

MAPPING LAI AND CHLOROPHYLL CONTENT FROM AT-SENSOR APEX DATA USING A BAYESIAN OPTIMISATION OF A COUPLED CANOPY-ATMOSPHERE MODEL

V.C.E Laurent¹, W. Verhoef², M.E. Schaepman¹, A. Damm¹, J.G.P.W. Clevers³

¹ Remote Sensing Laboratories, University of Zurich, Switzerland

² Faculty of Geo-Information Science and Earth Observation, University of Twente, the Netherlands

³ Centre for Geo-Information, Wageningen University, the Netherlands

ABSTRACT

This contribution proposes a methodological approach based on a coupled canopy-atmosphere radiative transfer model and a Bayesian optimization algorithm, which allows the use of a priori data in the retrieval. This approach was used to estimate LAI and leaf chlorophyll content (Cab) in the agricultural test site Oensingen, Switzerland, from at-sensor radiance data of the new airborne APEX imaging spectrometer. The Bayesian optimization allowed having up to 7 free variables in the optimization. The obtained maps of estimated LAI and Cab values at the field level show a good agreement with our expectations in terms of the values themselves, but also their variation range and spread.

Index Terms—canopy-atmosphere coupling, radiative transfer, Bayesian optimization, at-sensor radiance, APEX

1. INTRODUCTION

Leaf area index (LAI) and leaf chlorophyll content (Cab) are physiological variables which are related to the plant growth and are therefore important inputs in agricultural and climate models. These variables can be estimated and monitored by means of remote sensing. Two main categories of methods can be used to obtain LAI and Cab: empirical approaches which rely on statistical relationships between the remote sensing data and the variable of interest, and physically-based approaches which mostly rely on canopy radiative transfer (RT) models. The second category of approaches is more general because it can easily be adapted to different acquisition geometry, remote sensor and vegetation type [1]. However, the RT model has to be inverted to obtain the estimates of LAI and Cab. This is an ill-posed problem which has to be regularized by using a priori information and/or spatio-temporal constraints [2, 3].

Currently, most estimation algorithms use top-of-canopy (TOC) reflectance data to invert the canopy RT model. The pre-processing of the data from at-sensor level to TOC level, which includes atmospheric correction, and, eventually, corrections for topographic, adjacency, and directional effects, is a complex process and is subject to limitations. Indeed the corrections for each individual effect are usually

applied independently and sequentially, which creates error propagation issues, and does not reflect the physical interactions between these effects [4]. Moreover, the main pre-processing step, the atmospheric correction, relies on the inversion of an atmospheric RT model, which creates a number of limitations. In practice, it is necessary to make assumptions about topographic, adjacency, and directional effects, and to interpolate between the entries of a pre-computed look-up table [5].

Coupling the canopy and atmospheric RT models, one obtains a coupled system able to simulate the radiance of the vegetation as seen by the sensor [6]. It is then possible to directly invert this model using the at-sensor radiance data. The same regularization methods can be used as when working with TOC reflectance data, and the same inversion algorithms can be applied. In addition, topographic, adjacency, and directional effects can be included in the forward model, which would be more accurate than the successive corrections applied to obtain the TOC reflectance data [7]. Therefore, working directly with the radiance data allows reducing the pre-processing and its associated uncertainties. Thanks to its minimal pre-processing (calibration, geometric correction, ortho-rectification), using a coupled canopy-atmosphere approach facilitates multi-sensor and data assimilation.

In this contribution, we propose a methodological approach based on a coupled canopy-atmosphere RT model and a Bayesian optimization algorithm, which allows the use of a priori data in the retrieval. The approach is used to invert at-sensor radiance data of the new airborne APEX imaging spectrometer for retrieving the biogeophysical vegetation variables LAI and Cab within the agricultural test site Oensingen, Switzerland.

2. MATERIALS

The Oensingen test site is a flat agricultural area located in the Swiss midlands, at +47° 16' 44" N, +7° 43' 53" E, 523 m above sea level (a.s.l.). The climate is temperate continental, with annual rainfall of 1100 mm and an average annual temperature of 9°C. The study area consists of large homogeneous and well cultivated fields of various crops.

Data of the Airborne Prism Experiment (APEX)

instrument were acquired on June 25th, 2010, with north heading and a flight height of 5000m a.s.l., resulting in a 2 m ground pixel size. APEX is a dispersive pushbroom imaging spectrometer covering the spectral region between 380 nm and 2500 nm in 313 spectral bands with a sampling interval varying between 0.4 and 10nm [8, 9]. The radiometric calibration was applied to the raw data, resulting in an at-sensor radiance image.

