This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Varvia, Petri; Rautiainen, Miina; Seppänen, Aku

Bayesian estimation of seasonal course of canopy leaf area index from hyperspectral satellite data

Published in:
Journal of Quantitative Spectroscopy and Radiative Transfer

DOI:
[10.1016/j.jqsrt.2018.01.008](https://doi.org/10.1016/j.jqsrt.2018.01.008)

Published: 01/01/2018

Document Version
Peer reviewed version

Published under the following license:
CC BY-NC-ND

Please cite the original version:
Varvia, P., Rautiainen, M., & Seppänen, A. (2018). Bayesian estimation of seasonal course of canopy leaf area index from hyperspectral satellite data. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 208, 19-28. <https://doi.org/10.1016/j.jqsrt.2018.01.008>

Bayesian estimation of seasonal course of canopy leaf area index from hyperspectral satellite data

Petri Varvia^a, Miina Rautiainen^{b,c}, Aku Seppänen^a

^a*Department of Applied Physics, University of Eastern Finland*

^b*Department of Built Environment, School of Engineering, Aalto University*

^c*Department of Electronics and Nanoengineering, School of Electrical Engineering, Aalto University*

Abstract

In this paper, Bayesian inversion of a physically-based forest reflectance model is investigated to estimate of boreal forest canopy leaf area index (LAI) from EO-1 Hyperion hyperspectral data. The data consist of multiple forest stands with different species compositions and structures, imaged in three phases of the growing season. The Bayesian estimates of canopy LAI are compared to reference estimates based on a spectral vegetation index. The forest reflectance model contains also other unknown variables in addition to LAI, for example leaf single scattering albedo and understory reflectance. In the Bayesian approach, these variables are estimated simultaneously with LAI. The feasibility and seasonal variation of these estimates is also examined. Credible intervals for the estimates are also calculated and evaluated. The results show that the Bayesian inversion approach is significantly better than using a comparable spectral vegetation index regression.

Keywords: leaf area index, spectral invariants, reflectance model, uncertainty quantification, seasonal dynamics

1. Introduction

Remote sensing of forest biophysical parameters, such as canopy leaf area index (LAI), has traditionally utilized data with low spectral resolution (i.e. multispectral measurements). Hyperspectral measurements (imaging spectroscopy) offer finer-grained spectral and radiometric information on the environment. Yet, the use of hyperspectral satellite measurements in estimation of forest parameters has been hampered by the larger data dimension compared to multispectral data and the relatively low number of operational hyperspectral satellite sensors. Several new satellite missions providing high spectral resolution data will be launched in the forthcoming years (including, for example, the EnMAP and PRISMA missions). Therefore, there is an urgent need to develop more efficient analysis methods to tackle the problem of high data dimensions.

Most of the existing methods for estimation of canopy LAI from hyperspectral measurements use only a few select spectral bands and thus do not utilize the full information content of the remotely sensed measurements. Of the existing approaches, empirical regression using narrowband vegetation indices (VI) has been the most widely studied (e.g. Gong et al., 2003; Le Maire et al., 2008; Heiskanen et al., 2013). The primary drawback of empirical VI regression is its high site, time and species specificity: a regression model trained in one forest area usually does not

generalize well to other locations. In theory, approaches based on forest reflectance model inversion can overcome this problem. While studies using reflectance model inversion with hyperspectral measurements have been done using airborne hyperspectral sensors and either band selection (Meroni et al., 2004; Schlerf and Atzberger, 2006) or the full hyperspectral measurement (Banskota et al., 2015), studies using the inversion approach with spaceborn sensors are still either scarce or nonexistent.

Recently, Varvia et al. (2017) proposed a Bayesian method to estimate canopy LAI from hyperspectral satellite measurements. The method is based on Bayesian inversion of a physically-based forest reflectance model. The simulation results in (Varvia et al., 2017) indicated improved estimation accuracy compared to the empirical vegetation index regression approach. Main advantages of the new method are that it also allows simultaneous estimation of other forest reflectance model parameters, such as leaf albedo, and produces uncertainty estimates for the model variables. In this article, the Bayesian approach is tested using EO-1 Hyperion satellite data in a Finnish boreal forest. The method is compared to a conventional VI regression using both field-measurements and reflectance model simulations as a training data. The performance of the uncertainty estimates produced by the Bayesian method is also evaluated. Moreover, the seasonal dynamics of the estimated LAI, leaf albedo and understory reflectance are examined.

2. Materials and methods

2.1. Study area

The study area is located next to Hyytiälä Forestry Field Station in southern Finland (61° 50' N, 24° 17' E). Dominant tree species in the area are Norway spruce (*Picea abies* (L.) Karst.), Scots pine (*Pinus sylvestris* L.) and birches (*Betula pubescens* Ehrh., *Betula pendula* Roth.). Understory vegetation is usually composed of two layers: a ground layer of mosses and lichens, and an upper understory layer which has dwarf shrubs, graminoids, and/or herbaceous species. The greening of vegetation after the winter starts in early May, peak growing season is typically reached by late June and the vegetation stays relatively stable until mid-August.

