

**Research Project:**  
**Retrieval of plant biophysical and  
biochemical variables from remote  
sensing data using a hybrid machine  
learning method**



 **Trier University**

Zavud Baghirov  
Environmental Sciences  
University of Trier

Summer Semester 2021

# **Abstract**

This will be the abstract at the end [TO BE UPDATED]

# Contents

<b>List of Figures</b>	<b>iv</b>
<b>List of Tables</b>	<b>v</b>
<b>List of Abbreviations</b>	<b>vi</b>
<b>1 Methods</b>	<b>1</b>
1.1 Local sensitivity analysis . . . . .	1
1.2 RTM simulation (INFORM) . . . . .	3
<b>2 Results</b>	<b>6</b>
2.1 Local sensitivity analysis . . . . .	6
2.2 RTM simulation (INFORM) . . . . .	8
<b>References</b>	<b>9</b>

## List of Figures

2.1	Effects of varying the chosen parameters on the simulated spectra .	7
-----	---	---

## List of Tables

1.1	INFORM Parameters varied in local sensitivity analysis (each parameter were varied 15 times) . . . . .	1
1.2	INFORM Parameters that were kept constant while one parameter was varied at a time . . . . .	2
1.3	Range of full input parameters that were used to create a LUT size of 316800 . . . . .	4

## List of Abbreviations

<b>3D</b>	Three-dimensional
<b>INFORM</b>	Invertable Forest Reflectance Model
<b>RTM</b>	Radiative Transfer Model
<b>SAIL</b>	Scattering by Arbitrary Inclined Leaves
<b>PROSAIL</b>	The combination of PROSPECT and SAIL models
<b>FLIM</b>	Forest Light Interaction Model
<b>LAI</b>	Leaf Area Index
<b>MLRA</b>	Machine Learning Regression Algorithms
<b>ML</b>	Machine Learning
<b>DT</b>	Decision Trees
<b>ANN</b>	Artificial Neural Networks
<b>KBMLRM</b>	Kernel-Based Machine Learning Regression Methods
<b>RF</b>	Random Forest
<b>RFR</b>	Random Forest Regression
<b>LUT</b>	Look-Up-Table
<b>NN</b>	Neural Networks
<b>SVR</b>	Support Vector Regression
<b>SVM</b>	Support Vector Machines
<b>GPR</b>	Gaussian Process Regression
<b>GP</b>	Gaussian Process
<b>VI</b>	Vegetation Index
<b>DR</b>	Dimensionality Reduction
<b>WT</b>	Wavelet Tranform
<b>PCA</b>	Principal Component Analysis
<b>AL</b>	Active Learning
<b>NIR</b>	Near Infrared
<b>SWIR</b>	Short Wave Infrared

# 1 Methods

This section explains the methods used in this research.

## 1.1 Local sensitivity analysis

Local sensitivity analysis was performed to assess the effect of each of the main 6 plant biochemical and biophysical variables on the PRISMA image bands. In the local sensitivity analysis simulation is performed by keeping all the variables constant at their determined fixed or default values except the parameter of interest. This way the effect of a specific parameter on the simulated spectra can be assessed. In this research the plant parameters  $C_{ab}$ ,  $C_w$ ,  $C_m$ ,  $LAI_s$ ,  $CD$  and  $SD$  were varied each 15 times (Table (1.1)), while keeping the rest of the variables at their default values (Table (1.2)). The default and varied values were chosen based on the literature (e.g. Darvishzadeh et al. (2019); Laurent et al. (2011); Schlerf and Atzberger (2012)) where similar RTM method used to simulate reflectance for Spruce trees.

Table 1.1 shows the 6 parameters that were varied, their units, minimum and maximum values. Each parameter was varied 15 times, meaning 15 different spectra were simulated for each variable.