Extensive field measurements were carried out in five fields concurrently to the APEX flight: corn (Co), winter wheat (W), pea (P), sugar beet (S), and clover (Cl). In each field, four areas of about 2 m² were chosen and characterized by measuring radiometric signals and biogeochemical canopy variables including LAI, Cab, leaf dry matter (Cdm) and water (Cw) content [10]. Other crop types in the area were bean (B), rapeseed (R), and grass (G).

3. METHODS

The hybrid canopy RT model Soil-Leaf-Canopy (SLC) [11] was coupled to the atmosphere RT model MODTRAN4 [12]. The SLC model couples the 4SOIL soil model with the PROSPECT leaf model and the 4SAIL2 canopy model. The coupling between canopy and atmosphere has been implemented according to the 4-stream theory, as described in [7], and allows making full use of the directional information contained in the four canopy reflectance components provided by SLC.

The variable estimation was performed for each field in the study area by Bayesian optimization of the coupled SCL-MODTRAN model. The Bayesian cost function χ is defined as [13]:

$$\chi = \frac{1}{2}(\mathbf{L} - \mathbf{L}_0)^T \mathbf{C}_0^{-1} (\mathbf{L} - \mathbf{L}_0) + \frac{1}{2}(\mathbf{v} - \mathbf{v}_a)^T \mathbf{C}_a^{-1} (\mathbf{v} - \mathbf{v}_a)$$

where \mathbf{L} is the vector of simulated radiance, \mathbf{L}_0 is the vector of observed radiance, \mathbf{C}_0 is the diagonal matrix containing the noise variance for each spectral band (here, the noise variance was taken equal to 1 [mW/(m².sr.nm)]² for all spectral bands), \mathbf{v} is the vector of variable values, \mathbf{v}_a is the vector of a priori variable values, and \mathbf{C}_a is the diagonal matrix containing the variance of the a priori parameters.

From the cost function, one can derive the following rule for updating the parameter values in of the optimization algorithm:

$$\Delta \mathbf{v} = [\mathbf{J}^T \mathbf{C}_0^{-1} \mathbf{J} + \mathbf{C}_a^{-1} + \mu \mathbf{J}^T \mathbf{J}]^{-1} [\mathbf{J}^T \mathbf{C}_0^{-1} \Delta \mathbf{L} + \mathbf{C}_a^{-1} (\mathbf{v}_a - \mathbf{v})]$$

where \mathbf{J} is the Jacobian matrix which contains in each column the derivative of the model output with respect to each parameter, and $\Delta \mathbf{L} = (\mathbf{L}_0 - \mathbf{L})$, and μ is a damping factor whose initial value is zero. If the variable step leads to a cost increase, μ is set to one, and then multiplied by 10 until a better point is found in the variable space.

The starting point for each variable was the middle of its

Table 1. Values used for the fixed SLC parameters

Parameter	W	S	P	Co	Cl	B	R	G
greenCs	0.05	0.05	0.05	0.2	0	0.05	0.05	0.05
brownCab	5 $\mu\text{g}/\text{cm}^2$							
brownCw	0.001 cm							
brownCdm	0.005 g/cm^2							
brownCs	1							
brownN	2							
D	0							
LIDFa	-0.7	0	-0.35	-0.35	0.8	1	-0.35	-0.35
LIDFb	0	-1	-0.15	-0.15	0	0	-0.15	-0.15
Zeta	0.5	1	1	1	1	0.5	0.5	1
Hot	0.05	0.2	0.05	0.05	0.05	0.05	0.05	0.05

a priori range, except for LAI whose starting point was 2. Throughout the iterations, transformed variables were used in order to constrain the variables within their a priori range. The maximum number of iterations allowed was set to 10. The free variables in the optimization were: LAI, canopy cover (Cv), Cab, Cdm, leaf mesophyll parameter (N), and fraction of brown material in the canopy (fB) for the wheat, grass, and rapeseed fields.