2.2. Field measurements

For this study, we used data from 18 stands which represented different species compositions and age classes (stand age varied from 25 to 100 years) typical to the region. Canopy gap fractions and effective leaf area index (LAI_{eff}) of all the plots were measured in early May, early June and early July in 2010 (which coincide with the acquisition of Hyperion images, see Section 2.3 and Table 1). The measurements were carried out in exactly the same locations using two units of the LAI-2000 Plant Canopy Analyzer (Li-Cor Inc.) to obtain simultaneous readings from above and below the canopy. The instruments optical sensor measures diffuse sky radiation (320- 490 nm) in five different zenith angle bands (centered at zenith angles: 7°, 23°, 38°, 53° and 68°). Simultaneous measurements with two LAI-2000 PCA units provide canopy transmittances (i.e. canopy gap fractions) for five zenith angles which can then be used to generate LAI based on inversion of Beer's law (Welles, 1990). A total of twelve points per stand were measured

according to a standard VALERI network (Validation of Land European Remote Sensing Instruments) sampling design (see e.g. Majasalmi et al., 2012). The measurement points were located as a cross and placed at four-meter intervals on a North-South transect (6 points) and on an East-West transect (6 points). The below-canopy measuring height was 1 m above the ground i.e. only trees were included in the field-of-view. The LAI-2000 measurements were made during standard overcast sky conditions or during clear sky conditions in late evening and early morning when the Sun was below the field-of-view of the LAI-2000 instruments optical sensor. The measurements are described in more detail in Heiskanen et al. (2012).

Concurrently with the LAI measurements, data on understory reflectance spectra was collected in four stands representing common site fertility types: mesic, xeric, sub-xeric and herb-rich sites. In each of the four sites, a 28-m long transect was measured at 70 cm intervals under diffuse light conditions using a FieldSpec Hand-Held UV/VNIR (325 – 1075 nm) Spectroradiometer manufactured by Analytical Spectral Devices (ASD). No fore-optics were attached to the instrument which means that the field-of-view was 25°. The raw measurement data were processed to hemispherical-directional reflectance factors (HDRF) and averaged for all measurement points in each transect. The measurements and data are described in more detail by Rautiainen et al. (2011).

In addition, we had access to regular stand inventory data which had been collected in all our study plots a year before the satellite images were acquired. In this study, the forest inventory data are used only to provide background information on site fertility type, stand structure and species composition (Table 1).

Table 1: A summary of study stands.

	Coniferous	Broadleaved
Number of stands	11	7
Mean tree height (m)	7.5 – 18.6	11.7 – 23.1
Stem volume (m ³ ha ⁻¹)	40 – 220	71 – 243
LAI _{eff} (May)	1.5 – 3.6	0.7 – 1.5
LAI _{eff} (June)	1.3 – 4.3	2.2 – 3.1
LAI _{eff} (July)	1.8 – 4.6	2.6 – 3.4

2.3. Satellite data

Three EO-1 Hyperion satellite images were acquired from our study area concurrently with the field data collection (Table 2). Hyperion was a narrowband imaging spectrometer onboard NASA’s Earth Observing-One (EO-1) with 242 spectral bands (356 – 2577 nm) and a 30 m × 30 m spatial resolution (Pearlman et al., 2003). The set of Hyperion images captures the main phenological changes occurring in the study area in 2010: the image from May corresponds to the time of bud burst, the image from June to the full leaf-out situation, and the image from July to the time of maximal leaf area.

First, striping in the Hyperion images (originally accessed as L1B products) was removed using spectral moment matching (Sun et al., 2008) and corrected for missing lines using local destriping methods (Goodenough et al., 2003).

Table 2: Summary of field campaigns and satellite data in year 2010.

Month	Phenological phase	Field data		Satellite data		
		LAI measurements	Understory spectra measurements	Date	View zenith angle (degrees)	Solar zenith angle (degrees)
May	Budburst and leaf out	3 – 11 May	4 – 13 May	5 May	0.5	46.7
June	Full leaf out	26 May – 7 June	31 May – 9 June	2 June	14.8	41.5
July	Maximal leaf size	28 June – 5 July	28 June – 6 July	3 July	13.8	41.0

The spectral smile was corrected for in standard method using interpolation and pre-launch calibration measurements (Barry, 2001). Finally, the Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm was used for atmospheric correction of the images (Matthew et al., 2000). The end product of the atmospheric correction process was hemispherical-directional reflectance factor (HDRF). More details on the preprocessing of this set of Hyperion images is available in Heiskanen et al. (2013). The Hyperion images were georeferenced using approximately 20 ground control points. The mean HDRFs for each study stand were extracted using a 3×3 pixel window which corresponds to the area covered by the field measurements in each stand.

3. Bayesian estimation of effective LAI

In this section, the Bayesian approach to LAI estimation is shortly summarized. Except for slight adjustments in certain hyperparameters, the methodology is identical to Varvia et al. (2017) and the reader is referred there for more detail.

The forest reflectance spectrum is modeled is adopted from Rautiainen and Stenberg (2005); in this so-called PARAS model, the bidirectional reflectance factor (BRF, $r(\theta_1, \theta_2, \lambda)$) of the forest for solar zenith angle θ_1 , viewing angle θ_2 , and wavelength λ is:

$$r(\theta_1, \theta_2, \lambda) = \rho_g(\theta_1, \theta_2, \lambda)t_c(\theta_1)t_c(\theta_2) + Qi_c(\theta_1)\frac{\omega_L(\lambda) - p\omega_L(\lambda)}{1 - p\omega_L(\lambda)}, \quad (1)$$

where ρ_g is the BRF of the understory layer, t_c is the canopy transmittance, $i_c = 1 - t_c$ the canopy interceptance, Q the approximative portion of upwards scattered radiation from the canopy (Möttus and Stenberg, 2008), ω_L the leaf single scattering albedo, and p is the photon recollision probability, defined as the probability that a photon scattered in the canopy will interact within the canopy again (Knyazikhin et al., 1998a,b). It should be noted that the satellite measurements correspond to the hemispherical-directional reflectance factor (HDRF), which is approximated here

using BRF under the implicit assumption that the incoming diffuse sky radiation is negligible compared to the direct sun component.