**Table 1.1:** INFORM Parameters varied in local sensitivity analysis (each parameter were varied 15 times)

Parameter	Abbrev.	Unit	Min	Max
Chlorophyll content	$C_{ab}$	$\frac{\mu g}{cm^2}$	20	60
Equivalent water thickness	$C_w$	$\frac{g}{cm^2}$	0.0035	0.035
Leaf dry matter content	$C_m$	$\frac{g}{cm^2}$	0.008	0.03
Leaf area index (single)	$LAI_s$	$\frac{m^2}{m^2}$	0	7
Stem density	$SD$	$ha^{-1}$	200	5000
Crown diameter	$CD$	$m$	1.5	8.5

## 1. Methods

Table 1.2 shows the determined default values for each INFORM parameter that were kept during the sensitivity simulation while one of the parameter was varied (Table 1.1).

**Table 1.2:** INFORM Parameters that were kept constant while one parameter was varied at a time

Parameter	Abbr	Unit	Value
Leaf structure parameter	$N$	—	3
Chlorophyll content	$C_{ab}$	$\frac{\mu g}{cm^2}$	40
Leaf carotenoid content	$C_{ar}$	$\frac{\mu g}{cm^2}$	8
Brown Pigment Content	$C_{brown}$	—	0.001
Equivalent water thickness	$C_w$	$\frac{g}{cm^2}$	0.0117
Leaf dry matter content	$C_m$	$\frac{g}{cm^2}$	0.03
Average leaf inclination angle	$ALIA$	$^\circ$	65
Leaf area index (single)	$LAI_s$	$\frac{m^2}{m^2}$	6
Leaf area index (understorey)	$LAI_u$	$\frac{m^2}{m^2}$	0.5
Hot spot parameter	$Hot$	$\frac{m}{m}$	0.02
Solar zenith angle	$tts$	$^\circ$	45.43
Observer zenith angle	$tto$	$^\circ$	0
Sun-sensor azimuth angle	$psi$	$^\circ$	181.41
Soil brightness	$\alpha_{soil}$	—	0.5
Stem density	$SD$	$ha^{-1}$	700
Crown diameter	$CD$	$m$	5
Mean Height	$H$	$m$	20
Fraction of diffuse incoming	$skyl$	—	0.1
Soil reflectance spectrum	$B_g$	—	default

*Solar zenith angle* and *Sun-sensor azimuth angle* were calculated based on the PRISMA image acquisition parameters (date, lat/long etc.) using the *solar position calculator* at <https://www.esrl.noaa.gov/gmd/grad/solcalc/azel.html>.

RTM models PROSPECT5, 4SAIL and FLIM were coupled (INFORM) in order to simulate canopy reflectance. Simulations were carried out using the *ccrtm* package (Visser, 2021) in *R* (R Core Team, 2021). The default soil spectra provided by the *ccrtm* package (Visser, 2021) was used for the simulations. Spectral resampling was performed in order to resample the INFORM output spectra (1nm



## 1. Methods

resolution) into PRISMA image bands. For spectral resampling the *R* package *hdsar* (Lehnert et al., 2019) was utilized.

### 1.2 RTM simulation (INFORM)

PROSPECT5, 4SAIL and FLIM RTM models were coupled (INFORM) to simulate forest canopy reflectance based on different values of plant biophysical and biochemical parameters. The 6 parameters that were mentioned in the previous chapter were varied and spectra was simulated based on each combination of these variables. The number of combinations increase exponentially, which in turn requires increased computational power. Therefore, the trade-off must be taken into account between computational power or time and accurate simulation.

Different authors suggest different number of LUT size for RTM simulation. For example, Danner et al. (2021) mention that LUT size of minimum 50,000 is recommended. Ali et al. (2020) and Darvishzadeh et al. (2019) created a LUT size of 100,000 and 500,000 respectively.

In this research, LUT size of 316,800 was created based on each combination of different plant biophysical and biochemical parameters. The range of the varied parameters and parameters that were kept constant were determined based on the suggestions of the studies that were mentioned in the previous chapter. These studies used similar methods to simulate canopy reflectance for mainly Spruce forests/trees.