The optimization algorithm was applied once for each field in the study area, using the average signature as the reference signature \mathbf{L}_0 . The fields were digitized by hand, avoiding mixed pixels at the field boundaries. The crop type of each field was determined by classifying the ten first bands of the minimum noise fraction (MNF) image and taking the dominant crop as the crop type. For each crop type, the set of fixed parameter values (Table 1), a priori values (dotted lines in Fig. 1, for wheat) and parameter ranges (boundary values of the y axes on Fig. 1, for wheat) was chosen based on expert knowledge. The atmospheric parameters were kept constant, with rural aerosol profile, 1 cm⁻¹ band model, visibility of 20 km, and water column of 2.070 g/cm².

4. RESULTS AND DISCUSSION

The Bayesian optimization converges to a stable solution in usually less than five iterations, as can be seen on Fig. 1, which presents the variation of the variables through the iteration steps for a winter field (id = 6).

The total cost χ consists almost entirely of the part of the cost due to mismatch between the simulation and the observed APEX radiance signature (model cost). Its high values (e.g. final value of 900) are due to the very high number of bands. The part of the cost due to the difference between the current variable values and the a priori values (apriori cost) is several orders of magnitude smaller than the model cost. Therefore, the a priori has virtually no influence on the optimization process, which is then almost entirely controlled by the radiometric match between model simulation and APEX data. The variations of the N parameter to the minimum and then to the maximum value shows that the variable transformation is an effective means of constraining the variables in their a priori ranges. The damping factor μ can be seen as an indicator of the degree

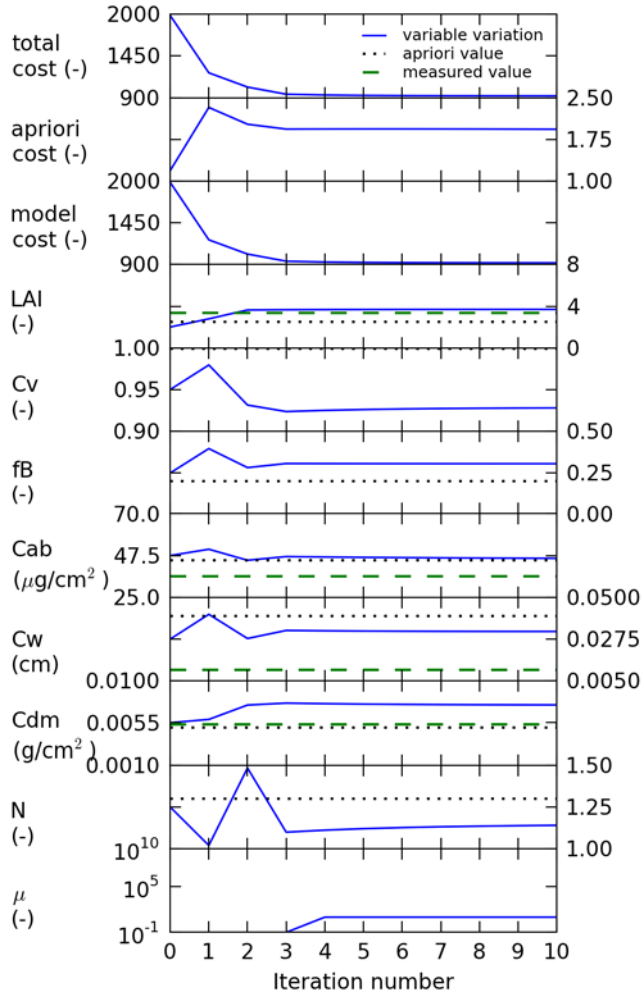


Figure 1. Variation of the costs, variables, and damping factor through the optimization process for a winter wheat field (id = 6).

of linearity of the model: the smaller the μ value, the more linear the model. A value of 10^{-1} indicates that no damping was used, as is the case in the first iterations.

The estimated values for the variables are the values obtained at the last iteration. For the winter wheat field used in Fig. 1, LAI was well estimated, whereas the estimates of Cab, Cdm, and Cw were less good. This may be due to the quality of the lab measurements, as the leaves may dry between the field and the lab, and coefficients for specific transfer functions applied to the SPAD measurements were taken from past field campaigns in France. The canopy cover was constrained in a realistic range, and the estimated value for fB is also plausible, as the wheat was very mature, with some brown leaves in the canopy and well-developed wheat ears. In addition, a value of 1.2 for the N parameter is common for leaves of monocot plants.