The wavelength dependent variables ω_L and ρ_g are approximated using splines following Varvia et al. (2017):

$$\omega_L = S(\lambda; \tilde{\lambda}, \tilde{\omega}_L), \quad (2)$$

$$\rho_g = S(\lambda; \tilde{\lambda}, \tilde{\rho}_g), \quad (3)$$

where $S(\cdot)$ is the cubic monotone Hermite spline function, $\tilde{\lambda}$ is the preselected nodal wavelengths of the spline (see Figure 1 in Varvia et al. (2017)), and $\tilde{\omega}_L \in \mathbb{R}^{27}$ and $\tilde{\rho}_g \in \mathbb{R}^{27}$ are the values of ω_L and ρ_g at those node points. The spline approximation is written in order to reduce the number of unknown variables and to produce the desired structure and smoothness for the spectral variables. The lower-dimensional vectors $\tilde{\omega}_L$ and $\tilde{\rho}_g$ are substituted for ω_L and ρ_g in the reflectance model.

In order to make the estimated quantities comparable with the field-measured effective LAI, the effective LAI is used in the reflectance model. The effective LAI is assumed to follow the model $\text{LAI}_{\text{eff}} = \beta \text{LAI}$, where β is the shoot clumping factor. The photon recollision probability p is approximated following Stenberg (2007) as

$$p = 1 - \frac{\beta(1 - t_d)}{\text{LAI}_{\text{eff}}}, \quad (4)$$

where t_d is the diffuse transmittance of tree canopy layer. The canopy transmittance is modeled using Beer-Lambert's law as

$$t_c(\theta) = \exp\left(-\frac{\beta \text{LAI}_{\text{eff}}}{2 \cos \theta}\right), \quad (5)$$

from which the diffuse canopy transmittance t_d is integrated following Manninen and Stenberg (2009).

3.1. Bayesian inversion

Let $r \in \mathbb{R}^{150}$ be the atmosphere corrected and calibrated Hyperion HDRF measurement and $x = [\text{LAI}_{\text{eff}} \quad \tilde{\omega}_L \quad \tilde{\rho}_g \quad \beta]^T \in \mathbb{R}^{56}$ the vector of unknown reflectance model variables defined above. Both r and x are modeled here as random variables. In Bayesian inference, the prior density of x (denoted $p(x)$) is updated using the new information gained from the measurements r . This is accomplished using the Bayes' theorem:

$$p(x|r) = \frac{p(r|x)p(x)}{p(r)} \propto p(r|x)p(x), \quad (6)$$

where $p(r|x)$ is the likelihood function containing the information from the measurements and $p(r)$ can be considered as a normalizing constant. The posterior density $p(x|r)$ describes the probability distribution of possible realizations x given the measured r and the prescribed prior formulation; the posterior density is the full solution of the inference problem.

The measurement r is modeled as

$$r = f(x) + e, \quad (7)$$

116 where $f(x)$ is the PARAS model (1), including the substituted approximations for ω_L , ρ_g , t_c , Q and p , and e is an
 117 additive Gaussian error term. With this model, the likelihood function $p(r|x)$ is of the form of a Gaussian function

$$p(r|x) \propto \exp\left(-\frac{1}{2}(r - f(x))^T \Gamma_e^{-1}(r - f(x))\right), \quad (8)$$

118 where Γ_e is the covariance matrix of e . In this paper, e is assumed to have zero mean and a standard deviation of
 119 5% of the average r in wavelengths 488–691 nm, 702–1346 nm, and 1477–1800 nm, and 10% of the average r in
 120 wavelengths 2032–2355 nm. The standard deviation values were chosen based on the EO-1 Hyperion radiometric
 121 accuracy (Pearlman et al., 2003) with some extra deviation to compensate for possible uncertainty resulting from the
 122 preprocessing and atmospheric correction (see Section 2.3).

123 3.1.1. Prior density

124 The prior density $p(x)$ in Equation (6) is formulated by approximating the parameters LAI_{eff} , $\tilde{\omega}_L$, $\tilde{\rho}_g$, and β as
 125 statistically uncorrelated (a priori), and constructing a prior distribution for each parameter.

126 As in Varvia et al. (2017), uniform prior distributions are set for LAI_{eff} and β : LAI_{eff} is by definition non-negative
 127 and unrealistically large values ($\text{LAI}_{\text{eff}} > 10$) are constrained out

$$p(\text{LAI}_{\text{eff}}) = \begin{cases} \frac{1}{10}, & 0 \leq \text{LAI}_{\text{eff}} \leq 10 \\ 0, & \text{otherwise,} \end{cases} \quad (9)$$

128 and β is constrained to the empirically observed range [0.4, 1] (Thrzien et al., 2007)

$$p(\beta) = \begin{cases} \frac{5}{3}, & 0.4 \leq \beta \leq 1 \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

