAIL; reconstructed spectra are afterwards pixel-wise compared to the HyMap spectra, RMSE: mean for all wavelengths and pixels

Both approaches did not differ very much with respect to the quality of reproducing the HyMap spectra (Table 2); thus, the differences in the obtained values of the canopy variables seem to have evolved from the ill-posed problem of model inversion, and could have been partly triggered by the slightly different coupling of C_w and C_m in both approaches.

In the PLS approach, we first identified a model based on 8 latent variables to predict $LAI \times C_w$. For the 5,000 calibration samples covering a range from 0.01 to 0.24 g H₂O per cm² canopy area, this model provided a value for r^2 of 0.981, and the RMSE (g cm⁻²) amounted to 0.006; these terms kept stable in the validation with another 5,000 test samples ($r^2 = 0.977$; RMSE = 0.007 g cm⁻²). Nevertheless, when applied to the HyMap image, results were not consistent, as $LAI \times C_w$ was estimated with negative values for 211 pixels. Thus, we decided to apply a model with only 3 latent variables that were still sufficient to explain more than 95 % of the $LAI \times C_w$ variation in the calibration. For the summer barley plot, this model provided a mean leaf water content (canopy level) of 0.1367 g cm⁻², the standard deviation was 0.0138 g cm⁻². Although these values seem to fit quite well to the results of Nelder Mead and ANN, the scatterplots reveal significant differences (Figure 4).

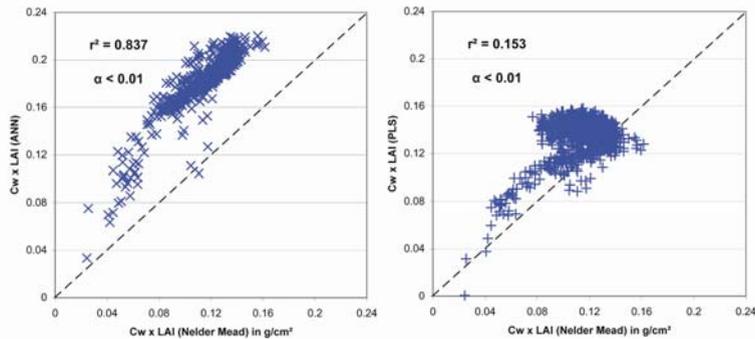


Figure 4: Scatterplots of pixel by pixel-results for $LAI \times C_w$ ($n = 1899$).

For the ANN results, there is a clear bias in the scatterplot, but the correlation with the Nelder Mead estimates is rather high. The retrievals of the PLS regression seem to be unconfident for the range beyond 0.12 g H₂O \times cm⁻², as results do not match neither to the Nelder Mead nor to the ANN-based estimates. However, for the pixels with medium to low water contents (PLS-values < 0.12 g cm⁻², $n = 168$) correlation with Nelder Mead is satisfactory ($r^2 = 0.69$). The limitation of the PLS approach for high values is probably due to its linearity, which restricts its generalization power in case of gradually saturating spectra paralleled by canopy variables still changing distinctly.

Beyond a purely pixel-based quantitative analysis, the inversion results were compared for their spatial distribution and the spatial details recognisable in the analysed image portion. This issue is highly relevant, as one benefit from remote sensing data is the spatial coverage they provide, so they can be used for calibrating or validating spatially distributed modelling approaches in hydrology or landscape ecology, for example. First, the data fields retrieved by the different approaches were interpreted visually, which reveals clear similarities between the Nelder Mead and ANN data, but also an obviously different spatial distribution provided by the PLS approach (Figure 5).

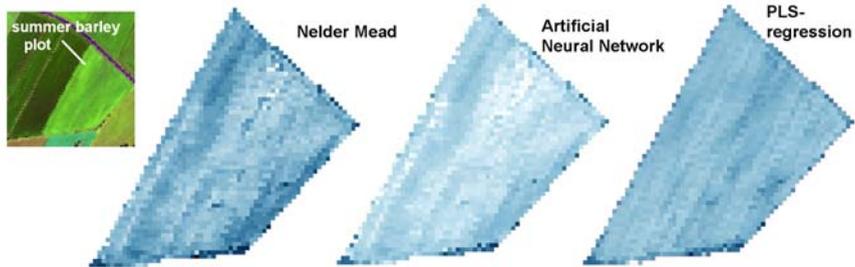


Figure 5: Spatially distributed results for $LAI \times C_w$ as retrieved by numerical minimisation, ANN and PLS regression for the selected summer barley plot (bright pixels indicate high values; images were scaled identically).

A more quantitative and normalized method to analyse the spatial variation is provided by the calculation of fractal dimensions (D), that were derived from the variograms of our data fields. In detail, the incremental slope (s) of the log-log plot of semivariance against sample interval (lag) was used to calculate D as $(3 - s/2)$ (XIA & CLARKE, 1997).

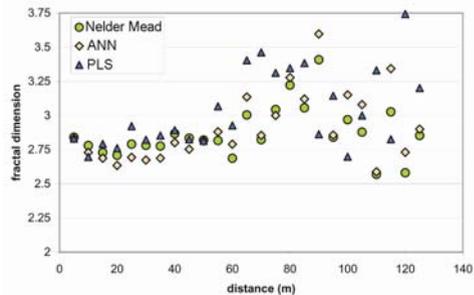


Figure 6: Piece-wise fractal dimensions of the retrieved data fields (all approaches) for $LAI \times C_w$.