The final simulated radiance signature for the winter wheat field (id = 6) is presented in Fig. 2, as well as the spectral profile of the absolute error. As can be seen on the plots, the Bayesian optimization achieved a very good

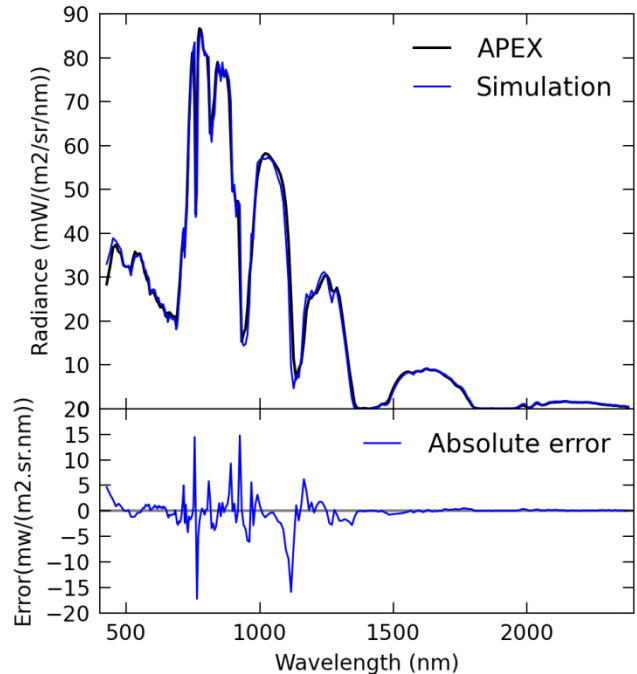


Figure 2. Optimized simulation and final absolute error for the winter wheat field (id = 6).

match with the APEX radiance signature, especially in the SWIR region (absolute error smaller than $2 \text{ mW}/(\text{m}^2.\text{sr}.\text{nm})$). A small discrepancy remained in the blue region, maybe due to atmospheric parameters. The highest errors remain in the NIR region, with values up to $\pm 15 \text{ mW}/(\text{m}^2.\text{sr}.\text{nm})$. This may be due to minor spectral shifts in the APEX data and to the spiky nature of the radiance curve in this domain due to atmospheric absorption features. These two factors combined can cause significant differences when convolving the simulated radiance (1 nm resolution) with the APEX spectral bands [14].

The map of the crop types used for selecting the a priori data to use in the Bayesian optimization is presented on the left of Fig. 3. It was obtained by selecting the dominant crop for each field, based on the pixel-based maximum likelihood classification of the first ten bands of MNF image. Bands 5-10 of the MNF image, however, had visible striping in the along-track direction, which caused visible vertical patterns in the pixel-based classification. This may have decreased the accuracy of the crop type selected for each field, and therefore the quality of the a priori data used in the Bayesian optimization. This can have very negative effect on the accuracy on the estimates, especially when a crop with a different leaf inclination distribution function (LIDF) has been selected.

The maps of the estimates of LAI and Cab obtained through the Bayesian optimization of all the fields in the study area are presented in the middle and right part of Fig. 3. For both variables, a very large range of values is present on the map, including very high LAI values above 6, mostly for winter wheat fields, but also grass and beans for