129 The spectral variables $\tilde{\omega}_L$ and $\tilde{\rho}_g$ are modeled by truncated multivariate Gaussian prior distributions:

$$p(\tilde{\omega}_L) \propto \begin{cases} \exp\left(-\frac{1}{2}(\tilde{\omega}_L - \mu_{\tilde{\omega}_L})^T \Gamma_{\tilde{\omega}_L}^{-1}(\tilde{\omega}_L - \mu_{\tilde{\omega}_L})\right), & 0 \leq \tilde{\omega}_L \leq 1 \\ 0, & \text{otherwise,} \end{cases} \quad (11)$$

$$p(\tilde{\rho}_g) \propto \begin{cases} \exp\left(-\frac{1}{2}(\tilde{\rho}_g - \mu_{\tilde{\rho}_g})^T \Gamma_{\tilde{\rho}_g}^{-1}(\tilde{\rho}_g - \mu_{\tilde{\rho}_g})\right), & 0 \leq \tilde{\rho}_g \leq 1 \\ 0, & \text{otherwise.} \end{cases} \quad (12)$$

130 The expected values $\mu_{\tilde{\omega}_L}$ and $\mu_{\tilde{\rho}_g}$ were derived from published measurement data: the leaf albedo data from Luke
 131 et al. (2013) and the understory reflectance data from Peltoniemi et al. (2008). The covariance matrices $\Gamma_{\tilde{\omega}_L}$ and
 132 $\Gamma_{\tilde{\rho}_g}$ were constructed by setting the standard deviation to 25% of the prior expectation on each channel and using a
 133 preconstructed correlation matrix, see Varvia et al. (2017).

134 Under the assumption of statistical independence, the (joint) prior density of x is

$$p(x) = p(\text{LAI}_{\text{eff}})p(\tilde{\omega}_L)p(\tilde{\rho}_g)p(\beta). \quad (13)$$

3.1.2. Estimates and computation

From the posterior density (6), various estimates for the variables x can be computed. In this work, we compute the posterior mean (i.e. $E(x|r)$) and 95% credible intervals (CI) for x .

Computation of both the posterior mean and CIs require integrating over the posterior density $p(x|r)$. The integration is done numerically using the delayed rejection adaptive Metropolis (DRAM) algorithm (Haario et al., 2006). In this study, total of 840000 Monte Carlo samples are computed for each stand, using 12 parallel sampling chains of 70000 samples each. The numbers do not include the 5000 burn-in period in the beginning of each chain. For a more detailed description, see Varvia et al. (2017).

3.2. Reference methods

The Bayesian approach was compared to a empirical linear regression with a narrow-band vegetation index (VI) following Heiskanen et al. (2013). Heiskanen et al. (2013) used exactly the same data set of processed Hyperion images and ground reference LAI measurements to compare the performance of different spectral vegetation indices in estimating boreal forest LAI. Therefore, our results can be directly compared to their results: in this study, we compare our new physically-based inversion method to the performance of the best VI reported by Heiskanen et al. (2013). They discovered that the best common narrowband VI for estimating boreal forest LAI from Hyperion data was the simple ratio water index:

$$SRWI = \frac{r_{\lambda=854 \text{ nm}}}{r_{\lambda=1235 \text{ nm}}} \quad (14)$$

In this study, two reference VI regressions were used: 1) regression where the real measurement data were used as a training set and 2) regression where the training set was simulated using the PARAS model.

In the VI regression based on real training data, SRWI was first calculated for each measured stand. Leave-one-out cross-validation was then done: each study stand was left out at a time and the rest of the data was used as a training set. Ordinary linear regression was done between field-measured LAI_{eff} and SRWI in the training set. The regression model was finally used to estimate LAI_{eff} for the left out stand.

In the simulation based VI regression, a set of synthetic training data was simulated using the PARAS model using the field-measured LAI_{eff} , and the known tree species composition and understory type of each stand in the data set. The aim was to construct a simulated training set as close as possible in composition to the data set used in this study. As in the real training data case, ordinary linear regression was done between LAI_{eff} and SRWI in the simulated training set, and the regression model was then used to predict LAI_{eff} for each stand.

4. Results and discussion

4.1. Performance of LAI_{eff} estimates

The total root mean square errors (RMSE) and biases of the Bayesian LAI_{eff} estimates and the two reference VI regressions are presented in Table 3. The VI regression using field-measured training data (in a leave-one-out cross-

validation setting) has the lowest RMSE value and bias. The Bayesian approach produces the second best results, with somewhat higher RMSE and bias. The VI regression using simulated training data performs the poorest. In analyzing these results, it must be kept in mind that only the Bayesian and simulation based VI regression are directly comparable, because neither method utilizes contemporary field measurement data in the estimation. The Bayesian method provides roughly 20% smaller RMSE and an improvement of roughly 25% units in bias over the simulation based VI regression.

Table 3: RMSE, relative RMSE and relative bias of effective LAI estimates for the Bayesian posterior mean and the reference VI regression estimates for all test cases.

	RMSE	RMSE%	bias%
Post. mean	0.83	31.77	-7.42
VI (real)	0.52	19.77	-0.57
VI (simul.)	1.00	38.19	-32.79

The RMSE and bias were also calculated by month (Table 4) and by the dominant species (Table 5) to examine possible seasonal and species-specific variation in the estimation error. In the by-month grouping, the RMSE and bias of the VI regression estimates do not show significant monthly variation. The Bayesian estimates, on the other hand, perform superiorly with the May measurements. In the other months, performance is only slightly better than the simulation based VI regression, performance on the other months. A possible cause for this behavior is that in May (beginning of the growing season), the understory is less "green" and is thus more easily separated from the canopy component. The difference in the view geometry between the May and June/July scenes (see Table 2) might also have a significant effect.