For the results of Nelder Mead and ANN, the fractal dimensions show a nearly perfect match in course (Figure 6), which verifies the analogy also found in the visual comparison. For the PLS regression results, differences of D are more pronounced, and the value patterns are less congruent. Beyond a distance of 60 m (12 pixels), all approaches show a clear scattering of D that can be traced back to the sills of the variograms that are reached at this lag.

6 Conclusions

The following conclusions can be drawn from this study:

- The validation by the data of 12 ground sub-plots proved the retrieval of the canopy water content (leaf fraction, $LAI \times C_w$) by the Nelder Mead algorithm to be reliable and relatively

precise. The neural network approach provided results with a clear shift to higher values. Nevertheless, the estimates of both approaches were highly correlated and showed a good match concerning their spatial distributions.

- A fast computation is feasible by using ANN (after being trained) and PLS regression. However, the latter did not provide dependable results when applied to the canopy reflectances contained in the image data. Shortcomings might result from the linearity of this approach, which cannot keep up with the more complex and – in our study – more efficient fitting capacities of properly trained neural networks.
- A larger image portion could be inverted rapidly by the neural network approach. As the ANN results were consistent in terms of the spatial distribution obtained, the resulting data field is most likely qualified to validate the output of e.g. spatially distributed process models that make use of the canopy water content (plant growth or SVAT models, for example).

Acknowledgements

This study was financially supported by the Forschungsfonds of Trier University. We would like to thank the landowner of the investigated plot, Matthias Mohn. Many thanks to Thomas Jarmer and Christoph Knotte for their substantial support. We would like to thank the following people who assisted in the field work: Henning Buddenbaum, Erik Mohr, Franz Ronellenfitsch and Martin Schlerf.

References

- ABDI, H., 2003: Partial least squares regression (PLS regression). In: Encyclopedia for research methods for the social sciences, M. Lewis-Beck, A. Bryman & T. Futing (eds.) (Sage Publications), 772-795.
- BARET, F. & FOURTY, T., 1997: Estimation of leaf water content and specific leaf weight from reflectance and transmittance. *Agronomie*, **17**: 455-464.
- COMBAL, B., BARET, F., WEISS, M., TRUBUIL, A., MACÉ, D., PRAGNÈRE, A., MYNENI, R., KNYAZIKHIN, Y. & WANG, L., 2002: Retrieval of canopy biophysical variables from bidirectional reflectance using prior information to solve the ill-posed inverse problem. *Remote Sens. Environ.*, **84**: 1-15.
- IGEL, C. & HÜSKEN, M., 2000: Improving the Rprop learning algorithm. In: Proc. Second International ICSC Symposium on Neural Computation, 115-121.
- KALMAN, B. L. & KWASNY, S. C., 1992: Why tanh: Choosing a sigmoidal function. In: Proc. International Joint Conference on Neural Networks, vol. IV, 578-581.
- MASTERS, T., 1993: Practical Neural Network Recipes in C++. Academic Press, 495 pp.
- MATTHEW, M. W., ADLER-GOLDEN, S. M., BERK, A., RICHTSMIEIER, S. C., LEVINE, R. Y., BERNSTEIN, L. S., ACHARYA, P. K., ANDERSON, G. P., FELDE, G. W., HOKE, M. P., RATKOWSKI, A., BURKE, H.-H., KAISER, R. D. & MILLER, D. P., 2000: Status of Atmospheric Correction Using a MODTRAN4-based Algorithm. In: Proc. SPIE 4049, 199-207.
- NELDER, J. A & MEAD, R., 1965: A Simplex method for function minimization. *Comput. J.*, **7**: 308-313.

- RIEDMILLER, M. & BRAUN, H., 1993: A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: Proc. of the IEEE Intl. Conf. on Neural Networks, 586–591.
- SCHLÄPFER, D. & RICHTER, R., 2002: Geoatmospheric Processing of Airborne Imaging Spectrometer Data Part1: Parametric Orthorectification. *Int. J. Rem. Sens.*, **23**: 2609-2630.
- VERHOEF, W., 1984: Light scattering by leaf layers with application to canopy reflectance modelling: the SAIL model. *Remote Sens. Environ.*, **17**: 125-141.
- VOHLAND, M. & JARMER, T., 2007: Estimating structural and biochemical parameters for grassland from spectroradiometer data by radiativ transfer modelling (PROSPECT + SAIL). *Int. J. Rem. Sens.*, in press.
- VOHLAND, M., JARMER, T. & MADER, S., 2006: Assessment of the leaf area index (LAI) for summer barley from field spectroradiometer and HyMap image data using the PROSPECT + SAIL models. In: Proc. 2nd Workshop of the EARSeL Special Interest Group on Land Use and Land Cover, 249-257.
- XIA, Z. G. & CLARKE, K. C., 1997: Approaches to Scaling of Geospatial Data. In: *Scale in Remote Sensing and GIS*, D. A. Quattrochi & M. F. Goodchild (eds.) (CRC Press), 309-360.