Table 1.3 shows the variables that were used to simulate forest canopy parameters. Table 1.3 also contains information about the range of the values and how many times each parameter was varied.

## 1. Methods

**Table 1.3:** Range of full input parameters that were used to create a LUT size of 316800

Parameter	Abbr	Unit	Min	Max	Steps
Leaf structure parameter	$N$	—	3	3	—
Chlorophyll content	$C_{ab}$	$\frac{\mu g}{cm^2}$	20	60	15
Leaf cartenoid content	$C_{ar}$	$\frac{\mu g}{cm^2}$	8	8	—
Brown Pigment Content	$C_{brown}$	—	0.001	0.001	—
Equivalent water thickness	$C_w$	$\frac{g}{cm^2}$	0.0035	0.035	10
Leaf dry matter content	$C_m$	$\frac{g}{cm^2}$	0.008	0.03	11
Average leaf inclination angle	$ALIA$	$^\circ$	65	65	—
Leaf area index (single)	$LAI_s$	$\frac{m^2}{m^2}$	0	6.5	16
Leaf area index (understorey)	$LAI_u$	$\frac{m^2}{m^2}$	0.5	0.5	—
Hot spot parameter	$Hot$	$\frac{m}{m}$	0.02	0.02	—
Solar zenith angle	$tts$	$^\circ$	45.43	45.43	—
Observer zenith angle	$tto$	$^\circ$	0	0	—
Sun-sensor azimuth angle	$psi$	$^\circ$	181.41	181.41	—
Soil brightness	$\alpha_{soil}$	—	0.5	0.5	—
Stem density	$SD$	$ha^{-1}$	200	5000	4
Crown diameter	$CD$	$m$	1.5	8.5	3
Mean Height	$H$	$m$	20	20	—
Fraction of diffuse radiation	$skyl$	—	0.1	0.1	—
Soil reflectance spectrum	$B_g$	—	default	default	—

All simulations were performed using the library *ccrtm* (Visser, 2021) in *R* programming language (R Core Team, 2021) using the most recent version 4.1.0. Generating a LUT size of 316,800 is an expensive process from a computational standpoint (depending on how much computer resources and time are available this might change). Also, all simulations are independent of each other, meaning simulation of one spectra has no effect on the other, as every simulated spectra is simulated based on a different combination of parameters. These two factors make the generation of such a large LUT good candidate for parallel computation. Therefore, the software packages *doParallel* (Corporation and Weston, 2020) and *foreach* (Microsoft and Weston, 2020) were utilized for parallel computation (using all the available cores) in *R* programming language (R Core Team, 2021). This significantly reduced the computational time. All of the simulations were computed on a Lenovo Thinkpad E480 running under Windows 10 operating system with a