In the by-species grouping, the difference between pine dominated and spruce dominated stands is insignificant for the Bayesian and the real training data based VI regression. The simulation based VI regression performs worse in spruce stands (in terms of the raw RMSE). The methods perform best in deciduous stands when measured by RMSE. With this further examination, the lower RMSE and bias of the Bayesian estimates in comparison to the simulation based VI regression is mostly the benefit of better performance with the May measurements, and in spruce dominated stands.

For the Bayesian approach, 95% percent credible intervals were computed for the LAI_{eff} estimates. The credible interval coverage percentage (CI%) and average CI width are shown in Table 6 for all the plots, by month, and by species. The CI coverage (CI%) is defined as the percentage of the field-measured LAI_{eff} values that are covered by the estimated credible interval; the ideal value is 95%. The realized CI coverage for all stands is 53.70%, which implies that the estimated intervals are significantly too narrow. The monthly and species-wise coverages range between 40% and 70%. The groupings with the best coverage values, May and Deciduous, correspond to the groups that also have the lowest RMSE values (Tables 4 and 5). The poor coverage percentages imply the presence of large additional errors that are not sufficiently modeled in the error e (see Equation (8)).

Table 4: RMSE, relative RMSE and relative bias of effective LAI estimates for the Bayesian posterior mean and the reference VI regression estimates by month.

	RMSE	RMSE%	bias%
May			
Post. mean	0.59	29.88	8.66
VI (real)	0.50	24.99	-7.07
VI (simul.)	1.03	51.70	-45.35
June			
Post. mean	0.97	34.04	-24.58
VI (real)	0.51	17.86	0.81
VI (simul.)	1.00	35.47	-30.56
July			
Post. mean	0.90	29.33	-1.93
VI (real)	0.55	18.09	3.79
VI (simul.)	0.98	31.93	-26.71

Figure 1 shows the scatter plots corresponding to the three LAI_{eff} estimation methods by species (different symbols) and by month (connecting line). The Bayesian posterior mean estimates are more scattered than both VI regressions, but have less bias than the simulation based VI regression (as in Table 3). The temporal behavior of the Bayesian estimates is less stable than the VI regressions: there are many more stands for which the estimated LAI for June is much smaller than the LAI in May. This is also indicated by the high negative bias in the Bayesian estimates corresponding to June (Table 4).

Figure 1: Estimated LAI_{eff} vs. the field-measured value for the Bayesian method and the reference VI regressions. Pine dominated stands are marked with circles, spruce stands with squares and deciduous stands with triangles. The estimates of the same stand during different months are connected with a line.

Table 5: RMSE, relative RMSE and relative bias of effective LAI estimates for the Bayesian posterior mean and the reference VI regression estimates by dominant species.

	RMSE	RMSE%	bias%
Pine			
Post. mean	0.86	39.03	7.98
VI (real)	0.60	27.52	-4.02
VI (simul.)	0.94	42.69	-33.34
Spruce			
Post. mean	0.93	30.18	-2.93
VI (real)	0.58	18.31	-6.90
VI (simul.)	1.23	38.79	-35.19
Deciduous			
Post. mean	0.71	29.40	-21.19
VI (real)	0.35	15.67	10.47
VI (simul.)	0.65	29.72	-27.84

Table 6: 95% credible interval coverage (CI%) and average interval width for the Bayesian LAI_{eff} estimates.

Group	CI%	width
All	53.70	1.57
May	66.67	1.32
June	38.89	1.39
July	55.56	2.02
Pine	41.67	1.36
Spruce	50.00	1.94
Deciduous	66.67	1.23

4.2. Seasonal dynamics in the Bayesian estimates

In this section, the seasonal behavior of the Bayesian estimates (and posterior densities) is examined further. First, two example stands were chosen to visualize some commonly occurring interesting effects. The first example demonstrates the effect of increasing LAI on the behavior of ρ_g and ω_L estimates, and the second example shows how the method works in a very dense canopy.

The first example is a pure broadleaved stand with a mean tree height of 14.1 m and a stem volume of 136 m³/ha. The understory vegetation is lush and herb-rich. The measured forest reflectance spectra and the estimates for the model parameters by month are shown in Figure 2. The measured reflectance changes drastically between May and June, and slightly between June and July. The shallow red edge in the May data implies the relative absence of green vegetation. In May, the estimated LAI_{eff} tends to zero, with a tight credible interval (the shaded area in the graph), while the field-measured LAI_{eff} is 0.91. Moving to June and July, the field-measured LAI_{eff} grows to around 2.7. The estimated LAI_{eff} follows suit, however, the Bayesian estimate underestimates in June. The credible interval grows in width as LAI_{eff} increases.

The spectral variables ω_L and ρ_g have an interesting behavior. In May, when LAI (and thus the canopy cover) is

low, the estimate for leaf albedo ω_L is both extremely low for a deciduous stand, and very uncertain, as is seen from the wide credible intervals. In June and July, after leaf out, the ω_L estimate has a sensible value and tight credible intervals. The behavior of the estimate for understory BRFR ρ_g is opposite. For May, ρ_g has been estimated with high certainty, but after leaf out, uncertainty significantly increases. The May ρ_g estimate has a very shallow red edge, which is realistic, as in the beginning of May, the herb-rich understory consists mostly of dry vegetation. In June and July, the estimated ρ_g is much more green, as is the actual understory. The uncertainty behavior of the spectral variables is logical: If LAI is close to zero (May), leaf albedo has no significant effect on the forest reflectance, which then consists nearly completely of the understory component. As LAI, and canopy cover, increases, less and less of the radiation penetrates the canopy and reflect from the understory, in which case the value of ρ_g in the model has negligible impact on the reflectance. Similar behavior was demonstrated in the simulation study by Varvia et al. (2017).