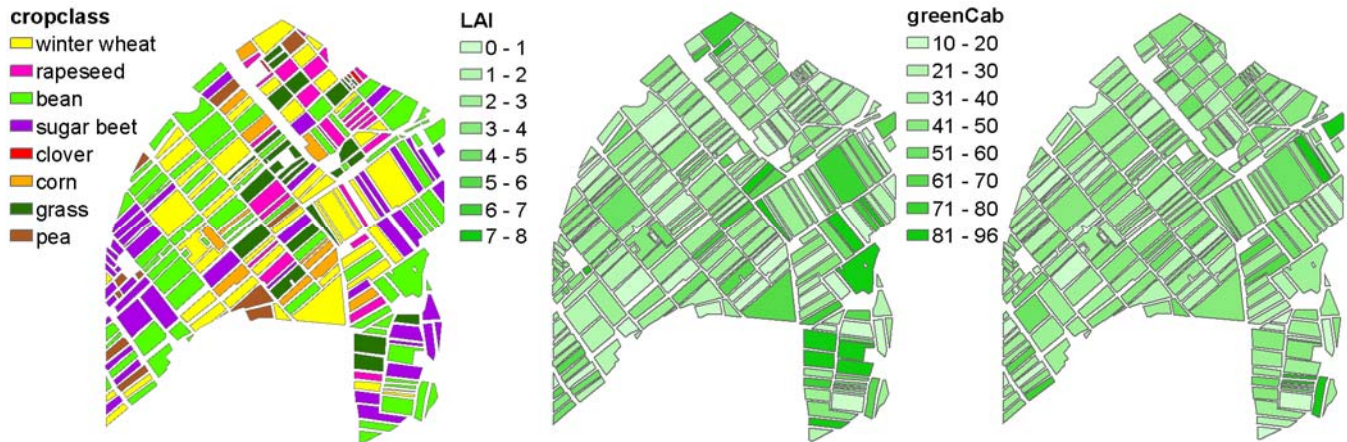


Figure 3. Maps of crop type, estimated LAI and Cab for the fields in the study area (raw image coordinates, north is to the top of the maps)

example. This shows the capacity of the estimation set-up to estimate very high LAI values. The Cab of a majority of fields was estimated between 20 and 50 $\mu\text{g}/\text{cm}^2$, which corresponds to the values usually found in the literature.

5. CONCLUSION

This study showed that the coupled SLC-MODTRAN model is able to simulate crop vegetation canopies. The Bayesian optimization of the coupled model was able to provide estimates for up to 7 free variables, including LAI and Cab. The maps of the estimates of LAI and Cab show a good coverage of the expected range for these parameters. Choosing appropriate values for the fixed parameters, especially the LIDFa and LIDFb parameters, is essential for accurate estimation of LAI, Cab and the other free variables. Based on the field results presented here, further research will focus on estimating LAI and Cab at the pixel level.

6. ACKNOWLEDGEMENT

The data collection was supported by a grant of the Swiss University Conference and ETH-Board in the framework of the HyperSwissNet project. The first author was employed at Wageningen University, the Netherlands when developing the Bayesian optimization algorithm.

7. REFERENCES

- [1] F. Gemmell, *et al.*, "Comparison of measured boreal forest characteristics with estimates from TM data and limited ancillary information using reflectance model inversion," *Remote Sensing of Environment*, vol. 81, pp. 365-377, 2002.
- [2] B. Combal, *et al.*, "Retrieval of canopy biophysical variables from bidirectional reflectance using prior information to solve the ill-posed inverse problem," *Remote Sensing of Environment*, vol. 84, pp. 1-15, 2002.
- [3] C. Atzberger, "Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models," *Remote Sensing of Environment*, vol. 93, pp. 53-67, 2004.
- [4] B. C. Gao, *et al.*, "Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean," *Remote Sensing of Environment*, vol. 113, 2009.
- [5] R. Richter, "Atmospheric / Topographic Correction for Satellite Imagery (ATCOR-2/3 User Guide, version 7.0, January 2009)," DLR - German Aerospace Center - Remote Sensing Data Center, Wessling, Germany 2009.
- [6] W. Verhoef and H. Bach, "Simulation of hyperspectral and directional radiance images using coupled biophysical and atmospheric radiative transfer models," *Remote Sensing of Environment*, vol. 87, pp. 23-41, 2003.
- [7] V. C. E. Laurent, *et al.*, "Estimating forest variables from top-of-atmosphere radiance satellite measurements using coupled radiative transfer models," *Remote Sensing of Environment*, vol. 115, pp. 1043-1052, 2011.
- [8] A. Hueni, *et al.*, "The structure of the APEX (Airborne Prism Experiment) Processing and Archiving Facility," Grenoble, 2009.
- [9] M. Jehle, *et al.*, "APEX - Current status, performance and validation concept," Waikoloa, HI, 2010, pp. 533-537.
- [10] A. Damm, *et al.*, "Data Acquisition Report 2011," Remote Sensing Laboratories, Zurich 2011.
- [11] W. Verhoef and H. Bach, "Coupled soil-leaf-canopy and atmosphere radiative transfer modeling to simulate hyperspectral multi-angular surface reflectance and TOA radiance data," *Remote Sensing of Environment*, vol. 109, pp. 166-182, 2007.
- [12] A. Berk, *et al.*, "MODTRAN4 Version 3 Revision 1 User's manual," Airforce Research Laboratory, Hanscom, MA, USA 2003.
- [13] A. Tarantola, *Inverse problem theory and methods for model parameter estimation: Society for Industrial and Applied Mathematics*, 2005.
- [14] P. D'Odorico, *et al.*, "Performance assessment of onboard and scene-based methods for Airborne Prism Experiment spectral characterization," *Applied Optics*, vol. 50, pp. 4755-4764, 2011.