## *1. Methods*

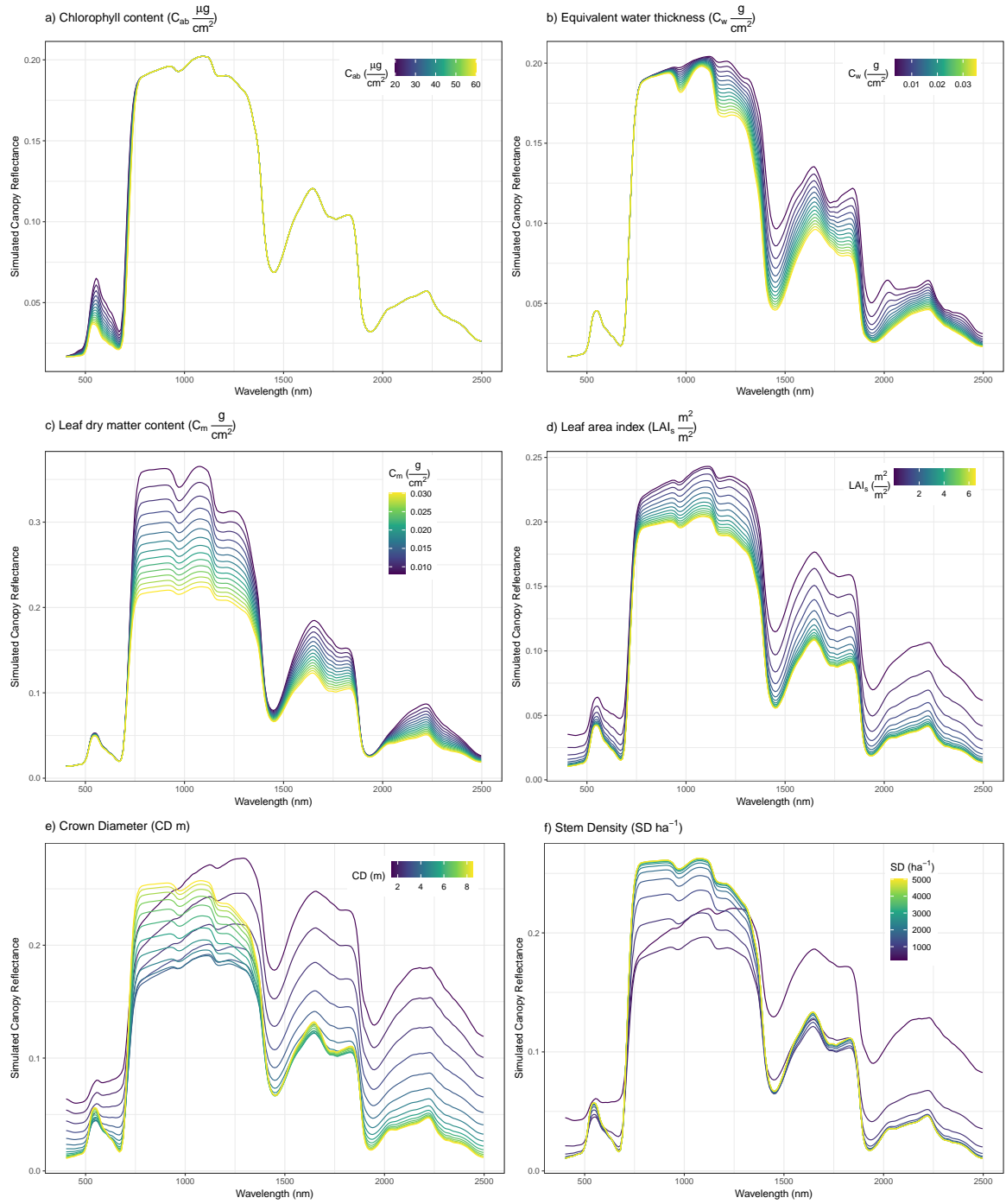
processor Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz, 2001 Mhz, 4 Core(s),  
8 logical processor(s).

## 2 Results

### 2.1 Local sensitivity analysis

Figure 2.1 shows the result of sensitivity analysis. Chlorophyll content ( $C_{ab}$ ) appears to almost exclusively impact the visible spectra. Some effect can also be noticed in the red-edge, but there is not a significant effect of varying  $C_{ab}$  on the simulated spectra within the near-infrared (NIR) and short wave infrared (SWIR) (Figure 2.1.a). Conversely, equivalent water thickness ( $C_w$ ) (Figure 2.1.b) and leaf dry matter content ( $C_m$ ) (Figure 2.1.c) both have large effects on simulated spectra within the NIR and SWIR but no significant effect within the visible spectra. Leaf Area Index (single) ( $LAI_s$ ) (Figure 2.1.d), Crown diameter ( $CD$ ) (Figure 2.1.e)) and Stem density (Figure 2.1.f) all have noticeable effect on the simulated canopy reflectance almost all over the spectra.