The last model parameter that is estimated is the clumping factor β . It should be noted, that while higher level clumping is not explicitly included in the model (i.e. the shoots/leaves are modeled as uniformly distributed in the canopy layer), the estimated β probably more closely corresponds to an effective total clumping factor and not to the actual shoot clumping factor. In this example, the posterior mean estimate of β in June and July is close to 0.5 and the posterior density tends to the low end of the allowed β range. In May, the posterior density of β is wider for the same reason as with ω_L : if LAI is close to zero, variation of β has no real effect.

The second example (Figure 3) is a pure spruce stand, where the mean tree height is 13.3 m and the stem volume is 211 m³/ha. The understory is mesic, consisting mostly of mosses and *Vaccinium* sp. dwarf shrubs. This example stand has one of the highest field-measured LAI values in the Hyytiälä data set. The seasonal variation of the measured reflectance is much more modest than in the previous broad leaved example: the shape of the spectra stays relatively same over the months, while reflectance somewhat increases between May and June (large part of this is caused by the different view geometry). The field measured LAI_{eff} increases slightly from 3.55 in May to around 3.90 in June and July. This value of LAI_{eff} is high and the stand can be considered to be in the saturation zone, where the effect of LAI on the forest reflectance saturates. The effect of LAI saturation is reflected by the width of the posterior density: as estimated LAI increases, so does the uncertainty. This is especially evident in the July estimate.

The estimates of the spectral variables stay fairly stable over the months and the estimated leaf albedo values are feasible for a spruce stand. There is a slight increase in ω_L from May to June, which might be caused by the seasonal variation of needle pigments (e.g. Linder, 1972). The behavior of estimate uncertainty of both spectral variables with respect to the variation of estimate LAI_{eff} is similar to the previous example, yet less visible. The posterior density of β again tends to the lower boundary in May and June.

In the posterior densities of May LAI_{eff} and β , multimodality is observed. Multimodality occurs on several other stands in the data set as well. In practical terms, this means that there are multiple "clusters" of feasible solutions. Computationally, the multimodality of the posterior density might cause error in the estimates if the MCMC algorithm does not sample all of the multiple modes.

Figures 4 and 5 show the overall seasonal behavior of estimated ω_L and ρ_g , respectively. Figure 4 shows that the estimated leaf albedo is higher for broad leaved stands, as it should be (Lukeš et al., 2013), and that leaf albedo has the tendency to increase from May to July. The ρ_g estimates (Figure 5) have less seasonal variation overall, and there is a less marked difference between coniferous and broad-leaved stands. The tendency towards the prior mean is stronger in July, this is explained by the increased LAI. In May, some ρ_g estimates are distinctly non-green, as are the corresponding understories.

4.3. Accuracy of the understory reflectance estimates

The estimated understory reflectance was compared with the field-measured values (see Section 2.2). Full comparison was not possible, because the field measurements of ρ_g were not done separately on every stand and have a different spectral range. However, the understory type of each stand is known and thus the ρ_g estimates corresponding to each understory grouping can be compared with the field-measurements on their common shared spectral range (c. 490 – 1080 nm). The results are shown in Figure 6. The performance of the estimated ρ_g is fairly lackluster. The results for xeric stands in May, mesic stands in June, and herb-rich and subxeric stands in July are close to the field-measured values. Yet, for most cases there is a tendency to overestimate the ρ_g , and to produce spectra that have a stronger characteristic structure of green vegetation (e.g. eminent red edge). The analysis is somewhat confounded by the interplay between the estimated LAI and ρ_g , described in the previous section.

The prior density for ρ_g was constructed using understory measurements mostly done relatively late in the growing season (see (Peltoniemi et al., 2008)) and thus it was not clear if the prior density was adequate in describing ρ_g in May. By examining the May results for herb-rich and subxeric stands in Figure 6, the prior density can, in the best case, work fairly sufficiently for even such non-green backgrounds.

4.4. General discussion

Overall, the results demonstrate that forest reflectance model inversion can be successfully done when using hyperspectral data and can provide a wealth of information on multiple forest parameters. In the Bayesian approach described in Varvia et al. (2017) and in this paper, there are two major aspects that can be feasibly developed further in the future to improve estimate quality: 1) model error, and 2) the prior density.

The reflectance model used in this study is fairly simple, which has the advantage that the model has only a few unknown variables. However, modeling of bark and branches, for example by inclusion of bark area index as was done by Stenberg et al. (2013), might improve the estimates. Additionally, having an angular dependent model for Q in the Equation (1) would be a further improvement. However, as Q depends crucially on the stand and canopy structure, constructing such a model would be difficult. Perhaps the most promising direction for future development is the error term e in equation (7). In this paper, we modeled e as an uncorrelated Gaussian random variable that mostly describes the radiometric error of the Hyperion instrument. However, the term e should also include model

error, that is, e is the full discrepancy between the model output and the measurement r . Unlike radiometric error, the model error is highly structured. If the magnitude and correlation of this error were known, even to some extent, and utilized, significant benefits would be expected. Especially the CI coverage would most probably see large gain in performance through the quantification of the model error's effect on the estimates.