## 2. Results



**Figure 2.1:** Effects of varying the chosen parameters on the simulated spectra

## 2.2 RTM simulation (INFORM)

Synthetic canopy reflectance data set were produced and stored in a LUT containing all 316,800 simulations. In this research, LUT was defined as a matrix. Each row of this matrix is a different simulated spectra and columns are simulated reflectance of wavelengths with the range of 400nm-2500nm with 1nm spectral resolution and 6 additional columns containing values of the corresponding variables  $C_{ab}$ ,  $C_w$ ,  $C_m$ ,  $LAI_s$ ,  $CD$  and  $SD$  that were used for each simulation. Hence the dimensions of the LUT matrix is 316,800 (number of simulations) rows by 2107 (2101 simulated “bands” + 6 INFORM variables) columns:

$$\begin{bmatrix} 400nm_1 & \dots & 2500nm_1 & C_{ab_1} & C_{w_1} & C_{m_1} & LAI_{s_1} & CD_1 & SD_1 \\ 400nm_2 & \dots & 2500nm_2 & C_{ab_2} & C_{w_2} & C_{m_2} & LAI_{s_2} & CD_2 & SD_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 400nm_{316,800} & \dots & 2500nm_{316,800} & C_{ab_{316,800}} & C_{w_{316,800}} & C_{m_{316,800}} & LAI_{s_{316,800}} & CD_{316,800} & SD_{316,800} \end{bmatrix}$$

In this matrix,  $400nm_n, \dots, 2500nm_n$  refer to the simulated reflectance for the corresponding wavelength in the simulation number  $n$ .  $C_{ab_n}$ ,  $C_{w_n}$ ,  $C_{m_n}$ ,  $LAI_{s_n}$ ,  $CD_n$  and  $SD_n$  are values of the INFORM parameters that were used in the  $n$ th simulation.

## References

- Ali, A. M., Darvishzadeh, R., Skidmore, A., Gara, T. W., and Heurich, M. Machine learning methods’ performance in radiative transfer model inversion to retrieve plant traits from sentinel-2 data of a mixed mountain forest. *International Journal of Digital Earth*, pages 1–15, 2020.
- Corporation, M. and Weston, S. *doParallel: Foreach Parallel Adaptor for the ‘parallel’ Package*, 2020. URL <https://CRAN.R-project.org/package=doParallel>. R package version 1.0.16.
- Danner, M., Berger, K., Wocher, M., Mauser, W., and Hank, T. Efficient rtm-based training of machine learning regression algorithms to quantify biophysical & biochemical traits of agricultural crops. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173:278–296, 2021.
- Darvishzadeh, R., Skidmore, A., Abdullah, H., Cherenet, E., Ali, A., Wang, T., Nieuwenhuis, W., Heurich, M., Vrieling, A., O’Connor, B., et al. Mapping leaf chlorophyll content from sentinel-2 and rapideye data in spruce stands using the invertible forest reflectance model. *International Journal of Applied Earth Observation and Geoinformation*, 79:58–70, 2019.
- Laurent, V. C., Verhoef, W., Clevers, J. G., and Schaepman, M. E. Inversion of a coupled canopy–atmosphere model using multi-angular top-of-atmosphere radiance data: A forest case study. *Remote Sensing of Environment*, 115(10):2603–2612, 2011.
- Lehnert, L. W., Meyer, H., Obermeier, W. A., Silva, B., Regeling, B., Thies, B., and Bendix, J. Hyperspectral data analysis in R: The hsdar package. *Journal of Statistical Software*, 89(12):1–23, 2019. doi: 10.18637/jss.v089.i12.
- Microsoft and Weston, S. *foreach: Provides Foreach Looping Construct*, 2020. URL <https://CRAN.R-project.org/package=foreach>. R package version 1.5.1.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL <https://www.R-project.org/>.
- Schlerf, M. and Atzberger, C. Vegetation structure retrieval in beech and spruce forests using spectrodirectional satellite data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(1):8–17, 2012.
- Visser, M. D. *ccrtm: Coupled Chain Radiative Transfer Models*, 2021. URL <https://CRAN.R-project.org/package=ccrtm>. R package version 0.2.