The prior density was constructed using available data on the variables of interest, and was found to work sufficiently well. Possible improvements are time-dependent seasonal priors, and priors that include correlation between different variables, for example ω_L and β . However, both of these potential improvements require further empirical data on these variables and their seasonal course.

The Bayesian approach was here used with hyperspectral data. Yet, the approach could be used as well with the more widely available multispectral data (e.g Landsat, Sentinel 2). Using a different instrument would require that the prior models for the spectral variables are rewritten to correspond to the spectral bands of the instrument.

5. Conclusions

In this article, the Bayesian approach was tested using EO-1 Hyperion satellite data in a Finnish boreal forest. The Bayesian estimates were compared to a conventional VI regression using both field-measured or simulated training data. Moreover, the performance of the uncertainty estimates (95% credible interval) produced by the Bayesian method was studied, and the seasonal behavior of the estimated LAI, leaf albedo and understory reflectance was evaluated.

The performance of the Bayesian LAI_{eff} estimates was superior in both RMSE and bias to the comparable simulation based VI regression. The improved estimation accuracy was most evident in the May data and on spruce dominated stands. VI regression using field-measured training data was superior to both methods, but has the significant drawback of requiring field measurements. The LAI_{eff} credible interval coverage was generally poor at roughly 40–70% range. Seasonality of the estimated leaf albedo and understory reflectance was examined. The ω_L estimates were feasible and showed a tendency to increase over the growing season. The ρ_g estimates were compared to monthly field-measurements grouped by understory type. The performance was not consistent, but in many cases promising. Several aspects of the simulation study Varvia et al. (2017), such as variation of uncertainty in estimated ω_L and ρ_g with varying LAI, were here reproduced using real data.

The results show that Bayesian forest reflectance model inversion is feasible in estimation of leaf area index, and other forest parameters, from hyperspectral satellite data. The method provides estimate uncertainty measures and the concurrent estimation of multiple forest parameters provides a potential wealth of information, not only in estimated values, but also in posterior covariances. For reliable uncertainty quantification, however, model error needs to be quantified. With further development, the presented approach might prove to be a valuable tool in retrieval of forest biophysical variables.

Acknowledgements

This work was supported in part by the Finnish Cultural Foundation, North Savo Regional fund and in part by the Academy of Finland (Finnish Centre of Excellence of Inverse Problems Research 2012-2017, project number 284715, and projects 286390, 135502, 270174 and 303801).

References

- Banskota, A., Serbin, S. P., Wynne, R. H., Thomas, V. A., Falkowski, M. J., Kayastha, N., Gastellu-Etchegorry, J.-P., Townsend, P. A., 2015. An LUT-based inversion of DART model to estimate forest LAI from hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (6), 3147–3160.
- Barry, P., November 2001. EO-1/ Hyperion Science Data User's Guide Level 1.B.
- Gong, P., Pu, R., Biging, G. S., Larrieu, M. R., 2003. Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing* 41 (6), 1355–1362.
- Goodenough, D. G., Dyk, A., Niemann, K. O., Pearlman, J. S., Chen, H., Han, T., Murdoch, M., West, C., 2003. Processing hyperion and ali for forest classification. *IEEE transactions on geoscience and remote sensing* 41 (6), 1321–1331.
- Haario, H., Laine, M., Mira, A., Saksman, E., 2006. DRAM: efficient adaptive MCMC. *Statistics and Computing* 16 (4), 339–354.
- Heiskanen, J., Rautiainen, M., Stenberg, P., Möttus, M., Vesanto, V.-H., 2013. Sensitivity of narrowband vegetation indices to boreal forest LAI, reflectance seasonality and species composition. *ISPRS Journal of Photogrammetry and Remote Sensing* 78 (0), 1–14.
- Heiskanen, J., Rautiainen, M., Stenberg, P., Möttus, M., Vesanto, V.-H., Korhonen, L., Majasalmi, T., 2012. Seasonal variation in modis lai for a boreal forest area in finland. *Remote Sensing of Environment* 126, 104–115.
- Knyazikhin, Y., Kranigk, J., Myneni, R. B., Panforyov, O., Gravenhorst, G., 1998a. Influence of small-scale structure on radiative transfer and photosynthesis in vegetation canopies. *J. Geophys. Res.* 103 (D6), 6133–6144.
- Knyazikhin, Y., Martonchik, J. V., Myneni, R. B., Diner, D. J., Running, S. W., 1998b. Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data. *J. Geophys. Res.* 103 (D24), 32257–32276.
- Le Maire, G., François, C., Soudani, K., Berveiller, D., Pontailier, J.-Y., Bréda, N., Genet, H., Davi, H., Dufrêne, E., 2008. Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. *Remote Sensing of Environment* 112 (10), 3846–3864.
- Linder, S., 1972. Seasonal variation of pigments in needles: A study of scots pine and norway spruce seedlings grown under different nursery conditions. *Studia Forestalia Suecica* 100, 1–37.
- Lukeš, P., Stenberg, P., Rautiainen, M., Möttus, M., Vanhatalo, K. M., 2013. Optical properties of leaves and needles for boreal tree species in Europe. *Remote Sensing Letters* 4 (7), 667–676.
- Möttus, M., Stenberg, P., 2008. A simple parameterization of canopy reflectance using photon recollision probability. *Remote Sensing of Environment* 112, 1545–1551.

344 Majasalmi, T., Rautiainen, M., Stenberg, P., Rita, H., 2012. Optimizing the sampling scheme for lai-2000 measurements in a boreal forest. *Agric-*
345 *cultural and forest meteorology* 154, 38–43.

346 Manninen, T., Stenberg, P., 2009. Simulation of the effect of snow covered forest floor on the total forest albedo. *Agricultural and Forest Meteorol-*
347 *ogy* 149, 303–319.

348 Matthew, M. W., Adler-Golden, S. M., Berk, A., Richtsmeier, S. C., Levine, R. Y., Bernstein, L. S., Acharya, P. K., Anderson, G. P., Felde, G. W.,
349 Hoke, M. P., 2000. Status of atmospheric correction using a MODTRAN4-based algorithm. Vol. 4049. pp. 199–207.

350 Meroni, M., Colombo, R., Panigada, C., 2004. Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar
351 plantations. *Remote Sensing of Environment* 92 (2), 195–206.

352 Pearlman, J. S., Barry, P. S., Segal, C. C., Shepanski, J., Beiso, D., Carman, S. L., 2003. Hyperion, a space-based imaging spectrometer. *IEEE*
353 *Transactions on Geoscience and Remote Sensing* 41 (6), 1160–1173.

354 Peltoniemi, J. I., Suomalainen, J., Puttonen, E., Näränen, J., Rautiainen, M., 2008. Reflectance properties of selected arctic-boreal land cover types:
355 field measurements and their application in remote sensing. *Biogeosciences Discuss.* 5, 1069–1095.

356 Rautiainen, M., Möttö, M., Heiskanen, J., Akujärvi, A., Majasalmi, T., Stenberg, P., 2011. Seasonal reflectance dynamics of common understory
357 types in a northern european boreal forest. *Remote Sensing of Environment* 115 (12), 3020–3028.

358 Rautiainen, M., Stenberg, P., 2005. Application of photon recollision probability in coniferous canopy reflectance simulations. *Remote Sensing of*
359 *Environment* 96, 98–107.

360 Schlerf, M., Atzberger, C., 2006. Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing
361 data. *Remote sensing of environment* 100 (3), 281–294.

362 Stenberg, P., 2007. Simple analytical formula for calculating average photon recollision probability in vegetation canopies. *Remote Sensing of*
363 *Environment* 109, 221–224.

364 Stenberg, P., Lukeš, P., Rautiainen, M., Manninen, T., 2013. A new approach for simulating forest albedo based on spectral invariants. *Remote*
365 *sensing of environment* 137, 12–16.

366 Sun, L., Neville, R., Staenz, K., White, H. P., 2008. Automatic destriping of hyperion imagery based on spectral moment matching. *Canadian*
367 *Journal of Remote Sensing* 34 (S1), S68–S81.

368 Thérézien, M., Palmroth, S., Brady, R., Oren, R., 2007. Estimation of light interception properties of conifer shoots by an improved photographic
369 method and a 3D model of shoot structure. *Tree physiology* 27 (10), 1375–1387.

370 Varvia, P., Rautiainen, M., Seppänen, A., 2017. Modeling uncertainties in estimation of canopy LAI from hyperspectral remote sensing data – A
371 Bayesian approach. *Journal of Quantitative Spectroscopy and Radiative Transfer* 191, 19–29.

372 Welles, J. M., 1990. Some indirect methods of estimating canopy structure. *Remote sensing reviews* 5 (1), 31–43.

Figure 2: Seasonal course of a broad leaved example stand. From top to bottom: 1) The measured forest reflectance spectrum. 2) Posterior marginal density of LAI_{eff} , posterior mean is marked with a circle, the field-measured value by a cross; 95% CI is shaded. 3) Estimated leaf albedo ω_L (black line), with 50%, 90% and 95% credible envelopes (shaded area from the darkest to the lightest). 4) Estimated understory BRF ρ_g (black line), with 50%, 90% and 95% credible envelopes (shaded area from the darkest to the lightest). 5) Posterior marginal density of β , posterior mean is marked with a circle, 95% CI is shaded.

Figure 3: Seasonal course of a coniferous example stand. From top to bottom: 1) The measured forest reflectance spectrum. 2) Posterior marginal density of LAI_{eff} , posterior mean is marked with a circle, the field-measured value by a cross; 95% CI is shaded. 3) Estimated leaf albedo ω_L (black line), with 50%, 90% and 95% credible envelopes (shaded area from the darkest to the lightest). 4) Estimated understory BRF ρ_g (black line), with 50%, 90% and 95% credible envelopes (shaded area from the darkest to the lightest). 5) Posterior marginal density of β , posterior mean is marked with a circle, 95% CI is shaded.

Figure 4: Seasonal behavior of the leaf albedo estimates. Coniferous stands have solid lines, deciduous stands are dashed. The shaded background corresponds to 95%, 90% and 50% credible intervals of the prior density.

Figure 5: Seasonal behavior of the understory reflectance estimates. Coniferous stands have solid lines, deciduous stands are dashed. The shaded background corresponds to 95%, 90% and 50% credible intervals of the prior density.

Figure 6: The field-measured understory reflectance (solid line) and the estimated understory reflectance (dashed line) on the common spectral interval by month and understory type. The estimated understory reflectance was computed as an average over the stand-wise estimates with the given month and understory type.